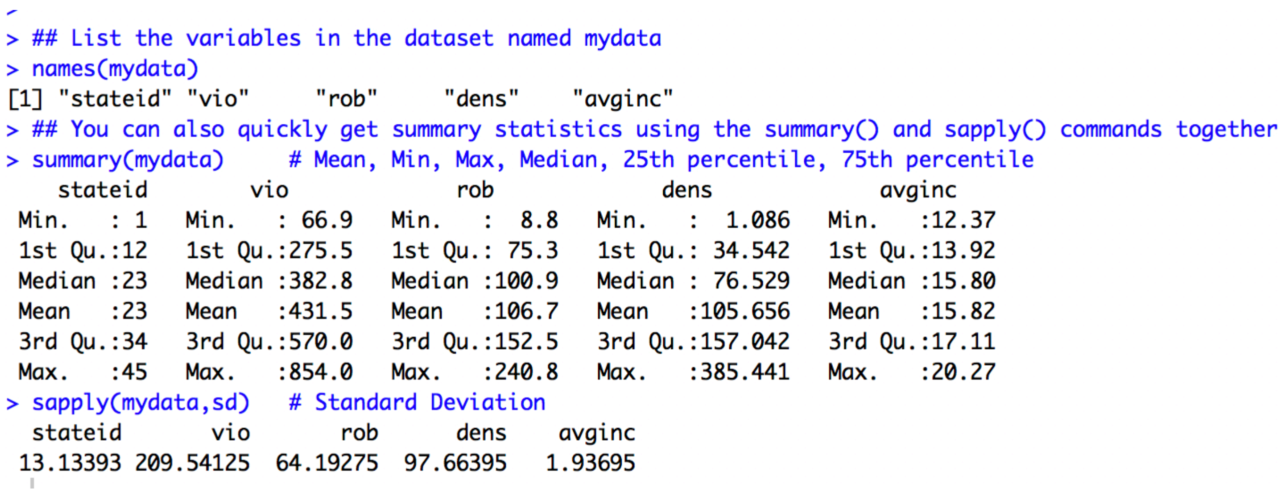
ECOM20001: Econometrics 1

Tutorial 2: Suggested Solutions

Part 1: Visualising and Describing Data in R

1. Descriptive statistics

R output from the summary statistics:



```
> ## List the variables in the dataset named mydata
> names(mydata)
[1] "stateid" "vio"     "rob"     "dens"    "avginc"
> ## You can also quickly get summary statistics using the summary() and sapply() commands together
> summary(mydata)     # Mean, Min, Max, Median, 25th percentile, 75th percentile
    stateid         vio             rob              dens              avginc
 Min.   : 1    Min.   : 66.9   Min.   :  8.8   Min.   :  1.086   Min.   :12.37
 1st Qu.:12    1st Qu.:275.5   1st Qu.: 75.3   1st Qu.: 34.542   1st Qu.:13.92
 Median :23    Median :382.8   Median :100.9   Median : 76.529   Median :15.80
 Mean   :23    Mean   :431.5   Mean   :106.7   Mean   :105.656   Mean   :15.82
 3rd Qu.:34    3rd Qu.:570.0   3rd Qu.:152.5   3rd Qu.:157.042   3rd Qu.:17.11
 Max.   :45    Max.   :854.0   Max.   :240.8   Max.   :385.441   Max.   :20.27
> sapply(mydata,sd)    # Standard Deviation
  stateid       vio       rob      dens    avginc
 13.13393 209.54125  64.19275  97.66395   1.93695
```

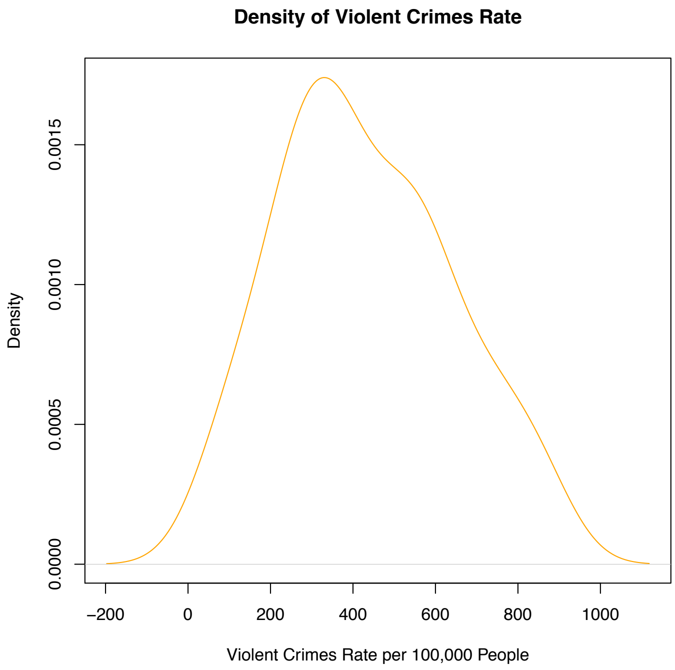
So we see a **typical state** has: 431 violent crimes per 100,000 people, 107 robberies per 100,000 people, an urban density of 106 people per square mile, and an average annual per-capita income of $15,8200 per year.

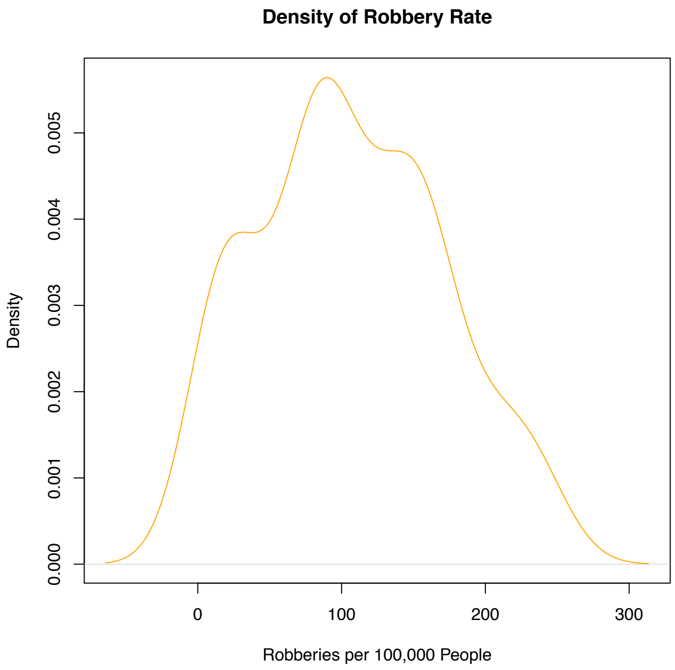
The range of robbery and violence rates is remarkable. Some states have only 67 violent crimes per 100,000 people per year, while others have up to 854 (!) violent crimes per 100,000 people per year. It's more than 10 times the difference between the least and most violent crime rates across states. Similarly, the robbery rate is as small as 9 robberies per 100,000 people year and goes up to 240 (!) per 100,000 people per year.

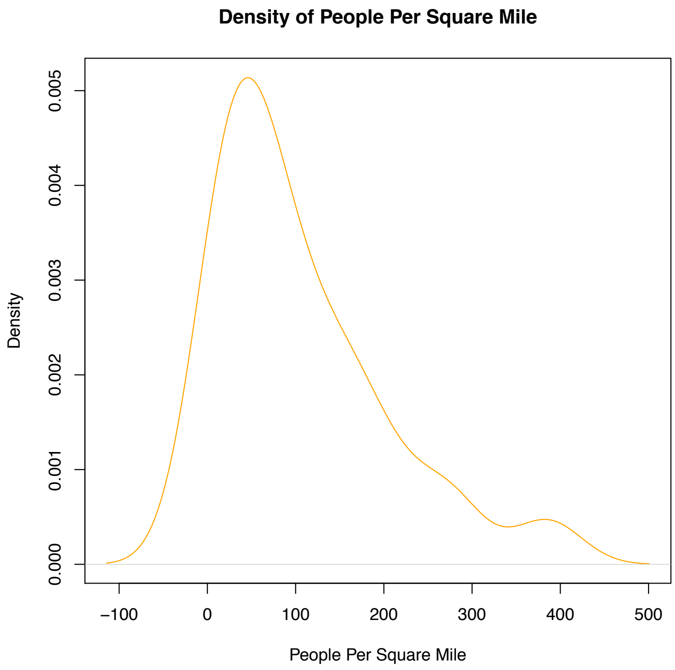
We also have some very rural (1 person per square mile) and urban (385 people per square mile) states. And per capital income similarly ranges from $12,370 to $20,270.

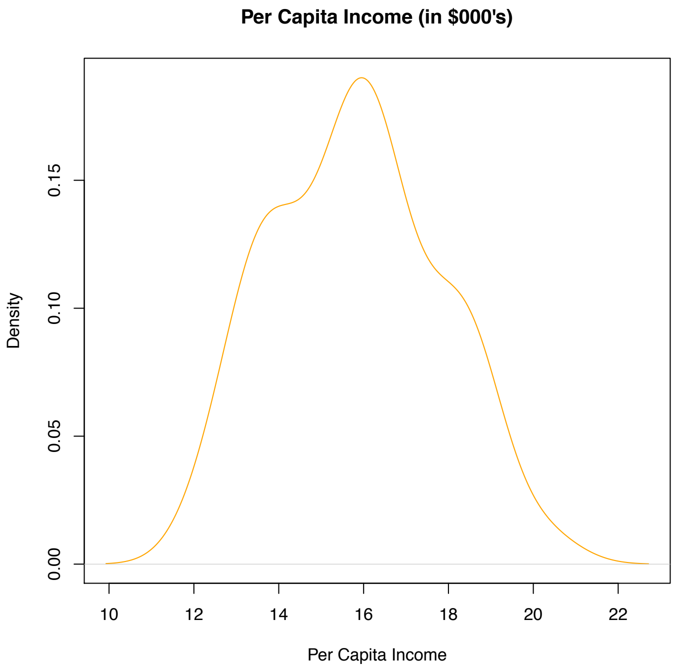
2. Probability densities

The probability densities of the four main variables are listed below. The urban density variable is right skewed, which means there are many similarly-dense US states, but a few in the right tail of the distribution that are very dense such as New York and California. The other distributions are relatively symmetric around their means.


Density of Violent Crimes Rate


Density of Robbery Rate

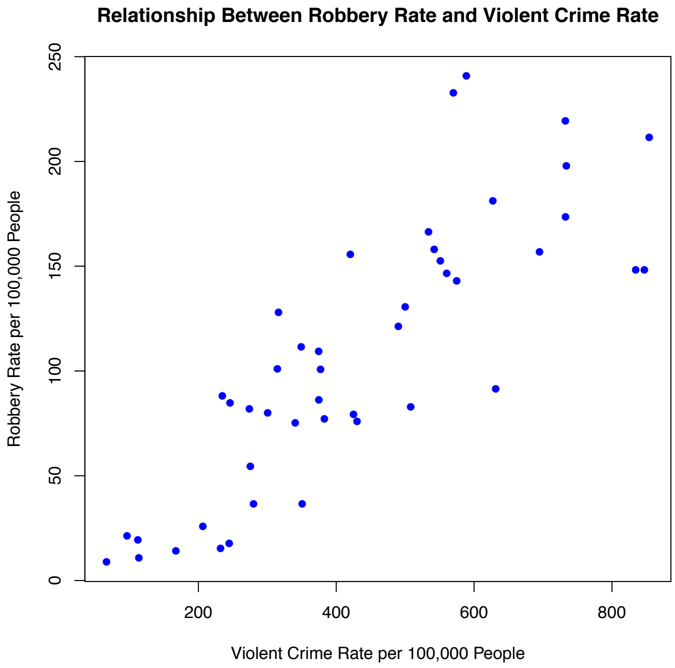

Density of People Per Square Mile

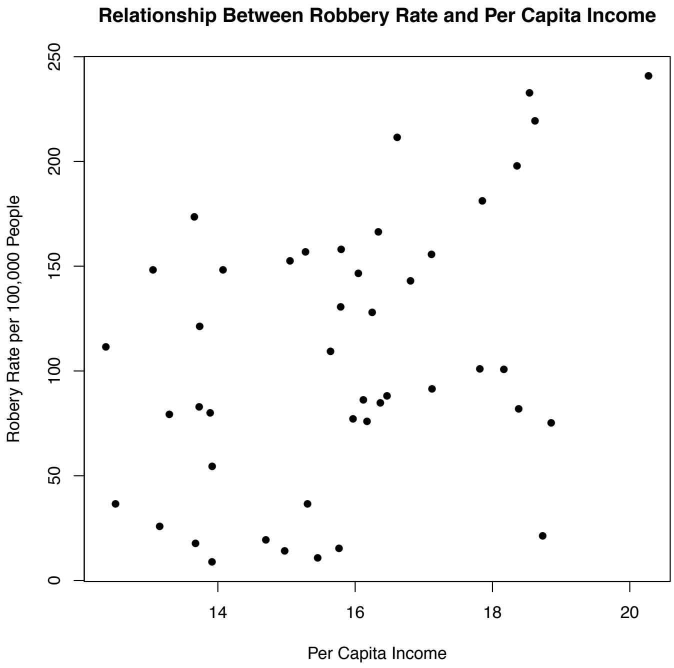

Per Capita Income (in $000's)

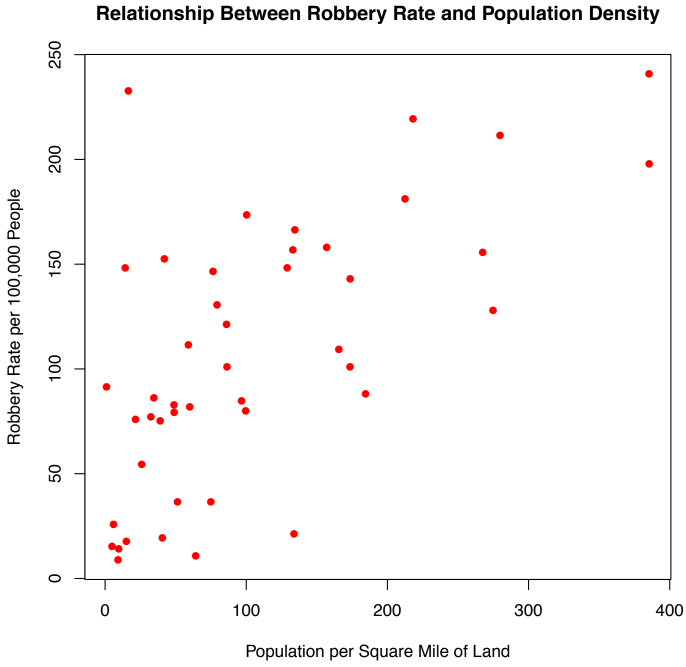
3. Scatter plots

The key scatter plots are listed on the next page below.

- Perhaps not surprisingly, there is a strong positive correlation between the robbery and violent crime rate. A natural interpretation is that similar types of people or demographics are likely to engage in robberies and violent crimes which can underly the strong correlation.
- The relationship between robberies and income is basically non-existent. Economists commonly refer to a lack of correlation in a scatter plot like this as a **cloud.** A cloud often results from economic explanations running into each other.
    - Higher income could mean more benefit on average from robbery, so we should expect a positive relationship between robberies and per capita income
    - However, higher income states may have higher tax bases to pay for more effective police forces which makes the cost of robbery higher. This would create a negative relationship between robberies and per capita income
    - This **collision of benefits and costs** associated with higher income locations can cause a cloud correlation, like we see in the second panel below.
- Finally, we see a positive relationship between robbery rates and urban density. As the footnote in the tutorial question alluded to, this could potentially reflect:
    - The cost of robbery being lower in more dense states as potential robbery targets are more plentiful in close proximity
    - The benefit of robbery being higher if more dense locations attract more retail shops and merchants (called “agglomeration” benefits of urban density), which provides more opportunities and hence benefit for robbery
    - More difficult for police to identify potential robbers in more crowded places, which again makes the expected costs of robbery lower since robbers are less likely to be caught
- To be clear: all of these “explanations” are just hypotheses and none of them are proved from a simple scatter plot. And there are potentially many other hypotheses. Later in ECOM20001, and throughout ECOM30002: Econometrics 2, we develop empirical approaches to unpack these various explanations for correlations found in scatter plots.


Relationship Between Robbery Rate and Violent Crime Rate
Robbery Rate per 100,000 People
0
50
100
150
200
250
200
400
600
800
Violent Crime Rate per 100,000 People


Relationship Between Robbery Rate and Per Capita Income
Robbery Rate per 100,000 People
0
50
100
150
200
250
14
16
18
20
Per Capita Income


Relationship Between Robbery Rate and Population Density
Robbery Rate per 100,000 People
0
50
100
150
200
250
0
100
200
300
400
Population per Square Mile of Land

Part 2: Summation Practice Problems

1. Show the following equality is true:

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0$$

Solution:

$$\begin{aligned}
\sum_{i=1}^{n}(x_i - \bar{x}) &= \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} \\
&= \sum_{i=1}^{n} x_i - n\bar{x} \\
&= \sum_{i=1}^{n} x_i - \cancel{n}\frac{\sum_{i=1}^{n} x_i}{\cancel{n}} \\
&= \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i \\
&= 0
\end{aligned}$$

2. Show the following equality is true:

$$n\bar{x} = \sum_{i=1}^{n} x_i$$

Solution:

$$\begin{aligned} n\bar{x} &= \cancel{n}\frac{\sum_{i=1}^{n} x_i}{\cancel{n}} \\ &= \sum_{i=1}^{n} x_i \end{aligned}$$

Notice how this means you can manipulate summations $\sum_{i=1}^{n} x_i$ and multiply by them by $\frac{n}{n} = 1$ to get means and sample sizes:

$$\sum_{i=1}^{n} x_i = \frac{n}{n}\sum_{i=1}^{n} x_i = n\bar{x}$$

3. Show the following equality is true:

$$\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}x_i^2-n\bar{x}^2$$

Solution:

$$\begin{aligned}
\sum_{i=1}^{n}(x_i-\bar{x})^2 &= \sum_{i=1}^{n}\left(x_i^2-2x_i\bar{x}+\bar{x}^2\right)\\
&= \sum_{i=1}^{n}x_i^2-\sum_{i=1}^{n}2x_i\bar{x}+\sum_{i=1}^{n}\bar{x}^2\\
&= \sum_{i=1}^{n}x_i^2-2\bar{x}\sum_{i=1}^{n}x_i+n\bar{x}^2\\
&= \sum_{i=1}^{n}x_i^2-2n\bar{x}\frac{\sum_{i=1}^{n}x_i}{n}+n\bar{x}^2; \quad (\text{multiply by } \frac{n}{n})\\
&= \sum_{i=1}^{n}x_i^2-\cancel{2}n\bar{x}\bar{x}+\cancel{n\bar{x}^2}\\
&= \sum_{i=1}^{n}x_i^2-n\bar{x}^2
\end{aligned}$$

4. Show the following equality is true:

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$$

Solution:

$$\begin{aligned}
\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) &= \sum_{i=1}^{n}(x_i y_i - \bar{x}y_i - \bar{y}x_i + \bar{x}\bar{y}) \\
&= \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n}\bar{x}y_i - \sum_{i=1}^{n}\bar{y}x_i + \sum_{i=1}^{n}\bar{x}\bar{y} \\
&= \sum_{i=1}^{n} x_i y_i - \bar{x}\sum_{i=1}^{n} y_i - \bar{y}\sum_{i=1}^{n} x_i + n\bar{x}\bar{y} \\
&= \sum_{i=1}^{n} x_i y_i - \bar{x}\sum_{i=1}^{n} y_i - \bar{y}\sum_{i=1}^{n} x_i + n\bar{x}\bar{y} \\
&= \sum_{i=1}^{n} x_i y_i - n\bar{x}\frac{\sum_{i=1}^{n} y_i}{n} - n\bar{y}\frac{\sum_{i=1}^{n} x_i}{n} + n\bar{x}\bar{y} \\
&= \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} - \cancel{n\bar{y}\bar{x}} + \cancel{n\bar{x}\bar{y}} \\
&= \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}
\end{aligned}$$