Question 1

Use the "USArrests" built-in dataset to plot beautiful graphs and find meaningful insight about the dataset. You may draft questions yourself and summarize the results. You will be marked for [10]

a. Creativity
b. Presentation
c. Originality
d. Summarization of results

Getting to know the dataset

Let us start by viewing and understanding the dataset.

```
data(USArrests)
help(USArrests)
```

Violent Crime Rates by US State

Description

This data set contains statistics, in arrests per 100,000 residents for assault, murder, and rape in each of the 50 US states in 1973. Also given is the percent of the population living in urban areas.

Usage

USArrests

Format

A data frame with 50 observations on 4 variables.

[,1] Murder   numeric Murder arrests (per 100,000)
[,2] Assault   numeric Assault arrests (per 100,000)
[,3] UrbanPop numeric Percent urban population
[,4] Rape   numeric Rape arrests (per 100,000)

Note

USArrests contains the data as in McNeil's monograph. For the UrbanPop percentages, a review of the table (No. 21) in the Statistical Abstracts 1975 reveals a transcription error for Maryland (and that McNeil used the same "round to even" rule that R's round() uses), as found by Daniel S Coven (Arizona).

See the example below on how to correct the error and improve accuracy for the '<n>.5' percentages.

Source

World Almanac and Book of facts 1975. (Crime rates).

Statistical Abstracts of the United States 1975, p. 20, (Urban rates), possibly available as https://books.google.ch/books?id=zl9gAAAAMAAJ&pg=PA20

```
names(USArrests)
```

> [1] "Murder" "Assault" "UrbanPop" "Rape"

```
dim(USArrests)
```

> [1] 50 4

```
View(USArrests)
```

Filter

	Murder	Assault	UrbanPop	Rape
Alabama	13.2	236	58	21.2
Alaska	10.0	263	48	44.5
Arizona	8.1	294	80	31.0
Arkansas	8.8	190	50	19.5
California	9.0	276	91	40.6
Colorado	7.9	204	78	38.7
Connecticut	3.3	110	77	11.1
Delaware	5.9	238	72	15.8
Florida	15.4	335	80	31.9
Georgia	17.4	211	60	25.8
Hawaii	5.3	46	83	20.2
Idaho	2.6	120	54	14.2
Illinois	10.4	249	83	24.0
Indiana	7.2	113	65	21.0
Iowa	2.2	56	57	11.3
Kansas	6.0	115	66	18.0
Kentucky	9.7	109	52	16.3
Louisiana	15.4	249	66	22.2
Maine	2.1	83	51	7.8
Maryland	11.3	300	67	27.8

Showing 1 to 21 of 50 entries, 4 total columns

So this is a dataset about the arrest reports in USA. There are four attributes and we can try to find the relations between them. We can also try to find the least crime filled states or the most crime filled states. The dataset provides us with a list of 50 rows (1 for each state) and its statistics for 4 types of crimes - Murder, Assault, UrbanPop and Rape. Let us try to have some closer look at the dataset by obtaining its summary.

```
summary(USArrests)
```

```
> summary(USArrests)
     Murder           Assault          UrbanPop          Rape
 Min.   : 0.800   Min.   : 45.0   Min.   :32.00   Min.   : 7.30
 1st Qu.: 4.075   1st Qu.:109.0   1st Qu.:54.50   1st Qu.:15.07
 Median : 7.250   Median :159.0   Median :66.00   Median :20.10
 Mean   : 7.788   Mean   :170.8   Mean   :65.54   Mean   :21.23
 3rd Qu.:11.250   3rd Qu.:249.0   3rd Qu.:77.75   3rd Qu.:26.18
 Max.   :17.400   Max.   :337.0   Max.   :91.00   Max.   :46.00
>
```

Let us also try to see if the four crimes have any correlation to each other.

```
cor(USArrests)
```

```
> cor(USArrests)
              Murder    Assault   UrbanPop        Rape
Murder    1.00000000 0.8018733 0.06957262 0.5635788
Assault   0.80187331 1.0000000 0.25887170 0.6652412
UrbanPop  0.06957262 0.2588717 1.00000000 0.4113412
Rape      0.56357883 0.6652412 0.41134124 1.0000000
>
```

Through the above data we infer that **Assault** are the most frequent crimes that are happening as it has the highest averages. Further **Rape** crimes are the least likely to happen (which is a good thing).

Visualising data

Running plot for these combinations, Murder and Assault do not appear to have a relation to UrbanPop. The distribution of plot points are scattered to the point that they do not appear to correlate to UrbanPop.

```
with(USArrests, plot(UrbanPop, Murder))
```
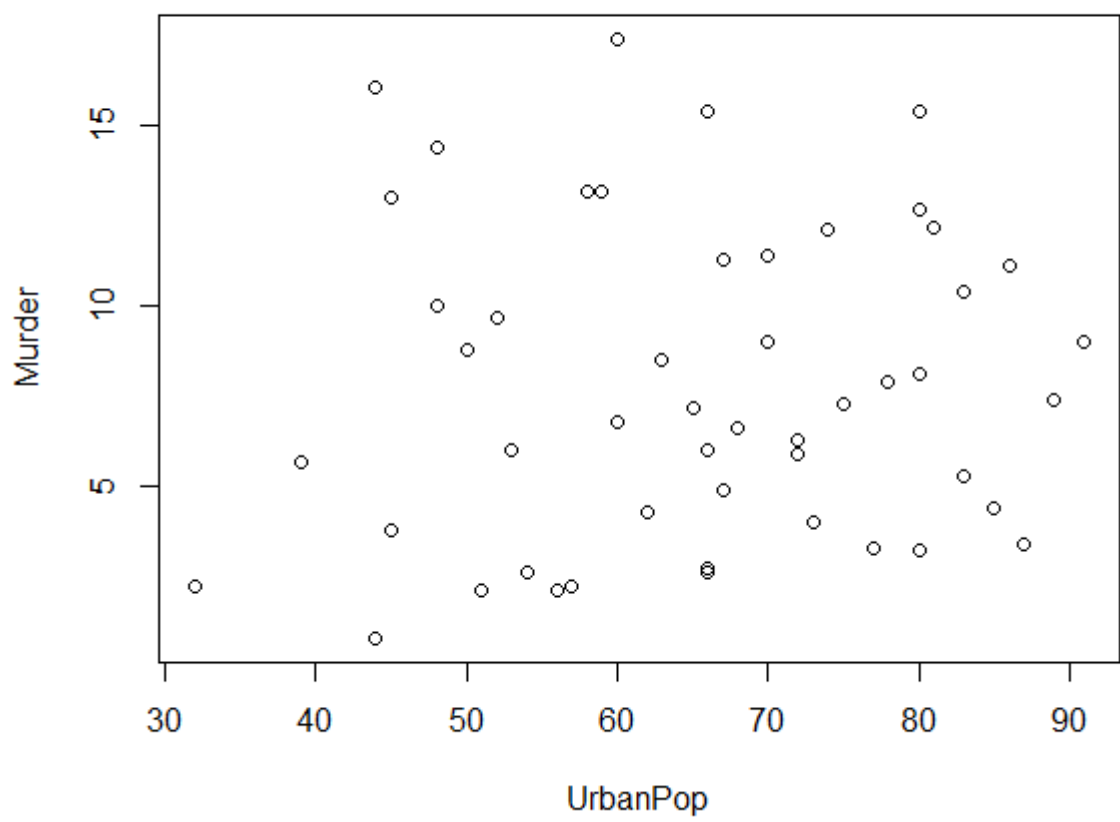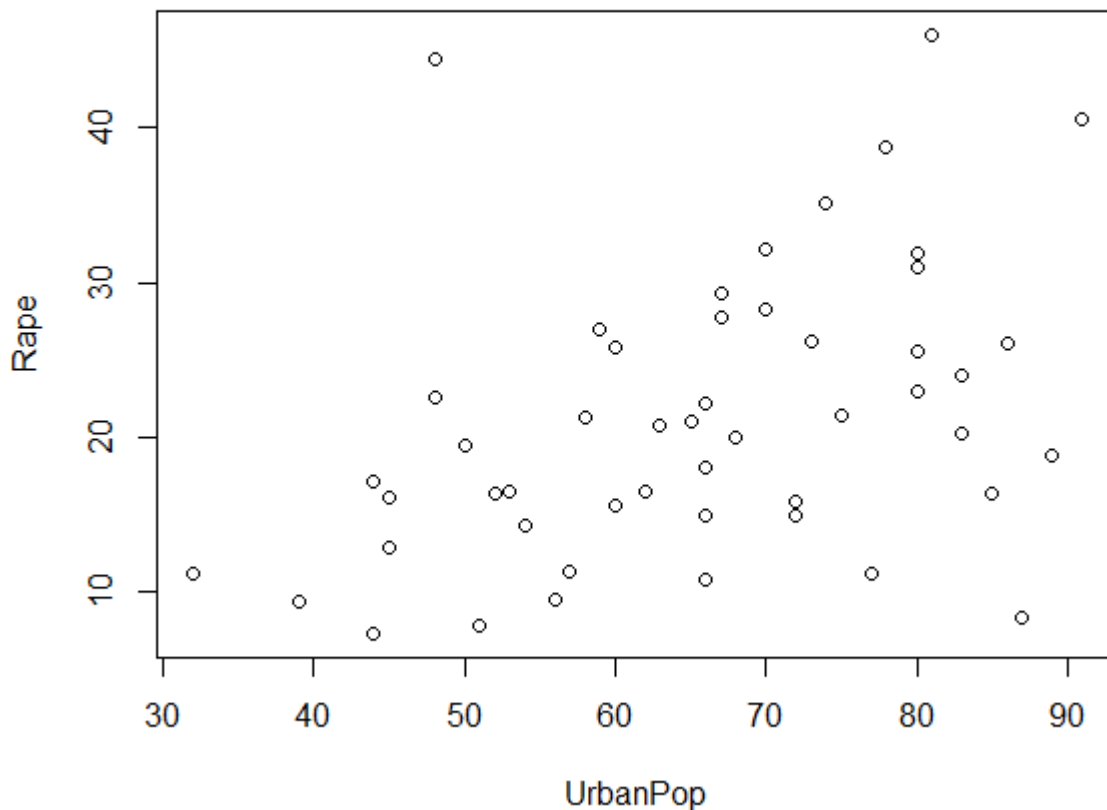

We see that most murders are likely to occur with higher Urban population. Although this isn't a fair measure since the plot seems to be very scattered. Let us try another one.

```
with(USArrests, plot(UrbanPop, Assault))
```
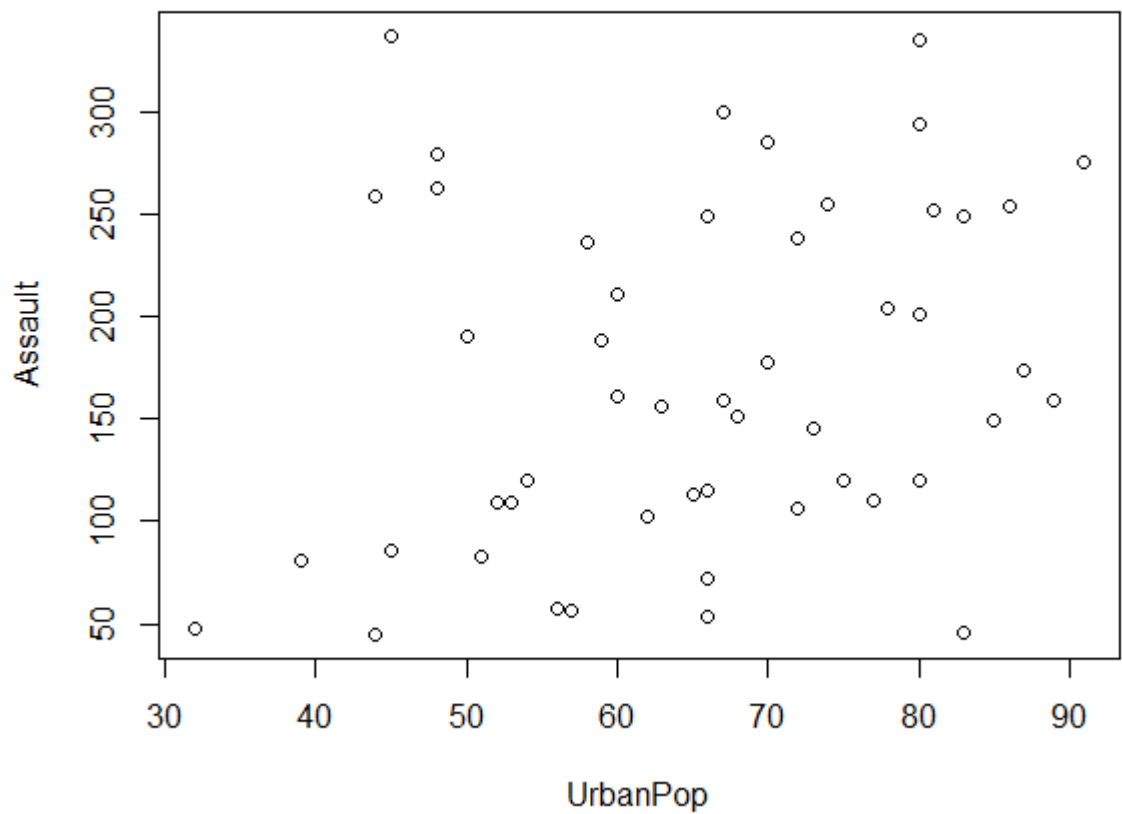
It is quite evident that more Assaults are likely to occur with more Urban Population.

```
with(USArrests, plot(UrbanPop, Rape))
```

Rape crimes are more likely to occur in states with average urban population. This goes to contrary to standard belief that more Urban population leads to more rapes.

Which states has most and least assault, murder, and rape arrests?

Let us try to figure out which states are more safer than the others so that if we ever plan a trip to US, we know where to steer clear off.

Most and Least assault

```
x <- which(USArrests$Assault == max(USArrests$Assault))
rownames(USArrests)[x]
```

[1] "North Carolina"

```
x <- which(USArrests$Assault == min(USArrests$Assault))
rownames(USArrests)[x]
```

[1] "North Dakota"

Most and Least murder

```
x <- which(USArrests$Murder == max(USArrests$Murder))
rownames(USArrests)[x]
```

[1] "Georgia"

```
x <- which(USArrests$Murder == min(USArrests$Murder))
rownames(USArrests)[x]
```

> [1] "North Dakota"

Most and least rape

```
x <- which(USArrests$Rape == max(USArrests$Rape))
rownames(USArrests)[x]
```

> [1] "Nevada"

```
x <- which(USArrests$Rape == min(USArrests$Rape))
rownames(USArrests)[x]
```

> [1] "North Dakota"

States which have assault arrests more than median of the country.

```
assault.median = median(USArrests$Assault)
assault.median
```

> [1] 159

```
subset(USArrests, Assault > assault.median, select= c(UrbanPop, Assault))
```

```
> subset(USArrests, Assault > assault.median, select= c(UrbanPop, Assault))
               UrbanPop Assault
Alabama              58     236
Alaska               48     263
Arizona              80     294
Arkansas             50     190
California           91     276
Colorado             78     204
Delaware             72     238
Florida              80     335
Georgia              60     211
Illinois             83     249
Louisiana            66     249
Maryland             67     300
Michigan             74     255
Mississippi          44     259
Missouri             70     178
Nevada               81     252
New Mexico           70     285
New York             86     254
North Carolina       45     337
Rhode Island         87     174
South Carolina       48     279
Tennessee            59     188
Texas                80     201
Wyoming              60     161
> |
```

States that are in the bottom 25% of murder

These are the safer states that I would prefer to go to.

```
bottomQuartileMurderRate <- quantile(USArrests$Murder)[2]
bottomQuartileMurderRate
```

```
subset(USArrests, Murder < bottomQuartileMurderRate, select= c(UrbanPop, Murder))
```

```
> subset(USArrests, Murder < bottomQuartileMurderRate, select= c(UrbanPop, Murder))
               UrbanPop Murder
Connecticut         77    3.3
Idaho               54    2.6
Iowa                57    2.2
Maine               51    2.1
Minnesota           66    2.7
New Hampshire       56    2.1
North Dakota        44    0.8
Rhode Island        87    3.4
South Dakota        45    3.8
Utah                80    3.2
Vermont             32    2.2
Washington          73    4.0
Wisconsin           66    2.6
> |
```

States which are in the top 25% of the murder.

Better stay away from these states for our own safety.

```
topQuartileMurderRate <- quantile(USArrests$Murder)[4]
topQuartileMurderRate
```

```
##    75%
## 11.25
```

```
subset(USArrests, Murder > topQuartileMurderRate, select= c(UrbanPop, Murder))
```

```
> subset(USArrests, Murder > topQuartileMurderRate, select= c(UrbanPop, Murder))
                UrbanPop Murder
Alabama              58   13.2
Florida              80   15.4
Georgia              60   17.4
Louisiana            66   15.4
Maryland             67   11.3
Michigan             74   12.1
Mississippi          44   16.1
Nevada               81   12.2
New Mexico           70   11.4
North Carolina       45   13.0
South Carolina       48   14.4
Tennessee            59   13.2
Texas                80   12.7
> |
```