

Assignments for Crim 250

This page will contain all the assignments you submit for the class.

This page will contain all the assignments you submit for the class.

Instructions for all assignments

I want you to submit your assignment as a PDF, so I can keep a record of what the code looked like that day. I also want you to include your answers on your personal GitHub website. This will be good practice for editing your website and it will help you produce something you can keep after the class is over.

1. Download the Assignment1.Rmd file from Canvas. You can use this as a template for writing your answers. It's the same as what you can see on my website in the Assignments tab. Once we're done with this I'll edit the text on the website to include the solutions.
2. On RStudio, open a new R script in RStudio (File > New File > R Script). This is where you can test out your R code. You'll write your R commands and draw plots here.
3. Once you have finalized your code, copy and paste your results into this template (Assignment 1.Rmd). For example, if you produced a plot as the solution to one of the problems, you can copy and paste the R code in R markdown by using the ```{r} ```` command. Answer the questions in full sentences and Save.
4. Produce a PDF file with your answers. To do this, knit to PDF (use Knit button at the top of RStudio), locate the PDF file in your docs folder (it's in the same folder as the Rproj), and submit that on on Canvas in Assignment 1.
5. Build Website, go to GitHub desktop, commit and push. Now your solutions should be on your website as well.

Assignment 1

Collaborators: none.

This assignment is due on Canvas on Monday 9/20 before class, at 10:15 am. Include the name of anyone with whom you collaborated at the top of the assignment.

Problem 1

Install the datasets package on the console below using `install.packages("datasets")`. Now load the library.

```
# install.packages("datasets")
library(datasets)
```

Load the USArrests dataset and rename it `dat`. Note that this dataset comes with R, in the package datasets, so there's no need to load data from your computer. Why is it useful to rename the dataset?

```
USArrests
```

##	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
## Massachusetts	4.4	149	85	16.3
## Michigan	12.1	255	74	35.1
## Minnesota	2.7	72	66	14.9
## Mississippi	16.1	259	44	17.1
## Missouri	9.0	178	70	28.2
## Montana	6.0	109	53	16.4
## Nebraska	4.3	102	62	16.5
## Nevada	12.2	252	81	46.0
## New Hampshire	2.1	57	56	9.5
## New Jersey	7.4	159	89	18.8
## New Mexico	11.4	285	70	32.1
## New York	11.1	254	86	26.1
## North Carolina	13.0	337	45	16.1
## North Dakota	0.8	45	44	7.3
## Ohio	7.3	120	75	21.4
## Oklahoma	6.6	151	68	20.0
## Oregon	4.9	159	67	29.3
## Pennsylvania	6.3	106	72	14.9
## Rhode Island	3.4	174	87	8.3
## South Carolina	14.4	279	48	22.5
## South Dakota	3.8	86	45	12.8
## Tennessee	13.2	188	59	26.9
## Texas	12.7	201	80	25.5
## Utah	3.2	120	80	22.9
## Vermont	2.2	48	32	11.2
## Virginia	8.5	156	63	20.7
## Washington	4.0	145	73	26.2
## West Virginia	5.7	81	39	9.3
## Wisconsin	2.6	53	66	10.8
## Wyoming	6.8	161	60	15.6

```
dat <- USArrests
```

Answer: It is useful to rename the data set because it makes it easier to remember the name we give it to use in code commands later on in the assignment. It helps to separate the base R package dataset and turn it into our own dataset to use with the following work.

Problem 2

Use this command to make the state names into a new variable called State.

```
dat$state <- tolower(rownames(USArrests))
```

This dataset has the state names as row names, so we just want to make them into a new variable. We also make them all lower case, because that will help us draw a map later - the map function requires the states to be lower case.

List the variables contained in the dataset USArrests.

```
list(dat)
```

```
## [[1]]
##           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
## Massachusetts 4.4    149      85 16.3
## Michigan      12.1    255      74 35.1
## Minnesota     2.7     72      66 14.9
## Mississippi   16.1    259      44 17.1
## Missouri      9.0    178      70 28.2
## Montana       6.0    109      53 16.4
## Nebraska      4.3    102      62 16.5
## Nevada        12.2    252      81 46.0
## New Hampshire 2.1     57      56  9.5
## New Jersey    7.4    159      89 18.8
## New Mexico    11.4    285      70 32.1
## New York      11.1    254      86 26.1
## North Carolina 13.0    337      45 16.1
## North Dakota  0.8     45      44  7.3
## Ohio          7.3    120      75 21.4
## Oklahoma      6.6    151      68 20.0
## Oregon        4.9    159      67 29.3
## Pennsylvania  6.3    106      72 14.9
## Rhode Island  3.4    174      87  8.3
## South Carolina 14.4    279      48 22.5
## South Dakota  3.8     86      45 12.8
```

```
## Tennessee      13.2      188      59 26.9
## Texas           12.7      201      80 25.5
## Utah            3.2      120      80 22.9
## Vermont         2.2       48      32 11.2
## Virginia        8.5      156      63 20.7
## Washington      4.0      145      73 26.2
## West Virginia   5.7       81      39  9.3
## Wisconsin       2.6       53      66 10.8
## Wyoming         6.8      161      60 15.6
```

```
head(dat)
```

```
##           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
```

```
names(dat)
```

```
## [1] "Murder"  "Assault" "UrbanPop" "Rape"
```

```
summary(dat)
```

```
##           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
```

Problem 3

What type of variable (from the DVB chapter) is **Murder**?

Answer: Murder is a quantitative variable, because there are measured numerical values representing murder rates for each state.

What R Type of variable is it?

Answer: Murder is a character variable, as seen in the summary (dat) description in the code above.

Problem 4

What information is contained in this dataset, in general? What do the numbers mean?

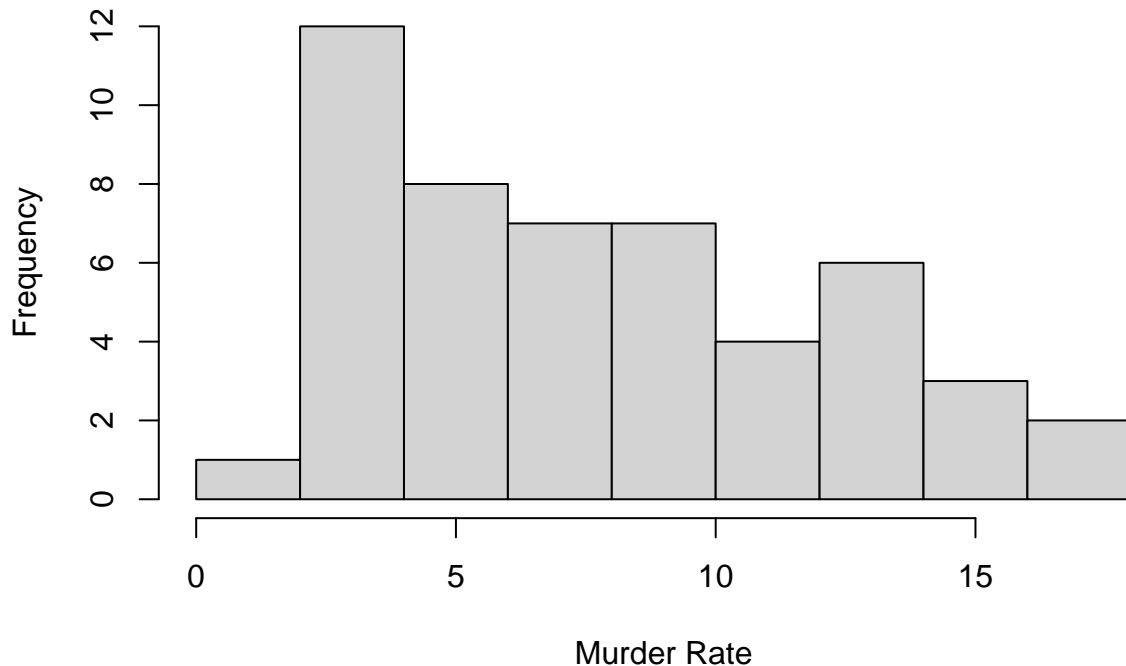
Answer: This dataset contains the number of Murders, Assaults, and Rapes in each of the 50 states in the year 1973. It also shows us how many people lived in an urban area each state in that year, demonstrated by "UrbanPop."

Problem 5

Draw a histogram of **Murder** with proper labels and title.

```
# histogram of Murder
hist(dat$Murder, main="Histogram of Murder", xlab="Murder Rate", ylab="Frequency")
```

Histogram of Murder



Problem 6

Please summarize **Murder** quantitatively. What are its mean and median? What is the difference between mean and median? What is a quartile, and why do you think R gives you the 1st Qu. and 3rd Qu.?

```
#Summary of Murder
summary(dat$Murder)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.800   4.075   7.250   7.788  11.250  17.400
```

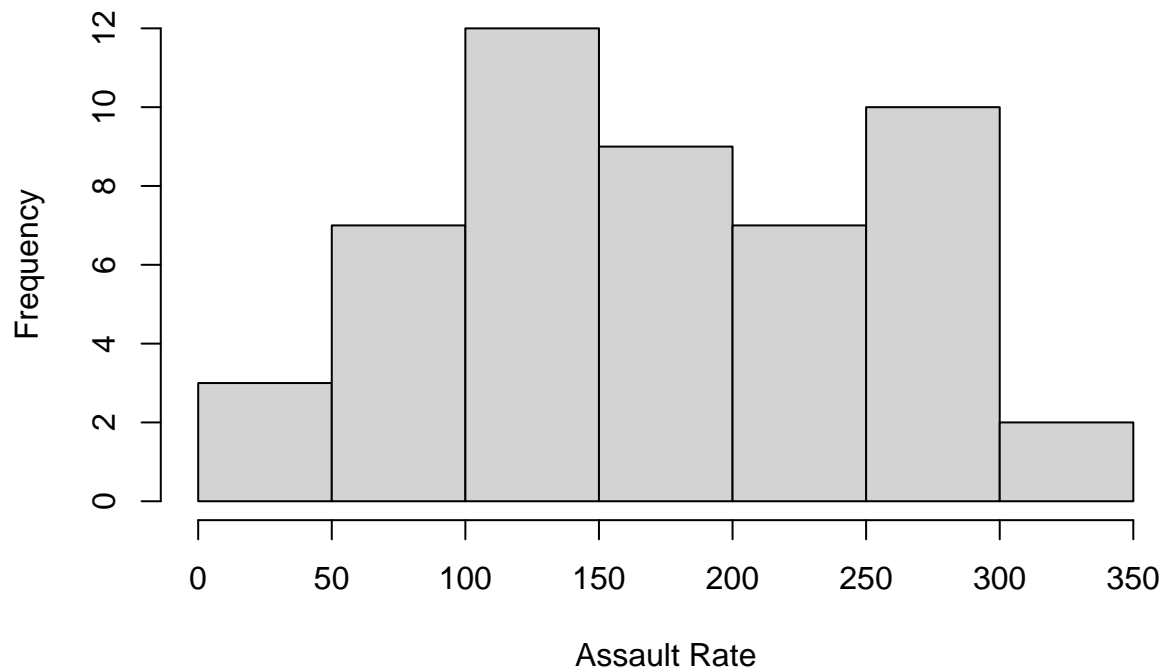
Answer: The mean of Murder is 7.788, the median of Murder is 7.250. Mean represents the average number of murders between the 50 states in 1973, whereas median is the middle number of murders if the quantities were lined up in numerical order. Quartiles are 4 approximately evenly sized groups with the data ordered from least to greatest. R most likely gives us Q1 and Q3 because the values between these two represent the middle 50% of the data, and provide more context for the spread of the mean.

Problem 7

Repeat the same steps you followed for **Murder**, for the variables **Assault** and **Rape**. Now plot all three histograms together. You can do this by using the command `par(mfrow=c(3,1))` and then plotting each of the three.

```
#histogram of Assault
hist(dat$Assault, main="Histogram of Assault", xlab="Assault Rate", ylab="Frequency")
```

Histogram of Assault

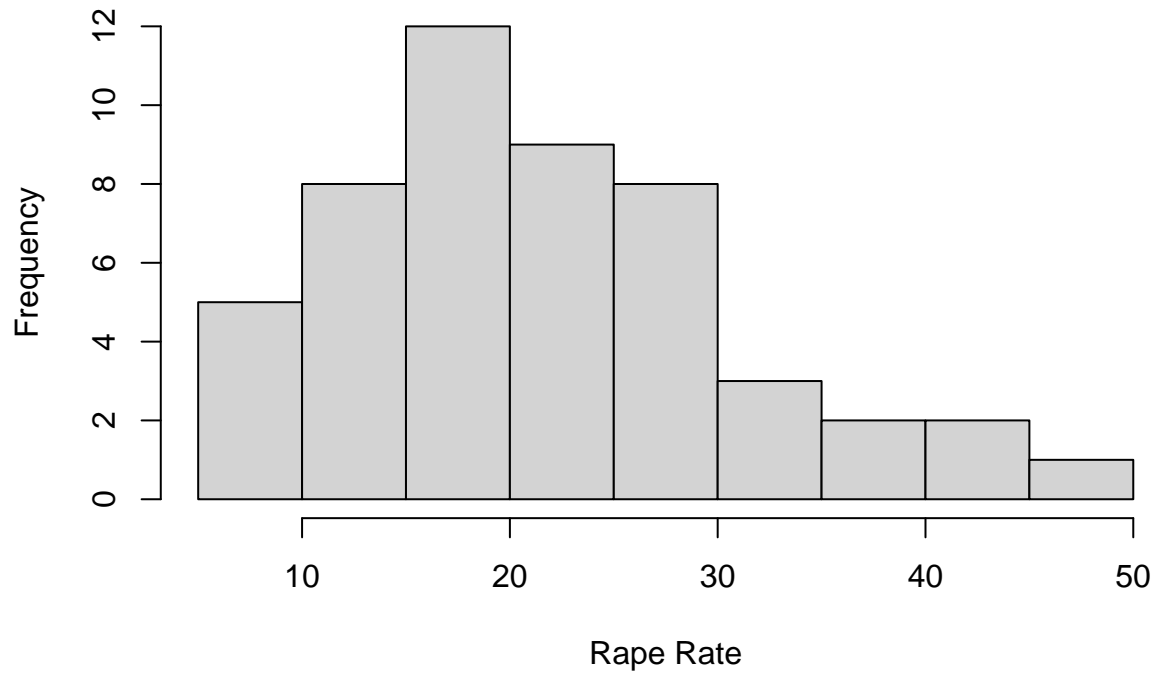


```
#Summary of Assault  
summary(dat$Assault)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
##      45.0   109.0   159.0   170.8   249.0   337.0
```

```
#Histogram of Rape  
hist(dat$Rape, main="Histogram of Rape", xlab="Rape Rate", ylab="Frequency")
```

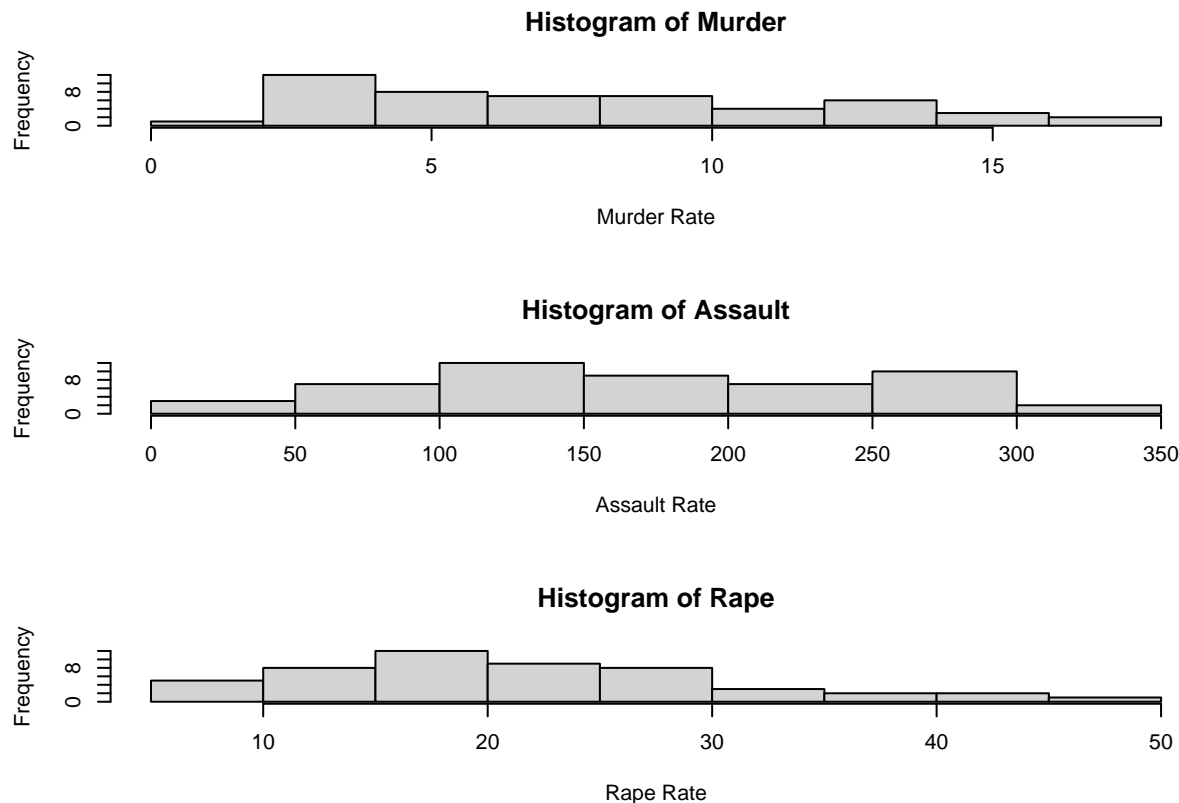
Histogram of Rape



```
#Summary of Rape  
summary(dat$Rape)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
##      7.30  15.07   20.10   21.23  26.18   46.00
```

```
#Histogram of Murder, Assault, Rape  
par(mfrow=c(3,1))  
hist(dat$Murder, main="Histogram of Murder", xlab="Murder Rate", ylab="Frequency")  
hist(dat$Assault, main="Histogram of Assault", xlab="Assault Rate", ylab="Frequency")  
hist(dat$Rape, main="Histogram of Rape", xlab="Rape Rate", ylab="Frequency")
```



What does the command `par` do, in your own words (you can look this up by asking R `?par`)?

Answer: the command `par` is used to set certain parameters within the data given to it. In this case, we used the `par` function to tell R that we wanted to plot 3 character vectors on histograms together in one plot.

What can you learn from plotting the histograms together?

Answer: Plotting the 3 histograms together allows us to easily visually compare the frequency of murders, assaults, and rapes to see which crime was most common among the 50 states in the year 1973.

Problem 8

In the console below (not in text), type `install.packages("maps")` and press Enter, and then type `install.packages("ggplot2")` and press Enter. This will install the packages so you can load the libraries.

Run this code:

```
library('maps')
library('ggplot2')

ggplot(dat, aes(map_id=state, fill=Murder)) +
  geom_map(map=map_data("state")) +
  expand_limits(x=map_data("state")$long, y=map_data("state")$lat)
```

What does this code do? Explain what each line is doing.

Answer: These lines of code are generating a dataframe that shows the map of the US, breaking up the information from the variable `Murder` into sections based on state. The lines also color-code the map based on the number of murders that occurred in that state.

Assignment 2

(Coming soon)