Assignments

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: Carmen Avery, Rachel Villari, and Halle Wasser.

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.

```
datasets::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
```

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

```
## Warning: package 'datasets' is in use and will not be installed
```

```
library("datasets")
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 # renames USArrests as dat
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
## 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
```

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?

Answer: It's useful to rename the dataset so that we know specifically which dataset we are working with. Additionally, when we edit the dataset it is helpful to have a differentiation (in name) from the original dataset. Additionally "dat" is shorter than the original dataset name.

Problem 2

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

```
dat$state <- tolower(rownames(USArrests))
names(dat)
# This line creates a new variable: states
```

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`. Answer: The variables are murder, assault, urbanpop, and rape.

Problem 3

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

Answer: According to the DVB chapter, murder is a quantitative variable.

What R Type of variable is it?

Answer:In R, Murder is a character variable.

Problem 4

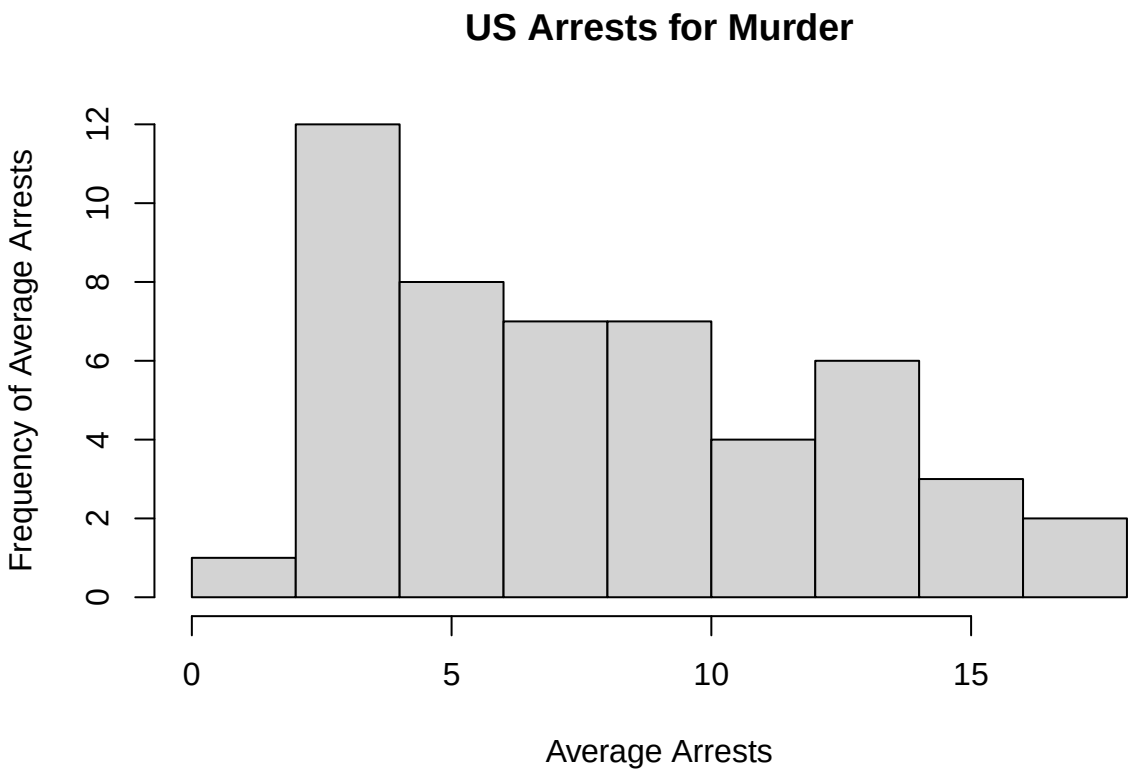
What information is contained in this dataset, in general? What do the numbers mean?

Answer: This dataset lists the average rate of arrests in each state for a specific type of crime. For example, in Iowa there has been an average of 2.2 arrests for murder. The fact that some of the data are listed as decimal numbers indicates that the dataset is reporting averages, as there cannot actually be 2.2 real arrests.

Problem 5

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

```
hist(dat$Murder, main="US Arrests for Murder", xlab="Average Arrests", ylab="Frequency of Average Arres
```



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.?

Answer: Murder's mean is 7.788. Its median is 7.25. The mean is the average of all values in the dataset, whereas the median is simply the middle value of those listed from smallest to largest. The median is also synonymous with the second quartile, Q2. A quartile is when the values of a dataset are divided into four groups. R gives the first and third quartiles because these are common values used in statistics, so it is helpful for R to calculate them automatically.

```
a<-c(13.2, 10.0, 8.1, 8.8, 9.0, 7.9, 3.3, 5.9, 15.4, 17.4, 5.3, 2.6, 10.4, 7.2, 2.2, 6.0, 9.7, 15.4, 2.
mean(a) # This code calculates the mean of a.
```

```
## [1] 7.788
```

```
median(a) # This code calculates the median of a.
```

```
## [1] 7.25
```

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.

```r
b<-c(236, 263, 294, 190, 276, 204, 110, 238, 335, 211, 46, 120, 249, 113, 56, 115, 109, 249, 83, 300, 14
mean(b) # calculates the mean of Assault
```

```
## [1] 170.76
```

```r
median(b) # calculates the median of Assault
```

```
## [1] 159
```

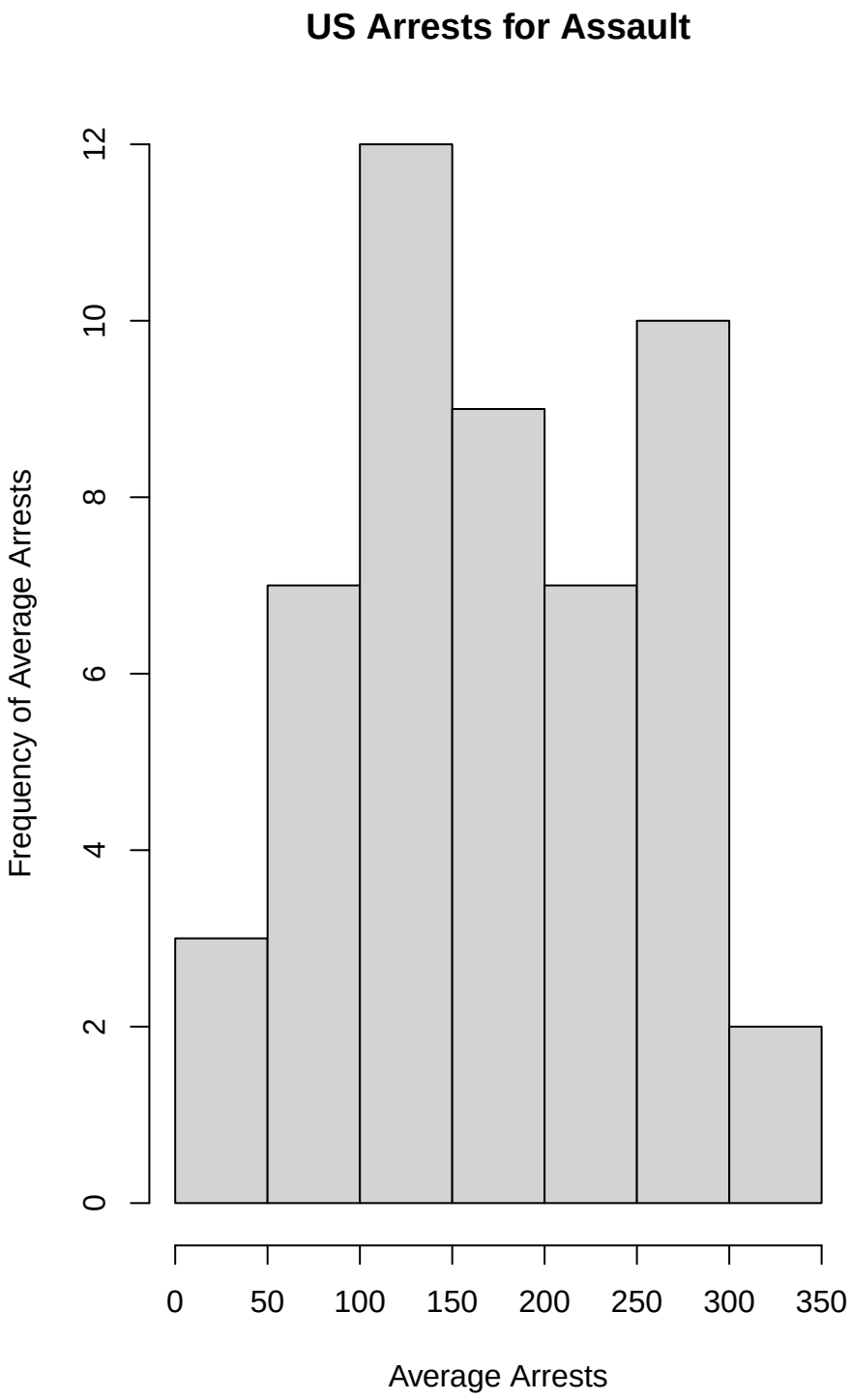
```r
b<-c(236, 263, 294, 190, 276, 204, 110, 238, 335, 211, 46, 120, 249, 113, 56, 115, 109, 249, 83, 300, 14
mean(b) # calculates the mean of Rape
```

```
## [1] 170.76
```

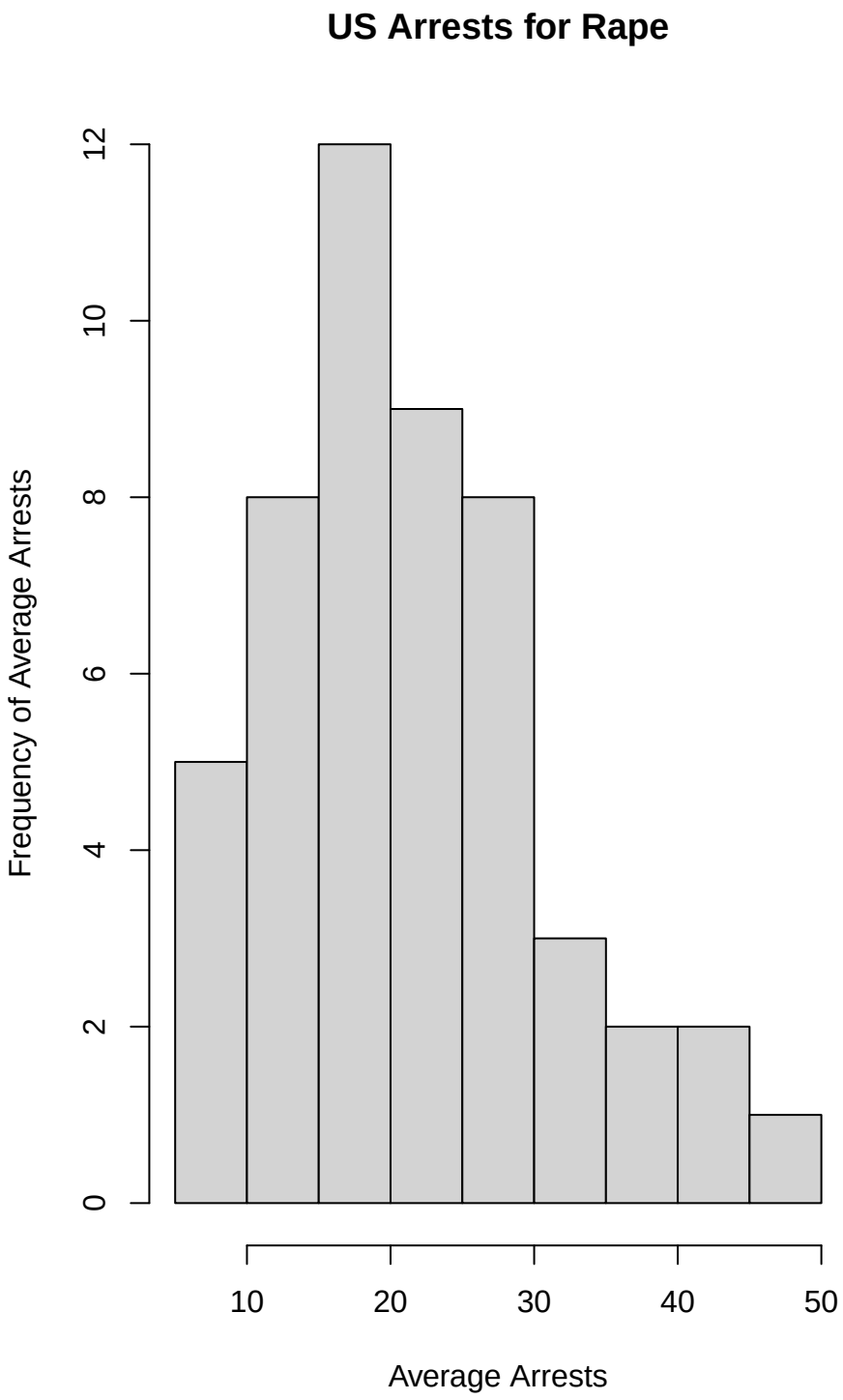
```r
median(b) # calculates the median of Rape
```

```
## [1] 159
```

```r
hist(dat$Assault, main="US Arrests for Assault", xlab="Average Arrests", ylab="Frequency of Average Arre
```



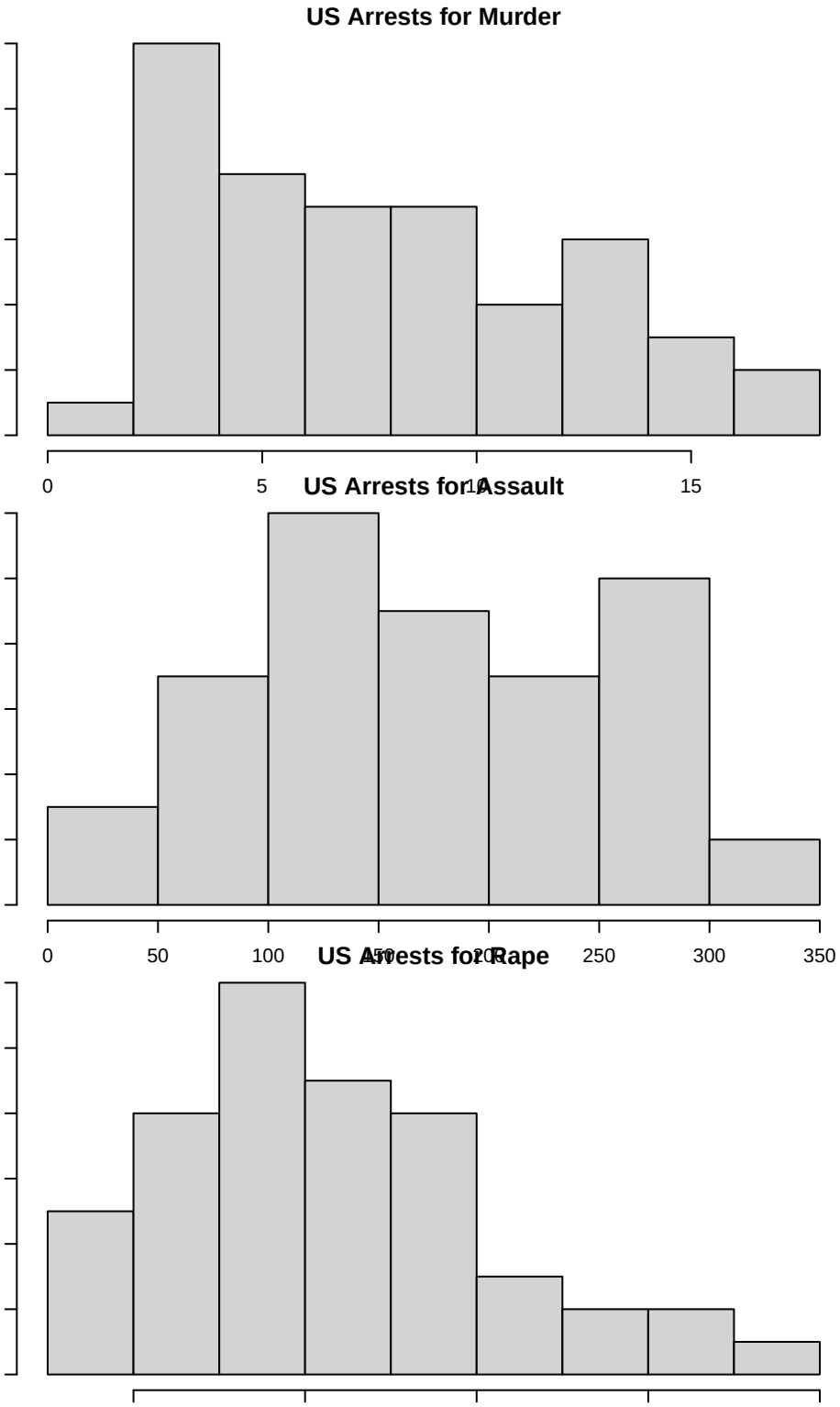

```
hist(dat$Rape, main="US Arrests for Rape", xlab="Average Arrests", ylab="Frequency of Average Arrests")
```




```
par("mar")
```

```
## [1] 5.1 4.1 4.1 2.1
```

```r
par(mar=c(1,1,1,1)) # command to fix the parameters to fix the fact that the figure margins were too la
par(mfrow=c(3,1)) # command to plot multiple histograms
hist(dat$Murder, main="US Arrests for Murder", xlab="Average Arrests", ylab="Frequency of Average Arres
hist(dat$Assault, main="US Arrests for Assault", xlab="Average Arrests", ylab="Frequency of Average Arr
hist(dat$Rape, main="US Arrests for Rape", xlab="Average Arrests", ylab="Frequency of Average Arrests")
```



Answer: The mean of Assault is 170.76 and the median is 159. The mean of Rape is 21.232 and the median is 20.1.

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

Answer: The command par allows you to set parameters when graphing, or to simply look up certain graphical parameters.

What can you learn from plotting the histograms together?

Answer:By plotting the histograms together, it becomes easier to see how the rates of arrests compare for the type of crimes committed. For example, it is evident from looking at the histograms that the rate of arrests for assault is much higher than that of murder. It is a helpful way to learn how the difrferent variables interact.

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: This code creates a map of the US displaying the arrest rates for murder by state. Lines one and two access the maps and ggplot functions via the library. Line three commands that the dataset for Murder be used in the plot, such that each state be colored by its corresponding arrest rate for murder. Line four finds the geographical information for the map, making each state distinct. The last line sets the limits of the axes and makes the x-axis longitude and the y-axis latitude.

Assignment 2

(Coming soon)