

Assignments

This page will contain all the assignments I 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: Jair Jean-Gilles

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.

```
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?

```
dat<-USArrests #This line is renaming the data.
```

Answer: It is useful to rename the dataset not only for ease of use, but also so that there is a new version to work off of that does not alter the original. ### 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`.

```
names (dat)
```

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

```
#The variables in USArrests are Murder, Assault, UrbanPop, and Rape (and now state).
```

Problem 3

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

Answer: Murder is a quantitative variable.

What R Type of variable is it?

Answer: Murder is an R type numeric variable.

Problem 4

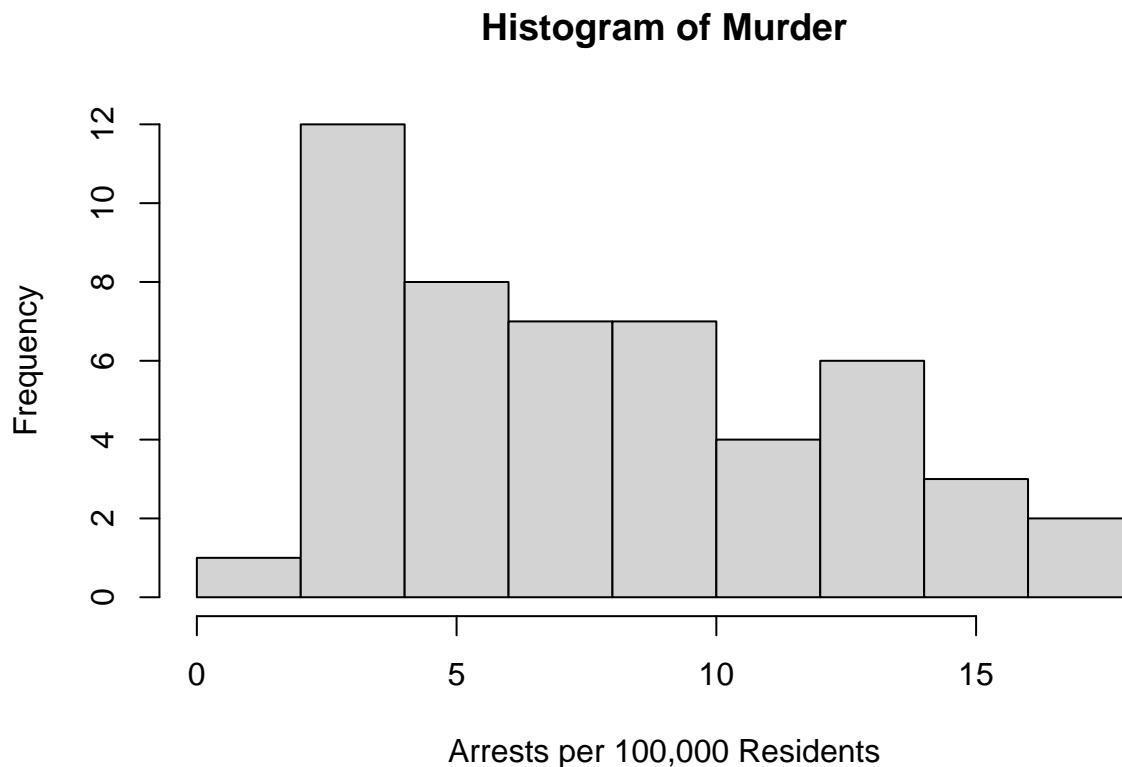
What information is contained in this dataset, in general? What do the numbers mean?

Answer: The information in the dataset is arrests per 100,000 residents for assault, murder, and rape in all 50 US states in 1973. Additionally, there is the percent of the population living in urban areas.

Problem 5

Draw a histogram of Murder with proper labels and title.

```
hist(dat$Murder, main="Histogram of Murder", xlab="Arrests per 100,000 Residents", ylab="Frequency")
```



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.?

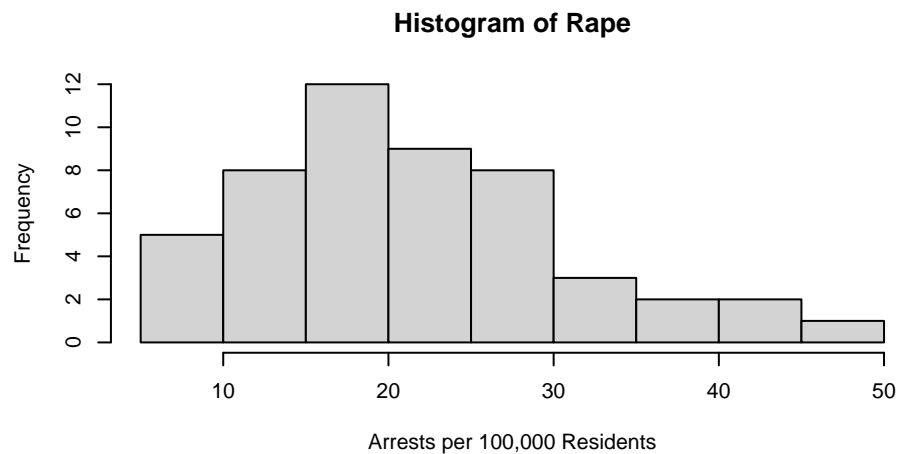
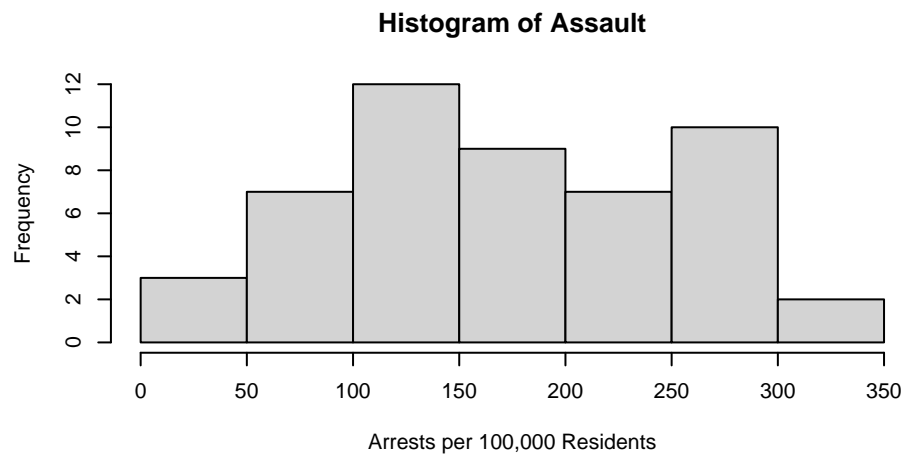
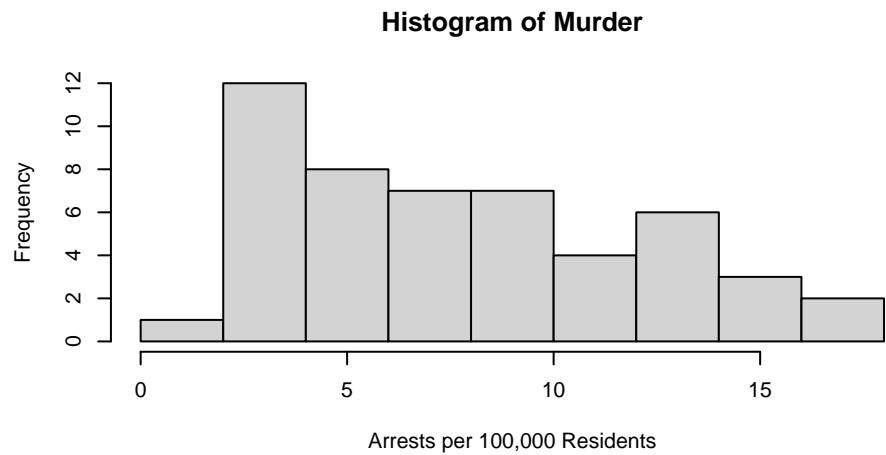
```
summary(dat$Murder)
```

```
##      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 7.778, the median 7.250, and the difference between them is 0.538. A quartile divides the data at 3 points, creating four groups of the dataset. R gives us the 1st and 3rd quartile because the 2nd quartile is simply the median. ### 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.

```
par(mfrow=c(3,1))
hist(dat$Murder, main="Histogram of Murder", xlab="Arrests per 100,000 Residents", ylab="Frequency")
hist(dat$Assault, main="Histogram of Assault", xlab="Arrests per 100,000 Residents", ylab="Frequency")
hist(dat$Rape, main="Histogram of Rape", xlab="Arrests per 100,000 Residents", ylab="Frequency")
```



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

Answer: `Par` can be used to set graphical parameters so that they appear in a specific format. For the previous problem, by using the `mfrow` we were able to set multiple rows of histograms, creating 3 in one image according to the next three functions created.

What can you learn from plotting the histograms together?

Answer: By plotting these three together, we can first learn more about the vastly different scales that the arrests occur, which may not have been clear with them on individual plots. Along with this, we can learn about how the frequency patterns differ, with murder and rape peaking at lower numbers, while assault sees a more spread distribution.

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: The first two lines are loading the maps and ggplot2 libraries so that we can use them. The ggplot line is using the original crime data set to make the `map_id` equal to the state variable and use Murder to fill in the subsequent map. The `geom_map` line is telling the program to pull the state map for use from the map library. Finally, the last line is expanding the limits to make the x and y equal to latitude and longitude.

Assignment 2

(Coming soon)