

Assignment 1-2

Steven Spiegel

2023-08-20

Module 1 Assignment 1-2

Assignment 1-2 details

- The goal of this assignment is to convert the code/answers from **Assignment 1-1** into **RMarkdown**.
 - The answerable questions from 1-1 were 3, 4, 5, and 6. Each question will be demarcated by a new heading.

Assignment 1-1, question 1

- Question: *Finish walking through this lab. (The actual lab starts on page 2 (right after question #7.)*
- Response: (There is no response for this question)

Assignment 1-1, question 2

- Question: Use the + button to create a new R script. Save it with a filename that looks like “lab0_yournamehere.R”. Put all the commands you use to answer the next questions into the new script file. Answers to questions that aren’t code can be set off by #, just as done in this file.
- Response: (There is no response for this question)

Assignment 1-1, question 3

- Question: Use the “summary” command to get a basic summary of each of the variables in the TitanicSurvival data set. Note that the type of summary you get depends on the data type of each variable. You need to install Titanic package first as below.
- Response:

```
library(carData)
```

```
summary(TitanicSurvival)
```

```
##   survived      sex       age   passengerClass
##   no :809  female:466   Min.   : 0.1667    1st:323
##   yes:500   male  :843   1st Qu.:21.0000   2nd:277
##                               Median :28.0000   3rd:709
##                               Mean   :29.8811
##                               3rd Qu.:39.0000
##                               Max.   :80.0000
##                               NA's   :263
```

Assignment 1-1, question 4

- Question: Use the ? operator to read about the “table” command. Then create a table that shows the number of people who survived and didn’t survive in each passenger class. What pattern do you see in the data?

- Response:

```
### First determine if there are any NAs in each of the pertinent columns
sum(is.na(TitanicSurvival$survived))

## [1] 0

sum(is.na(TitanicSurvival$passengerClass))

## [1] 0

#### No NAs in these columns
surv_table <- table(TitanicSurvival$passengerClass, TitanicSurvival$survived)
surv_table

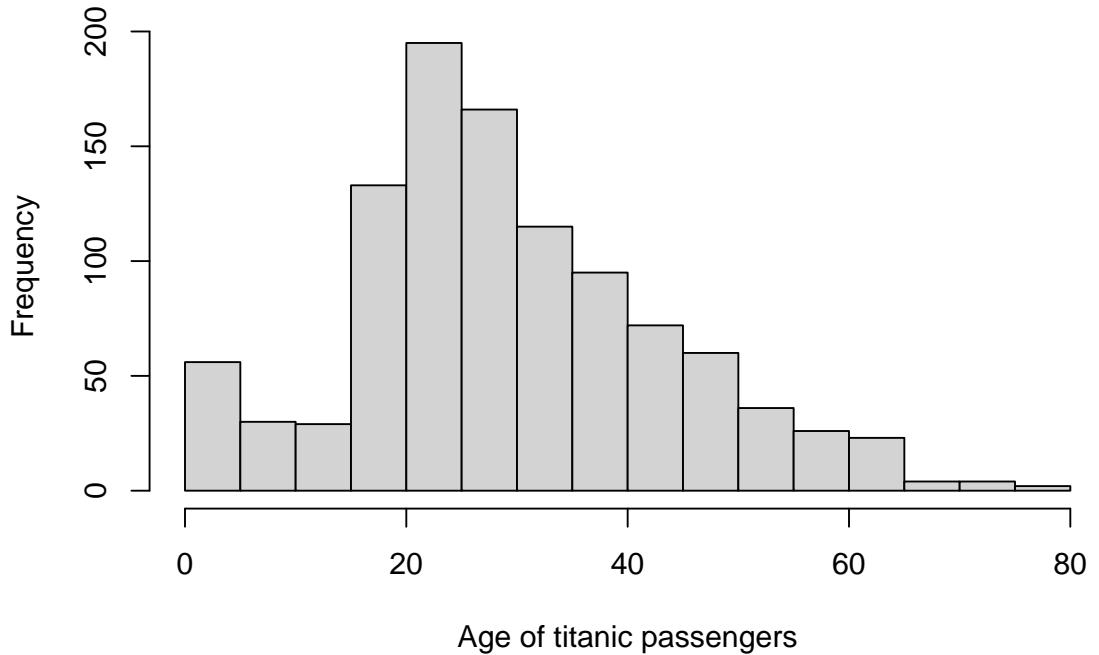
##
##          no yes
## 1st    123 200
## 2nd    158 119
## 3rd    528 181
```

The number of people in 3rd class have many more people who did not survive (25% survival rate) compared to first and second (62% and 43%, respectively)

Assignment 1-1, question 5

- Question: Use the “hist” command to make a histogram of ages. Do you notice anything interesting in the shape of the histogram?

Histogram of Passenger Ages



- Response:

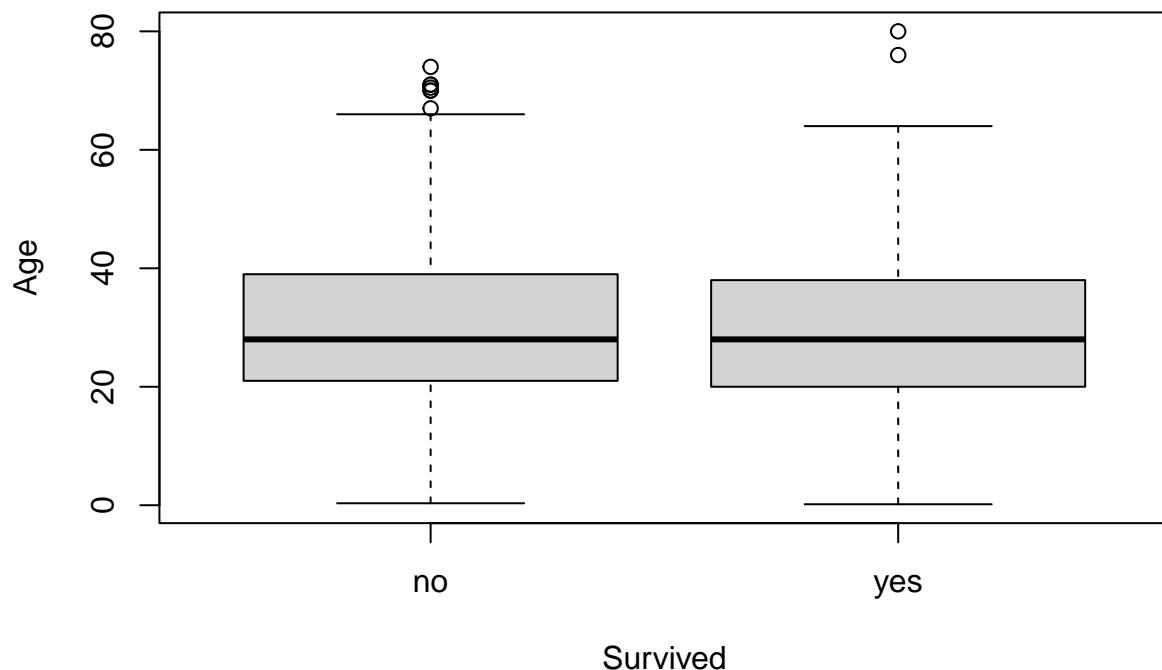
The histogram appears to be skewed right

Assignment 1-1, question 6

- Question: Use the ? operator to learn about the “boxplot” command. Hint: Often the examples at the end of the help file are especially helpful. Make side-by-side boxplots that show the distribution of ages for survivors and for those who didn’t survive. Does it appear that there is a difference in the age distribution of the two groups?
- Response:

```
boxplot(TitanicSurvival$age ~ TitanicSurvival$survived, xlab = "Survived", ylab = "Age", main= "Age Dis-
```

Age Distribution of passengers by survival



There seems to be very little difference in the age distribution. The Age distribution of those who didn't survive has more outlier values.

Assignment 1-1, question 7

- Question: Save your script file with only the commands you used and the answers to the questions that were asked (question 3,4,5, and 6). Copy-and-paste that RScript onto the word file and submit it as your assignment 1 on Blackboard. Submit the output for question 3,4,5, and 6 as well.
- Response: (There is no response for this question)