Lab 2 - R Markdown and Basic Objects

Code

AUTHOR
Maxim Dokukin

In this lab, you will familiarize yourself with R Markdown, file management, and basic objects and functions. **Follow the instructions below and use R Markdown to create a pdf document with your code and answers to the following questions on Gradescope.** You may find a template file by clicking “Code” in the top right corner of this page.

Your final submission should clearly include all code needed to generate your answers and should be formatted according to the guidelines outlined in class. In particular, make sure:

1. Code and output are clearly organized by question.
2. Unnecessary messages, warning, and output are removed.

You may collaborate with your classmates and consult external resources, but you should write and submit your own answer. **Any classmates with whom you collaborate should be credited at the top of your submission. Similarly, if you consult any external references, you should cite them clearly and explicitly.**

A. File paths and Introduction to R Markdown

1. Create a `MATH167R` directory somewhere on your computer that will contain all your work for this class. Within the `MATH167R` folder, create subdirectories for `labs`, `checkins`, and `data`. Save this .Rmd document in your `data` folder and then Render/Knit your Rmd document. The following code chunk should print the location of your Rmd. document:

```r
getwd()
```

```
[1] "/Users/xewe/Documents/Education/BS SJSU/Classes/MATH 167R
- R Programming/RStudio/Data"
```

2. In the console, run the command `getwd()`. What is the output? If it is different from your answer to Question 1, why? Explain, in your own words, what the working directory is.

Different. because default working directory is /Users/xewe/Documents/Education/BS SJSU/Classes/MATH 167R - R Programming/RStudio

Working dir- is the current directory that terminal accesses

3. Every time you Render/Knit your Rmd document, **all** of the code is run in a **new** R session. Any code run in the interactive session **does not affect** your rendered html file. Create a code chunk below that sets the value of `x` to be your favorite number and then uses `print(x)` to print it out.

```
x <- -1
print(x)
```

```
[1] -1
```

4. Download [this dataset]() to your `data` folder, then use a code chunk with the code `cars <- read.csv("YOUR PATH HERE")` to open the file. You will need to replace the text `YOUR PATH HERE` with the correct location of the dataset to open this successfully.

```
cars <- read.csv("cars.csv")
```

5. Run the code `head(cars)` in a code chunk to display the first few rows of this dataset.

```
head(cars)
```

```
  speed dist
1     4    2
2     4   10
3     7    4
4     7   22
5     8   16
6     9   10
```

B. Vectors and Matrices

6. Write a line of code to generate a vector of length 100 whose entries are 1, 2, ..., 100.

```
v <- 1:100
length(v)
```

```
[1] 100
```

```
print(v)
```

```
  [1]   1   2   3   4   5   6   7   8   9  10  11  12  13  14
15  16  17  18
 [19]  19  20  21  22  23  24  25  26  27  28  29  30  31  32
33  34  35  36
 [37]  37  38  39  40  41  42  43  44  45  46  47  48  49  50
51  52  53  54
 [55]  55  56  57  58  59  60  61  62  63  64  65  66  67  68
69  70  71  72
 [73]  73  74  75  76  77  78  79  80  81  82  83  84  85  86
87  88  89  90
 [91]  91  92  93  94  95  96  97  98  99 100
```

7. Write a line of code to generate the following matrix:

`1 2 3`

`4 5 6`

`7 8 9`

```
mat <- matrix(1:9, nrow = 3, ncol = 3, byrow = TRUE)
mat
```

```
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    4    5    6
[3,]    7    8    9
```

C. Working with Data Frames

Use the following code to load the `penguins` data.

```
# load palmer penguins package
library(palmerpenguins)

# open penguins data as a data frame
data(penguins)
penguins <- as.data.frame(penguins)
```

8. Using the `mean()` function, compute the mean body mass of penguins in the dataset, dropping any missing values.

```
mean(penguins$body_mass_g, na.rm=T)
```

```
[1] 4201.754
```

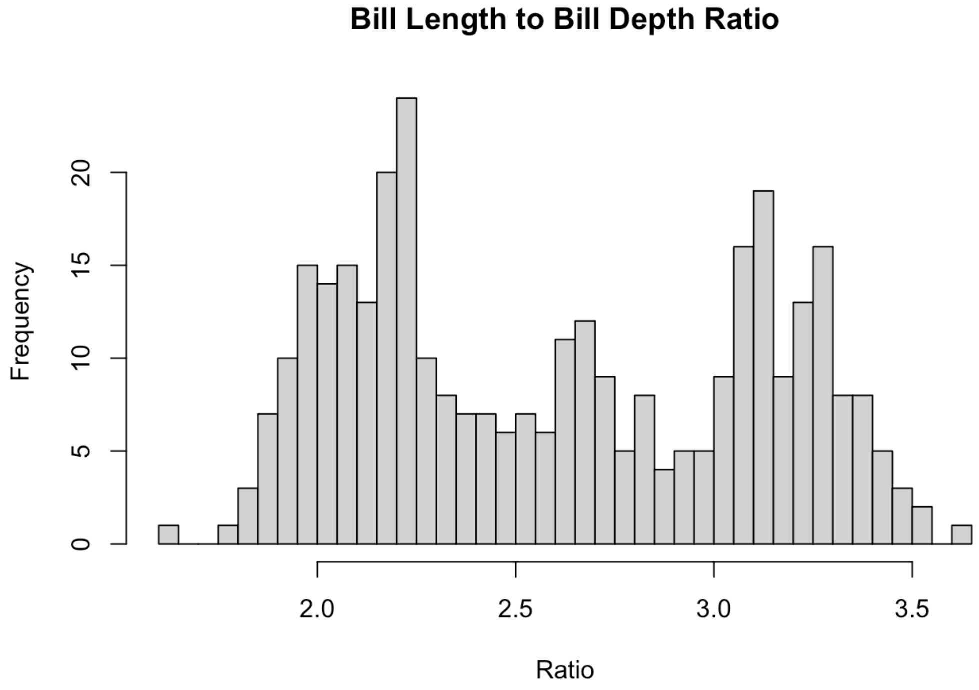
9. Using the `max` function, compute the maximum flipper length of penguins in the dataset, dropping any missing values.

```
mean(penguins$flipper_length_mm, na.rm=T)
```

```
[1] 200.9152
```

10. Using the `hist` function, create a histogram of the ratio of the penguins' bill length to bill depth, dropping any missing values. What is the shape of the reuslting distribution?

```
penguins$bill_ratio <- penguins$bill_length_mm / penguins$bill_
penguins_hist_data <- na.omit(penguins$bill_ratio)
hist(penguins_hist_data, main = "Bill Length to Bill Depth Rati
     xlab = "Ratio", ylab = "Frequency", breaks = 50)
```



Trimodal distribution? Breaking data down by species would likely produce 3 distinct bell shapes for each species.