Stat 120

Deepak Bastola

2023-01-02

Contents

About

This is a *sample* book written in **Markdown**.

Answer

Chapter 1

Class Activity 1

1.1 Your Turn 1

---

a. Run the following chunk. Comment on the output.

```r
example_data = data.frame(ID = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10),
                          Greeting = c(rep("Hello", 5), rep("Goodbye",5)),
                          Male = rep(c(TRUE, FALSE), 5),
                          Age = runif(n=10, 20, 60))
```

Click for answer

```r
example_data
```

```
   ID Greeting  Male      Age
1   1    Hello  TRUE 32.62417
2   2    Hello FALSE 47.20118
3   3    Hello  TRUE 45.16701
4   4    Hello FALSE 26.66707
5   5    Hello  TRUE 54.56103
6   6  Goodbye FALSE 58.43057
7   7  Goodbye  TRUE 23.98081
8   8  Goodbye FALSE 47.57207
9   9  Goodbye  TRUE 33.91807
10 10  Goodbye FALSE 37.23701
```

Answer: We see a data frame with four columns, where the first column is an `identifier` for the cases. We have information on the greeting types, gender, and age on these cases in the remaining columns.

   b. What is the dimension of the dataset called 'example_data'?

Click for answer

```
dim(example_data)
[1] 10  4
nrow(example_data)
[1] 10
ncol(example_data)
[1] 4
```

Answer: There are 10 rows and 4 columns.

1.2   Your Turn 2

   a. Read the dataset `EducationLiteracy` from the Lock5 second edition book.

Click for answer

```
# read in the data
education_lock5 <- read.csv("https://www.lock5stat.com/datasets2e/EducationLiteracy.cs
```

   b. Print the header (i.e. first 6 cases by default) of the dataset in part a.

Click for answer

```
head(education_lock5)
```

```
            Country EducationExpenditure Literacy
1        Afghanistan                 3.1     31.7
2            Albania                 3.2     96.8
3            Algeria                 4.3       NA
4            Andorra                 3.2       NA
5             Angola                 3.5     70.6
6 Antigua and Barbuda                 2.6     99.0
```

    c. What is the dimension of the dataset in a?

Click for answer

```
dim(education_lock5)
```

```
[1] 188   3
```

Answer: There are 188 rows and 3 columns.

    d. What type of variables are `Country`, `EducationExpenditure`, and `Literacy`?

Click for answer

Answer: `Country` is a categorical variable. `EducationExpenditure` and `Literacy` are both quantitative variables.

    e. If we would like to use education expenditure to predict the literacy rate of each countries, which variable is the explanatory variable and which one is the response?

Click for answer

Answer: The education expenditure is the explanatory variable, and the literacy rate is the response.

Chapter 2

(PART*) Basics R

Chapter 3

What is R?

R is a free and open source statistical programming language that facilitates statistical computation. There are a myriad of application that can be done in R, thanks to a huge online support community and dedicated packages. However, R has no graphical user interface and it has to be run by typing commands into a text interface.

3.1 What is RStudio?

RStudio provides graphical interface to R! You can think of RStudio as a graphical front-end to R that that provides extra functionality. The use of the R programming language with the RStudio interface is an essential component of this course.

3.2 R Studio Server

The quickest way to get started is to go to https://maize.mathcs.carleton.edu, which opens an R Studio window in your web browser. Once logged in, I recommend that you do the following:

- Step 1: Create a folder for this course where you can save all of your work. In the Files window, click on New Folder.
- Step 2: Click on Tools -> Global Options -> R Markdown. Then uncheck the box that says "Show output inline…"

(It is also possible to download RStudio on your own laptop. Instructions may be found at the end of this document.)

3.3   R Markdown Basics

An R Markdown file (.Rmd file) combines R commands and written analyses, which are 'knit' together into an HTML, PDF, or Microsoft Word document.

An R Markdown file contains three essential elements:

- Header: The header (top) of the file contains information like the document title, author, date and your preferred output format (pdf_document, word_document, or html_document).

- Written analysis: You write up your analysis after the header and embed R code where needed. The online help below shows ways to add formatting details like bold words, lists, section labels, etc to your final pdf/word/html document. For example, adding ** before and after a word will bold that word in your compiled document.

- R chunks: R chunks contain the R commands that you want evaluated. You embed these chunks within your written analysis and they are evaluated when you compile the document.

3.3.1   R Markdown example:

- Simple R Markdown example
    - compiled pdf

The following handouts, written by Prof Katie St Clair, contain useful information for making the figured and tables in your compiled documents look nice:

- Graph Formatting: Markdown .Rmd file and pdf
- Table Formatting: Markdown .Rmd file and pdf

3.4   Installing R/RStudio (not needed if you are using the maize server)

- Download the latest version of R:
    - Windows: http://cran.r-project.org/bin/windows/base/
    - Mac: http://cran.r-project.org/bin/macosx/
- Download the free Rstudio desktop version (Windows or Mac): https://www.rstudio.com/products/rstudio/download/

Use the default download and install options for each.

3.5 Install LaTeX (for knitting R Markdown documents to PDF):

If you want to compile R Markdown to .pdf files, you also need a LaTeX distribution (Note: this is not necessary if you choose to compile as a Word document.) Click instructions for Windows or instructions for Mac, depending on your operating system to complete the installation.

3.6 Updating R/RStudio (not needed if you are using the maize server)

If you have used a local version of R/RStudio before and it is still installed on your machine, then you should make sure that you have the most recent versions of each program.

- To check your version of R, run the command `getRversion()` and compare your version to the newest version posted on https://cran.r-project.org/. If you need an update, then install the newer version using the installation directions above.

- In RStudio, check for updates with the menu option `Help > Check for updates`. Follow directions if an update is needed.

3.7 Instructions

If using Rstudio on your computer, using the **File>Open File** menu to find and open this .Rmd file.

If using Maize Rstudio from your browser:

- In the Files tab, select **Upload** and **Choose File** to find the .Rmd that you downloaded. Click *OK* to upload to your course folder/location in the maize server account.

- Click on the .Rmd file in the appropriate folder to open the file.

Extra notes:

- You can run a line of code by placing your cursor in the line of code and clicking **Run Selected Line(s)**

- You can run an entire chunk by clicking the green triangle on the right side of the code chunk.

- After each small edit or code addition, **Knit** your Markdown. If you wait until the end to Knit, it will be harder to find errors in your work.

- Format output type: You can use any of pdf_document, html_document type, or word_document type.

- **Maize users**: You may also need to allow for "pop-up" in your web browser when knitting documents.

3.8   Few Instructions

The default setting in Rstudio when you are running chunks is that the "output" (numbers, graphs) are shown **inline** within the Markdown Rmd. If you prefer to have your plots appear on the right of the console and not below the chunk, then change the settings as follows:

1. Select Tools > Global Options.
2. Click the R Markdown section and uncheck (if needed) the option Show output inline for all R Markdown documents.
3. Click OK.

Now try running R chunks in the .Rmd file to see the difference. You can recheck this box if you prefer the default setting.

Chapter 4

R Markdown

This is a R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

You can use asterisk mark to provide emphasis, such as `*italics*` or `**bold**`.

You can create lists with a dash:

```
- Item 1
- Item 2
- Item 3
  + Subitem 1
* Item 4
```

- Item 1
- Item 2
- Item 3
    - Subitem 1
- Item 4

You can embed Latex equations in-line, $\frac{1}{n}\sum_{i=1}^{n} x_i$ or in a new line as

$$\text{Var}(X) = \frac{1}{n-1}\sum_{i-1}^{n}(x_i - \bar{x})^2$$

Embed an R code chunk:

Use

```r
```r
Use back ticks to
create a block of code
```
```

to produce:

```
Use back ticks to
create a block of code
```

You can also evaluate and display the results of R code. Each tasks can be accomplished in a suitably labeled chunk like the following:

```r
summary(cars)
```

```
     speed           dist
 Min.   : 4.0   Min.   :  2.00
 1st Qu.:12.0   1st Qu.: 26.00
 Median :15.0   Median : 36.00
 Mean   :15.4   Mean   : 42.98
 3rd Qu.:19.0   3rd Qu.: 56.00
 Max.   :25.0   Max.   :120.00
```

```r
fit <- lm(dist ~ speed, data = cars)
fit
```

```
Call:
lm(formula = dist ~ speed, data = cars)

Coefficients:
(Intercept)        speed
    -17.579        3.932
```

4.1   Including Plots

You can also embed plots. See Figure 4.1 for example:

```r
par(mar = c(0, 1, 0, 1))
pie(
  c(280, 60, 20),
  c('Sky', 'Sunny side of pyramid', 'Shady side of pyramid'),
```

```
  col = c('#0292D8', '#F7EA39', '#C4B632'),
  init.angle = -50, border = NA
)
```
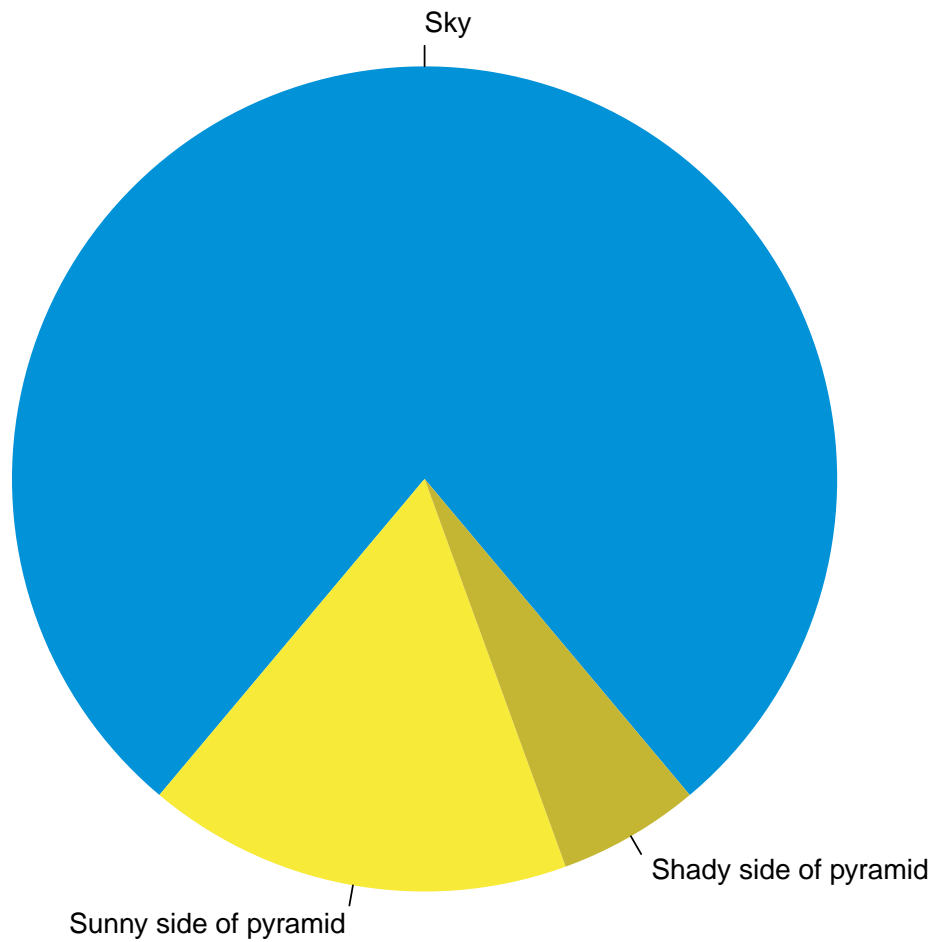

Figure 4.1: A fancy pie chart.

(Credit: Yihui Xie)

4.2   Read in data files

```
simple_data <- read.csv("https://deepbas.io/data/simple-1.dat", )
summary(simple_data)
```

```
   initials              state                    age
Length:3            Length:3            Min.    :45.0
Class :character    Class :character    1st Qu.:47.5
Mode  :character    Mode  :character    Median :50.0
                                        Mean    :52.0
                                        3rd Qu.:55.5
                                        Max.    :61.0

     time
Length:3
Class :character
Mode  :character
```

```
knitr::kable(simple_data)
```

initials	state	age	time
vib	MA	61	6:01
adc	TX	45	5:45
kme	CT	50	4:19

4.3   Hide the code

If we enter the `echo = FALSE` option in the R chunk (see the .Rmd file). This prevents the R code from being printed to your document; you just see the results.

initials	state	age	time
vib	MA	61	6:01
adc	TX	45	5:45
kme	CT	50	4:19