

R Stats Bootcamp

Megan Lewis

2024-11-07

How the R stats bootcamp works

- Practical, open instructional materials for learning R
- Traditional statistics in R
- Reproducible research and collaboration tools
- Self-guided learning and self-assessment
- Friendly community



R stats bootcamp - Module 1

Schedule:

- Session 1: An introduction and script workflow
- Session 2: R language
- Session 3: R functions
- Session 4: Data objects
- Session 5: Data frames
- Session 6: Data subsetting



Session 1 objectives:

- R motivation
- Installing R and R studio
- Rstudio components and set up
- Workflow for scripts in R
- Practice exercises

R motivation: why R?

- Objectively the best statistical software available
- Can perform sophisticated statistical analysis with minimum effort
- Commonly used
- Free and open source
- Works well on all computers and OSes, old and new

R motivation: why R?

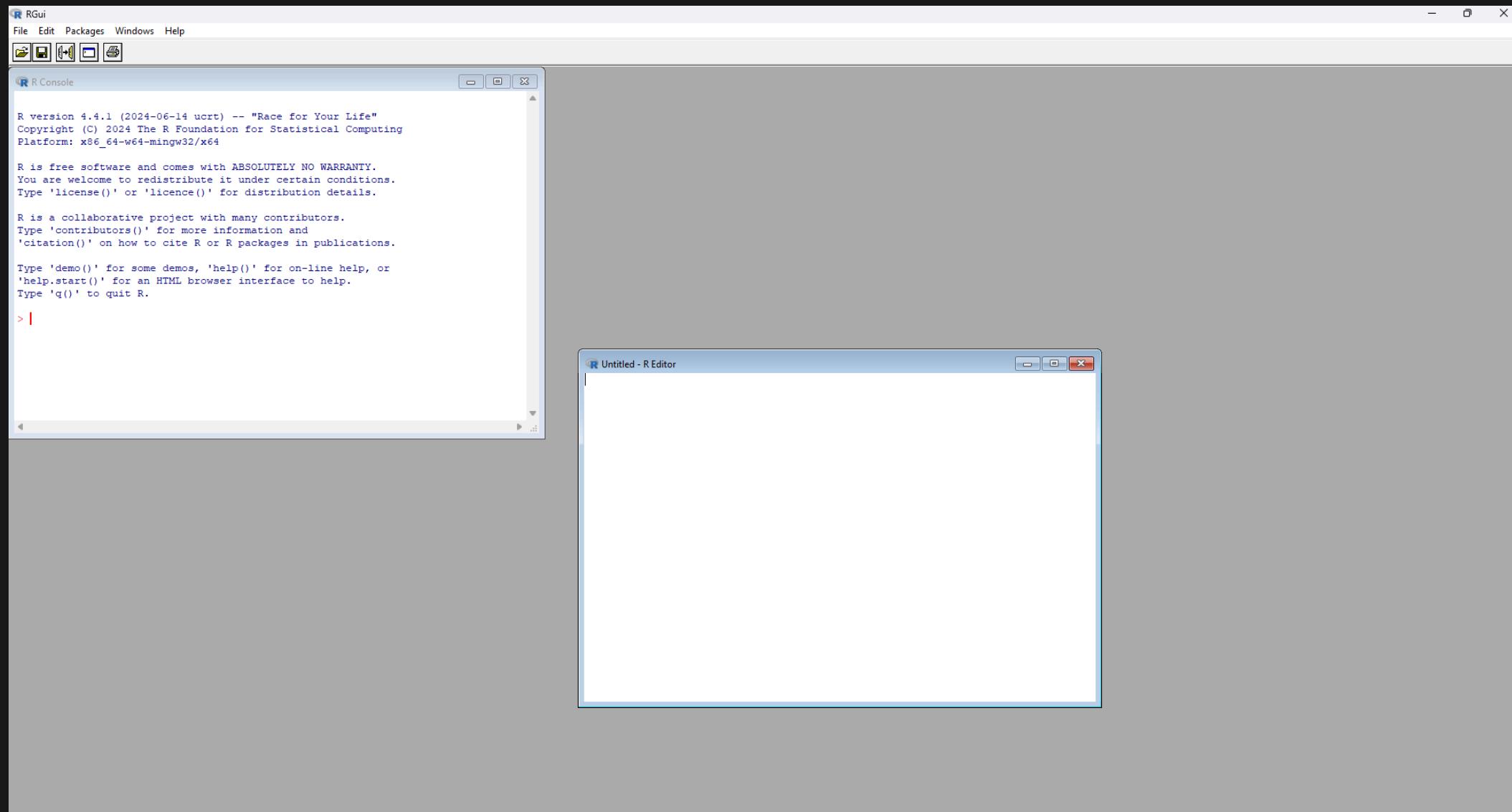
- Designed for people with no programming experience
- Very large community of users
- R is easy to learn (as you will see)



Using R

- Install R and R studio locally
- Posit cloud (Free account limited)
- Google colab

What does R look like?



What does R Studio look like?

The screenshot displays the R Studio interface with the following components:

- Code Editor:** Shows an R script named "script-1.R" containing code related to the "iris" dataset and a boxplot.
- Console:** Shows the R command-line interface with the same code being run, including the output of the boxplot command.
- Data View:** Shows the "iris" dataset in the Global Environment, indicating 150 observations and 5 variables.
- Plots:** A boxplot titled "SepalLength" versus "Species" (setosa, versicolor, virginica) is displayed, showing the distribution of Sepal Length for each Iris species.

What does R Studio look like?

The screenshot displays the RStudio interface with the following components:

- Script Editor:** Shows the file `script-1.R` containing R code for setting up a Bootcamp, loading the "iris" dataset, printing its head, and creating a boxplot of Sepal.Length by Species.
- Console:** Shows the R session history, including the execution of the R code from the script, resulting in the output of the iris dataset's first six rows and the generated boxplot.
- Environment:** Shows the "iris" dataset in the Global Environment, with 150 observations and 5 variables.
- Plots:** Displays a boxplot titled "SepalLength" with the x-axis labeled "Species" and the y-axis labeled "SepalLength". The plot shows three groups: setosa, versicolor, and virginica, with distinct box and whisker characteristics.

What does R Studio look like?

The screenshot displays the RStudio interface with the following components:

- Script Editor (left pane):** Shows the script file "script-1.R" containing R code for setting up the environment, loading the "iris" dataset, printing its head, and creating a boxplot.
- Console (bottom-left pane):** Displays the R command history and the resulting output, including the first six rows of the "iris" dataset and the generated boxplot command.
- Environment (top-right pane):** Shows the "iris" dataset in the Global Environment, with 150 observations and 5 variables.
- Plots (bottom-right pane):** A boxplot titled "Sepal.Length ~ Species" comparing the Sepal.Length for three iris species: setosa, versicolor, and virginica.

```
script-1.R
3 ## What: I Bootcamp Setup
4 ## Last edited: 2022-10-09
5 ####
6
7
8 ## CONTENTS ####
9 ## 00 Look at the "iris" data
10 ## 01 Make a simple graph
11
12
13 ## 00 Look at the "iris" data #####
14 # Here we are "telling R" we want to use a dataset called "iris"
15 data(iris)
16
17 # Print the "head" (first 6 lines) if iris data
18 head(iris)
19
20 # The code on the next line will display the help page for the iris data
21 ?iris
22
23
24 ## 01 Make a simple graph #####
25
26 # Make a boxplot showing the Sepal.Length for each iris Species
27 boxplot(formula = Sepal.Length ~ Species,
28         data = iris)
29 |
```

```
> ## 00 Look at the "iris" data #####
> # Here we are "telling R" we want to use a dataset called "iris"
> data(iris)
> # Print the "head" (first 6 lines) if iris data
> head(iris)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2   setosa
2          4.9         3.0          1.4         0.2   setosa
3          4.7         3.2          1.3         0.2   setosa
4          4.6         3.1          1.5         0.2   setosa
5          5.0         3.6          1.4         0.2   setosa
6          5.4         3.9          1.7         0.4   setosa
> # The code on the next line will display the help page for the iris data
> ?iris
> # Make a boxplot showing the Sepal.Length for each iris Species
> boxplot(formula = Sepal.Length ~ Species,
+         data = iris)
```

What does R Studio look like?

The screenshot displays the R Studio interface with the following components:

- Code Editor:** Shows the script-1.R file containing R code for setting up a bootcamp, loading the iris dataset, printing its head, and creating a boxplot of Sepal.Length by Species.
- Console:** Shows the R command-line interface (CLI) output corresponding to the script's execution.
- Environment:** Shows the Global Environment pane with the iris dataset loaded, containing 150 observations and 5 variables.
- Plots:** Shows a boxplot titled "SepalLength" versus "Species". The x-axis categories are setosa, versicolor, and virginica. The y-axis ranges from 4.5 to 8.0. The boxplot shows that virginica has the highest median Sepal.Length, followed by versicolor, and then setosa.

What does R Studio look like?

The screenshot displays the RStudio interface with the following components:

- Script Editor:** Shows the file `script-1.R` containing R code for setting up a bootcamp, loading the "iris" dataset, printing its head, and creating a boxplot of Sepal.Length by Species.
- Console:** Shows the R command history corresponding to the code in the script editor.
- Environment Browser:** Displays the "iris" dataset in the Global Environment, showing 150 observations and 5 variables.
- Plots:** A boxplot titled "SepalLength" versus "Species" (setosa, versicolor, virginica) is displayed in the bottom right corner.

What is an R script?

- An R script is a plain text file
- Ends in “dot R” (.R) by default
- Interface between your commands and R software
- Documents your analysis
- Organized and logical, has meaning for other people

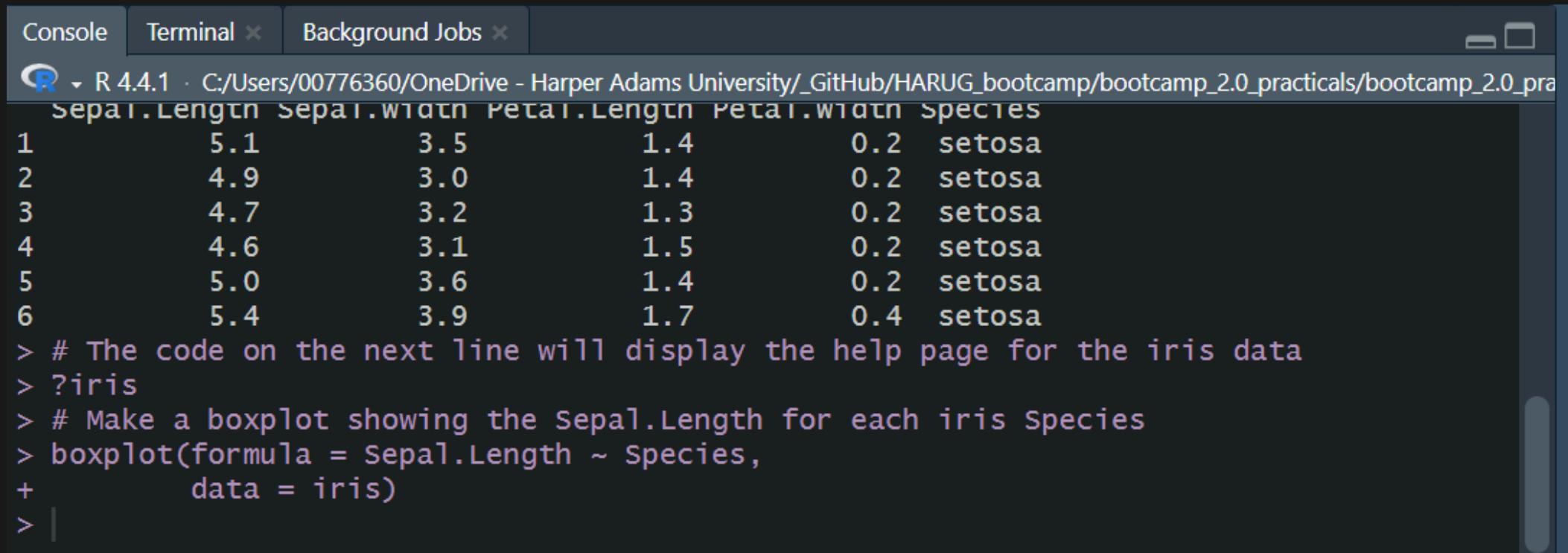
What is an R script?

The screenshot shows the RStudio interface with an R script file open. The title bar says "script-1.R". The main pane displays the following R code:

```
1 ## HEADER ####
2 ## Who: R Stats Bootcamp
3 ## What: 1 Bootcamp Setup
4 ## Last edited: 2022-10-09
5 #####
6
7
8 ## CONTENTS ####
9 ## 00 Look at the "iris" data
10 ## 01 Make a simple graph
11
12
13 ## 00 Look at the "iris" data ####
14 # Here we are "telling R" we want to use a dataset called "iris"
15 data(iris)
16
17 # Print the "head" (first 6 lines) if iris data
18
```

The right sidebar shows a tree view with nodes: HEADER, CONTENTS, 00 Look at the "iris" d..., and 01 Make a simple graph. The status bar at the bottom shows "16:1" and "# 00 Look at the \"iris\" data".

The R Console



The screenshot shows the RStudio interface with the 'Console' tab selected. The R console window displays the following content:

```
R 4.4.1 · C:/Users/00776360/OneDrive - Harper Adams University/_GitHub/HARUG_bootcamp/bootcamp_2.0_practicals/bootcamp_2.0_pr  
Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
1          5.1         3.5          1.4         0.2   setosa  
2          4.9         3.0          1.4         0.2   setosa  
3          4.7         3.2          1.3         0.2   setosa  
4          4.6         3.1          1.5         0.2   setosa  
5          5.0         3.6          1.4         0.2   setosa  
6          5.4         3.9          1.7         0.4   setosa  
> # The code on the next line will display the help page for the iris data  
> ?iris  
> # Make a boxplot showing the Sepal.Length for each iris Species  
> boxplot(formula = Sepal.Length ~ Species,  
+           data = iris)  
> |
```

The Global Environment

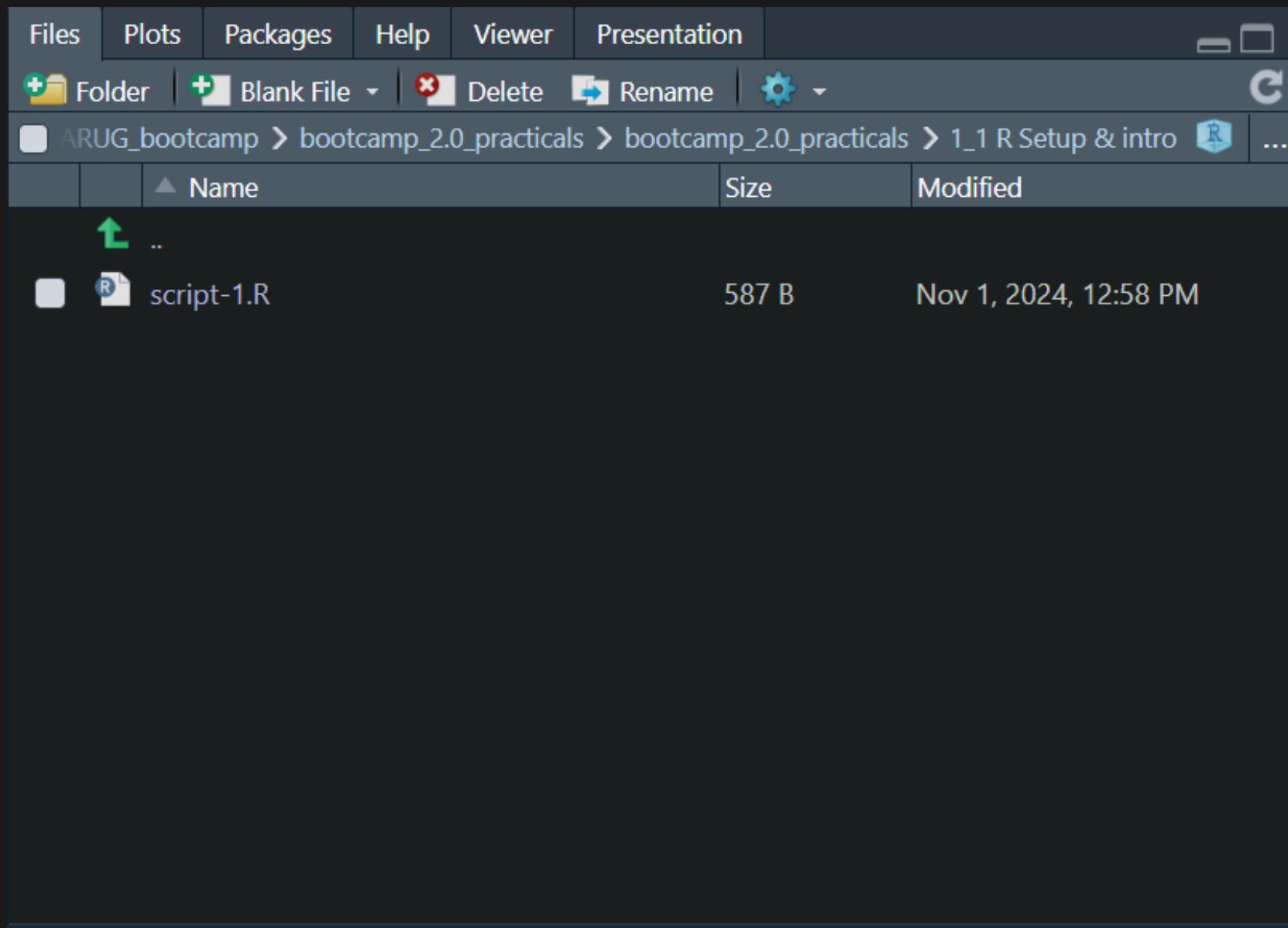
The screenshot shows the RStudio interface with the Global Environment tab selected in the top navigation bar. The main pane displays the 'iris' dataset, which contains 150 observations and 5 variables: Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, and Species. The Species variable is a factor with three levels: "setosa", "versicolor", and "virginica". The RStudio interface includes a toolbar with various icons for file operations like Import Dataset, Git, and Tutorial, and a search bar at the bottom.

		150 obs. of 5 variables	
\$ Sepal.Length:	num	5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...	
\$ Sepal.Width :	num	3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1...	
\$ Petal.Length:	num	1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1...	
\$ Petal.Width :	num	0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0...	
\$ Species	:	Factor w/ 3 levels "setosa","versicolor",...	

Files, plots and viewer pane



Files, plots and viewer pane



Workflow for R Scripts



Your R script should be good enough to show a friend (like a supervisor, or even your future self), a friend you respect.

Workflow for R Scripts



Your R script should be good enough to show a friend (like a supervisor, or even your future self), a friend you respect.

R Projects

- Automatically associate project and project files with working directory

Else:

```
1 # setwd("your_working_directory_here")
```

R script organisation

- Header
- Contents
- Sections

R script Header

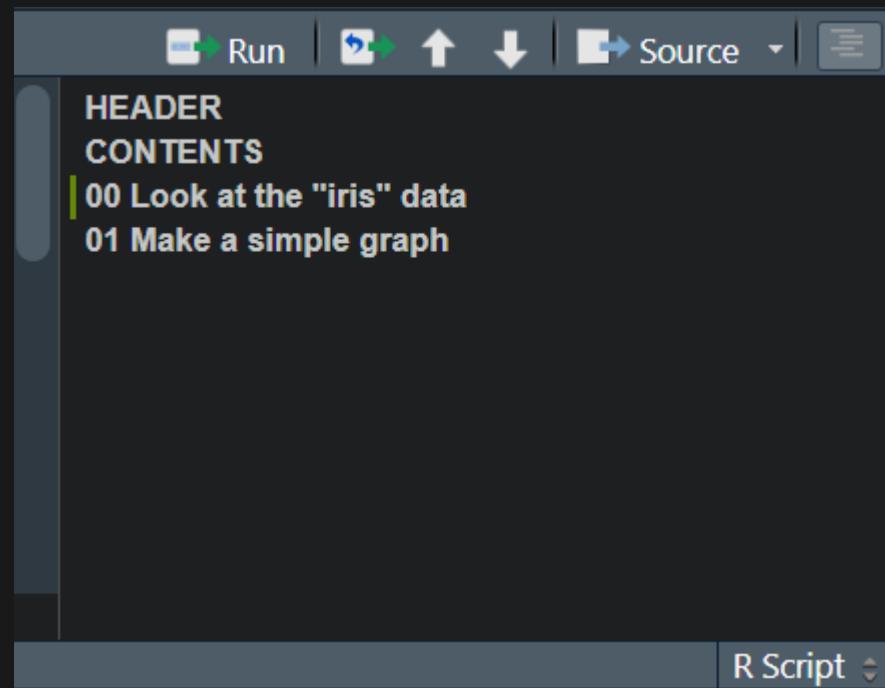
```
1 ## ----  
2 ## Author: <your name>  
3 ## Title: My first R script  
4 ## Last edited: yyyy-mm-dd (ISO 8601 date format)  
5 ## ----
```

R script Contents

```
1 # A typical script Contents section  
2  
3 ## Contents #####  
4 ## 00 Setup  
5 ## 01 Exploratory data analysis  
6 ## 02 Analysis  
7 ## 03 Etc...
```

`{< fa lightbulb >}` Script outline

Ending a comment line with four hastags (`#####`) or four dashes (`----`) adds comment to script outline



R script code chunks

- Start with at least one has sign
- Descriptive title
- End with four hashtags
- Consecutive numbering can be useful

R script comments

```
1 # A comment in an R script
2
3 # To make a comment you need to put a hashtag at the start
4 # if you don't R won't like it :(
```

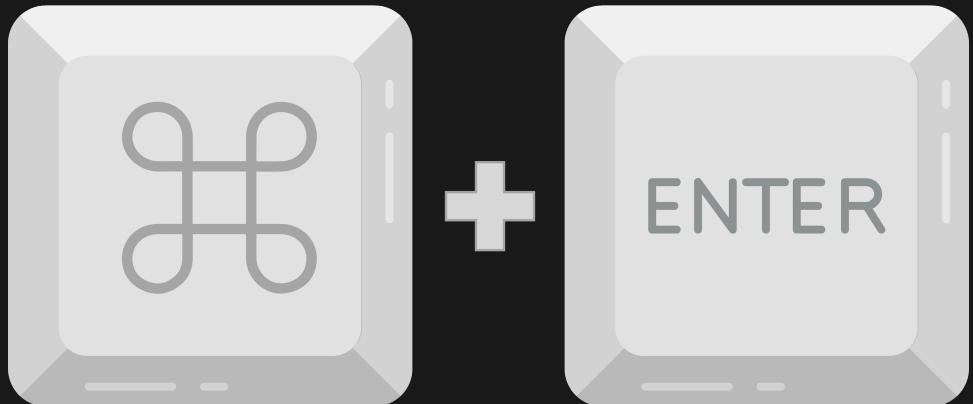
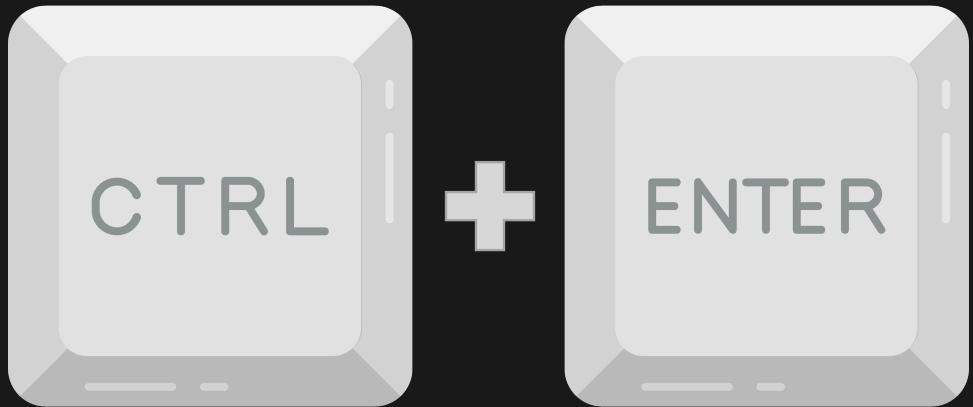
R script comments

```
1 # A comment in an R script
2
3 # To make a comment you need to put a hashtag at the start
4 # if you don't R won't like it :(
5
6 # Vector of numbers << comment
7 my_variable <- c(2,5,3,6,3,4,7)
8
9 # Calculate the mean << another comment
10 mean(my_variable)
```

```
[1] 4.285714
```

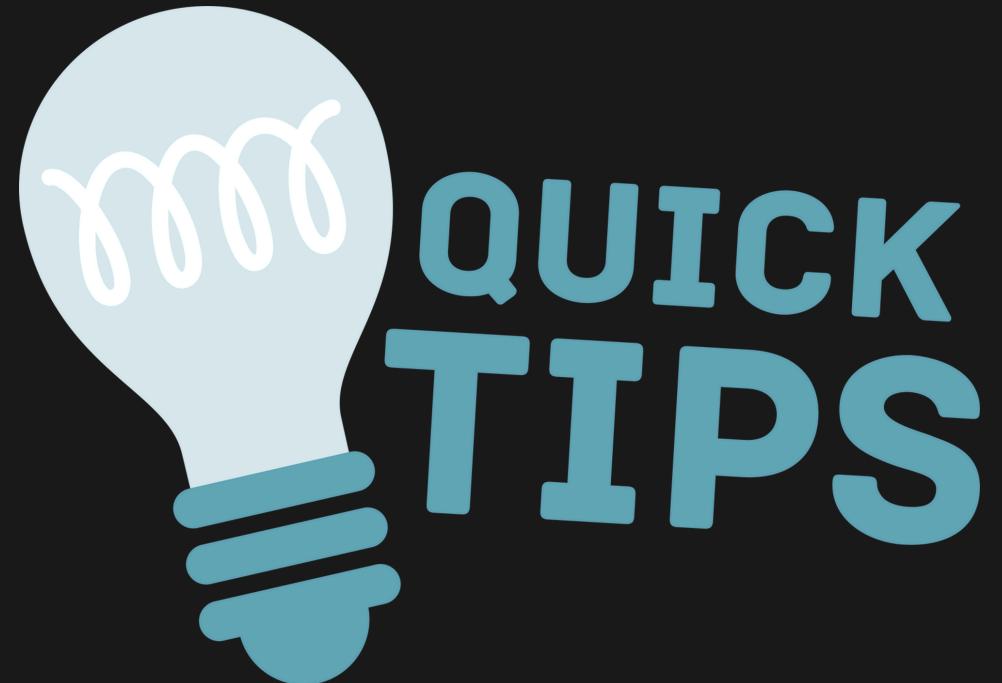
R script “submitting” commands

- Windows: **Ctrl + Return**
- Macs: **Cmd + Return**
- Run button



Other helpful and fun RStudio tips

- Appearance
- Rainbow parentheses
- Line wrapping
- Keyboard shortcuts



Practice exercises

- Opening R studio and exploring the IDE
- Starting a project
- Setting up an R script
- Running an R script
- Fun and helpful tips

