

Econ 106

Data Analysis in

Economics

Fall 2024

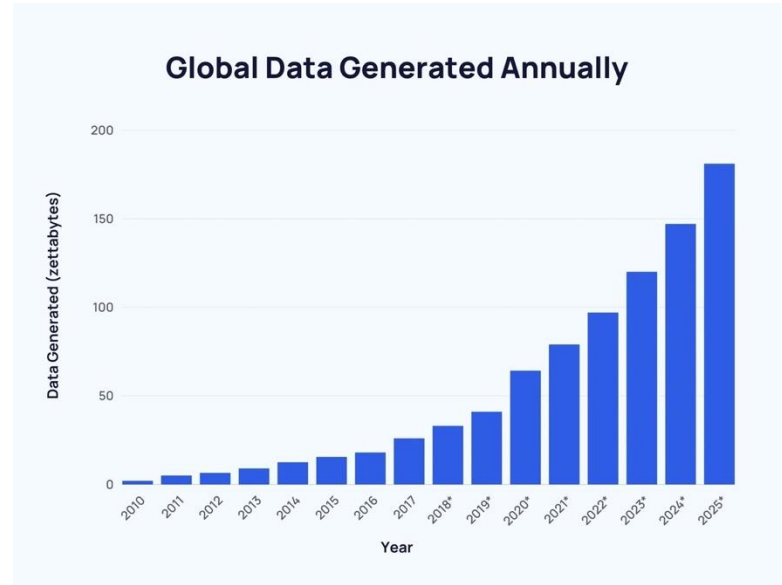
Based on: <https://datacarpentry.org/R-ecology-lesson/introduction-r-rstudio.html>

Outline

- Class Overview
- Intro to R:
 - Rstudio Console
 - Operators
 - Functions
 - Objects

Data is everywhere

- In 2015 the healthcare industry alone produced *1.2 billion clinical documents*.
- The amount of text data generated annually (digital forms, social media, online portals, pdf reports, emails, text messages) is ever increasing



Holmes, Dawn E., 'The data explosion', Big Data: A Very Short Introduction, Very Short Introductions (Oxford, 2017; online edn, Oxford Academic, 23 Nov. 2017), <https://doi.org/10.1093/actrade/9780198779575.003.0001>)

Big Data in economics

New sources of data create challenges that may require new skills



Matthew Harding
University of California, Irvine, USA



Jonathan Hersh
Chapman University, USA

ONE-PAGER

FULL ARTICLE

Deutsch

Español

Elevator pitch

Big Data refers to data sets of much larger size, higher frequency, and often more personalized information. Examples include data collected by smart sensors in homes or aggregation of tweets on Twitter. In small data sets, traditional econometric methods tend to outperform more complex techniques. In large data sets, however, machine learning methods shine. New analytic approaches are needed to make the most of Big Data in economics. Researchers and policymakers should thus pay close attention to recent developments in machine learning techniques if they want to fully take advantage of these new sources of Big Data.

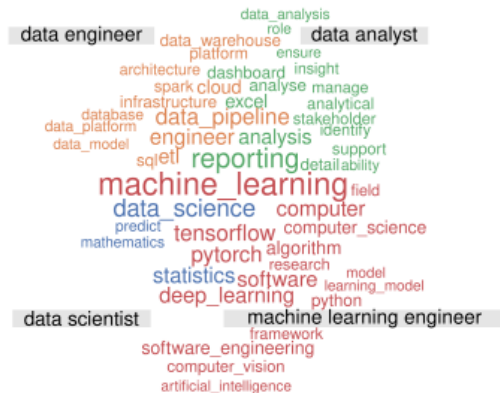
<https://wol.iza.org/articles/big-data-in-economics>

Need a job? Learn how to work with data

- Data analyst: emphasis on being able to interpret and communicate findings (insight, reporting)
- Data literacy is the ability of an individual to read, understand, and *work* with data to create meaningful insights

Comparison Cloud: English Language Job Descriptions

Words that are more distinct for a category appear in larger print



(N = 3,869)

<https://www.r-bloggers.com/2022/04/text-analysis-of-job-descriptions-for-data-scientists-data-engineers-machine-learning-engineers-and-data-analysts/>

Intended Audience

- Students who want to:
 - serve as research assistants
 - conduct empirical analyses
 - are interested in Data Science careers.
- Prerequisites: Econ 101 (Statistics)
- No prior programming experience is needed

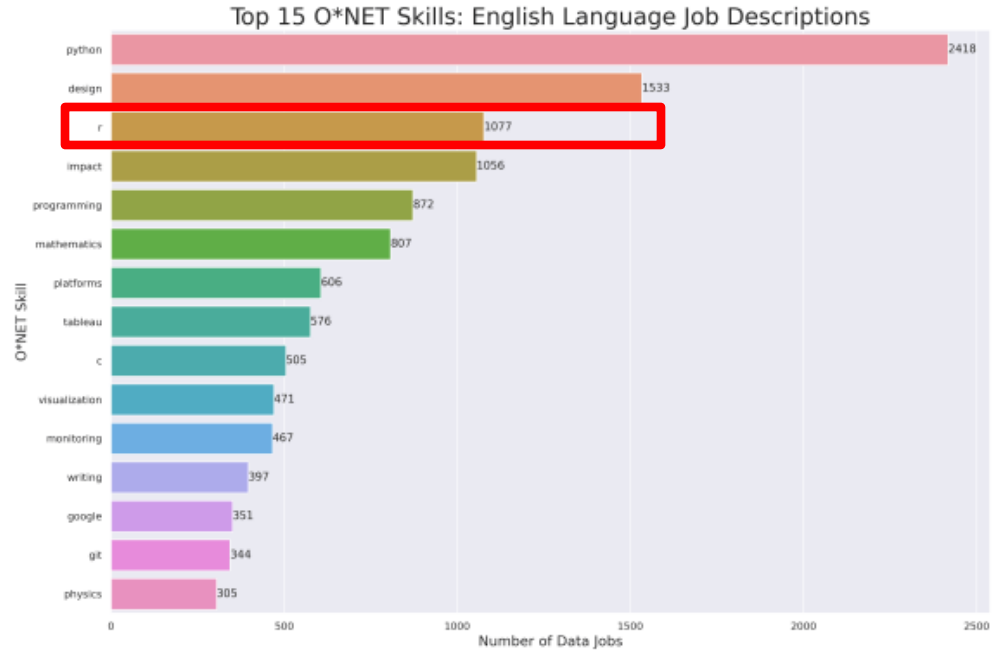
Course Objectives

- Master the ability to collect, process, interpret, analyze *raw* data using R.
- Develop visualization skills necessary to summarize data findings
- Communicate findings to various audiences and in various formats.

Why are we using R in this course?

Advantages of R:

- It's free and open source
- It's widely used in industry
- It has a large and active user community
- More of a true programming language compared to Stata



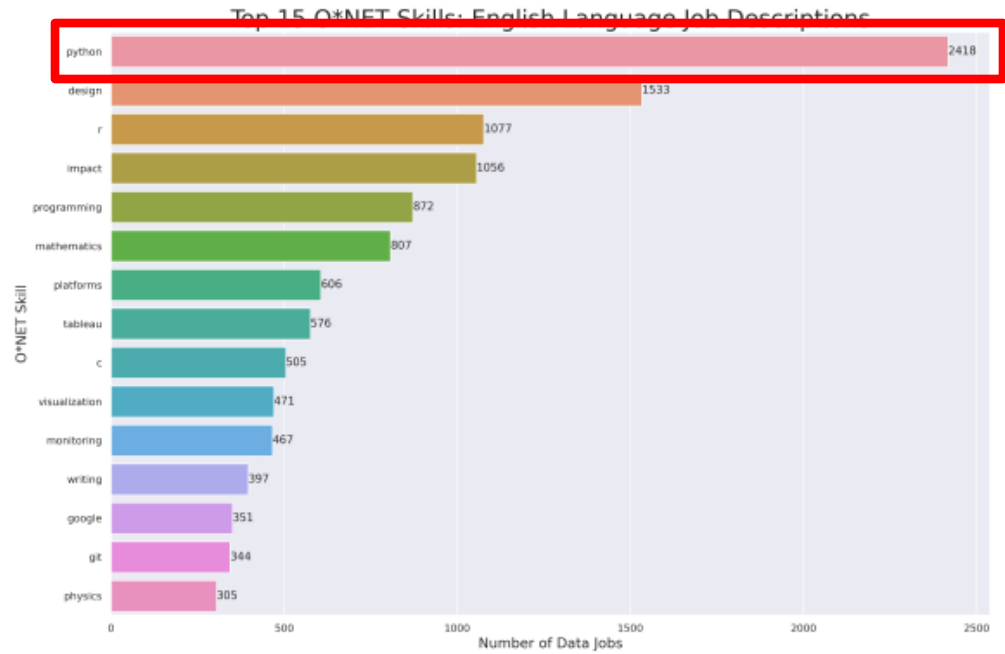
R vs. Python

R:

- Built for statistics and data analysis
- Better at econometrics and data visualization

Python:

- Built for general-purpose programming and software development
- Better at machine learning



R is a means, not an end

- The goals of this course are **platform-agnostic**
 - It's not about the syntax of specific packages
 - It's about the concepts, logic, and thought processes underlying what we're doing and why
- Your eventual goal: **Use the right tool for the job**
- You will likely learn all these tools at some point.

Grading

Assignment	Points
Lab Assignments	40
Poll Everywhere	10
Project Milestones	30
Final Project	20
TOTAL	100

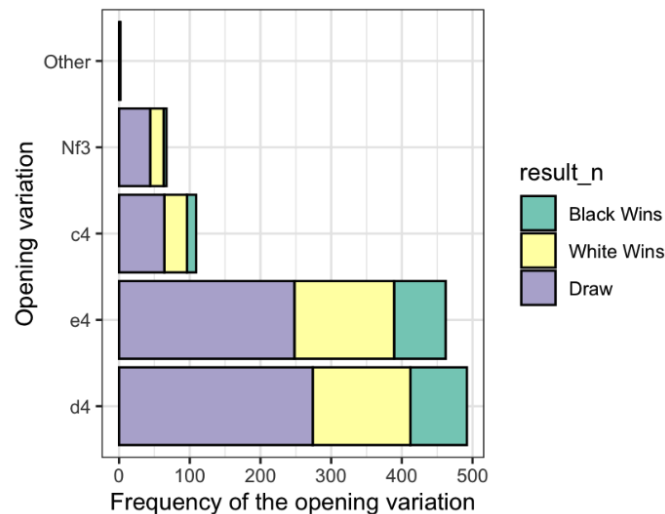
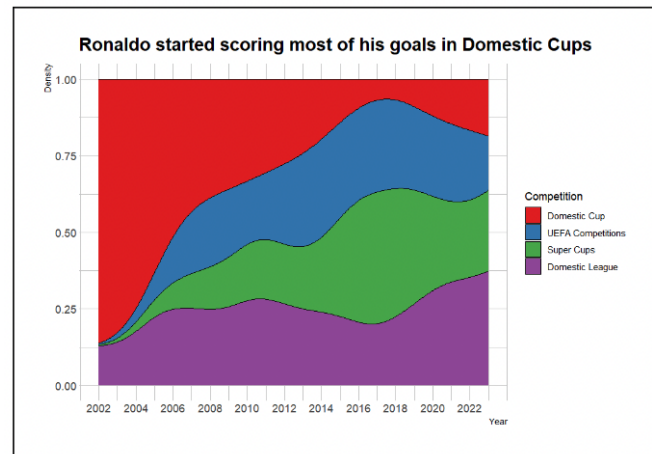
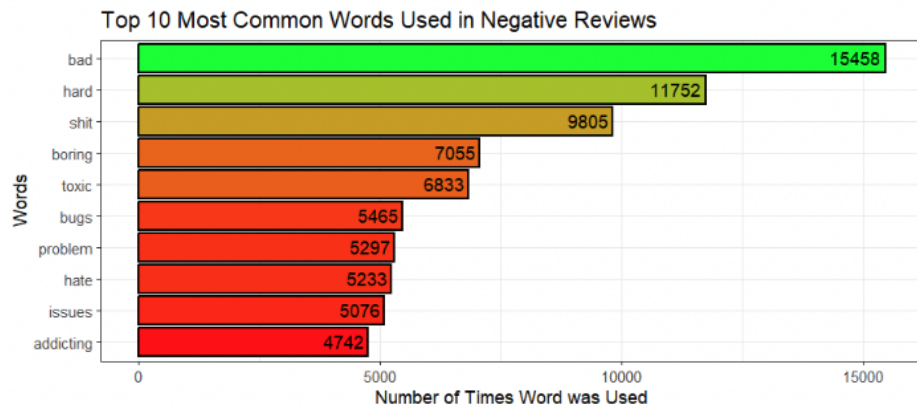
Lab Assignments

- Lab assignments: gives you the opportunity to practice the data skills covered during lecture
- Two components: writeup and R script
- Due on Sundays at 11:59pm
- I will drop the lowest lab score at the end of the quarter

Research Project

- The goal of this class is to give you the skills to conduct your own *small* research project.
- Three project milestones will walk you through the research process
- All three will count toward your final grade (no drops)
- Final product:
 - Writeup (~1500 words)
 - documentation (r scripts)

Examples of Student Work



ChatGPT/Generative AI

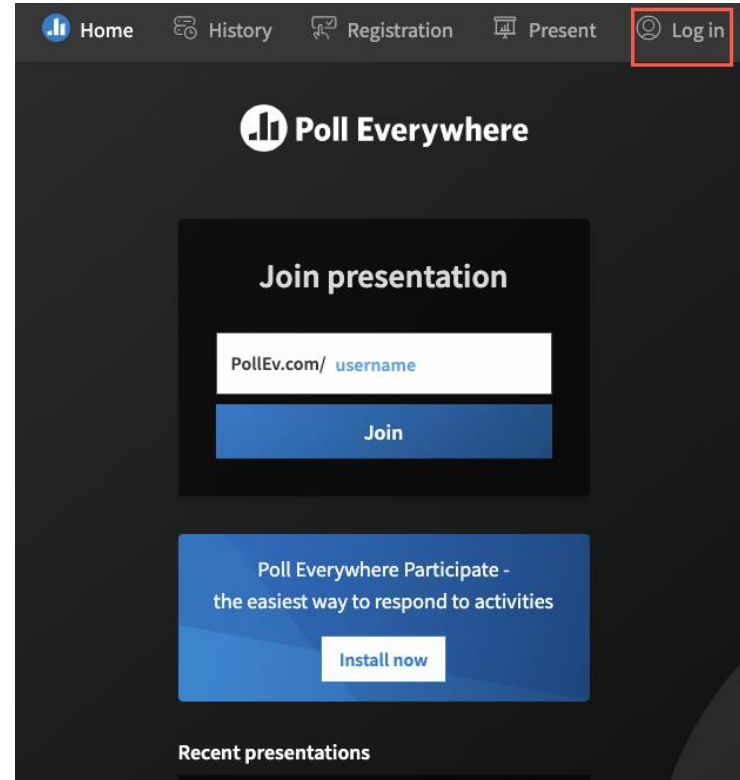
- It is up to you whether you want to use generative AI to help you with your code
- Guidelines:
 - The code has to use the tools covered in lecture
 - You cannot use it for the written portion of lab assignments or research milestones
 - You need to state whether you used it for your code and how.

Poll Everywhere

I use Poll Everywhere to encourage active learning during lecture

My polls will be available here when they go live:

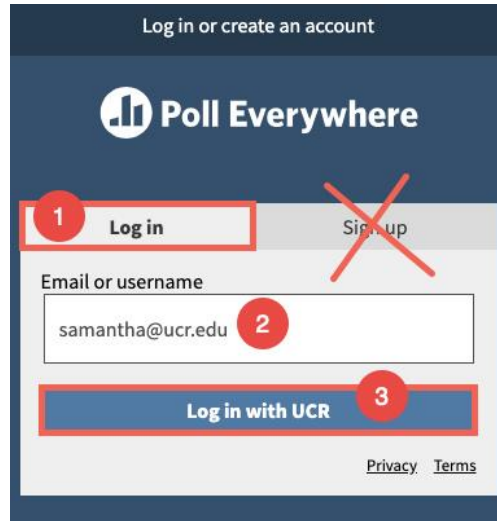
Pollev.com/vsovero



Poll Everywhere

Please use your ucr email address to sign in (do not sign up for an account)

You will be redirected to the UCR SSO



Log in or create an account

Poll Everywhere

1 **Log in** ~~Sign up~~

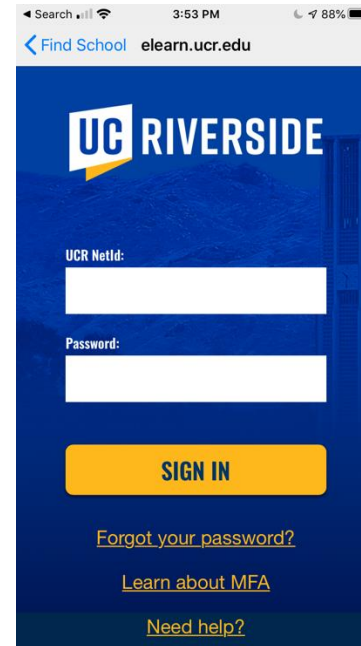
Email or username

samantha@ucr.edu 2

3 **Log in with UCR**

[Privacy](#) [Terms](#)

This image shows the Poll Everywhere login interface. A red box highlights the 'Log in' button, which is labeled with a red circle containing the number 1. The 'Sign up' button is crossed out with a large red X. Below the buttons is a text input field labeled 'Email or username' containing the email address 'samantha@ucr.edu', which is labeled with a red circle containing the number 2. At the bottom, a blue button labeled 'Log in with UCR' is highlighted with a red box and labeled with a red circle containing the number 3. Links for 'Privacy' and 'Terms' are at the bottom right.



Search 3:53 PM 88%

[Find School](#) elearn.ucr.edu

UC RIVERSIDE

UCR NetId:

Password:

SIGN IN

[Forgot your password?](#)

[Learn about MFA](#)

[Need help?](#)

This image shows the UCR Riverside login page. At the top, there is a header with 'Search', the time '3:53 PM', and the battery level '88%'. Below this is a navigation bar with a back arrow, 'Find School', and the URL 'elearn.ucr.edu'. The main content area has a blue background with the 'UC RIVERSIDE' logo. There are two white input fields for 'UCR NetId:' and 'Password:'. A large yellow button labeled 'SIGN IN' is centered below the fields. At the bottom, there are three links: 'Forgot your password?', 'Learn about MFA', and 'Need help?'.

Lecture Quizzes (ungraded)

- Posted on Canvas for you to practice after lecture
- Does not count towards your grade
- Feel free to ask Fan to review these questions during lab section

Lecture Scripts

- I post all of the code used during lecture as an R script on Canvas
- Running the code during or after lecture will help you gain more familiarity with Rstudio

This Week: Introduction to R Basics

- Before we cover more broad topics in data analysis, we need to understand some R basics
- We need these building blocks before moving onto data wrangling, visualization, etc.

Coding Goals

1. Run code without errors (pretty doable with ChatGPT)
2. Run code that produces what you were expecting
3. Run code that produces what you were expecting and you understand

Getting help

1. Read the error message.
2. Try some online resources:
 - Stack Overflow
 - ChatGPT (type in your code and the error message)
 - Slack
3. Still stuck?
 - Please reach out to myself or your TA (Fan Bu)
 - Share your code, data, screenshots of the error message, etc.

Example: Be Careful with Typos

- Implicit contract with the computer / scripting language:
 - Computer will do tedious computation for you.
 - In return, you will be completely precise in your instructions
- Typos matter. Case matters.

```
data_rocks <- 2 ^ 3
```

```
datarocks
```

```
#> Error in eval(expr, envir, enclos): object 'datarocks' not found
```

```
Data_rocks
```

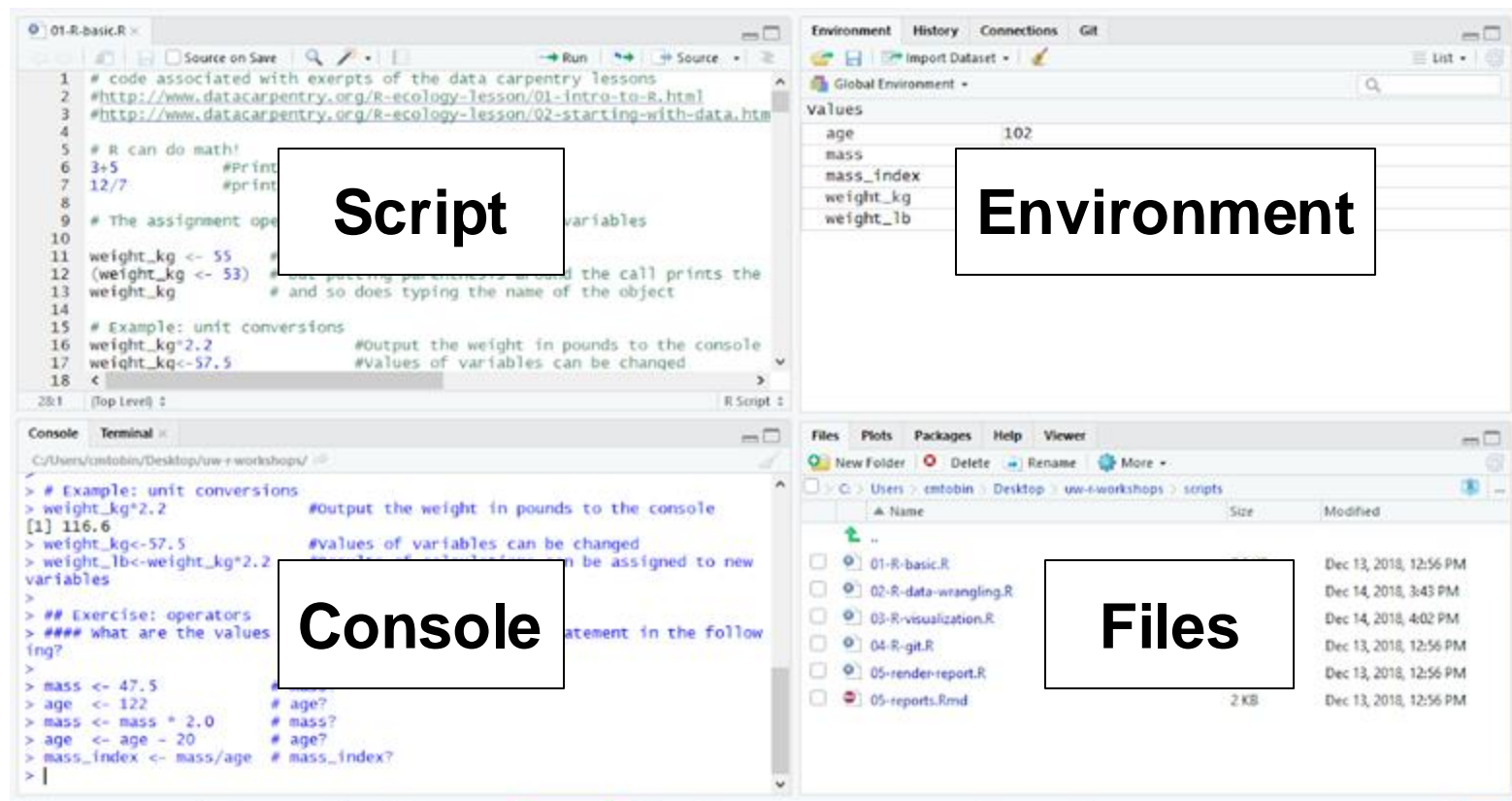
```
#> Error in eval(expr, envir, enclos): object 'Data_rocks' not found
```

What is R? RStudio?

- **R** – a programming language + software that interprets it
- **RStudio** – popular software to write R scripts and interact with the R software
- We will be using Rstudio in this course (however you need to install R in order for it to run correctly)

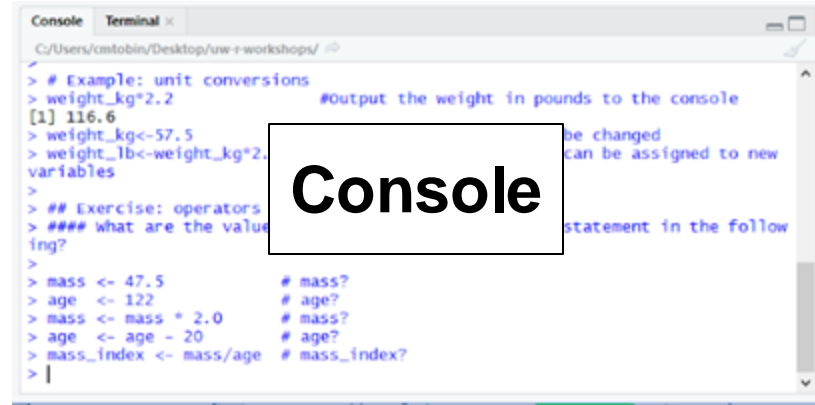


R Studio Interface



RStudio Console

- You can input and execute commands directly in the console.
- Output of certain commands will be displayed in the console.



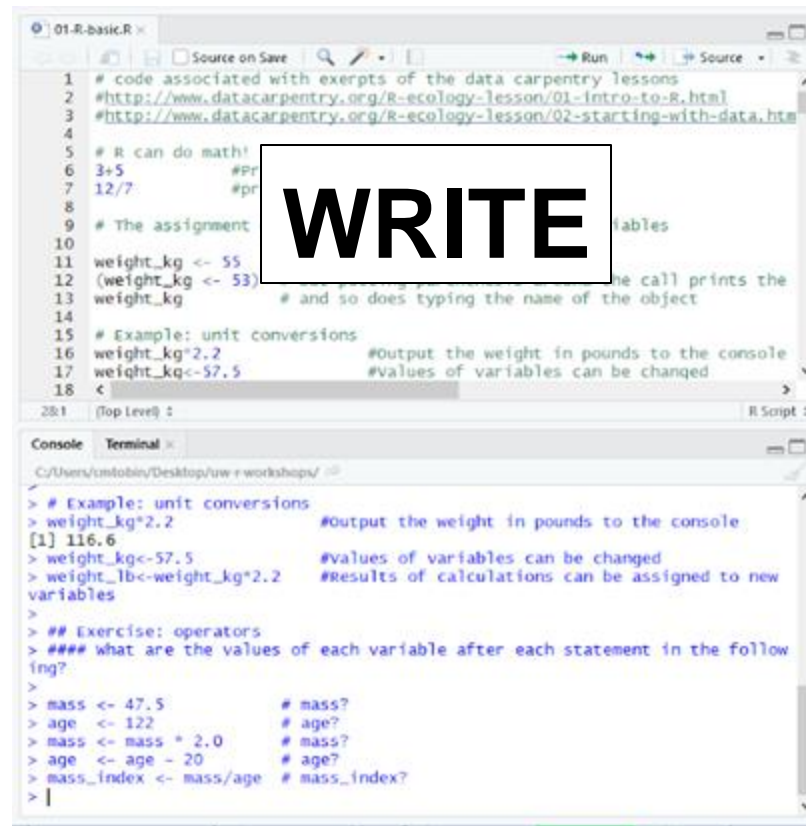
The screenshot shows the RStudio Console window with the following content:

```
Console Terminal x
C:/Users/cmlabin/Desktop/uw-r-workshops/ >
> # Example: unit conversions
> weight_kg*2.2 #Output the weight in pounds to the console
[1] 116.6
> weight_kg<-57.5
> weight_lb<-weight_kg*2.2
> # Exercise: operators
> ### what are the value
ing?
>
> mass <- 47.5 # mass?
> age <- 122 # age?
> mass <- mass * 2.0 # mass?
> age <- age - 20 # age?
> mass_index <- mass/age # mass_index?
> |
```

A white box with the word **Console** is overlaid on the right side of the console window.

Rstudio Script Editor

- Most data analysis will require several steps
- The script editor is used to record the steps (commands) we take to analyze our data



The screenshot displays the RStudio interface. The top pane is the Script Editor, showing an R script with comments and code. A white box with the word "WRITE" in bold black letters is overlaid on the script. The bottom pane is the Console, showing the output of the code executed in the script.

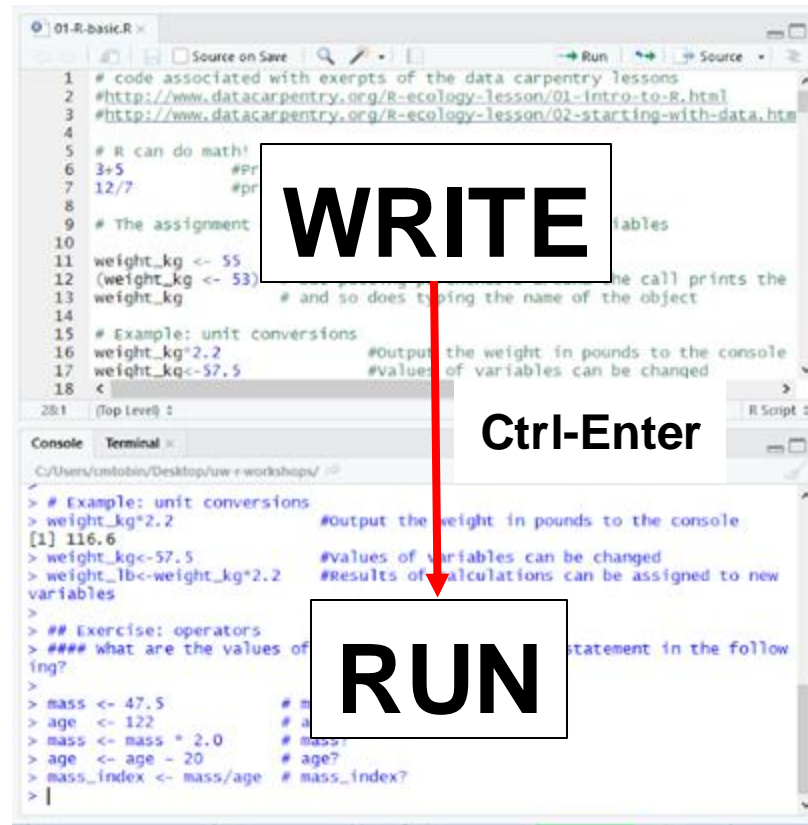
```
# 01-R-basic.R
1 # code associated with excerpts of the data carpentry lessons
2 #http://www.datacarpentry.org/R-ecology-lesson/01-intro-to-r.html
3 #http://www.datacarpentry.org/R-ecology-lesson/02-starting-with-data.htm
4
5 # R can do math!
6 3+5 #Pr
7 12/7 #pr
8
9 # The assignment
10
11 weight_kg <- 55
12 (weight_kg <- 53)
13 weight_kg # and so does typing the name of the object
14
15 # Example: unit conversions
16 weight_kg*2.2 #output the weight in pounds to the console
17 weight_kg<-57.5 #values of variables can be changed
18 <
19
20:1 [Top Level] :
```

Console

```
> # Example: unit conversions
> weight_kg*2.2 #output the weight in pounds to the console
[1] 116.6
> weight_kg<-57.5 #values of variables can be changed
> weight_lb<-weight_kg*2.2 #Results of calculations can be assigned to new
variables
>
> ## Exercise: operators
> #### what are the values of each variable after each statement in the follow
ing?
>
> mass <- 47.5 # mass?
> age <- 122 # age?
> mass <- mass * 2.0 # mass?
> age <- age - 20 # age?
> mass_index <- mass/age # mass_index?
> |
```

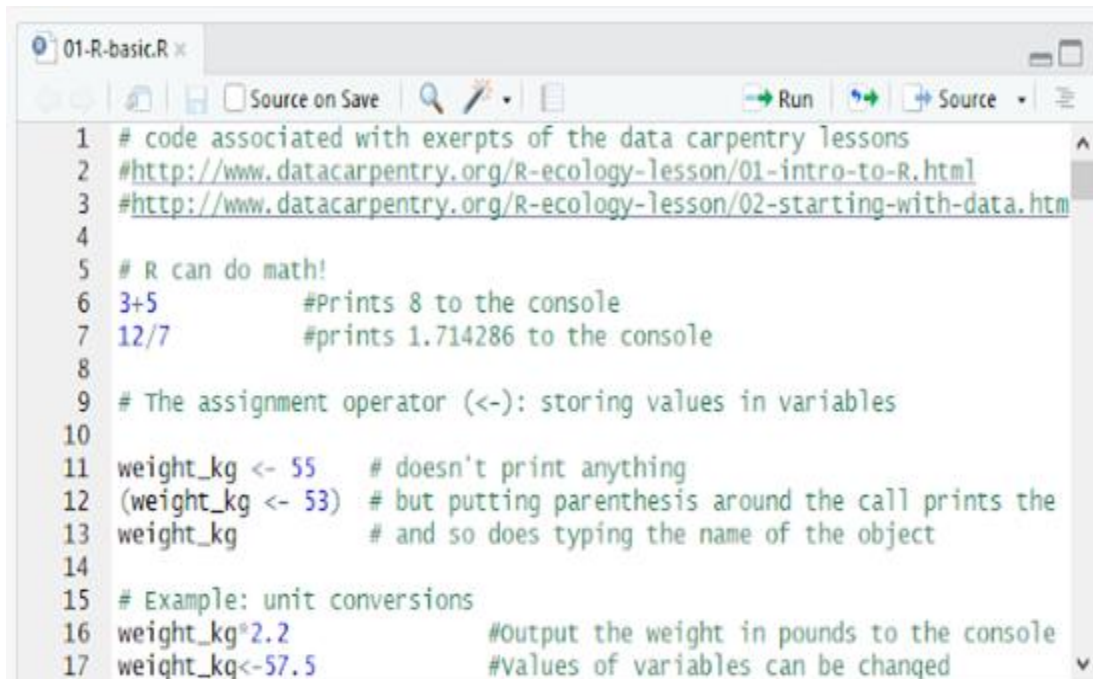
Rstudio Script Editor

- Once you type them in the script, you still need to send them to the console to be run.
- Scripts are saved with the .R extension (“lecture1.R”)



Commenting in Scripts

- R ignores the rest of a line after a #
- write notes to yourself about what your code is doing
- comments are automatically in green



```
01-R-basic.R x
Source on Save
Run
Source

1 # code associated with excerpts of the data carpentry lessons
2 #http://www.datacarpentry.org/R-ecology-lesson/01-intro-to-R.html
3 #http://www.datacarpentry.org/R-ecology-lesson/02-starting-with-data.htm
4
5 # R can do math!
6 3+5           #Prints 8 to the console
7 12/7          #prints 1.714286 to the console
8
9 # The assignment operator (<-): storing values in variables
10
11 weight_kg <- 55   # doesn't print anything
12 (weight_kg <- 53) # but putting parenthesis around the call prints the
13 weight_kg        # and so does typing the name of the object
14
15 # Example: unit conversions
16 weight_kg*2.2     #Output the weight in pounds to the console
17 weight_kq<-57.5   #Values of variables can be changed
```

Let's start coding!

- An operator is a **symbol** that tells R to perform some sort of operation.
- **Types of Operators:**
 - math (arithmetic)
 - logical (compare values)
 - assignment (saving values to objects)
 - extraction (take out/replace parts of objects)

Operators

<https://www.statmethods.net/management/operators.html>

Type	What it does	Symbol(s)
Arithmetic	Math on numbers	+ - * / ^
Logical	Compares values	> < == != &
Assignment	Assigns values to objects	<-
Extraction	Take out or replace part of an object	[] \$

Arithmetic operators

- Does math:
 - Add: $2 + 2$
 - Subtract: $3 - 1$
 - Multiply: $4 * 4$
 - Divide: $5 / 2$
 - Exponent: 5^2
- Ctrl/Cmd+Enter to run
- Results in the Console

Arithmetic Operators

Operator	Description
+	addition
-	subtraction
*	multiplication
/	division
^ or **	exponentiation

Example: Who left a larger Tip?

- Louberto's bill at Shake Shack came out to \$25
 - He tips 15%
- Louie's bill at Shake Shack came out to \$32
 - He tips 12%
- Use operators to calculate how much Louie and Louberto tipped

Functions

- For more complicated tasks, R also comes pre-installed with a variety of functions.
- Functions are a sequence of instructions that perform a task
 - Have **names**
 - Accept **arguments** (input)
 - Return a **value** (output)

Input	Output
round (x = 3.14159, digits = 0)	3

Arguments

- To find out what **arguments** a function takes, look up its help file (type `?round`)
- Some arguments are required, some are optional.
- You can see that `digits` is optional because you don't need to include it and you will get the same output

Input	Output
<code>round(x = 3.14159, digits = 0)</code>	3
<code>round(3.14159)</code>	3

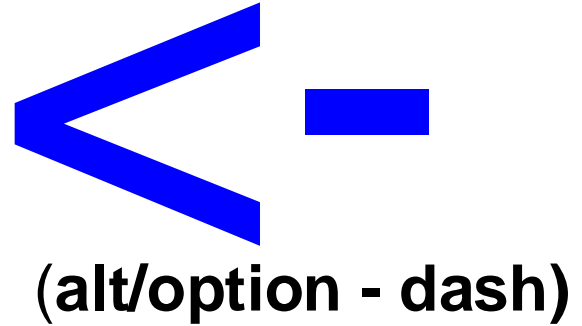
Arguments

- You technically don't even need to write out the names of the arguments (`x`, `digits`) if values are in the correct order
- However, it's much easier to understand your code when you write out the name of the argument

Input	Output
<code>round(x = 3.14159, digits = 2)</code>	<code>3.14</code>
<code>round(3.14159, 2)</code>	<code>3.14</code>

Assignment operator

- Saves **values** into **objects**
 - **object** <- **value**
 - **weight_kg** <- **55**
- Overwrites previous values
- Can be combined with other operators
 - **weight_lb** <- **2.2*weight_kg**



Naming Objects

- Object names:
 - cannot start with a digit
 - cannot contain certain other characters such as a comma or a space.
- Tips:
 - Make it as descriptive as possible
 - Use underscores to mark spaces

#this is good

weight_kg <- 55

#not these

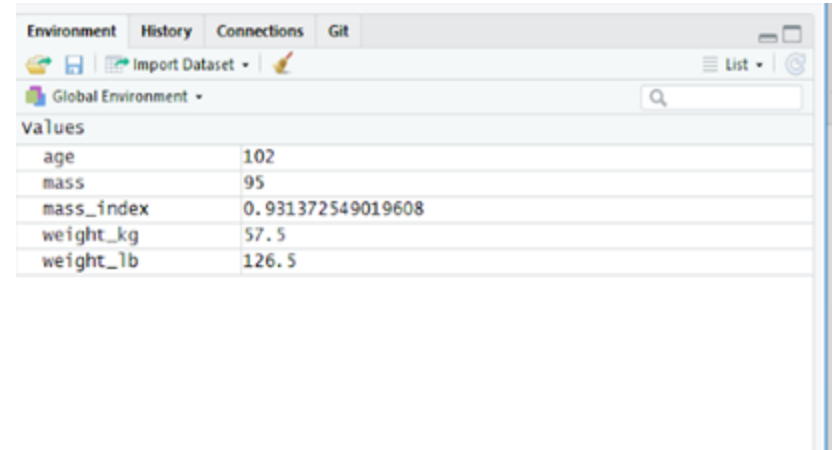
weightkg! <- 55

weight in kg <- 55

blah <- 55

Where are my objects?

- You can see your objects in the workspace (top right pane of Rstudio)
- You can also get a listing of these objects with commands:
 - `ls()`
 - `objects()`



The screenshot shows the RStudio Environment pane. At the top, there are tabs for 'Environment', 'History', 'Connections', and 'Git'. Below the tabs, there is a toolbar with icons for 'Import Dataset' and a search icon. The main area displays 'Global Environment' with a search bar. Below this, a table titled 'values' lists the objects in the workspace and their corresponding values.

values	
age	102
mass	95
mass_index	0.931372549019608
weight_kg	57.5
weight_lb	126.5

Checking Values of an Object

- If you want to inspect the values of an object, type the name of the object in the console

```
weight_kg <- 55
```

```
weight_kg
```


Example: Who left a larger Tip?

- Louberto's bill at Shake Shack came out to \$25
 - He tips 15%
- Louie's bill at Shake Shack came out to \$32
 - He tips 12%
- Use the assignment operator to save how much Louie and Louberto tipped (create 6 objects)

Removing Objects

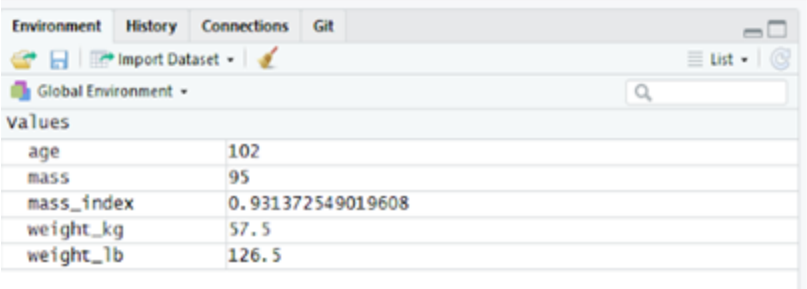
- Sometimes when you have Rstudio open for a while, your workspace gets cluttered up with objects
- You can remove objects with commands (right) or by clicking the broom icon in Rstudio's Environment pane

#remove object called y

```
rm(y)
```

#remove everything in your workspace

```
rm(list = ls())
```



The screenshot shows the RStudio Environment pane with tabs for Environment, History, Connections, and Git. The Environment pane displays a table of variables in the Global Environment. The variables and their values are:

Values	
age	102
mass	95
mass_index	0.931372549019608
weight_kg	57.5
weight_lb	126.5

Objects: Not Just for Storing Numbers

- Basic Data types:
 - Character
 - Numeric

	type
<code>x <- 32</code>	numeric
<code>y <- "hi"</code>	character