Resources for learning R

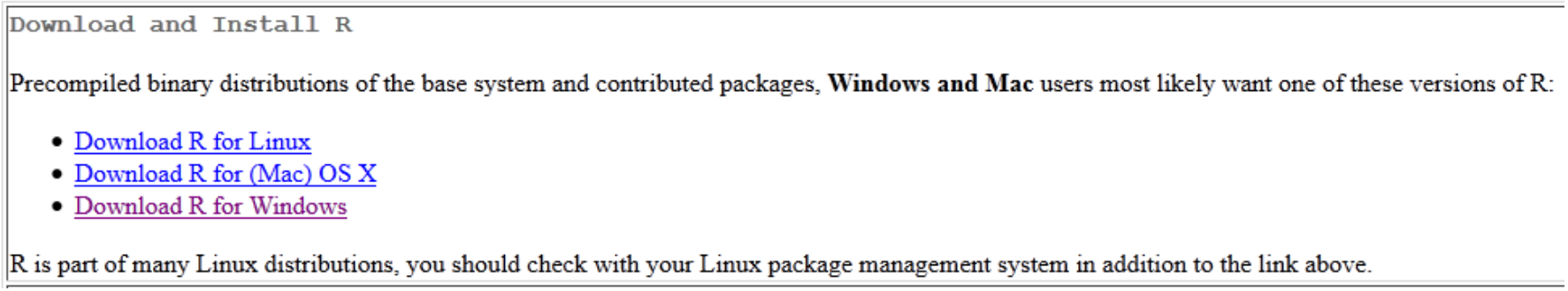
Tutorial for installation, recommended resources for learning R, and links to reference sheets (cheat sheets).

Install R

- Go to the R website, select “download R”, and choose your preferred CRAN mirror: https://www.r-project.org/

Easier:

- Or simply go to this page: https://cran.cnr.berkeley.edu/
- And choose the link for your operating system:


Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- Download R for Linux
- Download R for (Mac) OS X
- Download R for Windows

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

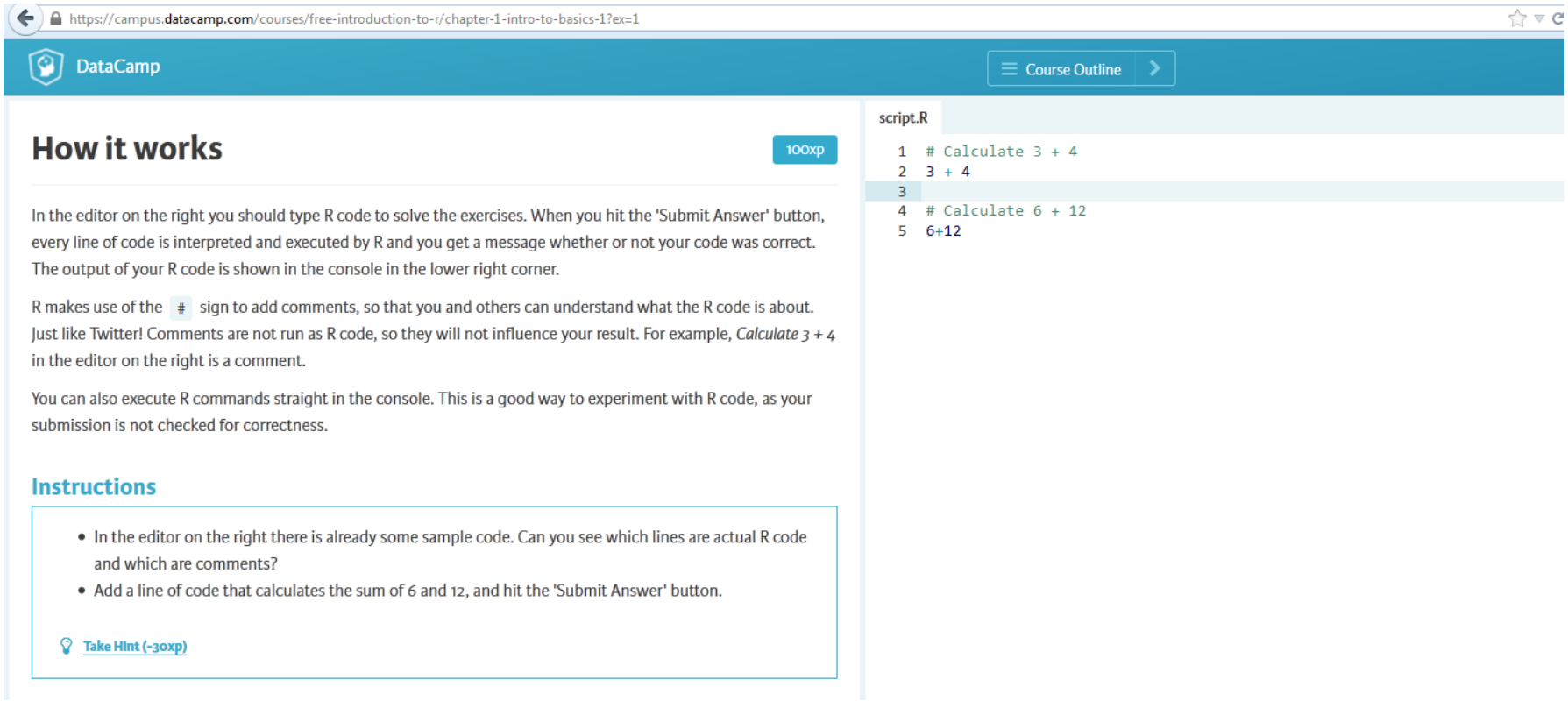
- Choose base R, or the “install R for the first time” link.
- Click download link at the top of the page, then install.

Install R studio

- Base R must be installed before Rstudio.
- https://www.rstudio.com/products/rstudio/download2/
- Choose RStudio Desktop Open Source License (the FREE one).
- Choose your operating system, download and install

DataCamp

- https://www.datacamp.com/courses/free-introduction-to-r
- Free, simple, interactive
- Highly recommended!



R for Data Science

- Quintessential text for data analysis in R
- http://r4ds.had.co.nz/
- Practice in DataCamp to learn the basic syntax, then read every word of this book.

Reference sheets

- https://www.rstudio.com/resources/cheatsheets/
- Recommended: Data import, Data transformation, Data visualization, Base R, Advanced R, Regular Expressions

Tidy Data with tidyr

Tidy data is a way to organize tabular data. It provides a consistent data structure across packages.

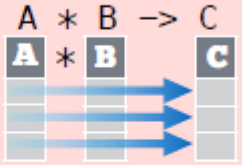
A table is tidy if:

Each **variable** is in its own **column**

Each **observation**, or **case**, is in its own **row**

Tidy data:

Makes variables easy to access as vectors



Preserves cases during vectorized operations

Reshape Data - change the layout of values in a table

Use **gather()** and **spread()** to reorganize the values of a table into a new layout. Each uses the idea of a key column: value column pair.

gather(data, key, value, ..., na.rm = FALSE, convert = FALSE, factor_key = FALSE)

Gather moves column names into a key column, gathering the column values into a single value column.

table4a

country	1999	2000
A	0.7K	2K
B	37K	80K
C	212K	213K

→

country	year	cases
A	1999	0.7K
B	1999	37K
C	1999	212K
A	2000	2K
B	2000	80K
C	2000	213K

key value

gather(table4a, `1999`, `2000`, key = "year", value = "cases")

spread(data, key, value, fill = NA, convert = FALSE, drop = TRUE, sep = NULL)

Spread moves the unique values of a key column into the column names, spreading the values of a value column across the new columns that result.

table2

country	year	type	count
A	1999	cases	0.7K
A	1999	pop	19M
A	2000	cases	2K
A	2000	pop	20M
B	1999	cases	37K
B	1999	pop	172M
B	2000	cases	80K
B	2000	pop	174M
C	1999	cases	212K
C	1999	pop	1T
C	2000	cases	213K
C	2000	pop	1T

→

country	year	cases	pop
A	1999	0.7K	19M
A	2000	2K	20M
B	1999	37K	172M
B	2000	80K	174M
C	1999	212K	1T
C	2000	213K	1T

key value

spread(table2, type, count)

Split and Combine Cells

Use these functions to split or combine cells into individual, isolated values.

separate(data, col, into, sep = "[^[:alnum:]]+", remove = TRUE, convert = FALSE, extra = "warn", fill = "warn", ...)

Separate each cell in a column to make several columns.

table3

country	year	rate
A	1999	0.7K/19M
A	2000	2K/20M
B	1999	37K/172M
B	2000	80K/174M
C	1999	212K/1T
C	2000	213K/1T

→

country	year	cases	pop
A	1999	0.7K	19M
A	2000	2K	20M
B	1999	37K	172
B	2000	80K	174
C	1999	212K	1T
C	2000	213K	1T

separate_rows(table3, rate, into = c("cases", "pop"))

separate_rows(data, ..., sep = "[^[:alnum:].]+", convert = FALSE)

Separate each cell in a column to make several rows. Also **separate_rows_()**.

table3

country	year	rate
A	1999	0.7K/19M
A	2000	2K/20M
B	1999	37K/172M
B	2000	80K/174M
C	1999	212K/1T
C	2000	213K/1T

→

country	year	rate
A	1999	0.7K
A	1999	19M
A	2000	2K
A	2000	20M
B	1999	37K
B	1999	172M
B	2000	80K

For my talk

- We will open Rstudio and run through the “R workshop demonstration.R” script
- To run the script on your computer/laptop:
- Install base R
- Install Rstudio
- Open Rstudio and load the “Tidyverse installation script.R” and run it. Or type the following into the Console:

install.packages(“tidyverse”)

- Then open the demonstration script.

Last thing…

- In the demonstration script, you will need to change the “setwd()” line to where the “evidence.txt” files are located.
- You can do this by pressing Control + Shift + H and selecting the folder.
- Beyond that, you should be able to run the script during the talk and follow along.

Questions

- If you have any R-related questions, e-mail me (Aaron Storey) at ajstorey@uams.edu
- Thanks for reading!