



HOW TO EXPLORE YOUR DATA

With EXCEL

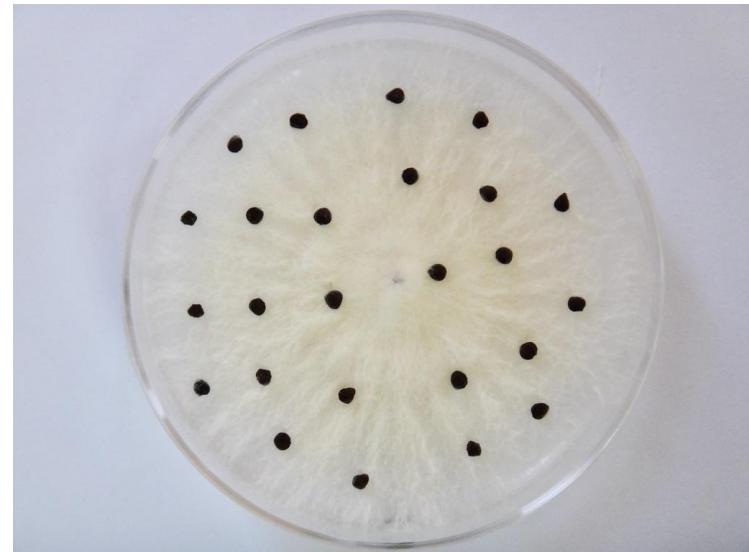
Carolina Sarmiento
USF
October 10, 2025

A little bit about me



- Microbial ecologist (tropical)
- Plant-fungal interactions
- Working mostly on seeds
(but also, some roots, leaves)
- I work with:

Dr. Luanna Prevost – Amgen Biotech Experience (Master Lab Tech – Tampa Site Coordinator)
Dr. Camilo Zalamea – Lab Manager/Research
Smithsonian Tropical Research Institute – Research Associate





Writing and Reviewing

Publications in:

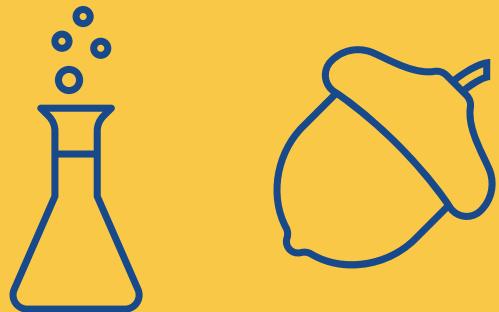
- Annual Review of Ecology, Evolution, and Systematics (AREES)
- PNAS
- Science
- Proceedings of the Royal Society B
- New Phytologist
- Ecology
- Functional Ecology
- Biotropica
- Others

Reviewer for:

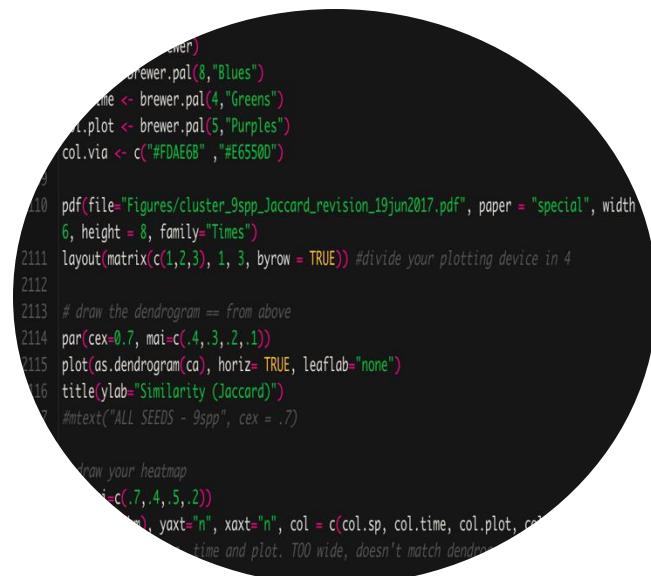
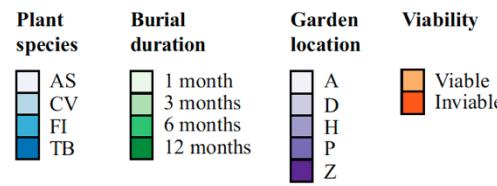
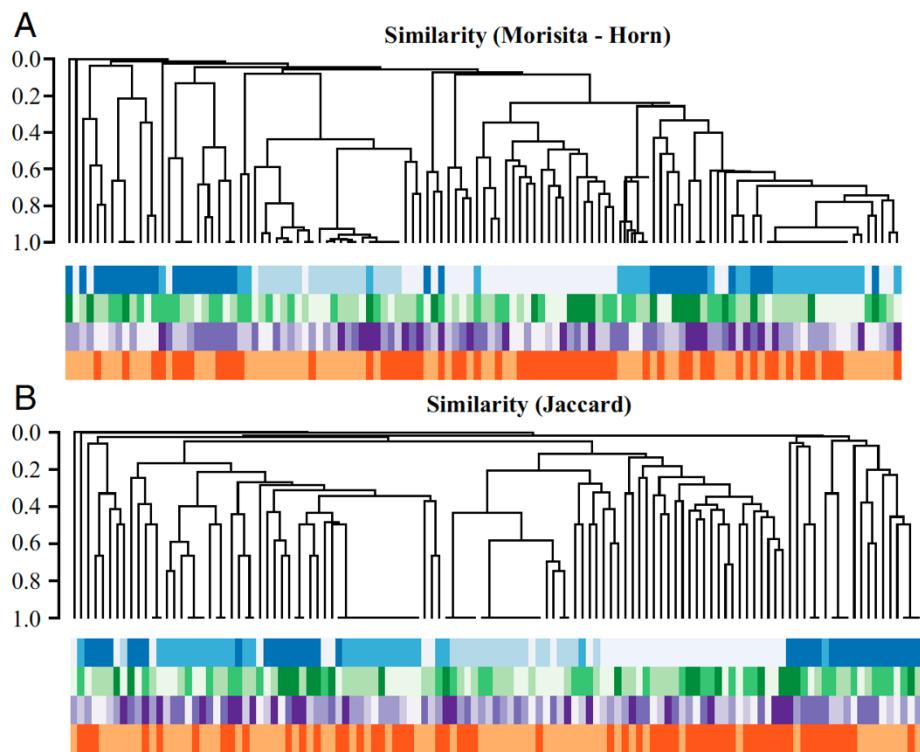
- New Phytologist
- Functional Ecology
- Fungal Ecology
- Ecology and Evolution
- Biotropica
- Journal of Ecology

[My profile on Scholar Google](#)

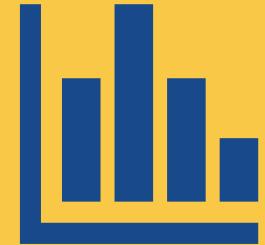
I enjoy ...



Lab work and data analysis and visualization!



I enjoy ...



Data analysis and data visualization!

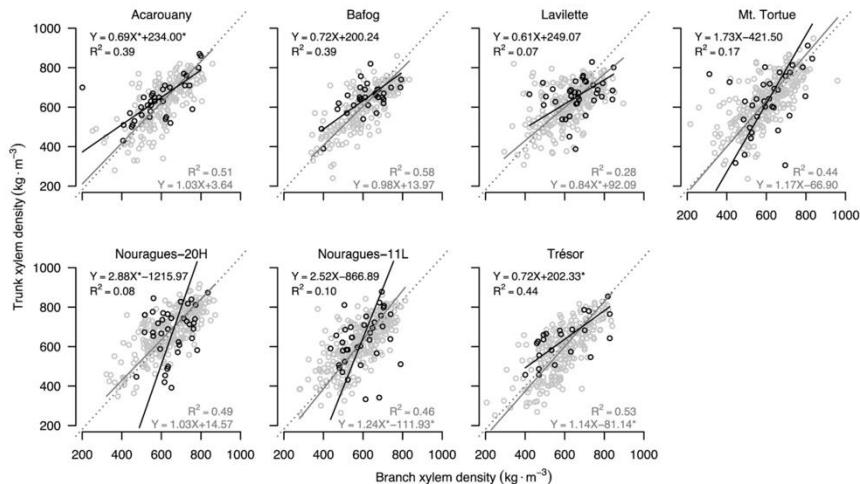
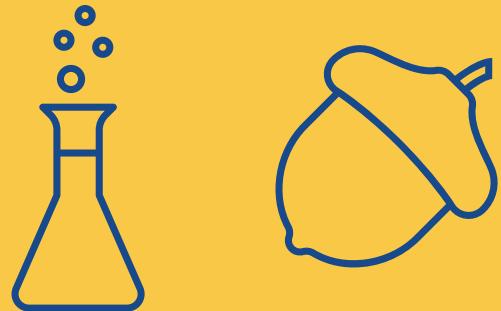


Fig. 3. The major-axis (MA) fit between trunk and branch xylem density in seven rain forest sites from French Guiana. The gray points and solid gray line indicate the site-level relationship, whereas the black points and lines indicate individuals of Fabaceae in each site. The dotted line represents the MA-fitted line for all 1909 samples together. Asterisks indicate significance as in Fig. 1.

I enjoy ...



Data analysis and data visualization!

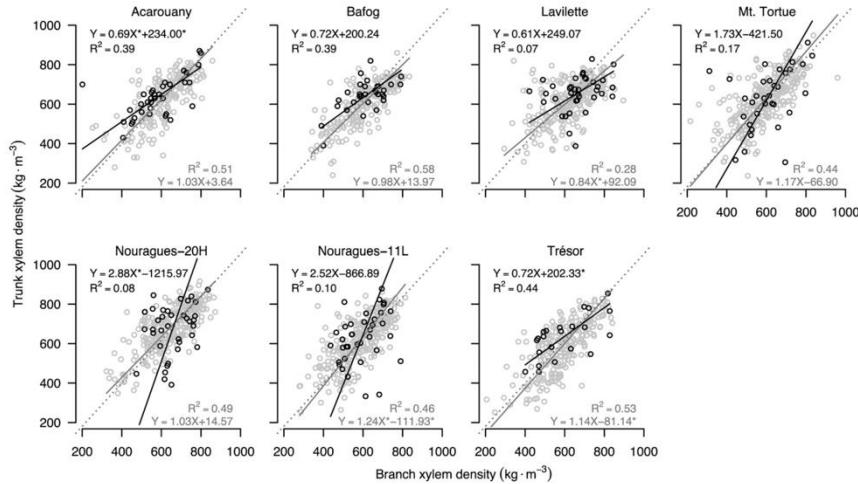
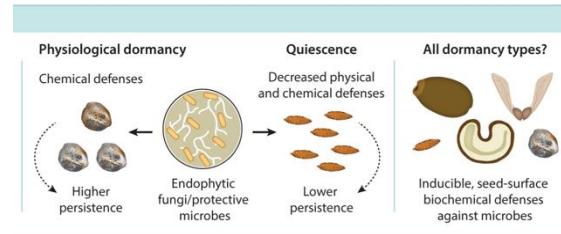
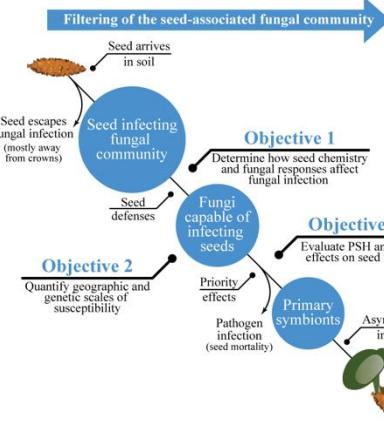


Fig. 3. The major-axis (MA) fit between trunk and branch xylem density in seven rain forest sites from French Guiana. The gray points and solid gray line indicate the site-level relationship, whereas the black points and lines indicate individuals of Fabaceae in each site. The dotted line represents the MA-fit line for all 1909 samples together. Asterisks indicate significance as in Fig. 1.

Sarmiento et al., 2011 (AJB)



Dalling et al., 2020 (AREES)



Zalamea et al., 2023 NSF grant (awarded)

Exploratory Data Analysis

“In Statistics [~Data Science], EDA is an **approach** to analyzing data sets to summarize their main **characteristics**, often with visual methods.”



- Discover patterns
- Extract important variables
- Detect outliers and anomalies
- Check assumptions



Start working now!

Data exploration starts before going to the field/lab.

Think about your questions:

- * Possible answers?
- * What are you going to measure?
- * How?
- * Statistical analysis?

Record your data (basics)

Excel spreadsheets are an easy way to enter and record your data



Very popular (lots of users)

Easy to access

Some useful/easy tools



Easy to mess up your data

Not a statistical software

Not very flexible (advanced analyses)

Record your data (basics)



- * Be consistent and organized
(with files, naming variables, etc.)

Record your data (basics)



- * Be consistent and organized
(with files, naming variables, etc.)
- * Keep your spreadsheets simple
(think saving as .CSV)

Be careful!

Record your data (basics)



- * Be consistent and organized
(with files, naming variables, etc.)

- * Keep your spreadsheets simple

- * Record one "thing" per column

Record your data (basics)



- * Be consistent and organized
(with files, naming variables, etc.)

- * Keep your spreadsheets simple

- * Record one "thing" per column

- * Do not leave empty cells (0?, NA?)

Record your data (basics)



- * Be consistent and organized
(with files, naming variables, etc.)
- * Keep your spreadsheets simple
- * Record one "thing" per column
- * Do not leave empty cells (0?, NA?)

Record your data (basics)



- * Be consistent and organized
(with files, naming variables, etc.)
- * Keep your spreadsheets simple
- * Record one "thing" per column
- * Do not leave empty cells (0?, NA?)
- * No special characters (or spaces)
- * Name variables in an easy,
clear way (Think R)

Record your data (basics)



- * Be consistent and organized
(with files, naming variables, etc.)
- * Keep your spreadsheets simple
- * Record one "thing" per column
- * Do not leave empty cells (0?, NA?)
- * No special characters (or spaces)
- * Name variables in an easy,
clear way (Think R)
- * Record metadata
(another spreadsheet)

	A	B	C	D	E	F	G	H	I	J	K	L
1	TANK	1	2	3	4	5	6	7	8	9	10	11
2	[ATZ]	50	500	5	50	5	0	50	0	0	500	5
3	rep	2	4	6	6	2	6	1	5	1	6	4
4	Date	APRIL 24, 2002										
5	MASS	0.1568	0.0827	0.0641	0.0739	0.1433	0.0695	0.1708	0.1456	0.1727	0.1427	0.0723
6	DATE	8-May-02										
7	mass	0.3938	0.2044	0.1873	0.2547	0.4641	0.1871	0.3750	0.3481	0.3660	0.3758	0.2256
8	date	22-May-02										
9	Mass/larva	0.7551	0.4876	0.4102	0.4898	0.9739	0.3160	0.7240	0.6678	0.6855	0.6832	0.4418
10												
11	Density	24										
12		48										

A	B	C	D	E	F
Tank #	Atrazine Conc. (ppb)	Density (48=48, 24=24)	Rep./Block	Date	Mass/larva
1	50	24	2	24-Apr-02	0.1568
2	500	48	4	24-Apr-02	0.0827
3	5	48	6	24-Apr-02	0.0641
4	50	48	6	24-Apr-02	0.0739
5	5	48	2	24-Apr-02	0.1433
6	0	48	6	24-Apr-02	0.0695
7	50	24	1	24-Apr-02	0.1708
8	0	24	5	24-Apr-02	0.1456
9	0	24	1	24-Apr-02	0.1727
10	500	24	6	24-Apr-02	0.1427
11	5	48	4	24-Apr-02	0.0723
12	50	48	5	24-Apr-02	0.0800
13	5	24	6	24-Apr-02	0.1308
14	0	48	4	24-Apr-02	0.0697
15	500	24	3	24-Apr-02	0.1511
16	500	24	1	24-Apr-02	0.1384

Which one is better?

	A	B	C	D	E	F	G	H	I	J	K	L
1	TANK	1	2	3	4	5	6	7	8	9	10	11
2	[ATZ]	50	500	5	50	5	0	50	0	0	500	5
3	rep	2	4	6	6	2	6	1	5	1	6	4
4	Date	APRIL 24, 2002										
5	MASS	0.1568	0.0827	0.0641	0.0739	0.1433	0.0695	0.1708	0.1456	0.1727	0.1427	0.0723
6	DATE	8-May-02										
7	mass	0.3938	0.2044	0.1873	0.2547	0.4641	0.1871	0.3750	0.3481	0.3660	0.3758	0.2256
8	date	22-May-02										
9	Mass/larva	0.7551	0.4876	0.4102	0.4898	0.9739	0.3160	0.7240	0.6678	0.6855	0.6832	0.4418
10												
11	Density	24										
12		48										

- ★ One “thing” per column
- ★ Do not use cell color as data
- ★ Consistent format
- ★ Record your metadata
(e.g. mass/larva? units?)

A	B	C	D	E	F
Tank #	Atrazine Conc. (ppb)	Density (48=48, 24=24)	Rep./Block	Date	Mass/larva
1	50	24	2	24-Apr-02	0.1568
2	500	48	4	24-Apr-02	0.0827
3	5	48	6	24-Apr-02	0.0641
4	50	48	6	24-Apr-02	0.0739
5	5	48	2	24-Apr-02	0.1433
6	0	48	6	24-Apr-02	0.0695
7	50	24	1	24-Apr-02	0.1708
8	0	24	5	24-Apr-02	0.1456
9	0	24	1	24-Apr-02	0.1727
10	500	24	6	24-Apr-02	0.1427
11	5	48	4	24-Apr-02	0.0723
12	50	48	5	24-Apr-02	0.0800
13	5	24	6	24-Apr-02	0.1308
14	0	48	4	24-Apr-02	0.0697
15	500	24	3	24-Apr-02	0.1511
16	500	24	1	24-Apr-02	0.1384

	A	B	C	D	E	F	G	H	I	J	K	L
1	TANK	1	2	3	4	5	6	7	8	9	10	11
2	[ATZ]	50	500	5	50	5	0	50	0	0	500	5
3	rep	2	4	6	6	2	6	1	5	1	6	4
4	Date	APRIL 24, 2002										
5	MASS	0.1568	0.0827	0.0641	0.0739	0.1433	0.0695	0.1708	0.1456	0.1727	0.1427	0.0723
	--	--	--	--	--	--	--	--	--	--	--	--

Metadata:

“a set of data that describes and gives information about other data”

Any information that is useful for understanding your variables
 (for others and for you in the future!)

- ★ One “thing” per column
- ★ Do not use cell color as data
- ★ Consistent format
- ★ Record your metadata
 (e.g. mass/larva? units?)

A	B	C	D	E	F
Tank #	Atrazine Conc. (ppb)	Density (48=48, 24=24)	Rep./Block	Date	Mass/larva
1	50	24	2	24-Apr-02	0.1568
2	500	48	4	24-Apr-02	0.0827
3	5	48	6	24-Apr-02	0.0641
4	50	48	6	24-Apr-02	0.0739
5	5	48	2	24-Apr-02	0.1433
6	0	48	6	24-Apr-02	0.0695
7	50	24	1	24-Apr-02	0.1708
8	0	24	5	24-Apr-02	0.1456
9	0	24	1	24-Apr-02	0.1727
10	500	24	6	24-Apr-02	0.1427
11	5	48	4	24-Apr-02	0.0723
12	50	48	5	24-Apr-02	0.0800
13	5	24	6	24-Apr-02	0.1308
14	0	48	4	24-Apr-02	0.0697
15	500	24	3	24-Apr-02	0.1511
16	500	24	1	24-Apr-02	0.1384
17	^	^	^	^	^

Record your data (basics)



Once your dataset is complete:

Save it and work on a copy

.csv only stores the current datasheet
(no colors, no filters, no formulas)

A real-life example:

The “old seeds” project (REU):
A sad story with a happy ending!

[See excel file](#)

A real-life example:

A real-life example:

The happy ending!
2015 - 2023



RESEARCH ARTICLE | Free Access

Decadal survival of tropical pioneer seeds in the soil seed bank is accompanied by fungal infection and dormancy release

Paul-Camilo Zalamea✉, Carolina Sarmiento, A. Elizabeth Arnold, Venus Kuo, Carolyn Delevich,
Adam S. Davis, Thomas A. Brown, James W. Dalling

First published: 14 December 2023 | <https://doi.org/10.1111/1365-2435.14476>

Take an initial look at your dataset



1

Dimensions

- Number of columns (**variables**) looks ok?
- Number of rows (**individuals/experimental units**) looks ok?

Take an initial look at your dataset



1

Dimensions

- Number of columns (**variables**) looks ok?
- Number of rows (**individuals/experimental units**) looks ok?

* Depends on your experiment *

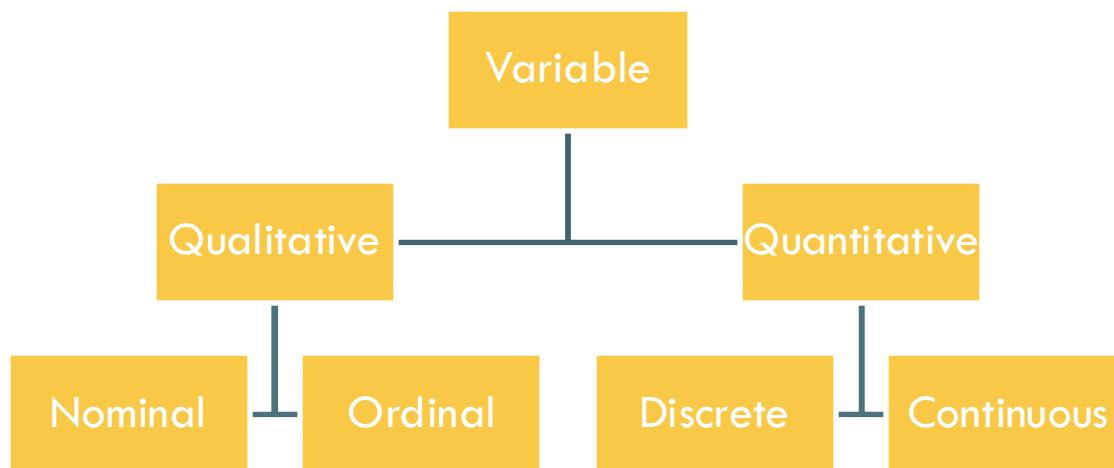
Take an initial look at your dataset



2

Variables

Categorical or continuous?



Take an initial look at your dataset



2 Variables

Qualitative

~ not countable, categorical

Ordinal: If there is an order (e.g. Low - Medium – High)

Nominal: no particular order (e.g. Female - Male)

Take an initial look at your dataset



2 Variables

Quantitative

~Things you can measure (numerical)

Discrete: Things you can count (e.g. species in a plot)

Continuous: ~ infinite (e.g. body mass)

Take an initial look at your dataset



2

Variables

Categorical or continuous?

How many in your dataset?

If categorical, do your categories look OK?

Take an initial look at your dataset



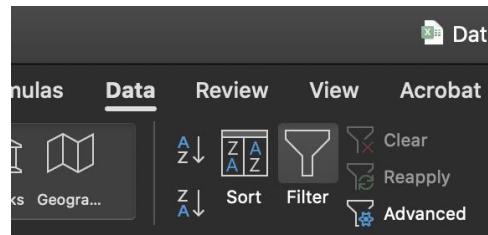
2

Variables

Categorical or continuous?

How many in your dataset?

If categorical, do your categories look OK?



Tip: Use the filter tool

(Mac: command + shift + F)

(Windows: control + shift + L) ?

Take an initial look at your dataset



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
1	fun_code	seed_id	sequence	plate	slant_set	sp_init	species	time	plot	rep					
2	PS0001	AM0_25	CGGGCAGCT	343	1	AM	Apeiba_membranacea	0	NA	NA					
3	PS0002	AM0_67	AGTTAACCA	343	1	AM	Apeiba_membranacea	0	NA	NA					
4	PS0003	AM0_74	GCTTGCGG	343	1	AM	Apeiba_membranacea	0	NA	NA					
5	PS0004	AM0_76	TACCGAGTT	343	1	AM	Apeiba_membranacea	0	NA	NA					
6	PS0005	AM0_77	CTTGGCGGG	398	1	AM	Apeiba_membranacea	0	NA	NA					
7	PS0006	AM0_85	TCGGCGCAG	343	1	AM	Apeiba_membranacea	0	NA	NA					
8	PS0007	AM0_90	GCTTGCGCG	343	1	AM	Apeiba_membranacea	0	NA	NA					
9	PS0008	AM0_93	CTTGGCGGG	343	1	AM	Apeiba_membranacea	0	NA	NA					
10	PS0009	AM0_99	TTGGCGGGC	343	1	AM	Apeiba_membranacea	0	NA	NA					
11	PS0010	AM0_103	ATCGATGAA	343	1	AM	Apeiba_membranacea	0	NA	NA					
12	PS0011	AM0_104	TACCGAGTT	343	1	AM	Apeiba_membranacea	0	NA	NA					
13	PS0012	AM0_108	TTGGCGGGC	343	1	AM	Apeiba_membranacea	0	NA	NA					
14	PS0013	AM0_109	TTATAGGTC	343	1	AM	Apeiba_membranacea	0	NA	NA					
15	PS0014	AM0_110	TACCGAGTT	343	1	AM	Apeiba_membranacea	0	NA	NA					
16	PS0017	AM0_123	TTTGGCGGGC	343	1	AM	Apeiba_membranacea	0	NA	NA					
17	PS0018	AM0_125	CCCGCTCCCG	343	1	AM	Apeiba_membranacea	0	NA	NA					
18	PS0019	AM0_126	CTTGGCGGG	343	1	AM	Apeiba_membranacea	0	NA	NA					
19	PS0020	AM0_138	TTATCGTGG	350	1	AM	Apeiba_membranacea	0	NA	NA					
20	PS0022	AM1A1_2	TACCGAGTT	343	1	AM	Apeiba_membranacea	1	A		1				
21	PS0024	AM1A3_1	AACTCCAACCG	350	1	AM	Apeiba_membranacea	1	A		3				
22	PS0025	AM1A3_5	GGATCATTA	343	1	AM	Apeiba_membranacea	1	A		3				
23	PS0028	AM1D2_2	TTCGGGCTT	343	1	AM	Apeiba_membranacea	1	D		2				
24	PS0031	AM1H1_2	GAGTTATCA	343	1	AM	Apeiba_membranacea	1	H		1				
25	PS0035	AM1H3_1	TATTACCCCTT	343	1	AM	Apeiba_membranacea	1	H		3				
26	PS0036	AM1H3_10	GTATAGGTT	343	1	AM	Apeiba_membranacea	1	H		3	NA	0	23	18
27	PS0037	AM1H3_4	TTACCCACTT	343	1	AM	Apeiba_membranacea	1	H		3	NA	0	24	19

Sort

By color: None

Filter

By color: None

Choose One

- (Select All)
- Annona_spraguei
- Apeiba_membranacea
- Apeiba_membranacea
- Apeiba_tibourbou
- Cecropia_insignis
- Cecropia_longipes
- Cecropia_noltii

Clear Filter

Take an initial look at your dataset



A screenshot of a spreadsheet application showing a dataset. The table has columns labeled A through N. Column G is titled 'species' and contains several entries. A red arrow points from the text below to the filter panel on the right, which shows a list of species names with checkboxes. Four checkboxes are checked and highlighted with a red box: 'Annona_spraguei', 'Apeiba_membranacea', 'Apeiba_membranacea', and 'Apeiba_tibourbou'. Other species listed include Cecropia_insignis, Cecropia_longipes, and Cecropia_noltii.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	fun_code	seed_id	sequence	plate	slant_set	sp_init	species	time	plot	rep				
2	PS0001	AM0_25	CGGGCAGCT	343	1	AM	Apeiba_men	0	NA	NA				
3	PS0002	AM0_67	AGTTAACCA	343	1	AM	Apeiba_men	0	NA	NA				
4	PS0003	AM0_74	GCTTGGCG	343	1	AM	Apeiba_men	0	NA	NA				
5	PS0004	AM0_76	TACCGAGTT	343	1	AM	Apeiba_men	0	NA	NA				
6	PS0005	AM0_77	CTTTGGCGG	398	1	AM	Apeiba_men	0	NA	NA				
7	PS0006	AM0_85	TCGGCGCAG	343	1	AM	Apeiba_men	0	NA	NA				
8	PS0007	AM0_90	GCTTGGCG	343	1	AM	Apeiba_men	0	NA	NA				
9	PS0008	AM0_93	CTTTGGCGG	343	1	AM	Apeiba_men	0	NA	NA				
10	PS0009	AM0_99	TTGGCGGC	343	1	AM	Apeiba_men	0	NA	NA				
11	PS0010	AM0_103	ATCGATGAA	343	1	AM	Apeiba_men	0	NA	NA				
12	PS0011	AM0_104	TACCGAGTT	343	1	AM	Apeiba_men	0	NA	NA				
13	PS0012	AM0_108	TTGGCGGC	343	1	AM	Apeiba_men	0	NA	NA				
14	PS0013	AM0_109	TTATAGTC	343	1	AM	Apeiba_men	0	NA	NA				
15	PS0014	AM0_110	TACCGAGTT	343	1	AM	Apeiba_men	0	NA	NA				
16	PS0017	AM0_123	TTGGCGGC	343	1	AM	Apeiba_men	0	NA	NA				
17	PS0018	AM0_125	CCCGCTCCCG	343	1	AM	Apeiba_men	0	NA	NA				
18	PS0019	AM0_126	CTTTGGCGG	343	1	AM	Apeiba_men	0	NA	NA				
19	PS0020	AM0_138	TTATCGTGG	350	1	AM	Apeiba_men	0	NA	NA				
20	PS0022	AM1A1_2	TACCGAGTT	343	1	AM	Apeiba_men	1	A	1				
21	PS0024	AM1A3_1	AACTCCAAC	350	1	AM	Apeiba_men	1	A	3				
22	PS0025	AM1A3_5	GGATCATTA	343	1	AM	Apeiba_men	1	A	3				
23	PS0028	AM1D2_2	TTCGGGCTT	343	1	AM	Apeiba_men	1	D	2				
24	PS0031	AM1H1_2	GAGTTATCA	343	1	AM	Apeiba_men	1	H	1				
25	PS0035	AM1H3_1	TATTACCCCTT	343	1	AM	Apeiba_men	1	H	3				
26	PS0036	AM1H3_10	GTTATAGGT	343	1	AM	Apeiba_men	1	H	3	NA	0	23	18
27	PS0037	AM1H3_4	TTACCCACT	343	1	AM	Aapeiba_men	1	II	2	NA	0	24	19

This typo will introduce two different categories for the same species!

Take an initial look at your dataset



3 Values

Do not leave empty cells

Take an initial look at your dataset



3 Values

Do not leave empty cells

time	plot	rep	mat_sou	viability	rep
0	NA	NA		1	1
0	NA	NA		2	0
0	NA	NA		2	1
0	NA	NA		2	1
0	NA	NA		0	0
0	NA	NA		2	0
0	NA	NA		2	0
0	NA	NA		2	0
0	NA	NA		2	0
0	NA	NA		2	0
0	NA	NA		2	0
0	NA	NA		2	1
0	NA	NA		2	1
0	NA	NA		2	1
0	NA	NA		2	0
0	NA	NA		2	0
0	NA	NA		2	0
0	NA	NA		2	1
1	A		1	NA	
1	A		3	NA	
1	A		3	NA	
1	D		2	NA	
1	H		1	NA	
1	H		3	NA	

Fill them with
“NA” if it is a
true missing
value

Take an initial look at your dataset



3 Values

Do not leave empty cells

Do your values make sense?

Take an initial look at your dataset



3 Values

Do not leave empty cells

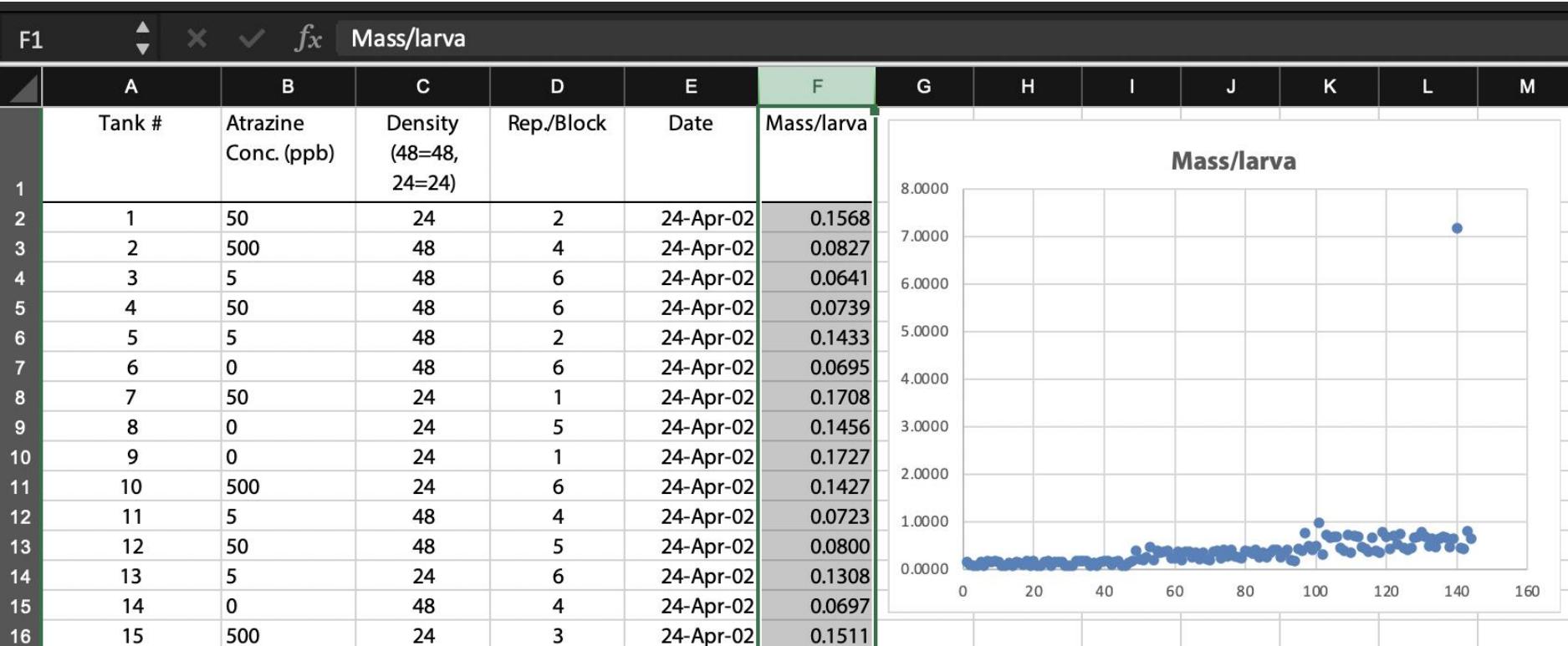
Do your values make sense?



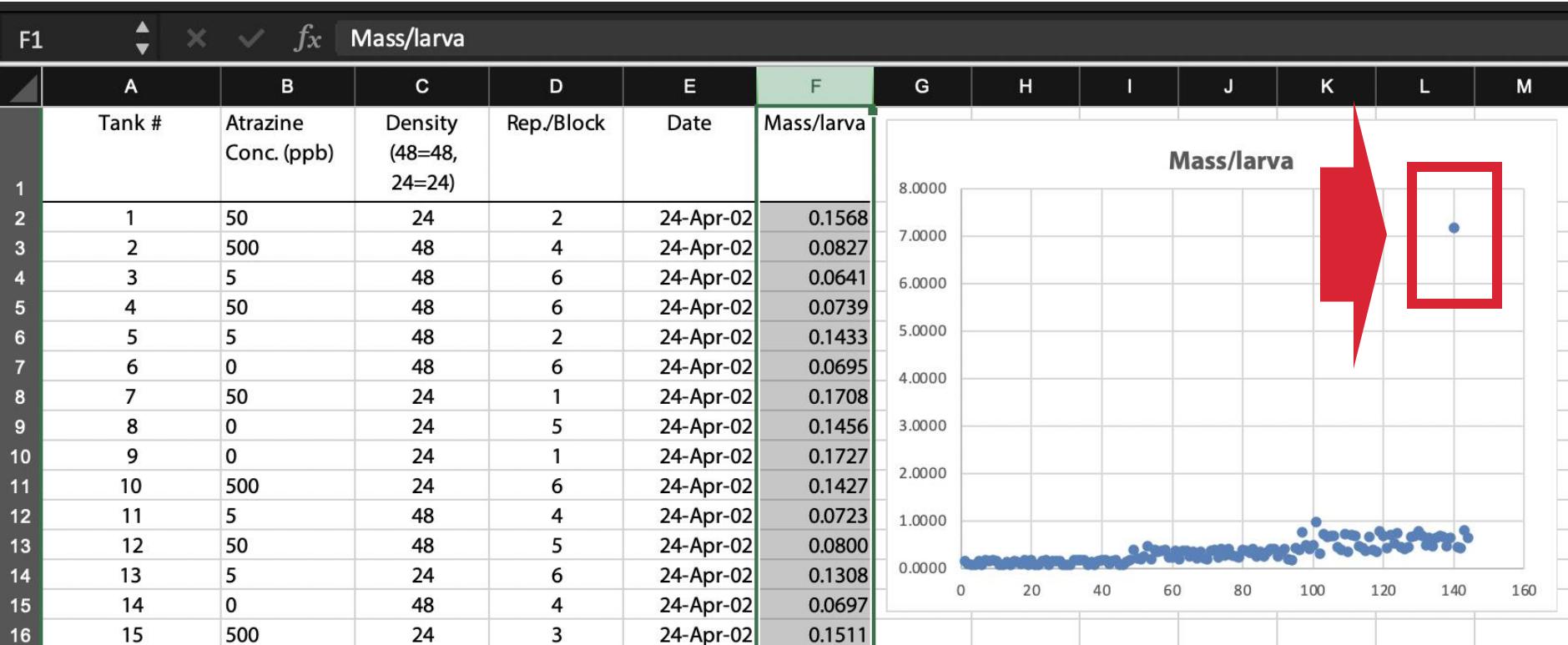
Tricky question!

Typing mistakes? Outliers? Duplicated values?

Simple visualizations are useful!

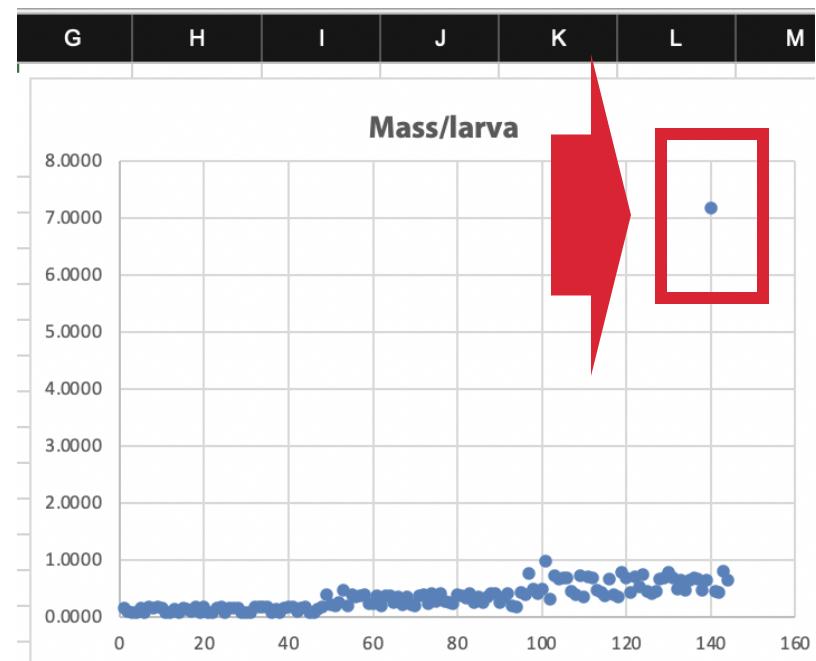


Simple visualizations are useful!



Simple visualizations are useful!

- Go back to your lab/field notes
- Correct – only if you are 100% sure it is a typo
- Do not change it if you are not sure
- Is it a measurement error?
Make a note, but do not erase it



Take an initial look at your dataset



3 Values

Do not leave empty cells

Do your values make sense?



Tricky question!

Duplicated values?

Tip: Use conditional formatting

Take an initial look at your dataset



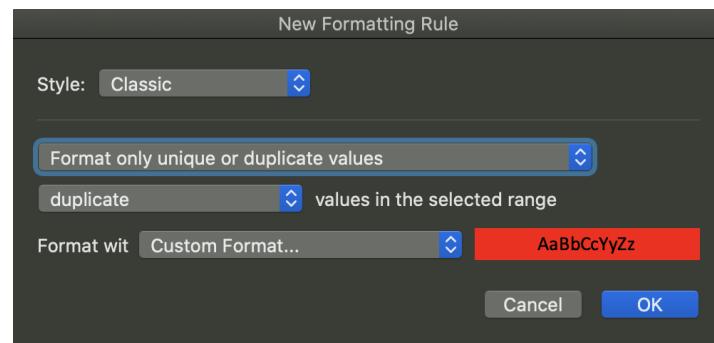
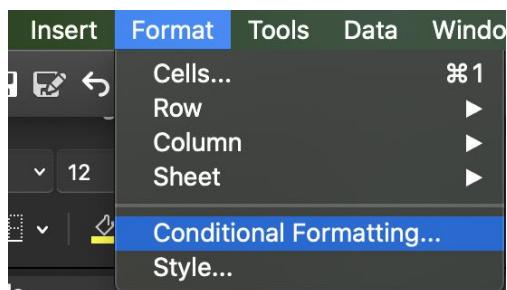
3 Values

Do not leave empty cells

Do your values make sense?



Tricky question!



PS0037	AM1H3_4	TTACCGAGT	343
PS0038	AM1H4_10	TTATCACAAC	394
PS0041	AM1P3_1	ACAACGGAT	402
PS0041	AM1P3_1	GC GGAGGG	625
PS0042	AM1P3_2	TGCTTGCG	343

Take an initial look at your dataset



3 Values

Do some basic calculations, to help you understand your data:

- Max. value [= `max()`]
- Min. value [= `min()`]
- Mean [= `average()`]
- Standard deviation [= `stdev()`]

Take an initial look at your dataset



4

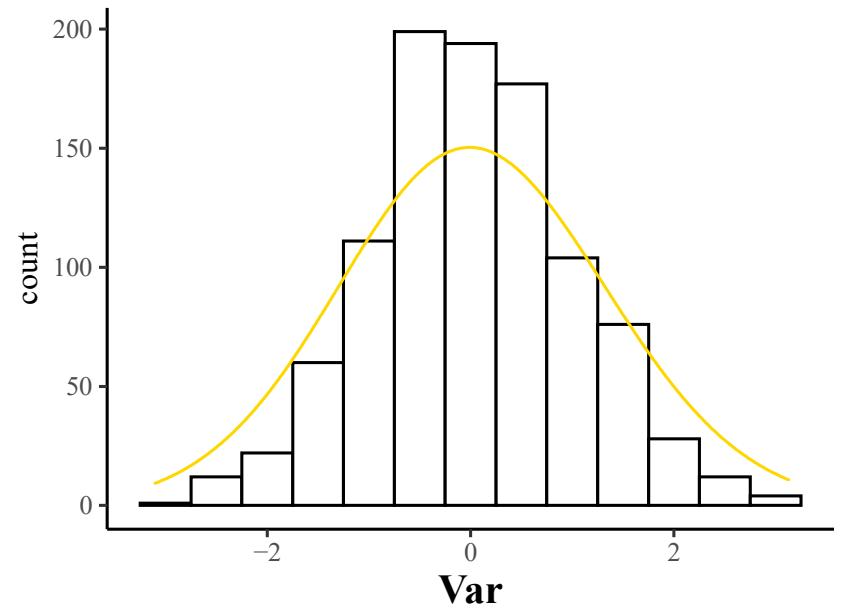
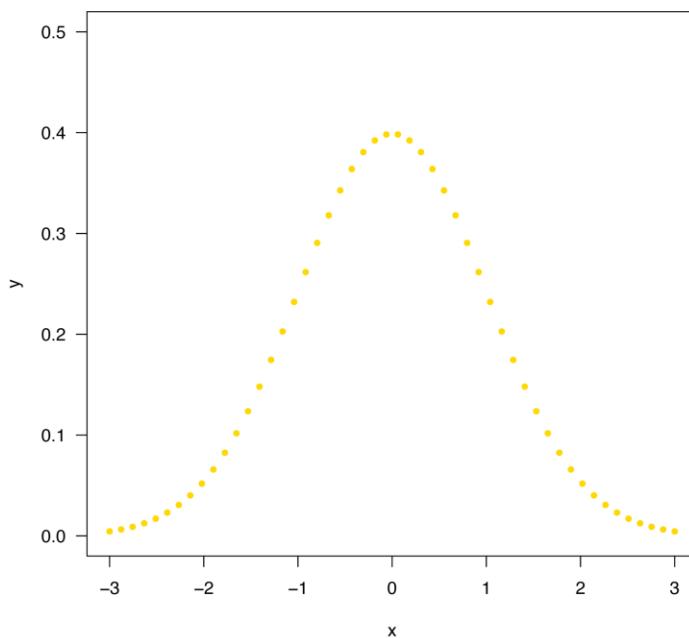
Distribution of data:

- Is a *function* that shows all the possible values of a variable - it also tells you how often each value occurs.
The shape of the graph when all possible values are plotted on a frequency graph or frequency distribution histogram
- Hypothesis tests are based on specific distributional assumptions (often, does the variable follow a normal distribution?)
- Needs to be verified so that the statistical technique yields valid conclusions.

Take an initial look at your dataset



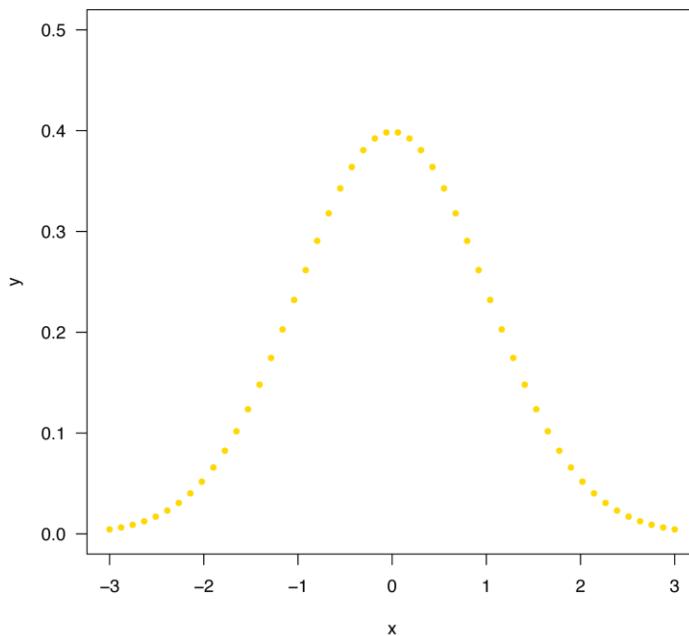
Normal Distribution (Gaussian):



Take an initial look at your dataset



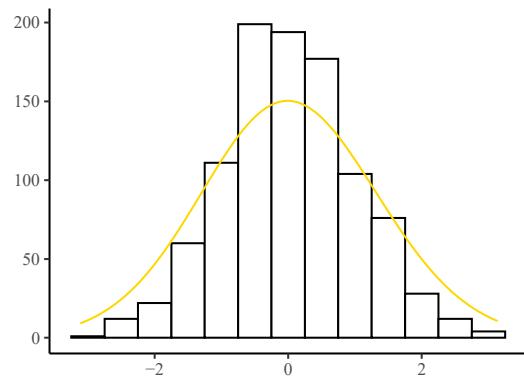
Normal Distribution (Gaussian):



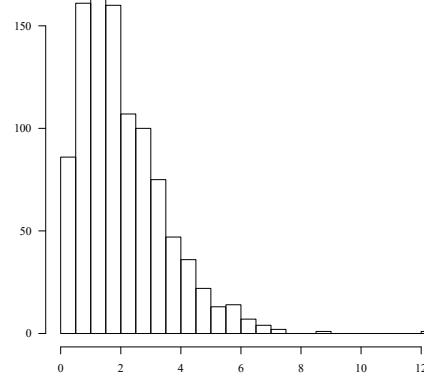
- Mean, mode and median coincide with each other
- The distribution has a bell-shaped distribution curve
- The distribution curve is symmetrical to the center
- The area under the curve is equal to 1

Continuous distributions:

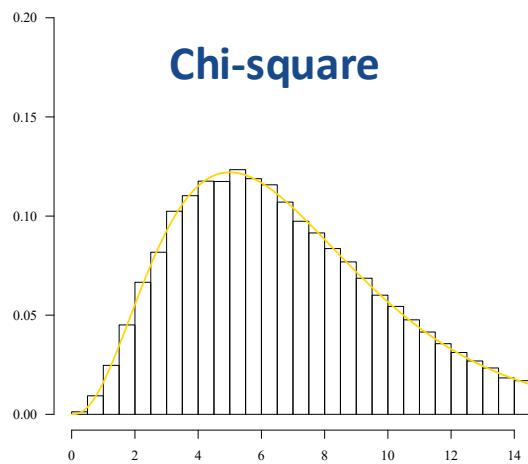
Normal



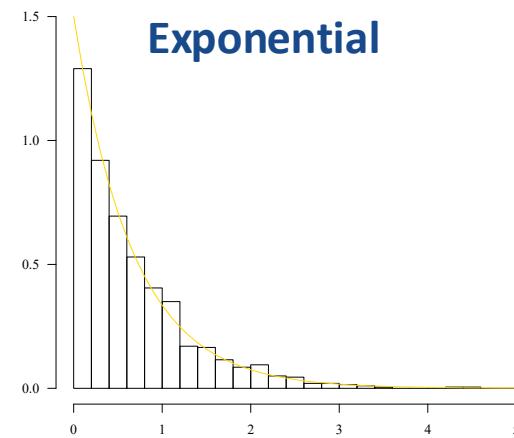
Gamma



Chi-square

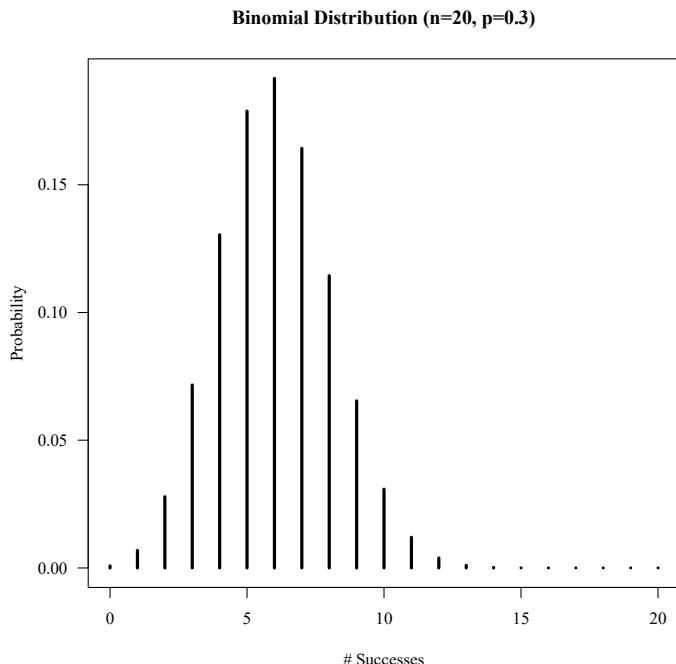


Exponential

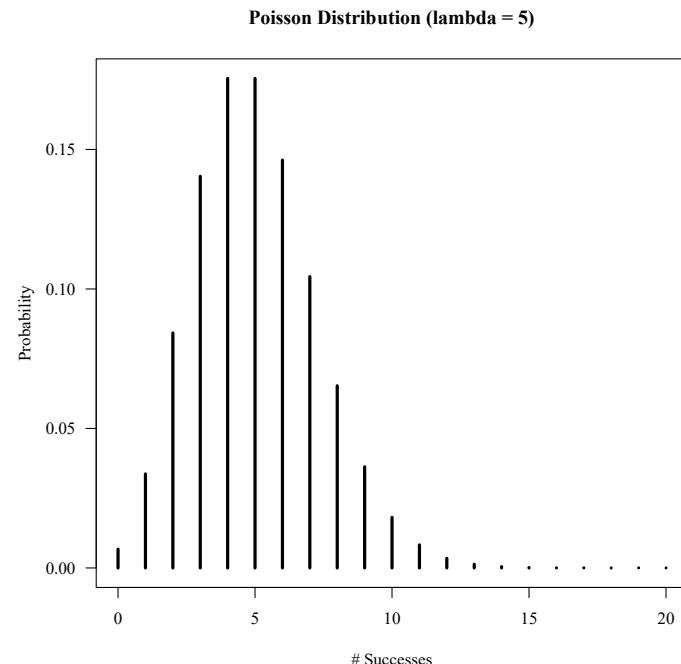


Discrete distributions:

Binomial



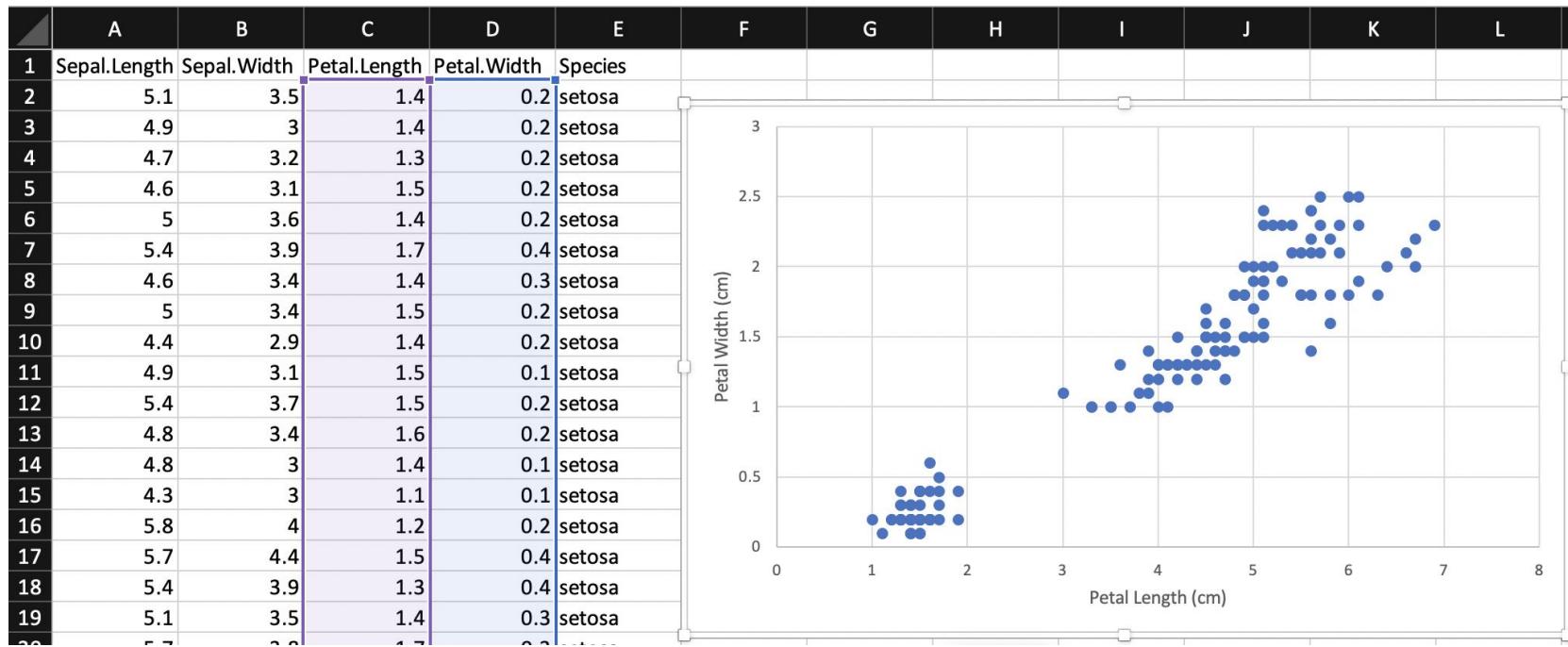
Poisson



Start looking for patterns



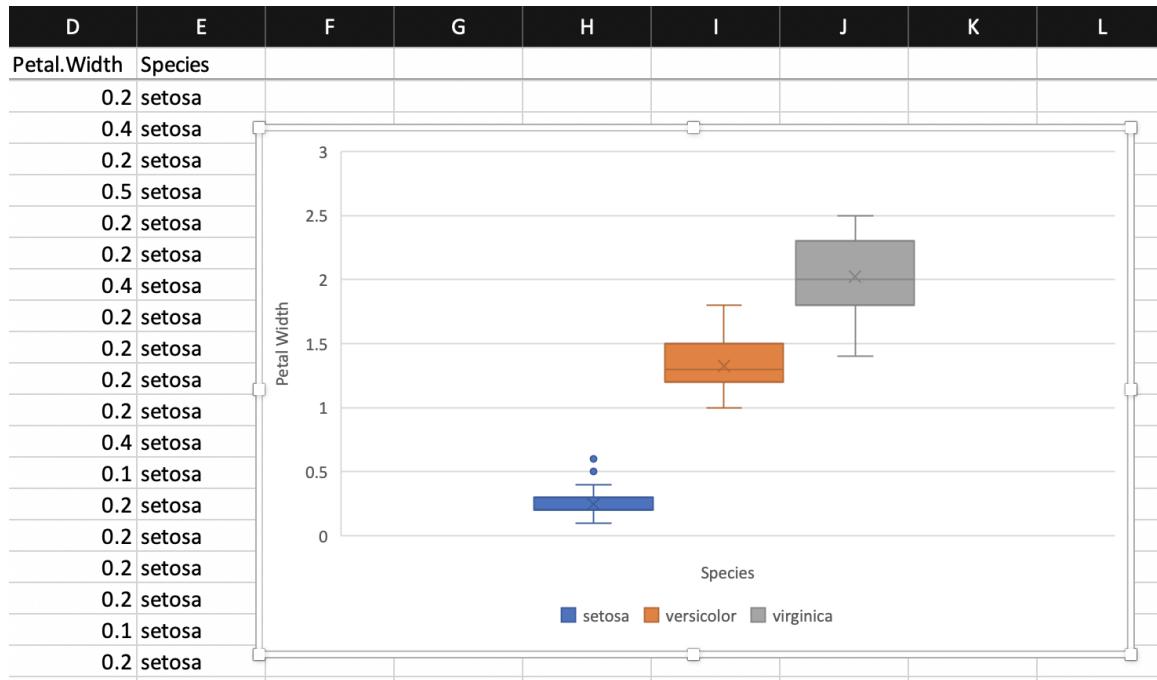
Are some of your variables correlated?



Start looking for patterns

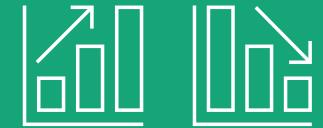


Are there (visual) differences among
your treatments/species?



Start looking for patterns

(practice at home 😊)



1. Are there changes in larval mass through time?
2. Does the atrazine concentration affect the size of the larvae?
3. Does the larval density in the tanks affect larval size?
4. Which variable looks more normal - larval mass or log-transformed larval mass? [Hint](#)

Use only plots (Excel charts)

Start looking for patterns

(practice at home 😊)



Iris versicolor



Iris setosa



Iris virginica

1. Which species has the largest flowers?
2. Which floral structure is smaller, petals or sepals?
3. Which variable looks more normal?

Use only Excel plots (charts)

Introduction to

Carolina Sarmiento
USF
October 10, 2025



What is R?

“R is a **language** and **environment** for statistical computing and graphics.”

- “To use it well, one must master its grammar and vocabulary.”
- It allows users to define new functions (= add additional functionality) – as opposed to very specific and inflexible tools offered by other data analysis software.
- “Provides a wide variety of statistical and graphical techniques and is highly extensible.”

It's a **free** and **open-source** software

Open-source?

Software for which the original source code is made freely available and may be redistributed and modified

- Provides full access to algorithms and their implementation
- Allows researchers to explore and expand the methods used to analyze data
- Promotes reproducible research by providing open and accessible tools
- Most of R is written in... R! This makes it easier to see what functions are actually doing

Why use R?

- It's **free** (and it won't change)
- State of the art – cutting edge **packages** are well ahead of commercial software
[Publication example](#) or [list of packages by date](#)
- Great **graphic** capabilities
- Active user **community** – Good online resources
- Forces you to **think** about analyses
- R is **stable**. It will continue to exist, even if programmers/users stop adding to it. Therefore, unlikely to go away.
- You can get it to **do exactly what you need** it to do:
Probably someone else has already done most of the work for you
- Almost every analytical **tool** you can think of is available (or will be soon)

But some disadvantages...

- **Not user friendly:** Steep learning curve,
minimal GUI (graphical user interface)
- Quality control – depends on **volunteer** programmers
- No commercial support: You must figure out how to use a function or correct
methods to implement it (it can be **frustrating**)
- Help files are **not user-friendly**: Learning how to read them is very
important

Base R and packages

What is a package?

A collection of functions, documentation, and data sets developed by the community. Packages increase the power of R by improving existing base R functionalities, or by adding new ones.

The R base package contains (many) *basic functions* for R to work:

- arithmetic
- input/output
- basic programming support
- etc.

Other packages uses R base to do many other things:

- Natural Language Processing (~analysis of speech and language) of Korean texts
- Extract weather data from the web
- Estimate actual evapotranspiration using land surface energy balance models

Packages in CRAN:

16,395 (2020) / 18,287 (2021) / 18,689 (2022) / 19,955 (2023) / 21,594 (2024) /
22,859 (2025)

Repositories

The location where the packages are stored
(typically, online and accessible to everyone)

The 3 most popular:

1. CRAN (Comprehensive R Archive Network)

- Network of servers maintained by the R community
- R foundation coordinates it
- For a package to be published here, it needs to pass several tests to ensure it follows CRAN policies

2. Bioconductor:

- Topic-specific repository, intended for open-source software for bioinformatics.
- It has its own submission and review processes
- Very active community (several conferences and meetings per year)

3. GitHub:

- Not R specific
- Probably the most popular repository for open-source projects
- Easy to share and collaborate with others
- No review process associated with it

Let's get into it!

Take a look:



R version 3.6.1 (2019-07-05) -- "Action of the Toes"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.70 (7684) x86_64-apple-darwin15.6.0]

[Workspace restored from /Users/carolinasmarmiento/.RData]
[History restored from /Users/carolinasmarmiento/.Rhistory]

> |

A screenshot of the RStudio IDE. The top menu bar includes File, Edit, Format, Workspace, Packages & Data, Misc, Window, and Help. The status bar shows the date (Sep 29), time (11:02), and system name (Carolina). The main window has two panes: the left pane is the R Console displaying the startup message and basic information about the R environment; the right pane is titled 'Untitled' and contains a 'functions' tab. The RStudio interface is set against a dark background.

Let's get into it!

Console



A screenshot of the RStudio interface. The top menu bar includes Apple, R, File, Edit, Format, Workspace, Packages & Data, Misc, Window, and Help. The bottom status bar shows the date (Sep 29), time (11:02), and user (Carolina). The left pane is titled "R Console" and displays the R startup message, including the version (3.6.1), license information, and help instructions. The right pane is titled "Untitled" and contains a single line of code: "1 <functions>". A magenta box highlights the "demo()", "help()", and "citation()" lines in the console, and another magenta box highlights the "demo()", "help()", and "citation()" lines in the text editor.

```
R version 3.6.1 (2019-07-05) -- "Action of the Toes"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.70 (7684) x86_64-apple-darwin15.6.0]

[Workspace restored from /Users/carolinasmarmiento/.RData]
[History restored from /Users/carolinasmarmiento/.Rhistory]
```

Text Editor



Let's get into it!

Console

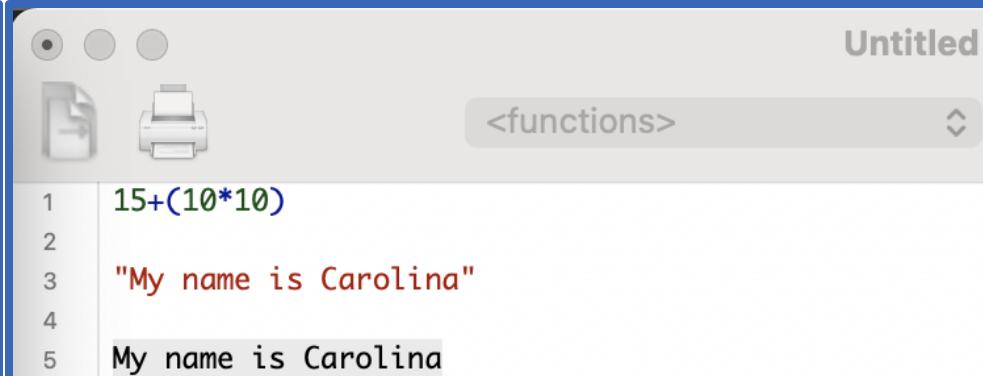


```
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.
```

```
[R.app GUI 1.70 (7735) x86_64-apple-darwin15.6.0]
```

```
> 2+2  
[1] 4  
> Carolina  
Error: object 'Carolina' not found  
> |
```

Text Editor



The screenshot shows a Text Editor window titled "Untitled". The toolbar includes icons for file, print, and other functions. The status bar shows "<functions>". The main area displays the following code and its results:

```
1 15+(10*10)  
2  
3 "My name is Carolina"  
4  
5 My name is Carolina
```



Run line by line (cursor)
or
Highlight text to run

Mac: command + Enter
Win: ctrl + Enter

1. Set up the Working Directory

R version 4.2.0 (2022-04-22) -- "Vigorous Calisthenics"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.

Change Working Directory... ⌘D
Reset Working Directory
Get Working Directory
Run X11 Server

Intro_R_GradSkills_7oct2022.R

```
1 # Introduction to R
2 #By Carolina Sarmiento
3 #USF - Grad Skills - Oct. 7, 2022
4
5 #####
6 # LET'S GET STARTED # -----
7 #####
```

2. Install packages

R version 4.2.0 (2022-04-22) -- "Vigorous Calisthenics"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.

Package Manager
Package Installer ⌘I
Data Manager

Help Search

Intro_R_GradSkills_7oct2022.R

```
1 # Introduction to R
2 #By Carolina Sarmiento
3 #USF - Grad Skills - Oct. 7, 2022
4
5 #####
6 # LET'S GET STARTED # -----
7 #####
```

3. Run your script

R version 4.2.0 (2022-04-22) -- "Vigorous Calisthenics"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.

R Console

```
> # Introduction to R
> #By Carolina Sarmiento
> #USF - Grad Skills - Oct. 7, 2022
>
> #####
> # LET'S GET STARTED # -----
> #####
>
> # 1. Set your Working Directory (WD) ----
> #You *ALWAYS* have to set your WD (i.e., tell R where - in your computer - are the files you will need and where it can store some of the objects when you ask R to do so)
>
> #A. Check current WD by running the following line
> getwd()
[1] "/Users/carolina"
```

Help Search

Intro_R_GradSkills_7oct2022.R

```
1 # Introduction to R
2 #By Carolina Sarmiento
3 #USF - Grad Skills - Oct. 7, 2022
4
5 #####
6 # LET'S GET STARTED # -----
7 #####
8
9 # 1. Set your Working Directory (WD) ----
10 #You *ALWAYS* have to set your WD (i.e., tell R where
11 #ask R to do so)
12
13 #A. Check current WD by running the following line
14 getwd()
15
16 #B. set your WD: Click on the drop-down menu "Session"
17
18 # **** I suggest you copy the line that just appeared
```



Open-source environment for working with R:

“RStudio integrates the tools you use with R into a single environment”

Includes graphical interface tools for many common R procedures

Includes:

- console
- syntax-highlighting editor
- tools for plotting
- R help
- history
- debugging
- workspace management



R Studio®

Project: (None)

Quiescent_germ_fungal_8Aug2018... x dada_phyllosea_JCexp.R x Intro_R_GradSkills_7Oct2022.R x

Source | Save | Run | Up | Down | Source

```
1 # Introduction to R-
2 #By Carolina Sarmiento-
3 #USF - Grad Skills - Oct. 7, 2022
4 -
5 #*****#
6 *# LET'S GET STARTED # -----
7 #*****#
8 -
9 *# 1. Set your Working Directory (WD) -----
10 #You *ALWAYS* have to set your WD (i.e., tell R where -- in your computer -- are the files you will need and where it can store some of the objects when you ask R to do so).
11 -
12 #A. Check current WD by running the following line-
13 getwd()-
14 -
15 #B. set your WD: Click on the drop-down menu "Session" --> Set Working Directory -> Choose Directory-
16 -
17 # **** I suggest you copy the line that just appeared in the Console to your script, for next time you open your script, just run that line.-
18 ****-
19 setwd("~/Desktop/Grad_Skills/R") # copy yours here!-
20 -
21 -
22 *# 2. Install a package -----
23 -
24 # A. Install the package-
25 # Use first the "Install" button on the Packages tab, look for the package of interest (dslabs) and make sure to check "install dependencies"-
```

3:34 (Top Level) : R Script

Console Terminal Background Jobs

R 4.2.0 · ~/

```
R version 4.2.0 (2022-04-22) -- "Vigorous Calisthenics"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

Warning: macOS is reporting that you have not yet agreed to the Xcode license.
This can occur if Xcode has been updated or reinstalled (e.g. as part of a macOS update).
Some features (e.g. Git / SVN) may be disabled.

Please run:
sudo xcodebuild -license accept

in a terminal to accept the Xcode license, and then restart RStudio.
```

>

Environment History Connections

Import Dataset 54 MB

R Global Environment

Environment is empty

Files Plots Packages Help Viewer Presentation

Home Find in Topic

R Resources

- Learning R Online
- CRAN Task Views
- R on StackOverflow
- Getting Help with R

RStudio

- RStudio IDE Support
- RStudio Community Forum
- RStudio Cheat Sheets
- RStudio Tip of the Day
- RStudio Packages
- RStudio Products

Manuals

- An Introduction to R
- Writing R Extensions
- R Data Import/Export

Reference

Packages

Miscellaneous Material

Search Engine & Keywords

About R License NEWS Authors FAQ User Manuals Resources Thanks Technical papers



Studio®

```
1 # Introduction to R
2 #By Carolina Sarmiento
3 #USF - Grad Skills - Oct. 7, 2022
4
5 #*****
6 # LET'S GET STARTED ! -----
7 #*****
8 -
9 # 1. Set your Working Directory (WD) -----
10 #You *ALWAYS* have to set your WD (i.e., tell R where - in your computer - are the files you will need and where it can store some of the objects when you ask R to do so).
11 -
12 #A. Check current WD by running the following line.
13 getwd()
14 -
15 #B. set your WD: Click on the drop-down menu "Session" and choose "Set Working Directory...". Choose Directory.
16 -
17 # **** I suggest you copy the line that just appeared in the Console to your script, for next time you open your script, just run that line.
18 ****
19 setwd("~/Desktop/Grad_Skills/R") # copy yours here!-
20
21
22 # 2. Install a package -----
23
24 # A. Install the package .
25 # Use first the "Install" button on the Packages tab, look for the package of interest (dslabs) and make sure to check "install dependencies"
26
```





Console

```
Console Terminal × Background Jobs ×
R 4.2.0 - ~/ →

R version 4.2.0 (2022-04-22) -- "Vigorous Calisthenics"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

Warning: macOS is reporting that you have not yet agreed to the Xcode license.
This can occur if Xcode has been updated or reinstalled (e.g. as part of a macOS update).
Some features (e.g. Git / SVN) may be disabled.

Please run:

    sudo xcodebuild -license accept

in a terminal to accept the Xcode license, and then restart RStudio.

>
```

A screenshot of the RStudio IDE interface. The top menu bar includes 'Environment', 'History', 'Connections', 'Import Dataset' (54 MB), 'Global Environment', and a search bar. Below the menu is a message 'Environment is empty'. The main workspace is dark gray. At the bottom, there's a navigation bar with tabs for 'Files', 'Plots', 'Packages', 'Help', 'Viewer', and 'Presentation'. A sidebar on the left contains sections for 'R Resources' (Learning R Online, CRAN Task Views, R on StackOverflow, Getting Help with R), 'Manuals' (An Introduction to R, Writing R Extensions, R Data Import/Export), 'Reference' (Search Engine & Keywords), 'Miscellaneous Material' (About R, License, NEWS, Authors, FAQ, User Manuals, Resources, Thanks, Technical papers), and 'Technical papers'. The bottom right corner shows the RStudio logo.



R Studio®

The screenshot shows the R Studio interface. On the left, the Source editor displays a script with various R code comments and instructions. A pink arrow points upwards from the bottom of this panel towards the main title area. On the right, the Data viewer panel shows a message stating "Environment is empty".

```
1 # Introduction to R-
2 #By Carolina Sarmiento-
3 #USF - Grad Skills - Oct. 7, 2022
4 -
5 #*****#
6 *# LET'S GET STARTED # -----
7 #*****#
8 -
9 *# 1. Set your Working Directory (WD) -----
10 #You *ALWAYS* have to set your WD (i.e., tell R where -- in your computer -- are the files you will need and where it can store some of the objects when you ask R to do so).
11 -
12 #A. Check current WD by running the following line-
13 getwd()-
14 -
15 #B. set your WD: Click on the drop-down menu "Session" --> Set Working Directory -> Choose Directory-
16 -
17 # **** I suggest you copy the line that just appeared in the Console to your script, for next time you open your script, just run that line.-
18 ****-
19 setwd("~/Desktop/Grad_Skills/R") # copy yours here!-
20 -
21 -
22 # 2. Install a package -----
23 -
24 # A. Install the package-
25 # Use first the "Install" button on the Packages tab, look for the package of interest (dslabs) and make sure to check "install dependencies"-
3:34 (Top Level) :
```

The screenshot shows the R Studio sidebar and footer. The sidebar includes links to R Resources (Learning R Online, CRAN Task Views, R on StackOverflow, Contributing to R, Citation Help), Manuals (An Introduction to R, Writing R Extensions, R Data Import/Export), Reference (The R Language Definition, R Installation and Administration, R Internals), Packages, and Miscellaneous Material. The footer contains links to About R, License, NEWS, Authors, FAQ, User Manuals, Resources, Thanks, and Technical papers.

Source editor (your script) and Data viewer



R Studio®

Environment



Shows the objects (i.e., data frames, arrays, values, and functions) you have in your workspace

```
Quiescent_germ_fungal_8Aug2018... R dada_phylose_JCexp.R Intro.R.GradSkills.7Oct2022.R
Source | Save | Run | Import Dataset | 54 MB | Source

1 # Introduction to R
2 #By Carolina Sarmiento
3 #USF - Grad Skills - Oct. 7, 2022
4 ...
5 #*****#
6 # LET'S GET STARTED # -----
7 #*****#
8 ...
9 # 1. Set your Working Directory (WD) ----
10 # You *ALWAYS* have to tell R where to look for files in order to run the code you will need and where to save some of the
11 # objects when you ask R to do so.
12 #A. Check current WD by running the following line:
13 getwd()
14 ...
15 #B. set your WD: Click on the drop-down menu "Session" --> Set Working Directory --> Choose Directory.
16 ...
17 # *** If you are running this script for the first time, just click "Set Working Directory" and you can run the next command in this script, just run that line.
18 ...
19 setwd("C:/Users/Carolina/Desktop/Grad_Skills/R") # copy yours here!
20 ...
21 ...
22 # 2. Install packages
23 ...
24 # A. Install the package
25 # Use first the "Install" button on the Packages tab, look for the package of interest (dsolve) and make sure to check "install dependencies".
334 (Top) R Script

Console Terminal Background Jobs
R 4.2.0 ~ / ...
R version 4.2.0 (2022-04-18 uc Berkeley Linux)
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

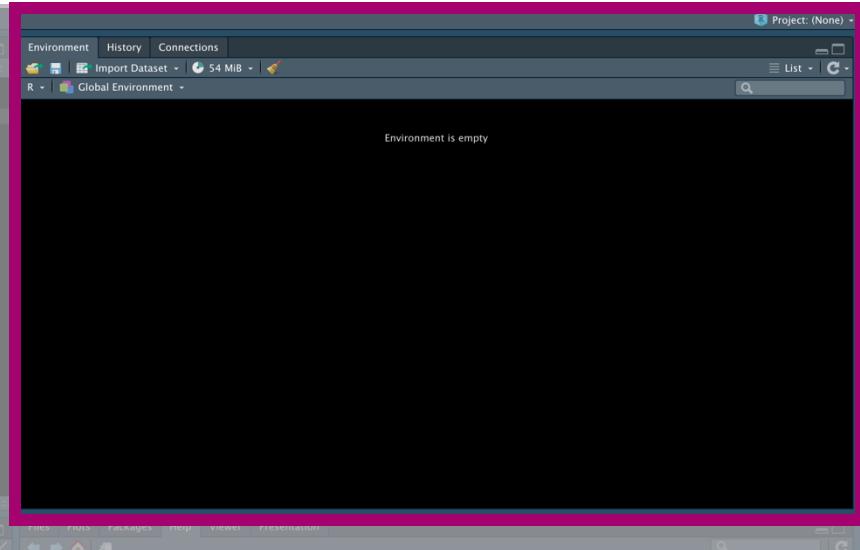
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

Warning: macOS is reporting that you have not yet agreed to the Xcode license.
This can occur if Xcode has been updated or reinstalled (e.g. as part of a macOS update).
Some features (e.g. Git / SVN) may be disabled.

Please run:
sudo xcodebuild -license accept

in a terminal to accept the Xcode license, and then restart RStudio.
```



Project: (None)

Environment History Connections

Import Dataset | 54 MB |

Global Environment

Environment is empty

Home Find in Topic

R Resources

- Learning R Online
- CRAN Task Views
- R on StackOverflow
- Getting Help with R

RStudio

- RStudio IDE Support
- RStudio Community Forum
- RStudio Cheat Sheets
- RStudio Tip of the Day
- RStudio Packages
- RStudio Products

Manuals

- An Introduction to R
- Writing R Extensions
- R Data Import/Export

Reference

Packages

Miscellaneous Material

Search Engine & Keywords

About R License NEWS Authors FAQ User Manuals Resources Thanks Technical papers

The R Language Definition R Installation and Administration R Internals



R Studio®

```
Quiescent_germ_fungal_8Aug2018.R  dada_phylose_JCexp.R  Intro.R.GradSkills_7Oct2022.R
Source | Save | Run | Source | Environment | History | Connections | Import Dataset | 54 MB | Global Environment | List | C |
1 # Introduction to R
2 #By Carolina Sarmiento-
3 #USF - Grad Skills - Oct. 7, 2022
4 ...
5 #*****#
6 # LET'S GET STARTED # -----
7 #*****#
8 ...
9 # 1. Set your Working Directory (WD) -----
10 #You *ALWAYS* have to set your WD (i.e., tell R where - in your computer - are the files you will need and where it can store some of the objects when you ask R to do so).
11 ...
12 #A. Check current WD by running the following line-
13 getwd()-
14 ...
15 #B. set your WD: Click on the drop-down-menu "Session" -> Set Working Directory -> Choose Directory-
16 ...
17 # **** I suggest you copy the line that just appeared in the Console to your script, for next time you open your script, just run that line.-
18 ...
19 setwd("~/Desktop/Grad_Skills/R") # copy yours here!-
20 ...
21 ...
22 # 2. Install a package -----
23 ...
24 # A. Install the package-
25 # Use first the "Install" button on the Packages tab, look for the package of interest (dsolve) and make sure to check "install dependencies".
3:34 (Top Level) : R Script
```

Console Terminal Background Jobs

R 4.2.0 ~/

R version 4.2.0 (2022-04-22) -- "Vigorous Calisthenics"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' or 'demo(package)' for help on a package
'help.start()' for an HTML browser interface to help pages
Type 'q()' to quit R.

Warning: macOS is reporting that you have not yet agreed to the Xcode license.
This can occur if Xcode has been updated after it was installed (e.g., as part of a macOS update).

Some features of Xcode may not be available.

Please run:

```
sudo xcodebuild -license accept
```

In a terminal to accept the Xcode license, and then restart RStudio.

Plots, Packages, and Help files



The screenshot shows the RStudio IDE interface. On the left, there's a code editor with several R scripts open. In the center, the environment pane is empty. At the bottom, a terminal window shows the R startup process and a warning about the Xcode license. To the right of the IDE is a large pink box containing the RStudio website navigation bar and various links to resources like R Resources, Manuals, Reference, Packages, and Miscellaneous Material.

Environment History Connections | Import Dataset | 54 MB | Global Environment | List | C |

Environment is empty

Console Terminal Background Jobs

R 4.2.0 ~/

R version 4.2.0 (2022-04-22) -- "Vigorous Calisthenics"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' or 'demo(package)' for help on a package
'help.start()' for an HTML browser interface to help pages
Type 'q()' to quit R.

Warning: macOS is reporting that you have not yet agreed to the Xcode license.
This can occur if Xcode has been updated after it was installed (e.g., as part of a macOS update).

Some features of Xcode may not be available.

Please run:

sudo xcodebuild -license accept

In a terminal to accept the Xcode license, and then restart RStudio.

Files Plots Packages Help Viewer Presentation | Home Find in Topic

RStudio

R Resources

- Learning R Online
- CRAN Task Views
- R on StackOverflow
- Getting Help with R

Manuals

- An Introduction to R
- Writing R Extensions
- R Data Import/Export

Reference

Packages

Miscellaneous Material

Search Engine & Keywords

About R
License
NEWS

Authors
FAQ
User Manuals

Resources
Thanks
Technical papers



R Studio®

```
1 # Introduction to R
2 # By Carolina Sarmiento
3 #USF - Grad Skills - Oct. 7, 2022
4 ...
5 #*****
6 # LET'S GET STARTED # -----
7 #*****
8 ...
9 # 1. Set your Working Directory (WD) -----
10 # You *ALWAYS* have to set your WD (i.e., tell R where - in your computer - are the files you will need and where it can store some of the objects when you ask R to do so).
11 ...
12 #A. Check current WD by running the following line-
13 getwd()-
14 ...
15 #B. set your WD: Click on the drop-down-menu "Session" -> Set Working Directory -> Choose Directory-
16 ...
17 # **** I suggest you copy the line that just appeared in the Console to your script, for next time you open your script, just run that line.-
18 ...
19 setwd("~/Desktop/Grad_Skills/R") # copy yours here!-
20 ...
21 ...
22 # 2. Install a package -----
23 ...
24 # A. Install the package-
25 # Use first the "Install" button on the Packages tab, look for the package of interest (dslabs) and make sure to check "install dependencies"-.
3:34 (Top Level) :  
R Script  
Console Terminal Background Jobs R 4.2.0 ~/  
R version 4.2.0 (2022-04-22) -- "Vigorous Calisthenics"  
Copyright (C) 2022 The R Foundation for Statistical Computing  
Platform: x86_64-apple-darwin17.0 (64-bit)  
R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.  
Natural language support but running in an English locale  
R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.  
Type 'q()' to quit R.  
Warning: macOS is reporting that you have not yet agreed to the Xcode license.  
This can occur if Xcode has been updated after R was installed (e.g., as part of a macOS update).  
Some features of R may not be available.  
Please run:  
sudo xcodebuild -license accept  
in a terminal to accept the Xcode license, and then restart RStudio.
```

Plots, Packages, and Help files

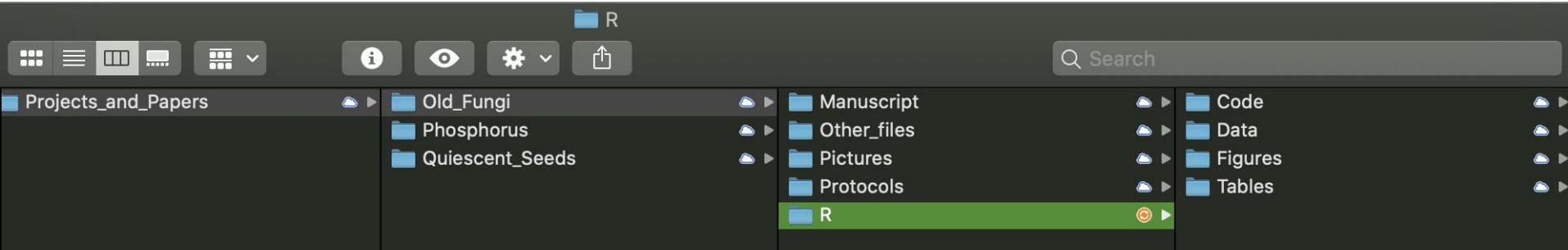


The RStudio website features a dark-themed header with a navigation bar including 'File', 'Edit', 'Packages', 'Help' (which is highlighted), 'Viewer', and 'Presentation'. Below the header is a search bar and a 'Home' dropdown menu. The main content area is divided into several sections:

- R Resources**: Includes links to 'Learning R Online', 'CRAN Task Views', 'R on StackOverflow', 'Getting Help with R' (which is underlined), 'RStudio IDE Support', 'RStudio Community Forum', 'RStudio Cheat Sheets', 'RStudio Tip of the Day', 'RStudio Packages', and 'RStudio Products'.
- Manuals**: Includes links to 'An Introduction to R', 'Writing R Extensions', and 'R Data Import/Export'.
- Reference**: Includes links to 'Packages' and 'Miscellaneous Material'.
- Search Engine & Keywords**: Includes links to 'About R', 'License', 'NEWS', 'Authors', 'FAQ', 'User Manuals', 'Resources', 'Thanks', and 'Technical papers'.

Before starting:

1. Create a folder, dedicated to store the “R things” for your Grad Skills Class:



2. Move the file "Intro_R_GradSkills_10oct2025.R" to your R folder (Code)

3. Open  R Studio[®]

4. Open the file "Intro_R_GradSkills_10oct2025.R"

1. Working Directory (WD)

Is a file path on your computer that sets the default location of any files you read into R, or save out of R:

If you ask R to import a dataset from a .csv file it will assume that the file is inside of your working directory.

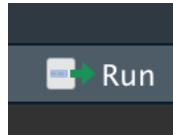
A. What is your current WD?

> **getwd()** *line 13*

Mac: command + Return

Win: ctrl + Enter

or click



at the top-right corner of
your text editor

* Note: R will ignore all the lines preceded by #

1. Working Directory (WD)

Is a file path on your computer that sets the default location of any files you read into R, or save out of R:

If you ask R to import a dataset from a .csv file it will assume that the file is inside of your working directory.

A. What is your current WD?

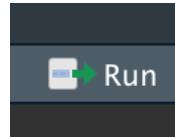
> `getwd()` *line 13*

Read all the text in the line and run each line at a time

Mac: command + Return

or click

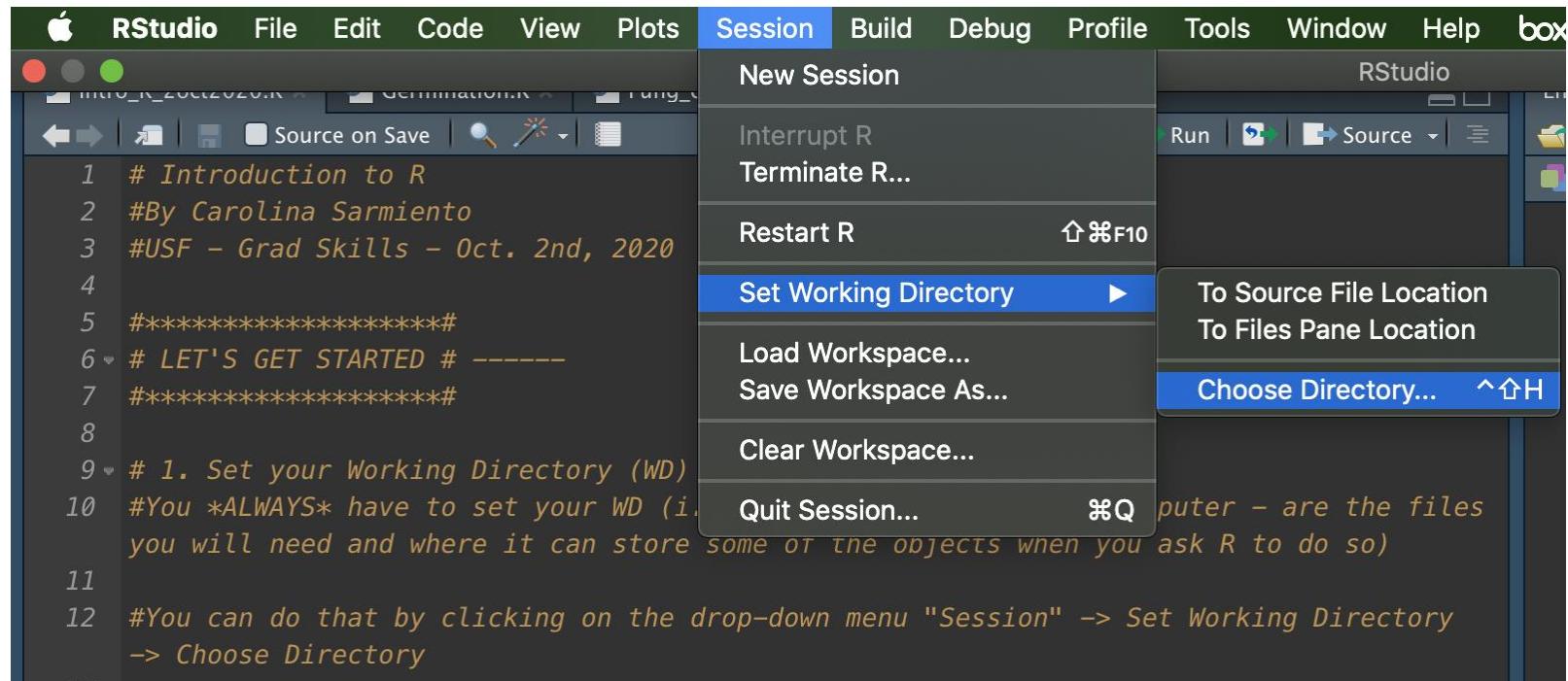
Win: ctrl + Enter



at the top-right corner of
your text editor

* Note: R will ignore all the lines preceded by #

1. Working Directory (WD)



B. Set your WD:

Or run the line:

`>setwd("path")` ---- > Check your console

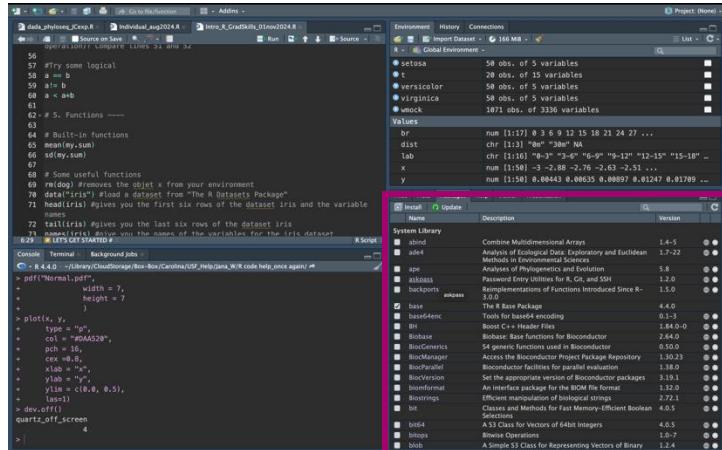
Here is mine:

`> setwd("~/Desktop/Grad_Skills/R")`

line 19

2. Install a package

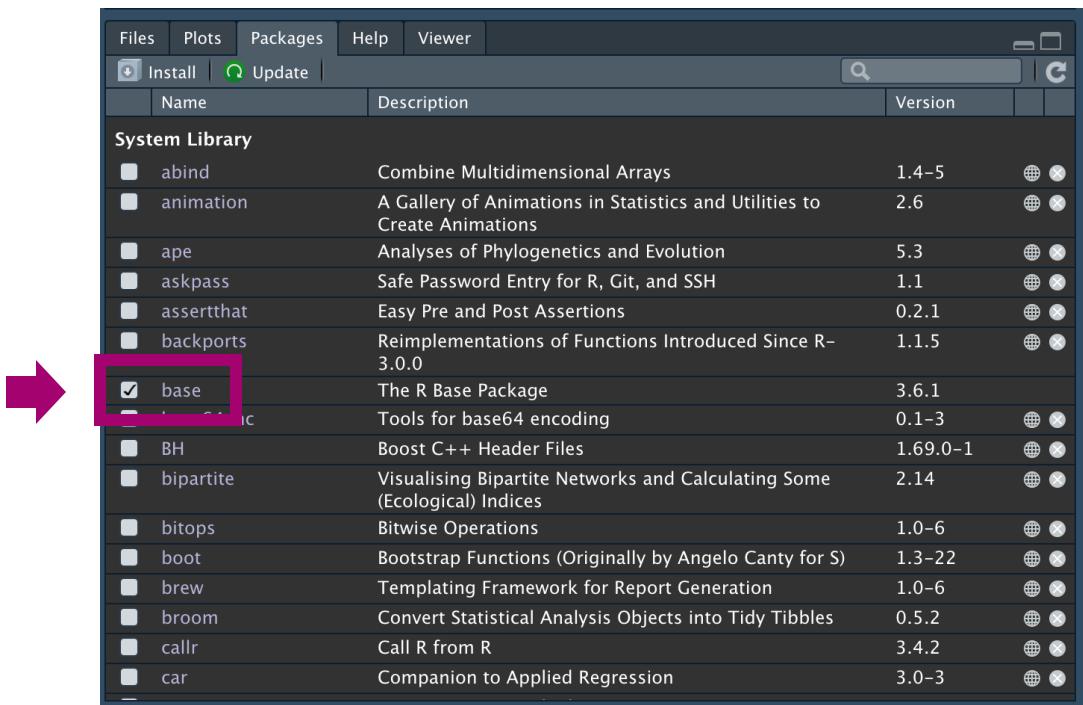
Let's take a moment to check your **Packages** tab on the bottom-right



```
56 # Try some logical
57 a == b
58 a != b
59 a <= b
60 a >= b
61
62 # Functions ----
63
64 # Built-in functions
65 mean(my.sum)
66 sd(my.sum)
67
68 # Some useful functions
69 rm(log) # removes the object x from your environment
70 rm(log) # removes the object x from "The R Dataframe Package"
71 head(iris) # gives you the first six rows of the dataset Iris
72 names(iris) # gives you the names of the variables in the Iris dataset
629 # LET'S GET STARTED #
```

R-4.4.0 - <https://cloud.r-project.org/R-4.4.0/> [Help/Dark Mode help once again]

```
> perfNormal.pdf
+ width = 7,
+ height = 7
>
> plot(x, y,
+       type = "p",
+       col = "#00A52B",
+       pch = 16,
+       cex = 0.8,
+       xlab = "x",
+       ylab = "y",
+       ylim = c(0, 0.5),
+       las=1)
> dev.off()
quartzOffScreen
> |
```

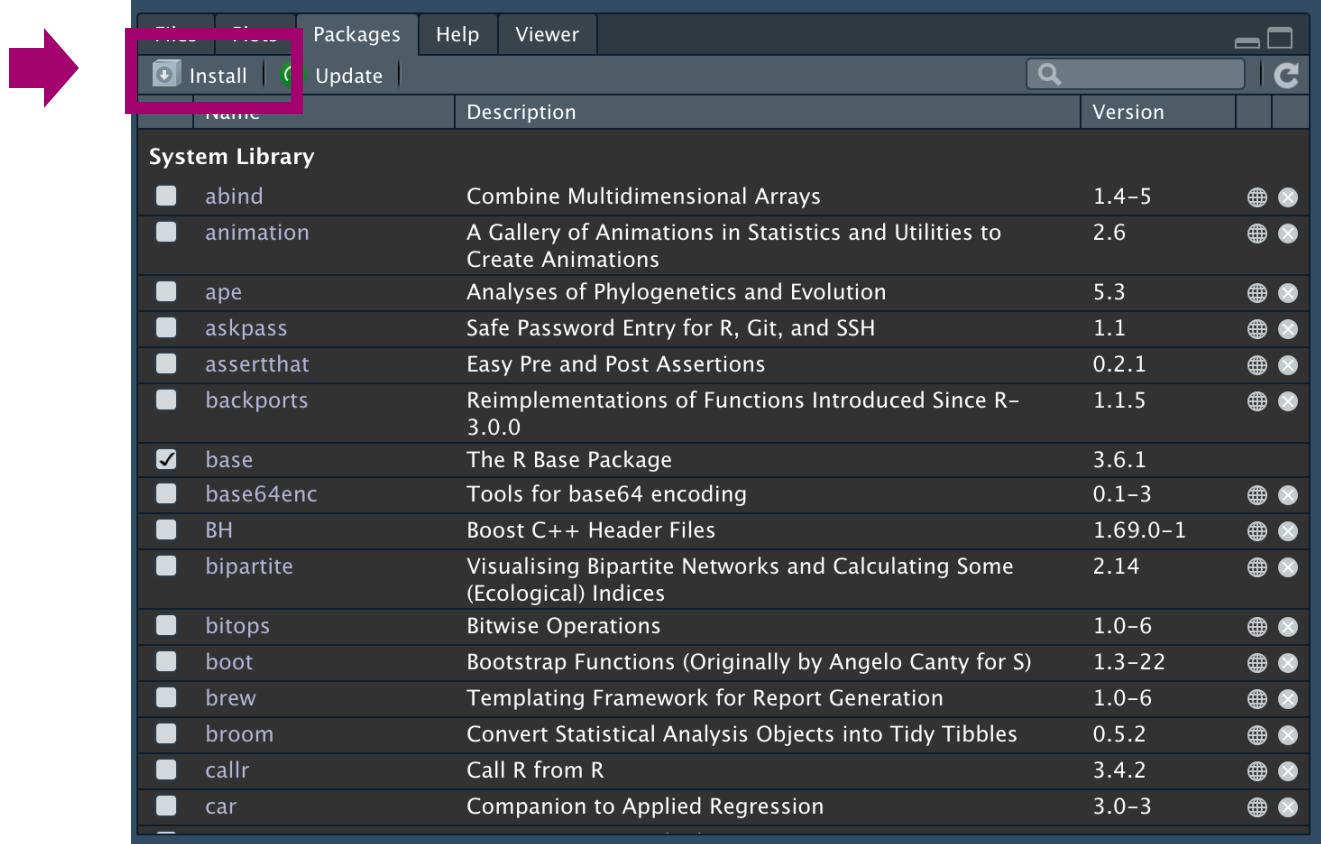


The screenshot shows the RStudio interface with the 'Packages' tab selected in the bottom navigation bar. The main area displays a list of packages from the 'System Library'. A purple arrow points to the 'base' package, which has a checked checkbox next to its name. Other packages listed include abind, animation, ape, askpass, assertthat, backports, base64enc, BH, bipartite, bitops, boot, brew, broom, callr, car, and carData.

Name	Description	Version
abind	Combine Multidimensional Arrays	1.4-5
animation	A Gallery of Animations in Statistics and Utilities to Create Animations	2.6
ape	Analyses of Phylogenetics and Evolution	5.3
askpass	Safe Password Entry for R, Git, and SSH	1.1
assertthat	Easy Pre and Post Assertions	0.2.1
backports	Reimplementations of Functions Introduced Since R-3.0.0	1.1.5
<input checked="" type="checkbox"/> base	The R Base Package	3.6.1
base64enc	Tools for base64 encoding	0.1-3
BH	Boost C++ Header Files	1.69.0-1
bipartite	Visualising Bipartite Networks and Calculating Some (Ecological) Indices	2.14
bitops	Bitwise Operations	1.0-6
boot	Bootstrap Functions (Originally by Angelo Canty for S)	1.3-22
brew	Templating Framework for Report Generation	1.0-6
broom	Convert Statistical Analysis Objects into Tidy Tibbles	0.5.2
callr	Call R from R	3.4.2
car	Companion to Applied Regression	3.0-3
carData	Companion to Applied Regression Data Sets	3.0-3

2. Install a package

We will install the package **dslabs**

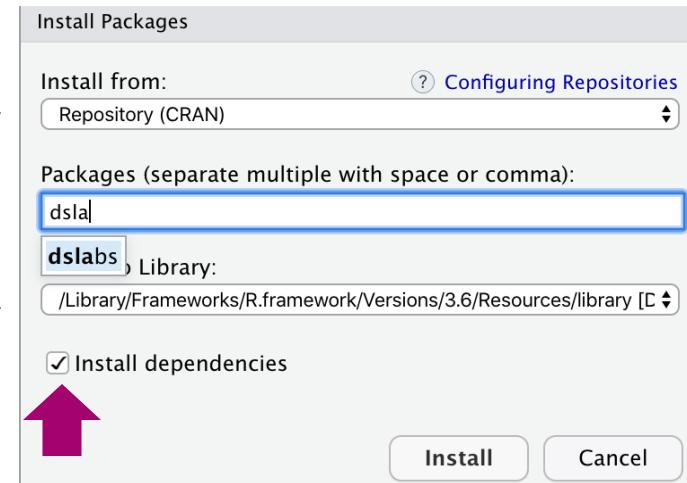


2. Install a package

We will install the package **dslabs**

From CRAN (default)

Path to the folder where it will be stored in your computer



Make sure this box is checked

What happened in the console?
`> install.packages("dslabs")`

2. Install a package

Now that we are here (Packages tab):



File	Plots	Packages	Help	Viewer	Presentation		
		<input checked="" type="checkbox"/> Install <input type="checkbox"/> Update				<input type="text"/>	
	Name	Description			Version		
<input checked="" type="checkbox"/>	dplyr	A Grammar of Data Manipulation			1.0.9		
<input checked="" type="checkbox"/>	dslabs	Data Science Labs			0.7.6		
<input checked="" type="checkbox"/>	DT	A Wrapper of the JavaScript Library 'DataTables'			0.23		
<input checked="" type="checkbox"/>	dtplyr	Data Table Back-End for 'dplyr'			1.2.1		

The package is stored in your "library" but to access it, you need to load the package (i.e., check that box!)

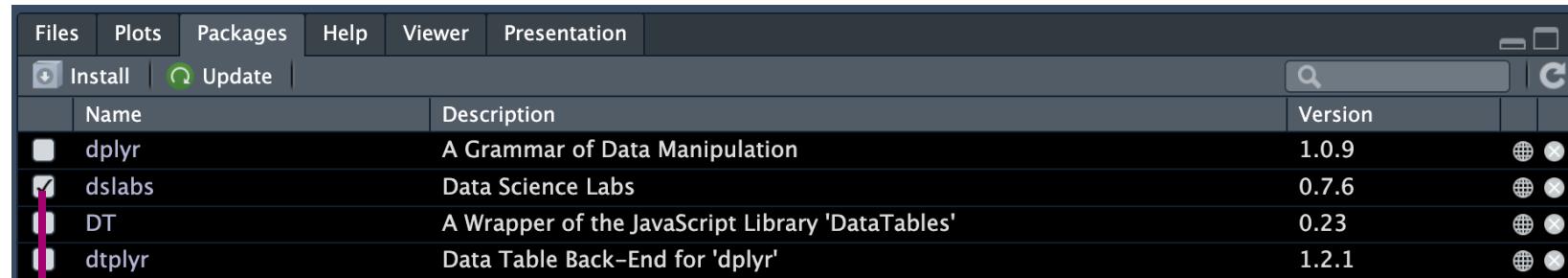
or

run the line with the function ***library()***

> *library(dslabs)*

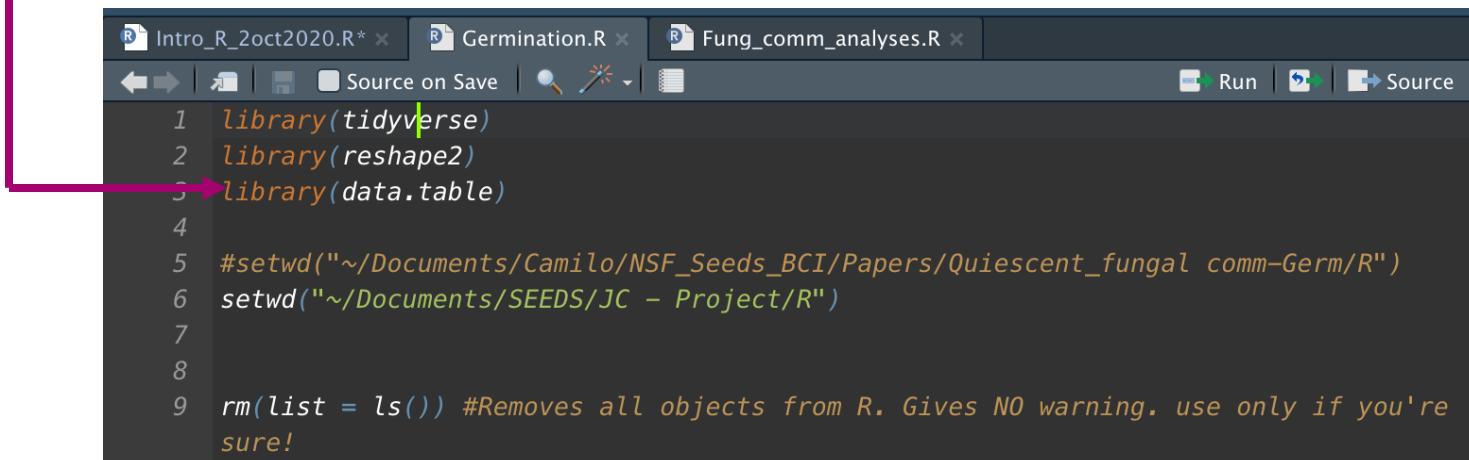
Tip: copy this line at the top of your script, so you don't forget to load all the packages you need for your code to run

2. Install a package



The screenshot shows the RStudio interface with the 'Packages' tab selected in the top menu bar. The 'Install' button is highlighted. A table lists four packages: dplyr, dslabs, DT, and dtplyr. The 'dplyr' package is unselected, while 'dslabs' has a checked checkbox and is highlighted in blue. The table includes columns for Name, Description, Version, and two circular icons.

Name	Description	Version	
dplyr	A Grammar of Data Manipulation	1.0.9	 
<input checked="" type="checkbox"/> dslabs	Data Science Labs	0.7.6	 
DT	A Wrapper of the JavaScript Library 'DataTables'	0.23	 
dtplyr	Data Table Back-End for 'dplyr'	1.2.1	 



The screenshot shows the RStudio code editor with three tabs open: 'Intro_R_2oct2020.R*', 'Germination.R', and 'Fung_comm_analyses.R'. The 'Fung_comm_analyses.R' tab is active. The code editor displays the following R script:

```
library(tidyverse)
library(reshape2)
library(data.table)

#setwd("~/Documents/Camilo/NSF_Seeds_BCI/Papers/Quiescent_fungal comm-Germ/R")
setwd("~/Documents/SEEDS/JC - Project/R")

rm(list = ls()) #Removes all objects from R. Gives NO warning. use only if you're sure!
```

2. Install a package

Now that we are here:

vectors				
<input type="checkbox"/>	dplyr	A Grammar of Data Manipulation	0.8.3	 
<input type="checkbox"/>	dslabs	Data Science Labs	0.7.3	 
<input type="checkbox"/>	DT	A Wrapper of the JavaScript Library 'DataTables'	0.9	 

If you click in the package name:

Full name



Description file for the package



Documentation for package 'dslabs' version 0.7.3

- [DESCRIPTION file](#).

Help Pages

admissions	Gender bias among graduate school admissions to UC Berkeley.
brca	Breast Cancer Wisconsin Diagnostic Dataset from UCI Machine Learning Repository
brexit_polls	Brexit Poll Data
death_prob	2015 US Period Life Table
divorce_margarine	Divorce rate and margarine consumption data
ds_theme_set	dslabs theme set
gapminder	Gapminder Data
greenhouse_gases	Greenhouse gas concentrations over 2000 years
heights	Self-Reported Heights
historic_co2	Atmospheric carbon dioxide concentration over 800,000 years
mnist_27	Useful example for illustrating machine learning algorithms based on MNIST data
movielens	Movie ratings

List of "objects" contained in the package
(usually functions and datasets) and their own
help files.

So far...

- Using a script (**as opposed to typing your instructions directly on the console**) is highly recommended
-  R Studio® makes easier some common tasks
- **#** is useful to comment your code
 - R “ignores” the text preceded by **#**
 - And I strongly suggest to use it (as much as you can!)
- Rstudio’s text editor recognize some elements and color them differently (**syntax-highlighting editor**) which is very useful

So far...

- If your console panel is active: Arrow up returns previously entered code
- R is case sensitive: “Anova” is different from “anova”
- Quotation marks must not be directional (“**NO**”; “**YES**”)
Do not use Microsoft Word as text editor!
- Click  at the bottom of your text editor:

```
5 #*****#
6 # LET'S GET STARTED # -----
7 #*****#
8
9 # 1. Set your Working Directory (WD) -----
10 #You *ALWAYS* have to set your WD (i.e. tell
you will need and where it can store some o
11
12 #A. check in what folder R is currently wor
22:1 # 2. Install a package ↓
```

So far...

- If your console panel is active: Arrow up returns previously entered code
- R is case sensitive: “Anova” is different from “anova”
- Quotation marks must not be directional (“**NO**”; “**YES**”)
Do not use Microsoft Word as text editor!
- Click  at the bottom of your text editor:

```
5 #*****  
6 # LET'S GET STARTED # -----  
7 #*****  
8  
9 # 1. Set your Working Directory (WD) -----  
10 #You *ALWAYS* have to set your WD (i.e. tell  
you will need and where it can store some o  
11  
12 #A. check in what folder R is currently wor  
13 # 2. Install a package ↓
```

It indexes your script by using
“----” Four or more dashes or #

3. Objects

R is an object-oriented language:

Variables, data, functions, results, etc., are stored in the active memory of the computer in the form of **objects**, which have a **name**.

`name <— object`

(To assign an object (value, data, function) to a name use `<-`)

Mac: option + -

Win: alt + -

3. Objects

R is an object-oriented language:

Variables, data, functions, results, etc., are stored in the active memory of the computer in the form of **objects**, which have a **name**.

name <— object

```
> a <- 15  
> b <- c(1:10)  
> c <- "My dog's name is Kona"
```

*Note that if you use the same name, R will replace the object without asking

*** e.g., the values 1 to 10 are stored in the object named “b”, to access it, type b in your console + return/enter ***

lines 32 - 49

*Rstudio will let you know when a line is “evidently” wrong (line 45)

3. Objects

R is an object-oriented language:

Variables, data, functions, results, etc., are stored in the active memory of the computer in the form of **objects**, which have a **name**.

name \leftarrow object

```
> a <- 15  
> b <- c(1:10)  
> c <- "My dog's name is Kona"
```

The user can create and modify these objects with
operators and **functions**

```
> a * b  
> my.sum <- a + b
```

4. Operators

■ Arithmetic

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

`> a * b`

■ Logical

Operator	Description
<	less than
<=	less than or equal to
>	greater than
>=	greater than or equal to
==	exactly equal to
!=	not equal to
!x	Not x
x y	x OR y
x & y	x AND y

`> a == b`

lines 52 - 64

5. Functions

Almost everything in R is done through functions

`function.name (arguments)`

- **Built-in Functions** (they can be numeric, character, statistical, probability, etc.)

Function	Description
<code>abs(x)</code>	absolute value
<code>sqrt(x)</code>	square root
<code>round(x, digits=n)</code>	<code>round(3.475, digits=2)</code> is 3.48
<code>log(x)</code>	natural logarithm
<code>log10(x)</code>	common logarithm
<code>exp(x)</code>	e^x
<code>sum(x)</code>	sum
<code>min(x)</code>	minimum
<code>max(x)</code>	maximum

We already know some others:

`getwd()`

`setwd()`

`install.packages("dslabs")`

`library(dslabs)`

`c()` *concatenate, useful to create vectors

<code>mean(x, na.rm=FALSE)</code>	mean of object x
<code>sd(x)</code>	standard deviation of object(x). also look at <code>var(x)</code> for variance and <code>mad(x)</code> for median absolute deviation.
<code>median(x)</code>	median

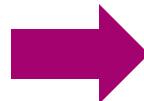
5. Functions

Almost everything in R is done through functions

- **User-written functions:** One of the great strengths of R is the user's ability to add functions.

```
myfunction <- function(arg1, arg2, ... ){  
  statements  
  return(object)  
}
```

```
> myfunction1 <- function(x, y=0){  
  print (x)  
  print (y)  
  return (sum ( (x - y) ^ 2 ))  
}
```



```
> myfunction1(10)  
[1] 10  
[1] 0  
[1] 100  
> myfunction1(10,5)  
[1] 10  
[1] 5  
[1] 25
```

lines 66 - 103

6. Object properties

- **Mode:** how objects are stored in R

character, numeric, logical, complex, & function

```
> mode(iris$Sepal.Length)  
[1] "numeric"
```

- **Class:** how objects are treated by functions

character, factor, numeric, matrix, data.frame, etc.

```
> class(iris$Species)  
[1] "factor"
```

- **Length:** Number of elements of an object

```
> length(b)  
[1] 10
```

7. Data structure

- **Vector:** list of elements of the same type (mode) 1-dimensional
> e <- c(TRUE, TRUE, FALSE, FALSE, FALSE)
> str(e)
logi [1:5] TRUE TRUE FALSE FALSE FALSE
- **Matrix:** elements in rows and columns, same type 2-dimensional
> b <- c(1:10)
> d <- c(11:20)
> g <- cbind(b, d) *function cbind() joins the two objects as columns
- **Data Frame:** like a matrix (rows,columns) but different columns can have different modes 2-dimensional
> x <- c(1:10)
> y <- rep(TRUE, 10)
> z <- c("a", "b", "c", "d", "e", "f", "g", "h", "i", "j")
> mydf <- data.frame(x, y, z)

lines 124 - 138

8. Accessing the values of an object

The indexing system

Matrices and Data frames have a similar structure
[rows , columns]

```
> A <- matrix(rnorm(100),10,10) *check >help("rnorm")
```

- >A[3,] shows 3rd row and all columns
- >A[4] shows all rows and the 4th column
- >A[3,4] shows element in row 3, column 4
- >A[c(1,9),] shows rows 1 and 9, all columns
- >A[,-1] shows all of A, except column 1
- >A[-c(1,9),] shows all of A, except rows 1 and 9
- >A[1:10,5:8] shows first 10 rows and columns 5 to 8
- >A[A[,1]>0,] shows only rows where first column is a positive number
- >A[10:1,] shows first 10 rows of A, in opposite order

8. Accessing the values of an object

The indexing system

**Matrices and Data frames have a similar structure
[rows , columns]**

A data frame has **column names** that can be used to access its elements

> mydf <- iris To access one column by its name use **\$**
(e.g., mydf\$Species)

A. Show all the rows that meet the condition ‘Species = setosa’:

>mydf [mydf\$Species == "setosa",]

B. Show all the rows with sepal lengths > 6cm and sepal widths < 4cm:

>mydf [mydf\$Sepal.Length > 6 & mydf\$Sepal.Width < 4,]

9. Import a dataset

There are several ways of importing data into R:

- A. The one that I use and recommend:

```
>mydata <- read.csv("Data/Iris_data.csv", header = TRUE)
```

**imports a .csv file stored in the Data sub folder of my WD and
recognizes the first row as column names**

- B.

```
>mydata <- read.table("Data/Iris_data.txt", sep= "\t", h = T)
```

Imports a text file and recognizes the first row as column names

**Also

```
>read.table("Data/Iris_data.csv", sep = ", ", header =TRUE)
```**

- C.

```
>mydata <- read.csv(file.choose( ) )
```

Imports a .csv file from anywhere in your computer

9. Import a dataset

Common problems when importing data:

- R doesn't recognize the path you gave or can't find the file:

```
> mydata <- read.table("Data/Iris_data.txt", sep= "\t", h = T)
Error in file(file, "rt") : cannot open the connection
In addition: Warning message:
In file(file, "rt") :
  cannot open file 'Data/Iris_data.txt': No such file or directory
```

9. Import a dataset

Common problems when importing data:

- Your data set has spaces (that you can't notice in Excel!)

The screenshot shows the RStudio interface. On the left, there is a code editor with several tabs open, including "Intro_R_2oct2020.R*", "mydata2", "Germination.R", and "Documentation for package 'datas...'. The main area displays a data frame titled "mydata2" with 153 observations and 8 variables. The variables are Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, Species, X, X.1, and X.2. The data consists of rows of numerical values and character strings. The "Species" column contains the value "setosa" for all entries. The "X", "X.1", and "X.2" columns contain the value "NA" for all entries. The "mydata2" data frame is shown in the "Data" pane on the right, along with its dimensions: 153 obs. of 8 variables.

| | Sepal.Length | Sepal.Width | Petal.Length | Petal.Width | Species | X | X.1 | X.2 |
|----|--------------|-------------|--------------|-------------|---------|----|-----|-----|
| 1 | 5.1 | 3.5 | 1.4 | 0.2 | setosa | NA | NA | NA |
| 2 | 4.9 | 3.0 | 1.4 | 0.2 | setosa | NA | NA | NA |
| 3 | 4.7 | 3.2 | 1.3 | 0.2 | setosa | NA | NA | NA |
| 4 | 4.6 | 3.1 | 1.5 | 0.2 | setosa | NA | NA | NA |
| 5 | 5.0 | 3.6 | 1.4 | 0.2 | setosa | NA | NA | NA |
| 6 | 5.4 | 3.9 | 1.7 | 0.4 | setosa | NA | NA | NA |
| 7 | 4.6 | 3.4 | 1.4 | 0.3 | setosa | NA | NA | NA |
| 8 | 5.0 | 3.4 | 1.5 | 0.2 | setosa | NA | NA | NA |
| 9 | 4.4 | 2.9 | 1.4 | 0.2 | setosa | NA | NA | NA |
| 10 | 4.9 | 3.1 | 1.5 | 0.1 | setosa | NA | NA | NA |
| 11 | 5.4 | 3.7 | 1.5 | 0.2 | setosa | NA | NA | NA |
| 12 | 4.8 | 3.4 | 1.6 | 0.2 | setosa | NA | NA | NA |
| 13 | 4.8 | 3.0 | 1.4 | 0.1 | setosa | NA | NA | NA |
| 14 | 4.3 | 3.0 | 1.1 | 0.1 | setosa | NA | NA | NA |
| 15 | 5.8 | 4.0 | 1.2 | 0.2 | setosa | NA | NA | NA |
| 16 | 5.7 | 4.4 | 1.5 | 0.4 | setosa | NA | NA | NA |
| 17 | 5.4 | 3.9 | 1.3 | 0.4 | setosa | NA | NA | NA |
| 18 | 5.1 | 3.5 | 1.4 | 0.2 | setosa | NA | NA | NA |
| 19 | 5.9 | 3.0 | 1.5 | 0.2 | setosa | NA | NA | NA |

Showing 1 to 19 of 153 entries, 8 total columns

9. Import a dataset

Common problems when importing data:

- Your data set has spaces (that you can't notice in Excel!!)

To fix this, I suggest to open your .csv file in Excel, select several columns and delete them

9. Import a dataset

Common problems when importing data:

- Your data set has spaces (that you can't notice in Excel!)

To fix this, I suggest to open your .csv file in Excel, select several columns and delete them

The screenshot shows a Microsoft Excel interface with the following details:

- Ribbon:** Home, Insert, New Tab, Draw, Page Layout, Formulas, Data, Review, View, Acrobat, Tell me.
- Clipboard:** Paste, Cut, Copy, Paste Special, Find & Select.
- Font:** Calibri (Body), Size 12, Bold, Italic, Underline, Font Color (red).
- Number Format:** General, Currency, Percentage, Text, Date, Time, Scientific, Number, Custom.
- Data Tools:** Conditional Formatting, Format as Table, Cell Styles.
- Cells Context Menu:** Insert, Delete, Delete Cells..., Delete Sheet Rows (highlighted in blue), Delete Sheet Columns, Delete Sheet.
- Worksheet:** A152 is the active cell. The data starts from row 143 and includes columns A through M. Row 143 contains values: 6.9, 3.1, 5.1, 2.3, virginica. Subsequent rows show similar data points for virginica.
- Message Bar:** Possible Data Loss: Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel Workbook (.xlsx) format.

9. Import a dataset

Common problems when importing data:

- Your data set has empty cells, or “weird” things that don’t let you get what you want:

The screenshot shows the RStudio interface with two panes. The left pane is a data viewer showing the 'iris' dataset with 150 rows and 5 columns. Row 4 has a value '4.6' highlighted with a pink box. The right pane is the Environment browser, which lists 'mydata2' as a global environment containing 150 observations of 8 variables. The 'Sepal.Length' variable is a factor with levels '4.3', '4.4', '4.5', etc. Below the environment browser is a console window displaying R code and an error message:

```
> View(mydata2)
> range(mydata2$Sepal.Length)
Error in Summary.factor(c(9L, 7L, 5L, 4L, 36L, 12L, 4L, 8L, 2L, 7L, 12L, :
  'range' not meaningful for factors
>
```

The rightmost part of the image shows a separate window for the 'read.table' function, which provides documentation for the algorithm used in R.

Depending on the size of your dataset, you can check carefully on excel or a text editor or sometimes you must clean your dataset in R

10. Export a dataset

Sometimes you need to export your objects.

For data you can use:

```
write.table(object, "path", sep=",")  
write.csv(object,"path/file.name")  
> write.csv(mydata, "Data/New_iris.csv")
```



This will save a new .csv file inside the folder “Data” on my WD

The *tidyverse*

**A collection of R packages for data science.
They share a design philosophy, grammar, and
data structures.**

Tools for:

- data import
- data wrangling
- visualization
- modeling

Makes code more readable and consistent
through features like data **pipelines** and a friendly **syntax**



The *tidyverse*

A different syntax: using the pipe operator (%>% or |>) to chain together a series of functions



Clear and readable sequence of operations that flow from one step to the next

Instead of nesting function calls: `function3(function2(function1(data)))`,

chain them using the pipe operator - more readable and intuitive
`data %>%`

```
function1( ) %>%
function2( ) %>%
function3( )
```

The *tidyverse*

A different syntax: using the pipe operator (%>% or |>) to chain together a series of functions



Creates a clear and readable sequence of operations that flow from one step to the next

In base R:

```
cars.6.cyl <- mtcars[mtcars$cyl == 6,]  
my.df <- cars.6.cyl[, c("mpg", "hp")]  
print(my.df)
```

Using the *tidyverse*:

```
my.df <- mtcars %>%  
  filter(cyl == 6) %>%  
  select(mpg, hp)
```

dplyr and *ggplot2*

***dplyr*:** Data manipulation – set of verbs that help you solve the most common challenges

- **`mutate()`:** add new variables
- **`select()`:** select variables (columns)
- **`filter()`:** select cases (rows)
- **`summarise()`:** reduce multiple values to a single summary
- **`arrange()`:** order rows

11. R Graphics

The graphical features in R are one of its major strengths

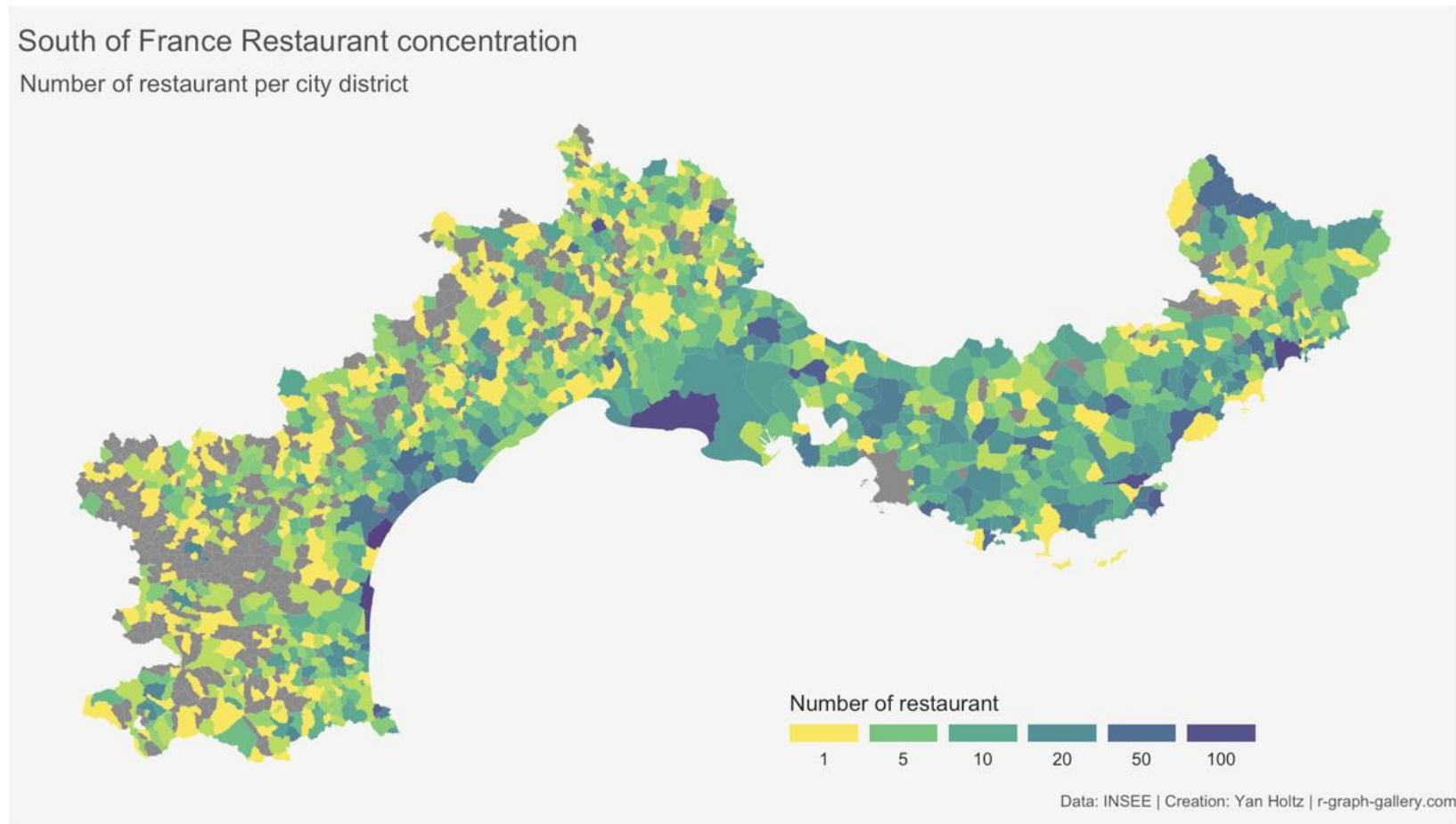
To get an idea what is possible with R run:

`>demo(graphics)`

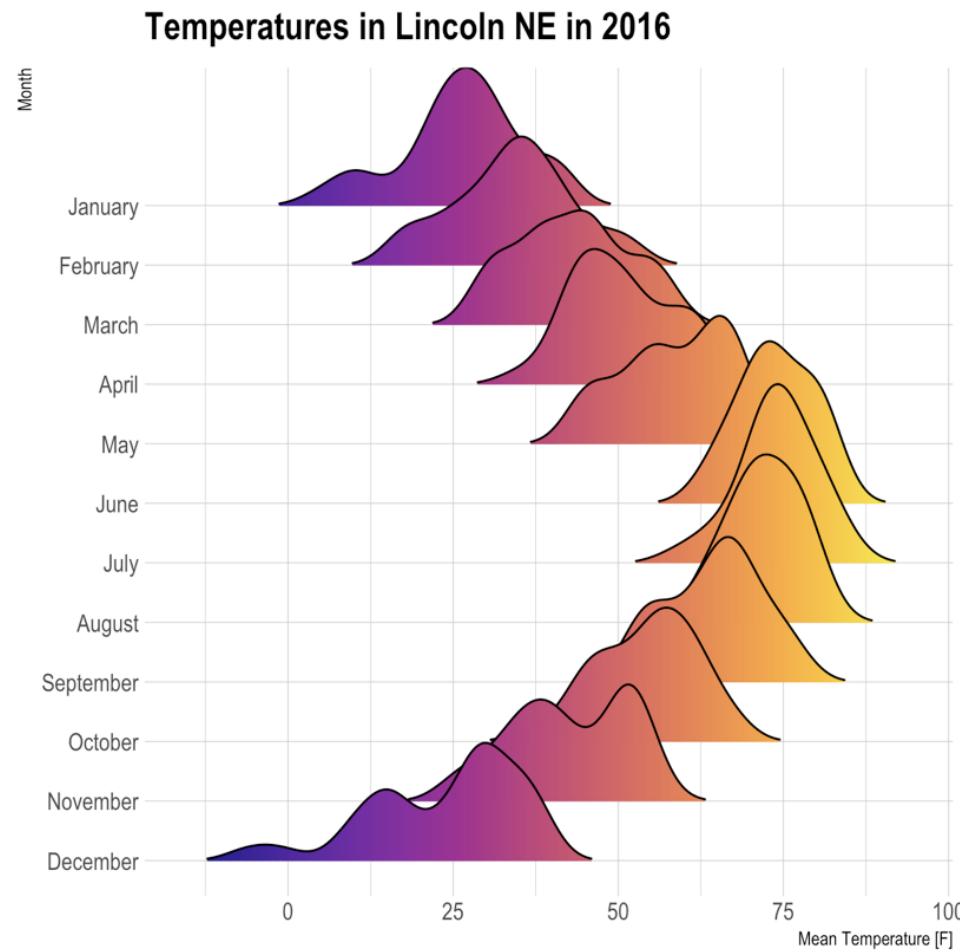
or visit:

<https://www.r-graph-gallery.com>

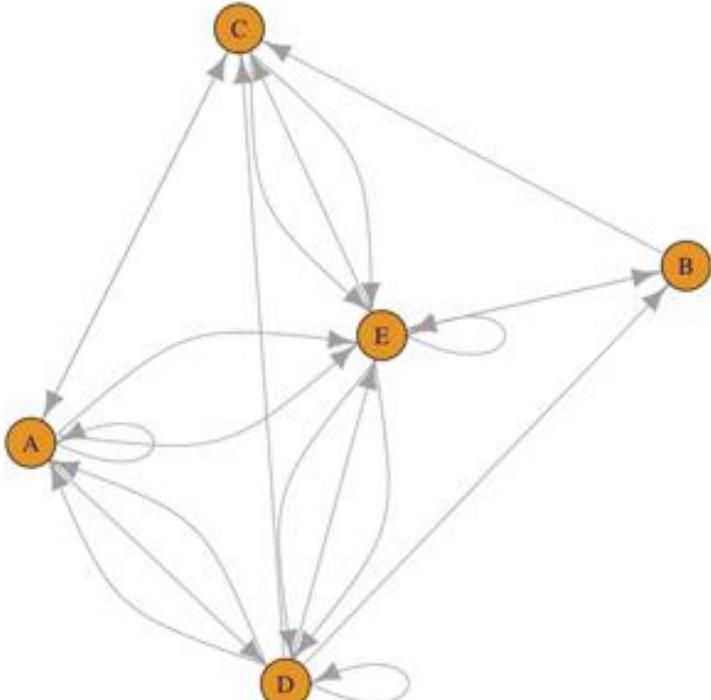
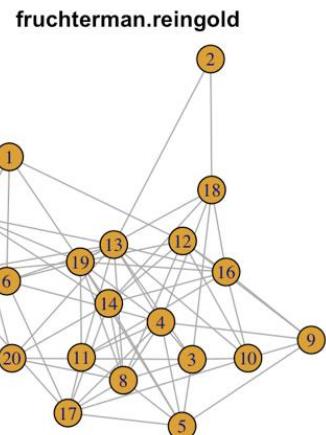
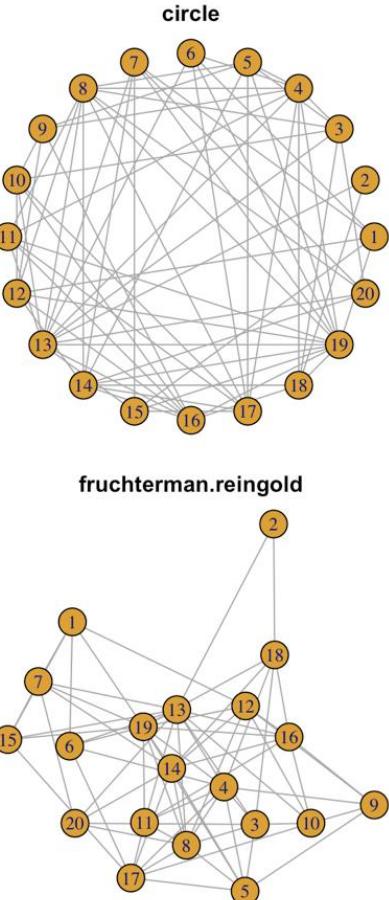
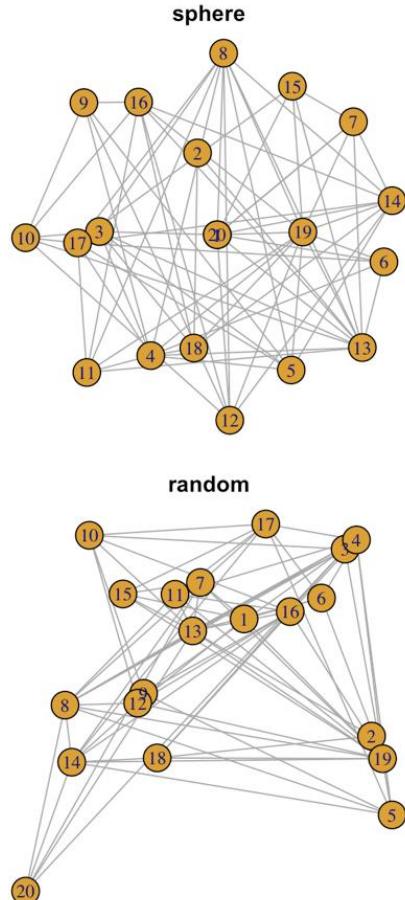
11. R Graphics



11. R Graphics



11. R Graphics



11. R Graphics

Generic “plot “ function – graphic type depends on object type:

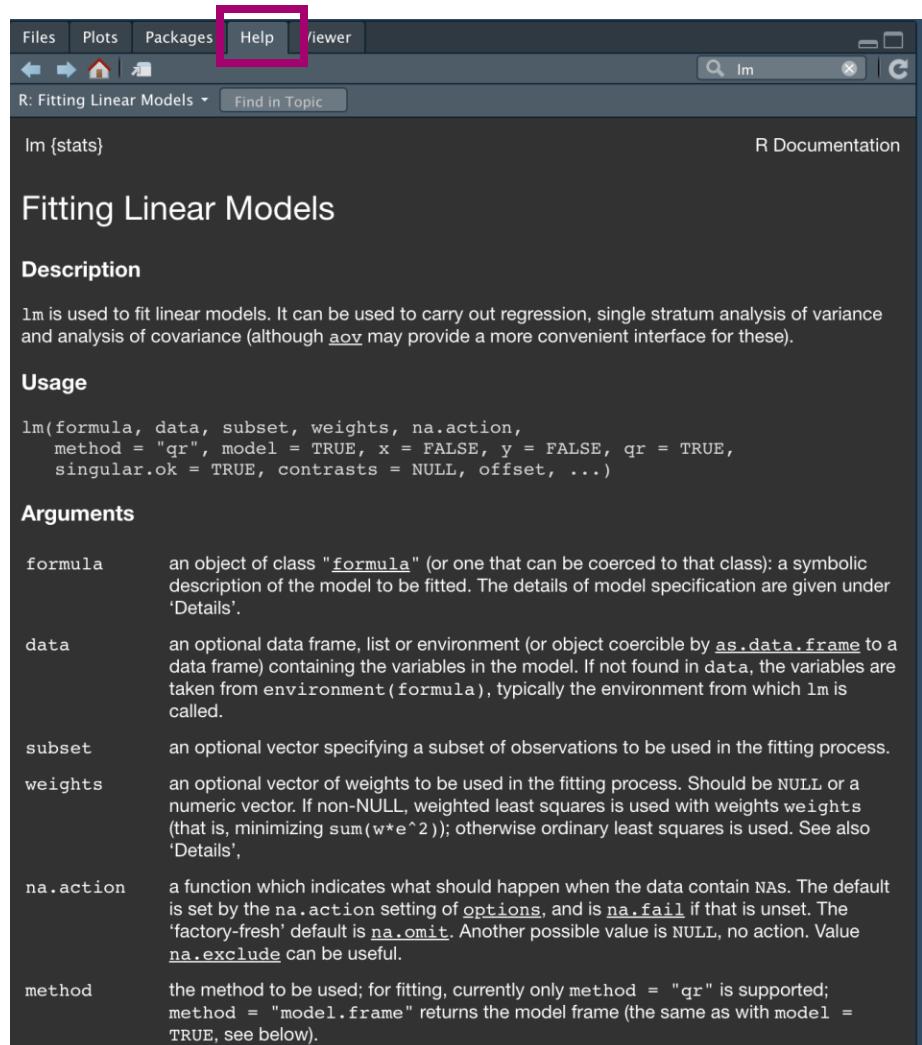
- plot(x,y)** if x, y are numeric vectors creates a scatterplot
- plot(x)** if x is a factor creates a barplot
- plot(x,y)** if x is a factor and y a numeric vector creates a boxplot of y at each level of x
- plot(x)** if x is a data frame, creates distributional plots for all components

12. Getting help

- To get the documentation file for a function:

```
> help("anova")
> help(anova)
> ?anova
```

- Easy in RStudio: just type in the function/package name in the 'help' tab



The screenshot shows the RStudio interface with the 'Help' tab highlighted by a red box in the top navigation bar. The main content area displays the documentation for the `lm` function from the `stats` package. The title is "Fitting Linear Models". The "Description" section states that `lm` is used to fit linear models. The "Usage" section shows the function signature: `lm(formula, data, subset, weights, na.action, method = "qr", model = TRUE, x = FALSE, y = FALSE, qr = TRUE, singular.ok = TRUE, contrasts = NULL, offset, ...)`. The "Arguments" section lists several parameters with their descriptions:

| Argument | Description |
|------------------------|---|
| <code>formula</code> | an object of class " <code>formula</code> " (or one that can be coerced to that class): a symbolic description of the model to be fitted. The details of model specification are given under 'Details'. |
| <code>data</code> | an optional data frame, list or environment (or object coercible by <code>as.data.frame</code> to a data frame) containing the variables in the model. If not found in <code>data</code> , the variables are taken from <code>environment(formula)</code> , typically the environment from which <code>lm</code> is called. |
| <code>subset</code> | an optional vector specifying a subset of observations to be used in the fitting process. |
| <code>weights</code> | an optional vector of weights to be used in the fitting process. Should be <code>NULL</code> or a numeric vector. If non- <code>NULL</code> , weighted least squares is used with weights <code>weights</code> (that is, minimizing <code>sum(w*e^2)</code>); otherwise ordinary least squares is used. See also 'Details'. |
| <code>na.action</code> | a function which indicates what should happen when the data contain NAs. The default is set by the <code>na.action</code> setting of <code>options</code> , and is <code>na.fail</code> if that is unset. The 'factory-fresh' default is <code>na.omit</code> . Another possible value is <code>NULL</code> , no action. Value <code>na.exclude</code> can be useful. |
| <code>method</code> | the method to be used; for fitting, currently only <code>method = "qr"</code> is supported; <code>method = "model.frame"</code> returns the model frame (the same as with <code>model = TRUE</code> , see below). |

12. Getting help

- If this fails, try searching the R website or
www.rseek.org
- Tons of online help
([StackOverflow](http://StackOverflow.com) and other forums)



Created and maintained by [Sasha Goodman](#).
Serving the R community since 2007. Version 2.0.

[Privacy Policy](#)

[Download and Install R](#)

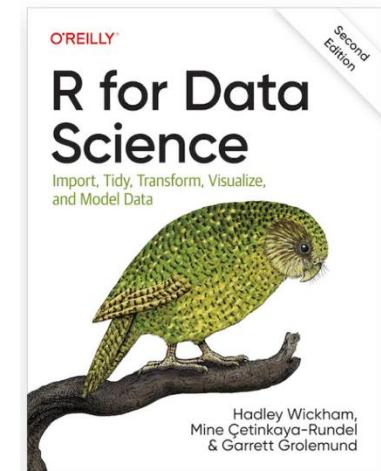


Learning R

- **Read through the CRAN website:**
lots of links to manuals, help files
- **QuickR website:** Excellent place to learn
<https://www.statmethods.net>
- **Learn R in R: Swirl**
<http://swirlstats.com>
- **Check the Book (tidyverse-focused):**
[R for Data Science](#)



swirl teaches you R programming and data science interactively, at your own pace, and right in the R console!



Some helpful links

- [Data Carpentry: Data Analysis and Visualization in R for Ecologists](#)
- [Tidyverse: R packages for Data Science](#)
- [Advanced R by Hadley Wickham](#)
- [ggplot2: Elegant graphics for data analysis by Hadley Wickham](#)
- [Posit Cheatsheets by Posit](#)
- [The R Gallery Book by Kyle W. Brown](#)
- [Introduction to R by Douglas, Roos, Mancini, Cuoto, & Lusseau](#)

Learning R

Five things to remember:

1. R always does exactly what you tell it to do
(But R only tells you what it did if you know how to ask)
2. R assumes that you know what you are saying
(But we often do not know what we are saying)
3. There are a lot of right ways to do everything
(But there are also a lot of wrong ways to do everything)
4. Spaces and capitals are your enemy
5. **BE PATIENT AND DON'T GIVE UP - the more time you spend using R, the more comfortable you become with it**

Assignment (Due Oct 17, 2025)

1. Use the file Iris_data.csv

Import the dataset and store it as an object:

- a. Calculate the mean value of sepal length and sepal width
- b. Calculate the Standard Deviation of sepal length and sepal width
- c. Make a plot (scatter plot) of sepal length (y-axis) against sepal width (x-axis) and save it as a .pdf file

* Remember to use what we learned today (as much as you can!)

2. Use the file Reported_heights.csv

Import the dataset as an object:

- a. Calculate the mean value of heights
- b. Calculate the standard deviation of heights
- c. Make a plot (boxplot) showing the differences between male and female heights and save it as a .pdf file

* Remember to use what we learned today (as much as you can!)

Once you are done, upload your script (.R – one file) and your plots (.pdf – two files)

Any questions? carolina.sarm@usf.edu