



Introduction to R for Plant Pathologists

APS Online Webinar

Sydney Everhart¹, Nikita Gambhir¹, and Katie Gold²

¹University of Nebraska-Lincoln

²Cornell University

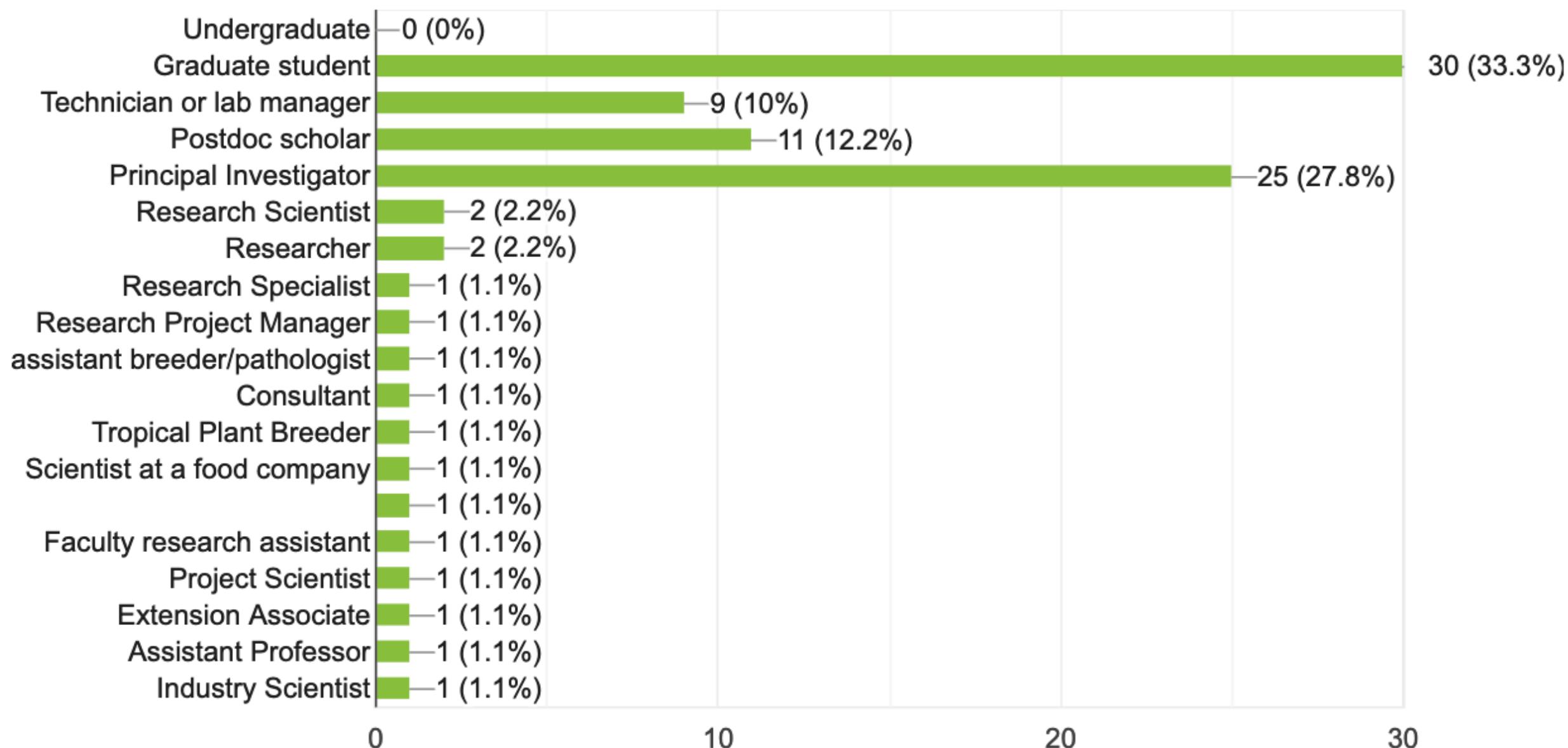
February 19th and 26th, 2020

Workshop Structure

- February 19th:
 - Part 1. Introduction to R (Sydney Everhart)
- February 26th:
 - Part 2a. Data Manipulation and Analysis (Nikita Gambhir)
 - Part 2b. Data Visualization (Katie Gold)
 - Part 2c. Starting a Project (Katie Gold)

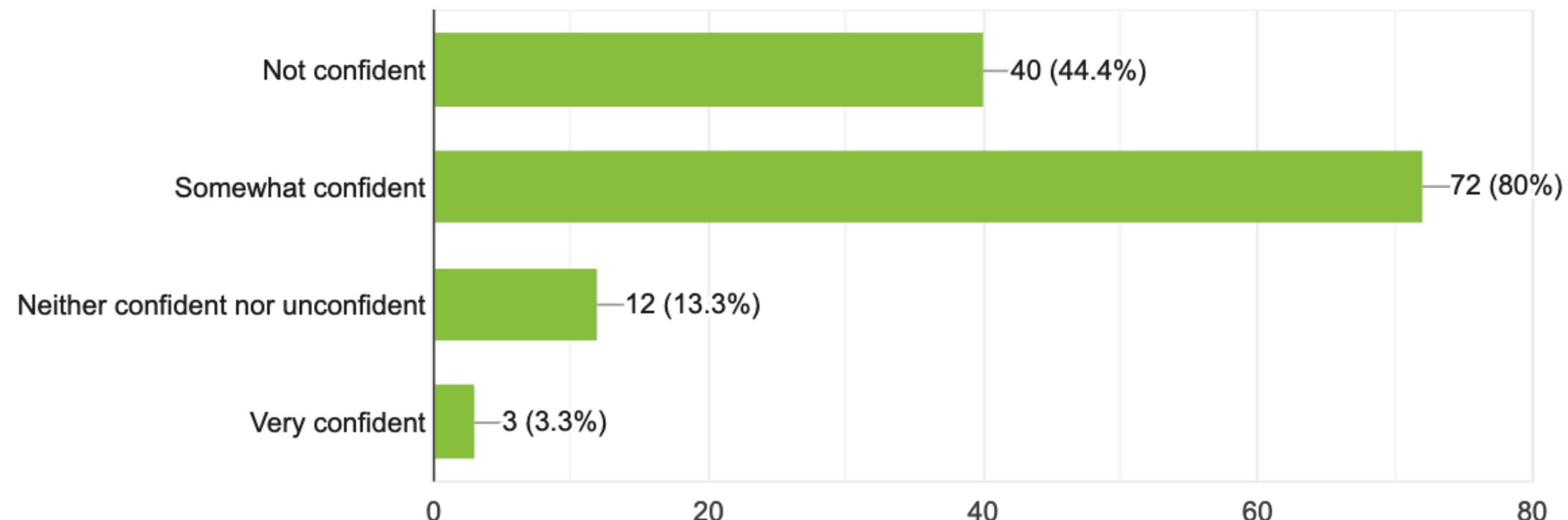
Your Position

90 responses



How confident do you feel using R?

90 responses



For what purpose do you want to learn R?



Getting started...

- R – programming language for statistical computing, data manipulation, and graphics
- RStudio – Integrated Development Environment (IDE) that assists to use R more effectively



R: Engine



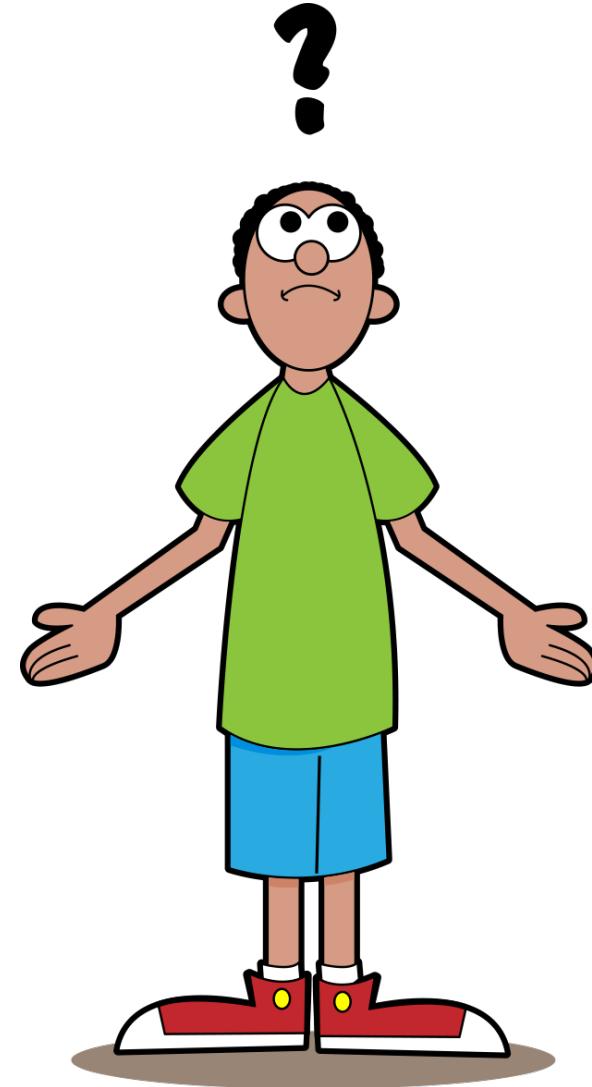
RStudio: Dashboard



<http://moderndrive.com/>

Why R?

- Available for all systems
- Free for everyone
- Open source
- Helpful community
- Flexible and powerful

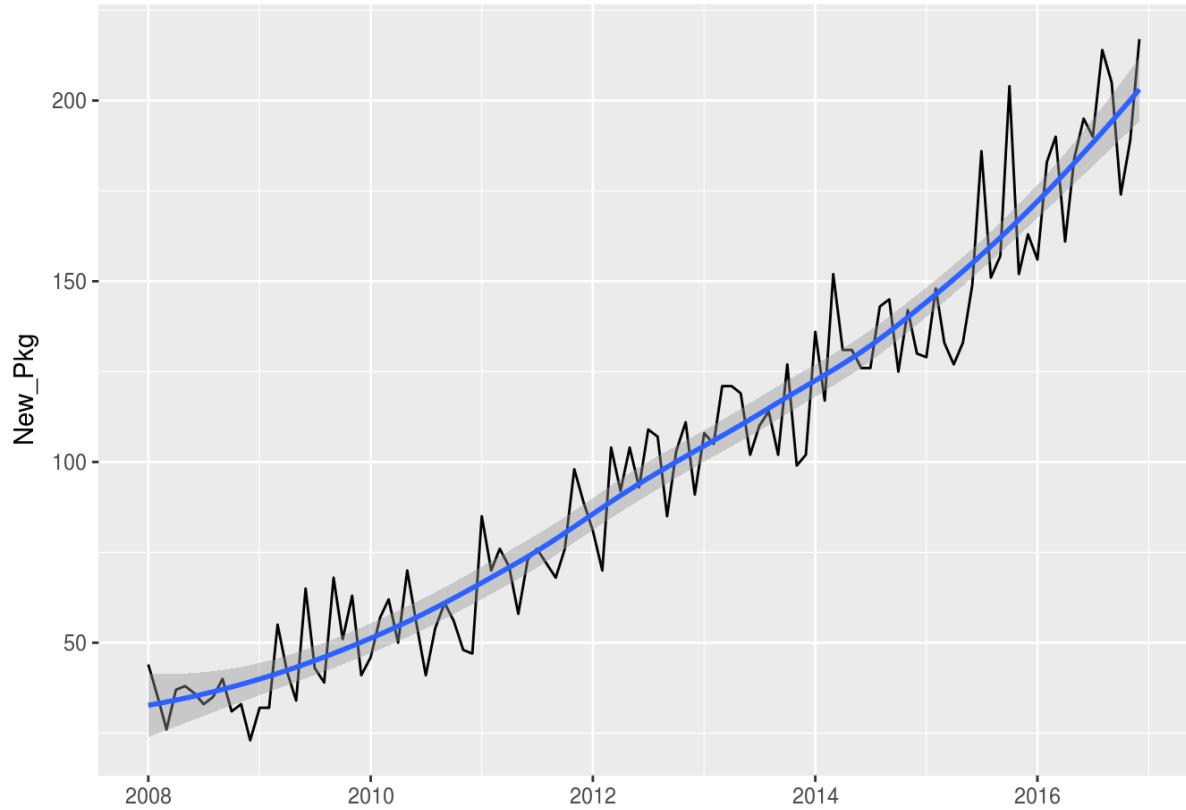


What is an R package?

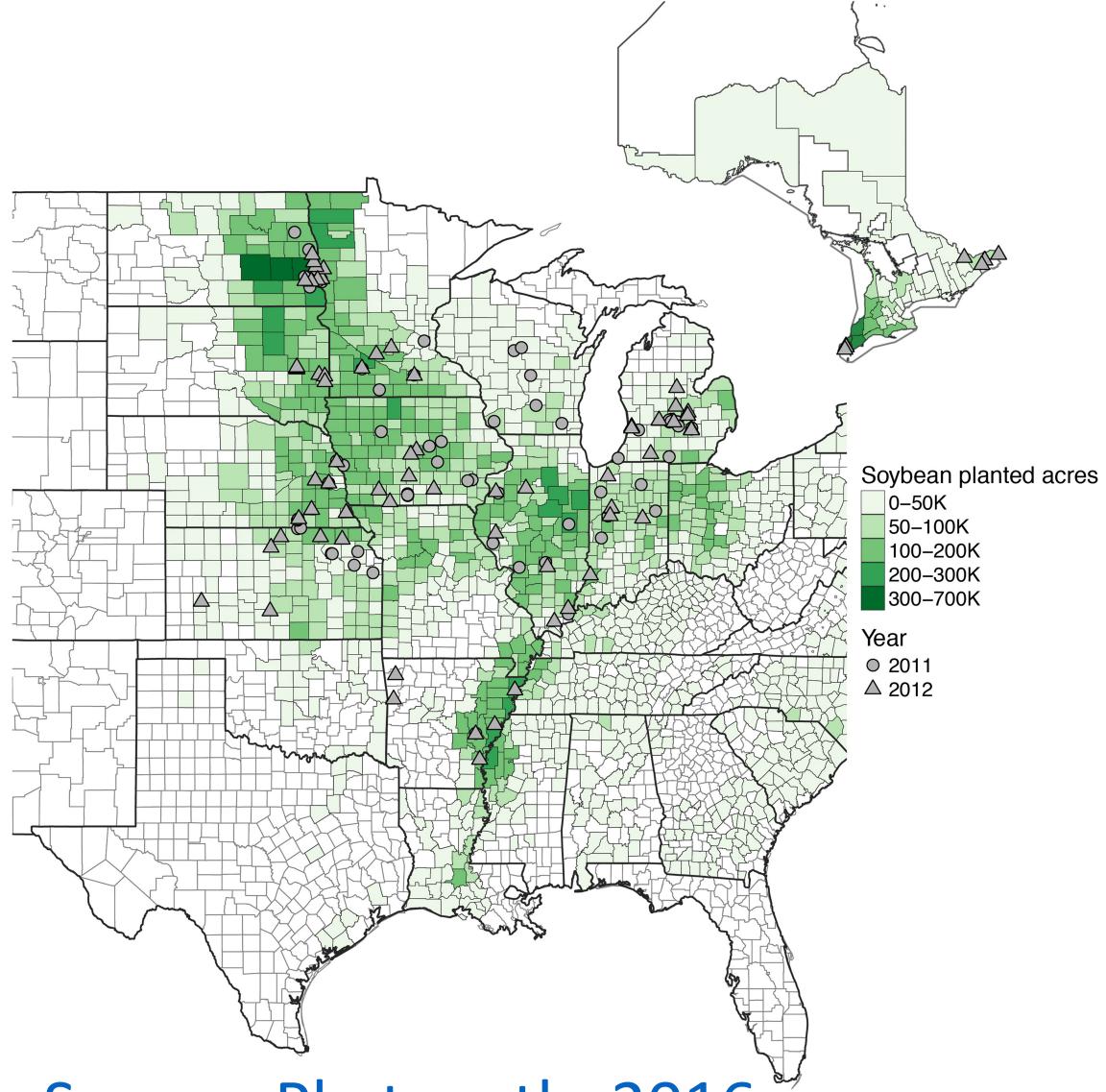
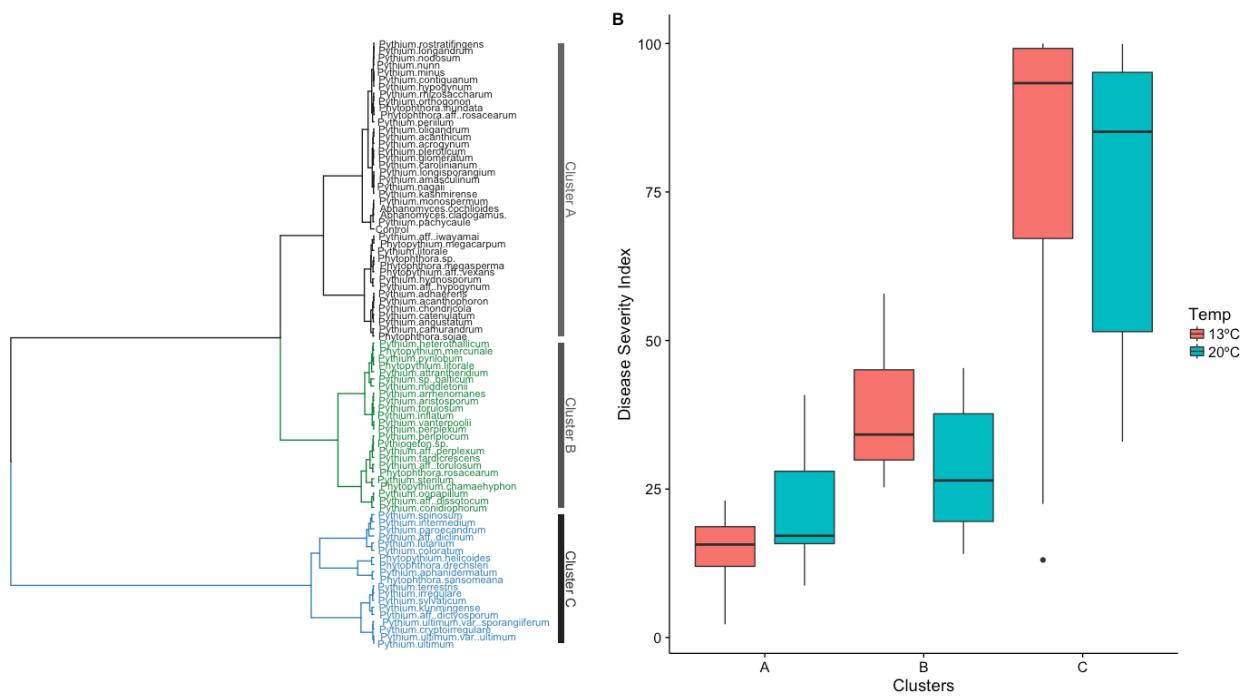
- In R, the fundamental unit of shareable code is the package
- Packages are located in a place called repository so that you can install them from it:
 - [CRAN](#) (**C**omprehensive **R** **A**rchive **N**etwork)
 - [Bioconductor](#)
 - [Github](#)

There are over 10,000 R packages

- Easy to install
`> install.packages("poppr")`
- Easy to use
`> library("poppr")`



<https://rviews.rstudio.com/2017/01/06/10000-cran-packages/>



- [https://github.com/alejorojas2/Rojas Survey Phytopath 2016](https://github.com/alejorojas2/Rojas_Survey_Phytopath_2016)

Fig. 1. Map of sampled soybean fields in 2011 and 2012 and of intensity of planted soybean acres demonstrated by color intensity at the county/parish/province level.

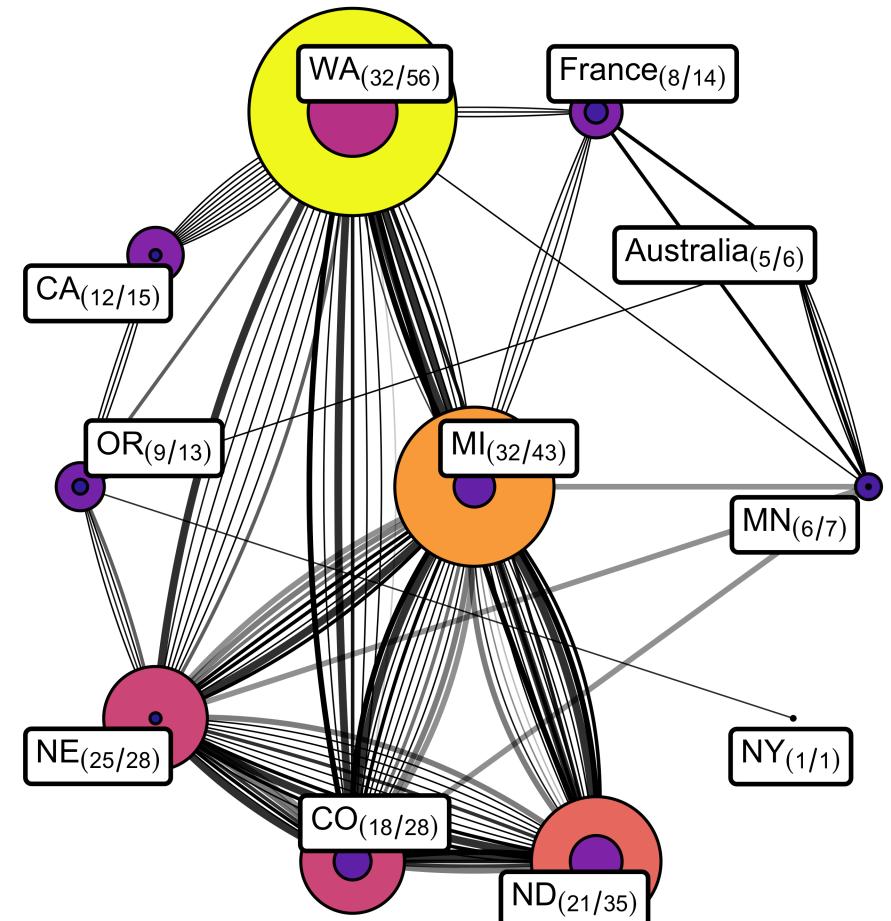
Published in: J. Alejandro Rojas; Janette L. Jacobs; Stephanie Napieralski; Behirda Karaj; Carl A. Bradley; Thomas Chase; Paul D. Esker; Loren J. Giesler; Doug J. Jardine; Dean K. Malvick; Samuel G. Markell; Berlin D. Nelson; Alison E. Robertson; John C. Rupe; Damon L. Smith; Laura E. Sweets; Albert U. Tenuta; Kiersten A. Wise; Martin I. Chilvers; *Phytopathology*. 2017, 107, 280-292.

Markell, Berlin B. Nelson, Allison E. T.
DOI: 10.1094/PHYTO-04-16-0177-R

Copyright © 2017 The American Phytopathological Society

Shared haplotypes across regions

(11 loci)



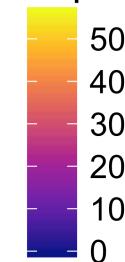
Populations per haplotype

- 2
- 3
- 4
- 5

Probability of second encounter

- 9.1e-02
- 1.7e-02
- 7.9e-04
- 7.1e-05
- 4.0e-08

Number of haplotypes



accidental aRt

Follow accidental-art

tumblr.

2017-01-10

14 notes



When data
visualization goes
beautifully wrong.
@accidental_aRt
Tweet Curated by
@kara_woo and
@ErikaMudrak

SUBMIT A POST

ARCHIVE / RSS

Too many lines make your plot look like Jackson Pollock's painting. via Tim.

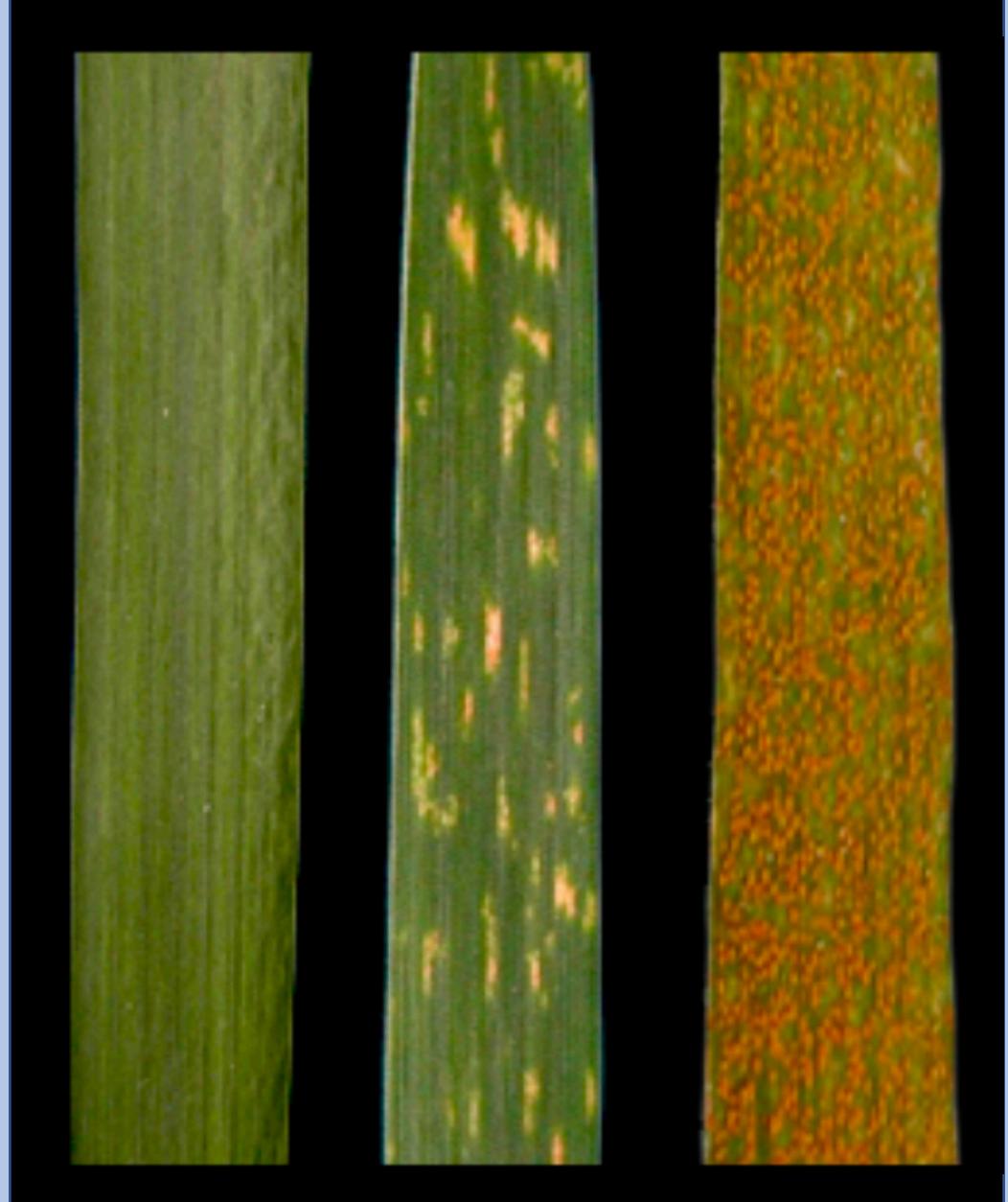
Today's Workshop

- Thinking about data
- How does R think about data?
- Field data example
- Troubleshooting problems
- *Coding at the command line!*
 - Part 1: Introduction to R
 - Part 2: Manipulating and analyzing field data
 - Part 3: Plotting data

QUESTION:

What types of data do we typically collect?

Example:



Example data

	A	B	C	D	E
1	Treatment	Yield_bu_per_acre	Severity	Fungicide	
2	Control	173.82	5.5	FALSE	
3	Control	174.23	5.6	FALSE	
4	Control	173.57	5.4	FALSE	
5	Control	173.61	5.4	FALSE	
6	Control	174.19	5.6	FALSE	
7	Control	173.8	5.5	FALSE	
8	Fungicide_A	173.98	5.1	TRUE	
9	Fungicide_A	174.27	5.2	TRUE	
10	Fungicide_A	173.61	5	TRUE	
11	Fungicide_A	173.88	5	TRUE	
12	Fungicide_A	174.17	5.2	TRUE	
13	Fungicide_A	173.49	5.1	TRUE	
14	Fungicide_B	175.98	4.1	TRUE	
15	Fungicide_B	175.58	3.9	TRUE	
16	Fungicide_B	175.75	4.2	TRUE	
17	Fungicide_B	175.88	4	TRUE	
18	Fungicide_B	175.68	4.1	TRUE	
19	Fungicide_B	175.95	4.2	TRUE	
20					

7 terms we want you to remember

- 1. Numeric
- 2. Character
- 3. Logical
- 4. Objects
- 5. Vectors
- 6. Data Frames
- 7. Functions

Types of data

Organization of data

1. Numeric

- Can be integer
- Can be decimal (aka double)

The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E
1	Treatment	Yield_bu_per_acre	Severity	Fungicide	
2	Control	173.82	5.5	FALSE	
3	Control	174.23	5.6	FALSE	
4	Control	173.57	5.4	FALSE	
5	Control	173.61	5.4	FALSE	
6	Control	174.19	5.6	FALSE	
7	Control	173.8	5.5	FALSE	
8	Fungicide_A	173.98	5.1	TRUE	
9	Fungicide_A	174.27	5.2	TRUE	
10	Fungicide_A	173.61	5	TRUE	
11	Fungicide_A	173.88	5	TRUE	
12	Fungicide_A	174.17	5.2	TRUE	
13	Fungicide_A	173.49	5.1	TRUE	
14	Fungicide_B	175.98	4.1	TRUE	
15	Fungicide_B	175.58	3.9	TRUE	
16	Fungicide_B	175.75	4.2	TRUE	
17	Fungicide_B	175.88	4	TRUE	
18	Fungicide_B	175.68	4.1	TRUE	
19	Fungicide_B	175.95	4.2	TRUE	
20					

2. Character

- Surrounded by quotations
- Can be single word text
- Can be multiple words

```
> print("Control")
```

```
[1] "Control"
```

The screenshot shows a Microsoft Excel spreadsheet with a green header bar. The table has columns labeled A through E. Column A is 'Treatment', B is 'Yield_bu_per_acre', C is 'Severity', D is 'Fungicide', and E is an empty column. Rows 1 through 7 show 'Control' treatments with yields of 173.82, 174.23, 173.57, 173.61, 174.19, 173.8, and 173.98 respectively. Rows 8 through 19 show 'Fungicide_A' treatments with yields of 173.98, 174.27, 173.61, 173.88, 174.17, 173.49, 175.98, 175.58, 175.75, 175.88, 175.68, and 175.95. Rows 20 through 23 are empty. The 'Fungicide' column for rows 8-19 is colored light green, while the rest of the table is white.

	A	B	C	D	E
1	Treatment	Yield_bu_per_acre	Severity	Fungicide	
2	Control	173.82	5.5	FALSE	
3	Control	174.23	5.6	FALSE	
4	Control	173.57	5.4	FALSE	
5	Control	173.61	5.4	FALSE	
6	Control	174.19	5.6	FALSE	
7	Control	173.8	5.5	FALSE	
8	Fungicide_A	173.98	5.1	TRUE	
9	Fungicide_A	174.27	5.2	TRUE	
10	Fungicide_A	173.61	5	TRUE	
11	Fungicide_A	173.88	5	TRUE	
12	Fungicide_A	174.17	5.2	TRUE	
13	Fungicide_A	173.49	5.1	TRUE	
14	Fungicide_B	175.98	4.1	TRUE	
15	Fungicide_B	175.58	3.9	TRUE	
16	Fungicide_B	175.75	4.2	TRUE	
17	Fungicide_B	175.88	4	TRUE	
18	Fungicide_B	175.68	4.1	TRUE	
19	Fungicide_B	175.95	4.2	TRUE	
20					
21					
22					
23					

3. Logical

- TRUE or FALSE
- TRUE equal to 1
- FALSE equal to 0

The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E
1	Treatment	Yield_bu_per_acre	Severity	Fungicide	
2	Control	173.82	5.5	FALSE	
3	Control	174.23	5.6	FALSE	
4	Control	173.57	5.4	FALSE	
5	Control	173.61	5.4	FALSE	
6	Control	174.19	5.6	FALSE	
7	Control	173.8	5.5	FALSE	
8	Fungicide_A	173.98	5.1	TRUE	
9	Fungicide_A	174.27	5.2	TRUE	
10	Fungicide_A	173.61	5	TRUE	
11	Fungicide_A	173.88	5	TRUE	
12	Fungicide_A	174.17	5.2	TRUE	
13	Fungicide_A	173.49	5.1	TRUE	
14	Fungicide_B	175.98	4.1	TRUE	
15	Fungicide_B	175.58	3.9	TRUE	
16	Fungicide_B	175.75	4.2	TRUE	
17	Fungicide_B	175.88	4	TRUE	
18	Fungicide_B	175.68	4.1	TRUE	
19	Fungicide_B	175.95	4.2	TRUE	
20					

4. Objects

At the most basic level, an object can be **anything** that is assigned to a variable

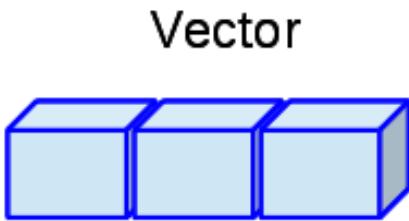
```
> x <- 1:10  
> day <- "Wednesday"  
> D <-read.csv("data.csv")
```



www.megachess.com

Hint: Chess pieces are like **objects** because they are just symbols that have a value (rook, pawn, queen, etc.), which determines how they can be used

5. Vectors

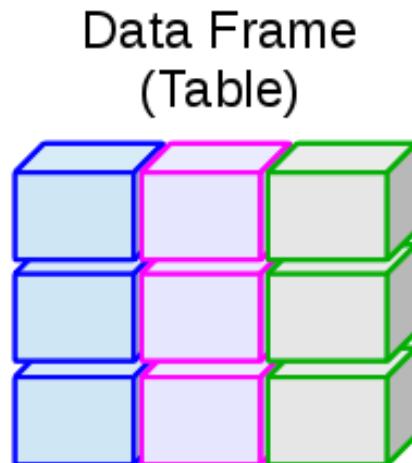
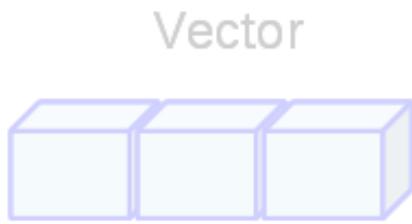


Vectors: one-dimensional arrays
to store data of the same type

Hint: This is like a single column of
data from an Excel spreadsheet
that has one type of data



6. Data Frame



Data Frames: a table where different columns can store different data types

7. Functions

- **Functions:** objects created by the user and reused to make specific operations

$$f(x) = x^2$$

```
> f <- function(x) x^2  
> f(2)  
> f(4)
```



Troubleshooting problems in R

Getting help with R

1. Help Page
2. Vignette
3. Google



help



Introduction to dplyr

When working with data you must:

- Figure out what you want to do.
- Describe those tasks in the form of a computer program.
- Execute the program.

The dplyr package makes these steps fast and easy:

- By constraining your options, it helps you think about your data manipulation challenges.
- It provides simple “verbs”, functions that correspond to the most common data manipulation tasks, to help you translate your thoughts into code.
- It uses efficient backends, so you spend less time waiting for the computer.

This document introduces you to dplyr’s basic set of tools, and shows you how to apply them to data frames. dplyr also supports databases via the dbplyr package, once you’ve installed, read `vignette("dbplyr")` to learn more.

Help Page

Description: brief description.

Usage: for a function, gives the name with all its arguments and the possible options (with the corresponding default values); for an operator gives the typical use.

Arguments: for a function, details each of its arguments.

Details: detailed description.

Value: if applicable, the type of object returned by the function or the operator.

See Also: other help pages close or similar to the present one.

Examples: some examples which can generally be executed without opening the help with the function `example`.

How to access a help page?

- For a function in a package that is
 - already loaded
 - > `?function_name`
 - > `help(function_name)`
 - not currently loaded
 - > `help(function_name, package="package_name")`
- For a package in your library
 - > `help(package="package_name")`

Vignette

- Tutorial demonstrating practical uses of a package along with discussion and interpretation of results
- To find vignettes for a specific package
 - > `browseVignettes("package_name")`
- To view a vignette
 - > `vignettes("vignette_name")`

Google

- Type your query along with the word “R”
- Paste an error message
- Ask your question on Stack Overflow

finding help for a function in R

All Videos Images News Shopping More Settings Tools

About 736,000,000 results (0.53 seconds)

Help function in R

Use the **help()** command. To find information for a particular **function**, such as the **function print**, type **help('print')** on the **R** command line and press enter (I recommend using quotes whenever you use this command, but there are some special cases when they are unnecessary). Apr 21, 2007

Getting Help in R - University of Warwick
https://warwick.ac.uk/fac/sci/moac/degrees/moac/ch923/r_introduction/help/

Search for: Help function in R

[About this result](#) [Feedback](#)

R: Getting Help with R
<https://www.r-project.org/help.html> ▾
The **help()** function and **? help** operator in **R** provide access to the ... engine (also see below) to search for information in **function help** pages and vignettes for all ...
You've visited this page 2 times. Last visit: 6/24/18

Quick-R: Getting Help
<https://www.statmethods.net/interface/help.html> ▾
apropos("foo") # list all **functions** containing string foo example(foo) # show an example of **function** foo. # **search for** foo in **help** manuals and archived mailing ...

r faq - How to get help in R? - Stack Overflow
<https://stackoverflow.com/questions/15289995/how-to-get-help-in-r> ▾
5 answers
Mar 8, 2013 - There are also **help** pages for datasets, general topics and some packages. ?iris ?Syntax



Let's Get Coding!

Summary of Part 1

- <- assigns value to an object
- c() – combines values into a vector
- str() – compactly displays structure of an object
- length() – gives length of an object
- head() – returns first part of an object
- [] – gives specified elements of an object
- \$ – returns specified columns of a data frame
- ? – provides help page for a function or package

Join us for Part 2 next week!

Recorded webinar available online in about 1 week at:

<https://www.apsnet.org/edcenter/resources/Webinars>

- **Intro to R for Plant Pathologists: Part 1**

Wednesday, February 19, 2020

11:30 a.m. - 1:30 p.m. Central

- **Intro to R for Plant Pathologists: Part 2**

Wednesday, February 26, 2020

11:30 a.m. - 1:30 p.m. Central



Data manipulation & analysis

Visualization and graphics

Starting your own R Project



Introduction to R for Plant Pathologists

APS Online Webinar

Sydney Everhart¹, Nikita Gambhir¹, and Katie Gold²

¹University of Nebraska-Lincoln

²Cornell University

February 19th and 26th, 2020