



R functions and tidy evaluation

Date: 10th February 2023

Presenter: Balogun Stephen

About me

- Public health expert and data scientist
- Certified Rstudio tidyverse and shiny instructor
- Co-organizer Abuja R user group
- Author of two R packages {tidyndr} and {checkndr}
- Social media
 - [LinkedIn](#): Stephen Balogun
 - [Twitter](#) : @eppydoc
 - [GitHub](#): stephenbalogun
 - [Website](#)



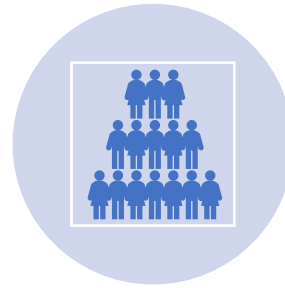
Outline

- Definition
- Function structure/components
- Why write functions
- Types of functions
- More on dataframe functions
- Practice
- Q & A

Assumptions



Basic understanding of R



However, anyone should be able to follow along

- Functions are reusable blocks of code that perform specific tasks and can be called upon to run those tasks multiple times.
- Functions allow you to automate common tasks in a more powerful and general way than copy-and-pasting.

Definition



Function structure

```
name <- function(arguments) {  
  body  
}
```

fx name

argument

arg name

value

filter

```
<- function (.data, ..., .by = NULL, .preserve = FALSE)  
{  
  check_by_typo(...)  
  by <- enquo(.by)  
  if (!quo_is_null(by) && !is_false(.preserve)) {  
    abort("Can't supply both `.by` and `.preserve`.")  
  }  
  UseMethod("filter")  
}
```

Body

fx

Why write functions?

- Functions allow you to automate common tasks. As requirements change, you only need to update code in one place, instead of many.
- Reduces code duplication and waste
- Functions can have names that makes your code easier to understand.
- Eliminate the mistakes that can happen due to copy and paste

Types of functions



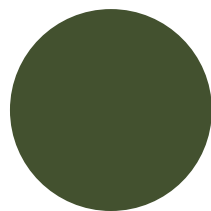
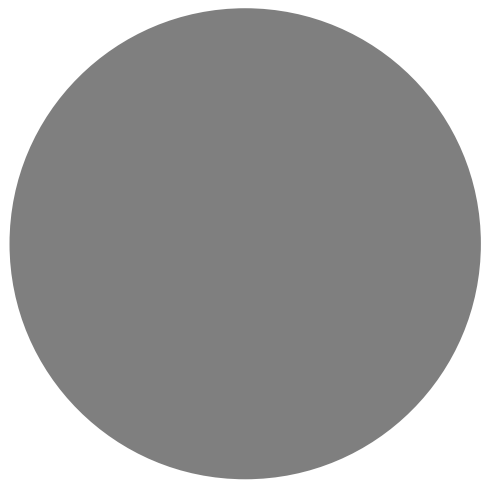
Vector functions



Dataframe functions



Plot functions



More on dataframe functions

Source

Console Terminal Background Jobs

Data variables

```
R 4.2.2 · ~/My R/Abuja_r/r functions and tidyeval/
```

```
> mpg
```

```
# A tibble: 234 × 11
```

	manufacturer	model	displ	year	cyl	trans	drv	cty	hwy	fl	class
	<chr>	<chr>	<dbl>	<int>	<int>	<chr>	<chr>	<int>	<int>	<chr>	<chr>
1	audi	a4	1.8	1999	4	auto(l5)	f	18	29	p	compact
2	audi	a4	1.8	1999	4	manual(m5)	f	21	29	p	compact
3	audi	a4	2	2008	4	manual(m6)	f	20	31	p	compact
4	audi	a4	2	2008	4	auto(av)	f	21	30	p	compact
5	audi	a4	2.8	1999	6	auto(l5)	f	16	26	p	compact
6	audi	a4	2.8	1999	6	manual(m5)	f	18	26	p	compact
7	audi	a4	3.1	2008	6	auto(av)	f	18	27	p	compact
8	audi	a4 quattro	1.8	1999	4	manual(m5)	4	18	26	p	compact
9	audi	a4 quattro	1.8	1999	4	auto(l5)	4	16	25	p	compact
10	audi	a4 quattro	2	2008	4	manual(m6)	4	20	28	p	compact

```
# ... with 224 more rows
```

```
# i Use `print(n = ...)` to see more rows
```

Environment History Connections Git Tutorial

Import Dataset 208 MiB

R Global Environment

Data

mpg 234 obs. of 11 variables

Values

x	2
y	int [1:7] 4 5 6 7 8 9 10

Environment variables

Variables

- Environment variables
- Dataframe variables

{tidyverse} and tidy evaluation

- Most base R functions are explicit in the type of variables referenced.
- {tidyverse} functions use tidy evaluation - a framework for writing R code that is both readable and flexible.
- Tidy evaluation makes code easy to read and understand by making variable referencing implicit
 - Data-masking: used by {tidyverse} functions that alter rows e.g., `filter()`, `group_by()`, `arrange()`
 - Tidy selection: used by {tidyverse} functions that alter columns e.g., `select()`, `rename()`, `relocate()`, `across()`
- However, this can create a problem when used in R functions - indirection

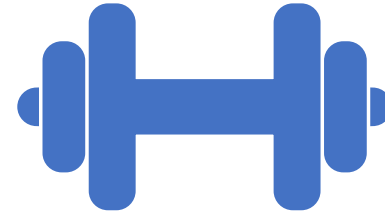


Practice

Scenario one: vector function



You are working with a data containing age of children (0-15). You would want to clean the data such that any data less than 0 or greater than 15 is recoded as missing



You are also aware that the weight should be between 0 and 45. So weights outside the range should be converted to missing

Scenario two: dataframe function for a specific task



summarise the diamonds dataset cut by price




summarise the diamonds dataset cut by depth



summarise the diamonds dataset color by price



summarise the diamonds dataset color by depth

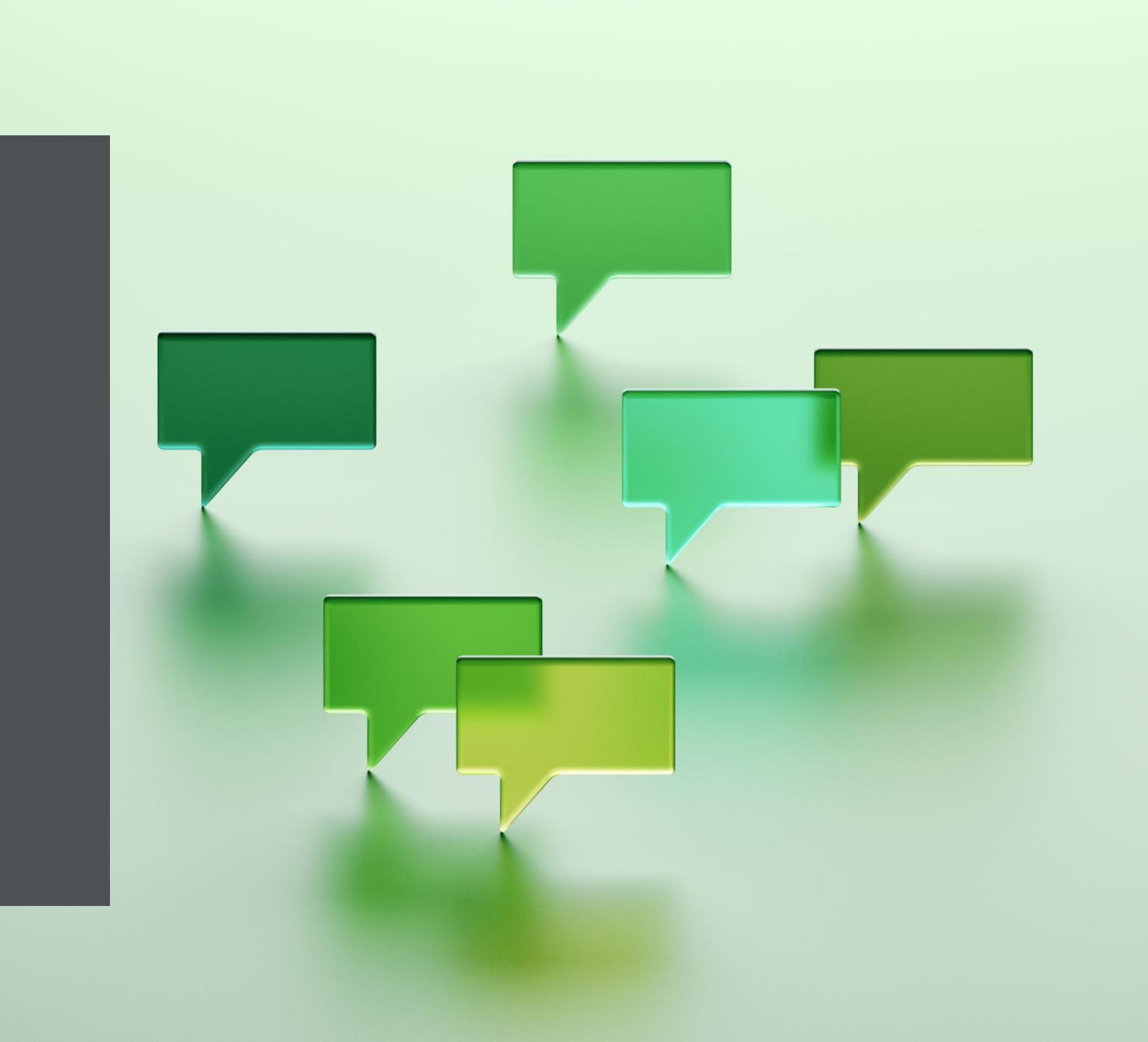
A photograph of a classical marble column and its base, with steps leading up to it. The image is partially obscured by a dark grey vertical bar on the left side.

Scenario three: make the function more generic

Reduce the four functions from previous slide
to only one



**Scenario four:
supply more
than one values
as arguments**

A decorative graphic on the left side of the slide. It features a dark grey vertical bar on the far left. To its right, several green speech bubbles of various shades (dark green, medium green, light green) are arranged in a scattered pattern. The background of the left half is a light green gradient.

**Scenario five:
customize the
variable names
in your output**



Additional resources

- <https://rlang.r-lib.org/reference/topic-data-mask.html>
- <https://dcl-prog.stanford.edu/tidy-eval-section.html>
- <https://dplyr.tidyverse.org/articles/programming.html>