

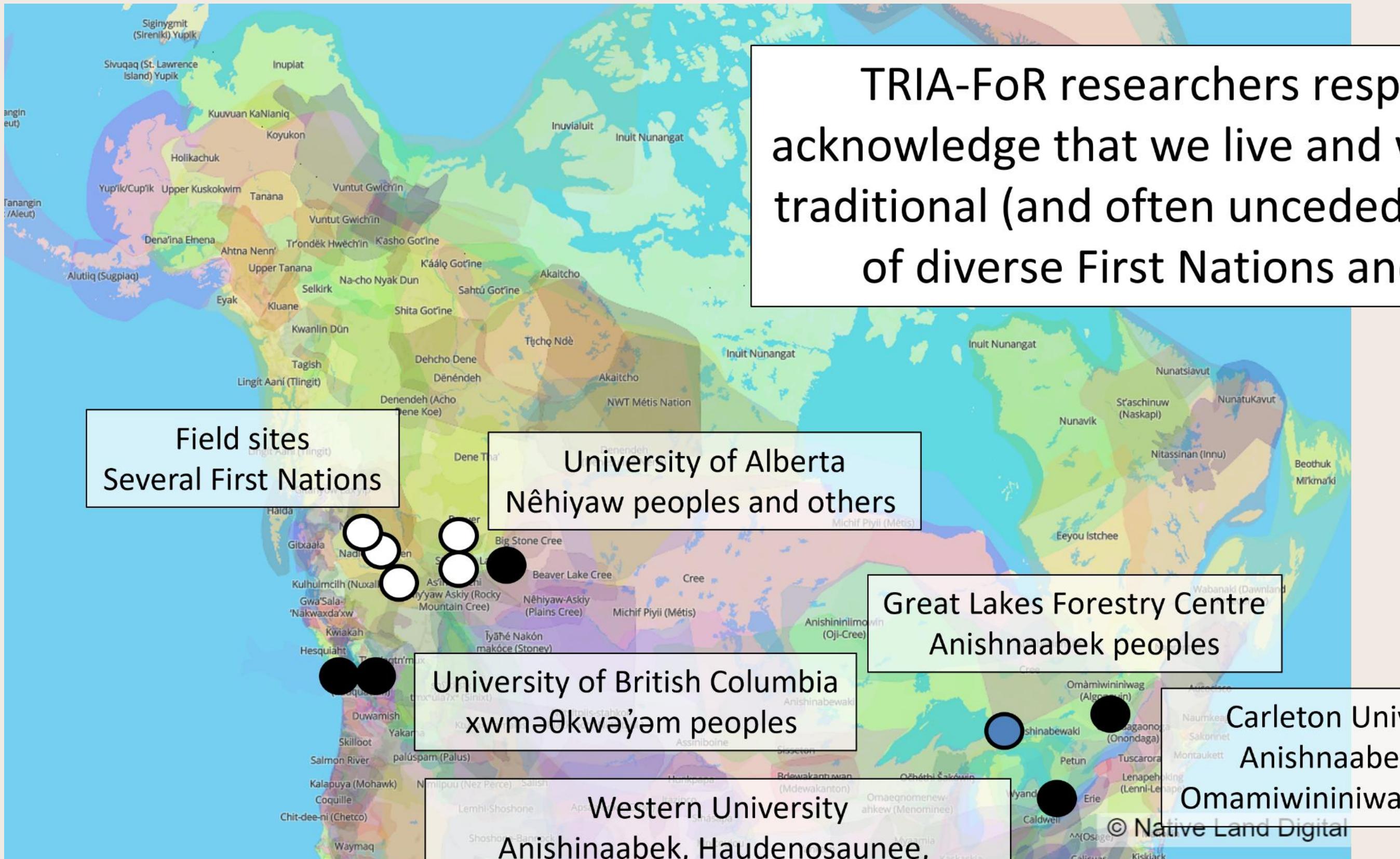


DATA WRANGLING

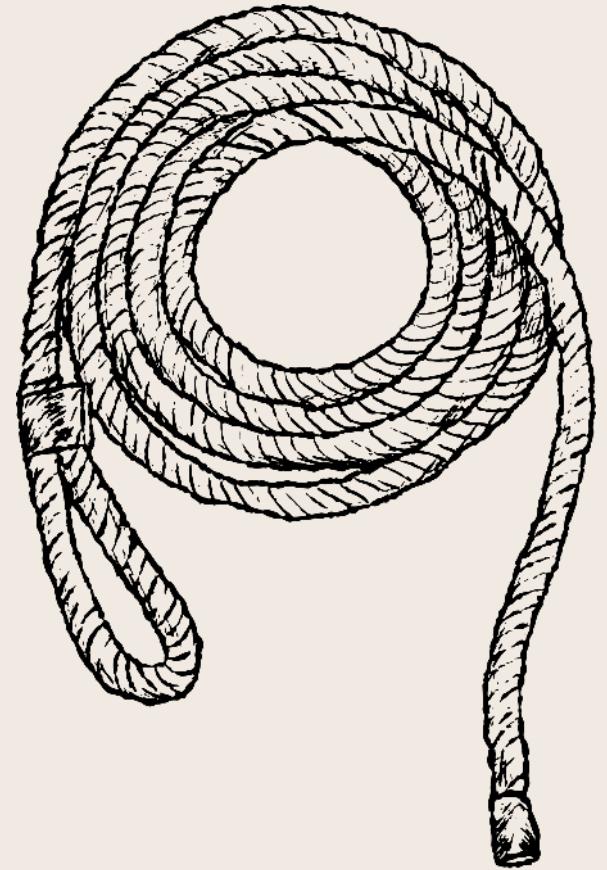
★ IN R WITH THE TIDYVERSE ★



LAND ACKNOWLEDGEMENT



TRIA-FoR researchers respectfully acknowledge that we live and work on the traditional (and often unceded) territories of diverse First Nations and Métis



WHAT IS “TIDY” DATA?

more info on
tidy data



click to be
taken to link!

1. Each variable must have its own column
2. Each observation must have its own row
3. Each value must have its own cell

country	year	cases	population
Afghanistan	1990	745	19087071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	21666	128042583

variables

country	year	cases	population
Afghanistan	1990	745	19087071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	216766	128042583

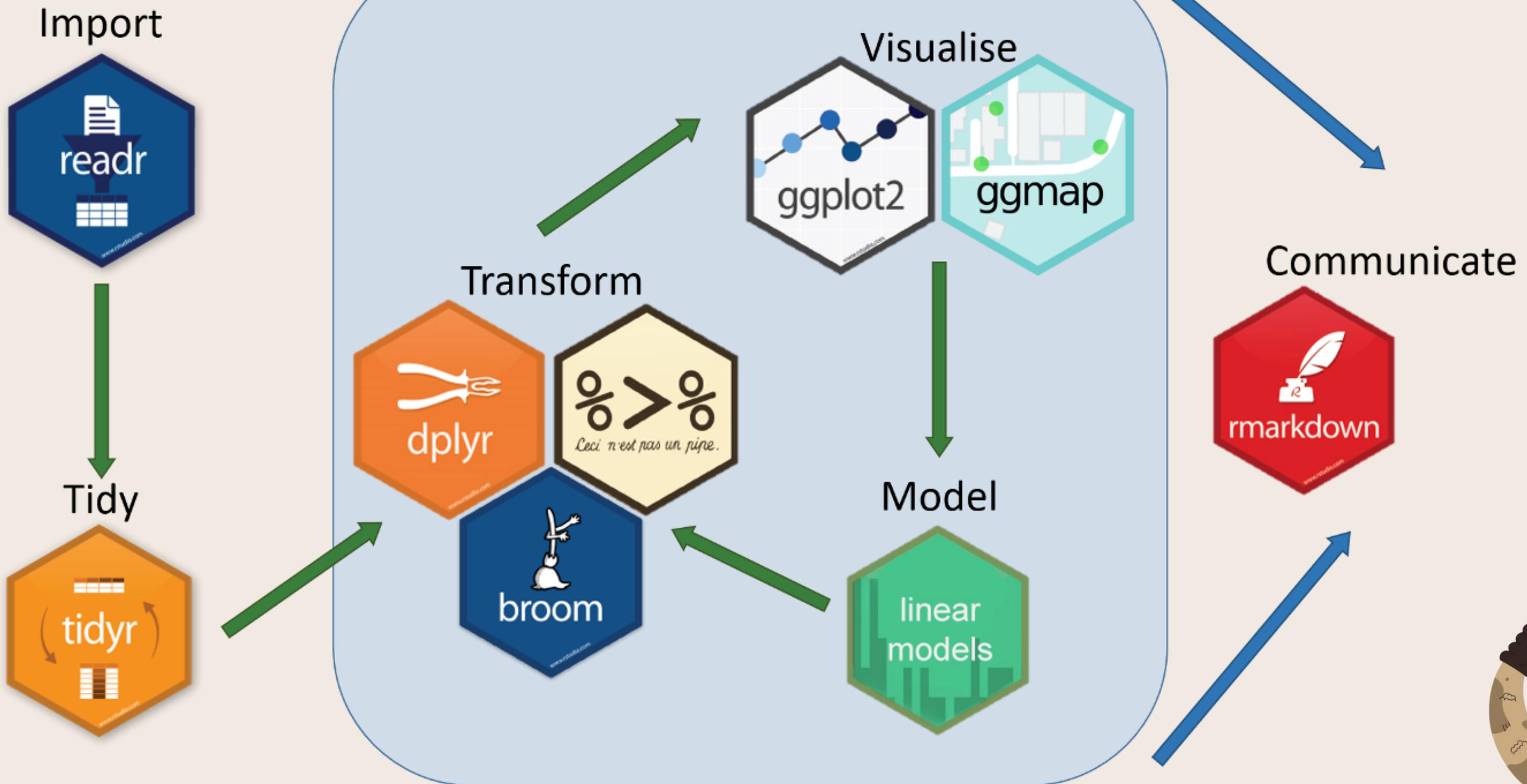
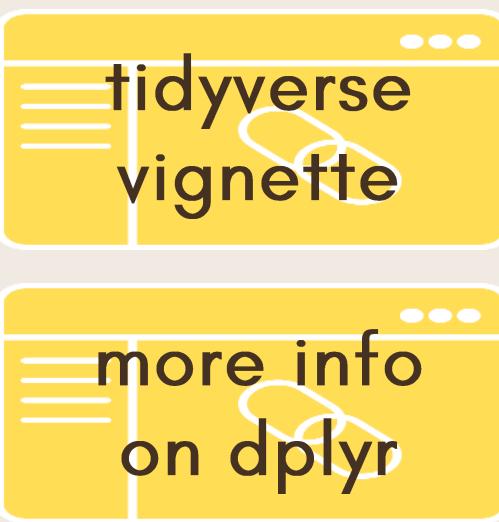
observations

country	year	cases	population
Afghanistan	1990	745	19087071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	216766	128042583

values



THE TIDYVERSE



HOW TO GET SEMINAR DATA

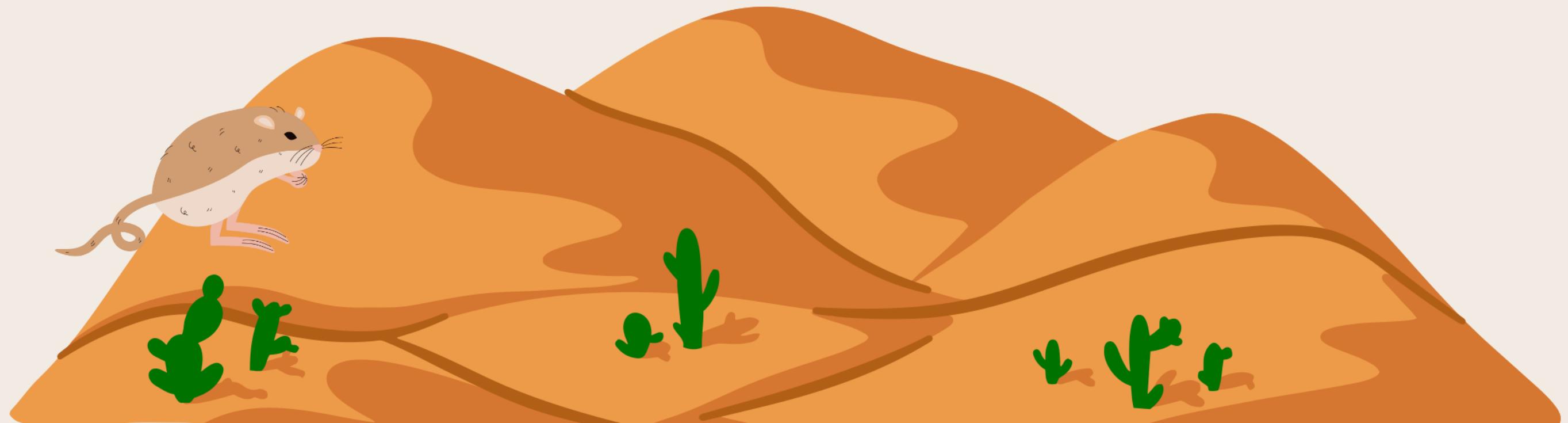
#Download Rproject.zip file from my [github](#)



#Unzip the folder and open the .rproj file

#Rprojects use relative file paths

#you only need to install packages, then you're good to run!



option for
reading excel
files

USING READR IN RSTUDIO



Tidy_seminar_Rproject - main - RStudio

Addins ▾

R tidyverse

From Text (base)...
From Text (readr)...
From Excel...
From SPSS...
From SAS...
From Stata...

Environment History Connections

Run

Source

Tidy_seminar_script_Apr17-2025.R*

```
1 #R.version: 4.4.1 last run by Colleen Apr. 15, 2025
2
3 #install.packages("tidyverse") #only run once, don't typically need to include in script
4 #repeat for any other packages you don't have installed already
5
6 library(tidyverse) #v2.0.0-
7 library(janitor) #v2.2.1, for clean_names()
8 library(scales) #v1.3.0, for scale_x_date() on graph
9
10 ¶
```

Files Plots Packages Help View

Console Background Jobs

R 4.4.1 ~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject/

R version 4.4.1 (2024-06-14) -- "Race for Your Life"
Copyright (C) 2024 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin20

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.

“Good Code” Tip:
include version info

USING READR IN RSTUDIO

Tidy_seminar_Rproject - main - RStudio

File/URL: ~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject/data/untidy_seminar_data.csv

Import Text Data

Data Preview:

Tree (double)	Date Collected (character)	Bud 1 Wax (mg) (character)	Bud 1 Fresh Weight (g) (character)	Bud 1 Dry Weight (mg) (character)	Bud 2 Wax (mg) (character)
57	Feb7-2025	0.235	Guess	6.142	0.194
56	Feb7-2025	0.713	Character	5.759	0.420
35	Feb7-2025	0.191	Double	6.070	0.149
97	Feb7-2025	0.519	Integer	6.029	0.338
45	Feb7-2025	0.524	Numeric	5.759	0.652
28	Feb7-2025	0.624	Logical	5.817	0.592
48	Feb7-2025	0.116	Date	5.943	0.216
89	Feb7-2025	0.229	Time	6.271	0.303

Previewing first 50 entries.

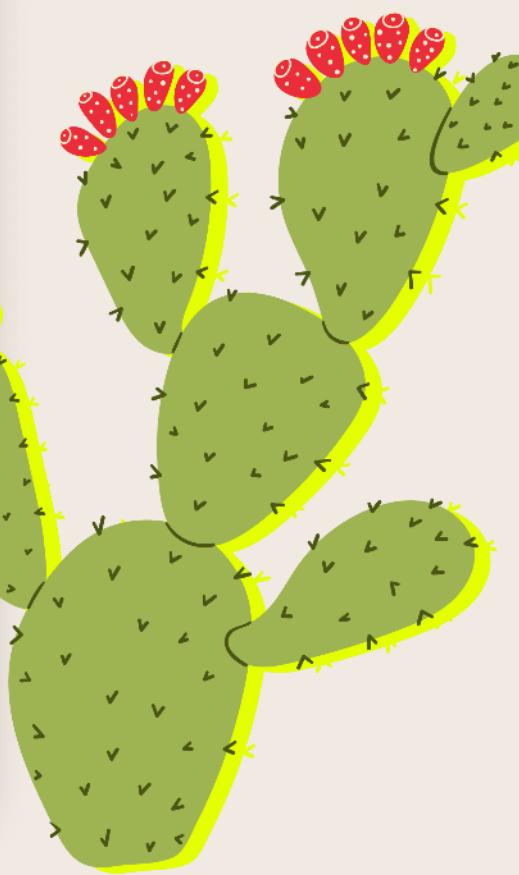
Import Options:

- Name: untidy_seminar_data
- First Row as Names
- Trim Spaces
- Open Data Viewer

Code Preview:

```
library(readr)
untidy_seminar_data
<- read_csv("data/untidy_seminar_data.csv")
View(untidy_seminar_
```

Import Cancel



USING READR IN RSTUDIO

Tidy_seminar_Rproject - main - RStudio

File/URL: ~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject/data/untidy_seminar_data.csv

Data Preview:

Tree (double)	Date Collected (character)	Bud 1 Wax (mg) (character)	Bud 1 Fresh Weight (g) (character)	Bud 1 Dry Weight (mg) (character)	Bud 2 Wax (mg) (character)
57	Feb7-2025	0.235	0.092704	6.142	0.194
56	Feb7-2025	0.713	0.113864	5.759	0.420
35	Feb7-2025	0.191	0.089647	6.070	0.149
97	Feb7-2025	0.519	0.143839	6.029	0.338
45	Feb7-2025	0.524	0.122457	5.759	0.652
28	Feb7-2025	0.624	0.122933	5.817	0.592
48	Feb7-2025	0.116	0.091300	5.943	0.216
89	Feb7-2025	0.229	0.149313	6.271	0.303

Previewing first 50 entries.

Import Options:

Name: untidy_seminar_data	<input checked="" type="checkbox"/> First Row as Names	Delimiter: Comma	Escape: None
Skip: 0	<input checked="" type="checkbox"/> Trim Spaces	Quotes: Default	Comment: Default
	<input checked="" type="checkbox"/> Open Data Viewer	Locale: Configure...	NA: Default

Code Preview:

```
library(readr)
untidy_seminar_data
<- read_csv("data/untidy_seminar_data.csv")
View(untidy_seminar_
```

Import Cancel

25

LET'S TAKE A LOOK AT OUR DATA

Tidy_seminar_Rproject - main - RStudio

Tidy_seminar_script_Apr17-2025.R* untidy_seminar_data

Tree Date Collected Bud 1 Wax (mg) Bud 1 Fresh Weight (g) Bud 1 Dry Weight (mg) Bud 2 Wax (mg) Bud 2

1	57	Feb7-2025	0.235	0.092704	6.142	0.194	0.1157
2	56	Feb7-2025	0.713	0.113864	5.759	0.420	0.0882
3	35	Feb7-2025	0.191	0.089647	6.070	0.149	0.1005
4	97	Feb7-2025	0.519	0.143839	6.029	0.338	0.0843
5	45	Feb7-2025	0.524	0.122457	5.759	0.652	0.0912
6	28	Feb7-2025	0.624	0.122933	5.817	0.592	0.1122
7	48	Feb7-2025	0.116	0.091300	5.943	0.216	0.1307
8	89	Feb7-2025	0.229	0.149313	6.271	0.303	0.1116
9	83	Feb7-2025	0.778	0.122321	5.750	0.781	0.0850
10	17	Feb7-2025	0.459	0.102999	5.972	0.359	0.1491
11	99	Feb7-2025	0.158	0.084239	5.875	0.160	0.1031
12	10	Feb23-2025	1.484	0.087159	6.585	1.129	0.1318
13	14	Feb23-2025	1.359	0.110853	6.118	1.083	0.1203
14	85	Feb23-2025	0.809	0.106154	6.493	1.473	0.1187

Showing 1 to 14 of 33 entries, 12 total columns

Console Background Jobs

R 4.4.1 · ~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject/

Column specification

Delimiter: ","

chr (8): Date Collected, Bud 1 Wax (mg), Bud 1 Fresh Weight (g), Bud 1 Dry Weight (mg), Bud 2 Wax (...
dbl (4): Tree, Bud 3 Wax (mg), Bud 3 Fresh Weight (g), Bud 3 Dry Weight (mg)

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.

> View(untidy_seminar_data)
>

Environment History Connections
190 MiB
Global Environment
Data
untidy_seminar_data 33 obs. of 12 variables
Files Plots Packages Help View
data figures r_scripts Tidy_seminar_Rproject.Rproj
Name
..
data
figures
r_scripts
Tidy_seminar_Rproject.Rproj



PIPES

more info
about pipes

#Old version (magrittr): %>%

#New version: |>

#RStudio shortcut: CMD+SHIFT+M (Mac)
CTRL+SHIFT+M (Windows)

#There's instructions on how to change the RStudio default at the end of the slides

PIPES

note the
relative file
path (b/c
Rproject)



Tidy_seminar_Rproject - main - RStudio

Tidy_seminar_script_Apr17-2025.R*

```
1 #R.version: 4.4.1 · last · run · by · Colleen · Apr · 15, · 2025
2 -
3 #install.packages("tidyverse") · #only · run · once, · don't · typically · need · to · include · in · script ·
4 #repeat · for · any · other · packages · you · don't · have · installed · already ·
5 -
6 library(tidyverse) · #v2.0.0 ·
7 library(janitor) · #v2.2.1, · for · clean_names() ·
8 library(scales) · #v1.3.0, · for · scale_x_date() · on · graph ·
9 -
10 #read · in · the · untidy · dataset ·
11 untidy_seminar_data · <- · read_csv("data/untidy_seminar_data.csv") · |> ·
```

Environment History Connections 177 MiB Global Environment

Environment is empty

Files Plots Packages Help View

Parse_seminar > Tidy_seminar_Rproject R ...

Name

- ..
- data
- figures
- r_scripts
- Tidy_seminar_Rproject.Rproj

11:68 (Top Level) R Script

Console Background Jobs

R 4.4.1 ~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject/

```
R version 4.4.1 (2024-06-14) -- "Race for Your Life"
Copyright (C) 2024 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin20

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.
```

"Good Code" Tip:
annotate using #



COMMON “TIDIES”: RENAME

#Base R:

```
>names(df)[names(df) == "old_name"] <- "new_name"
```

#tidyR:

```
>rename(df, new_name = old_name)
```

#if your old name has any spaces, make sure to include “ ”

#NOTE: df only needed when not using pipes, so won't be included going forward

more info
about janitor

COMMON “TIDIES”: RENAMING



#One more package (not included in the tidyverse):
janitor
`>clean_names()`

COMMON “TIDIES”: RENAME

The screenshot shows the RStudio interface with a project titled "Tidy_seminar_Rproject". The left pane displays a script named "Tidy_seminar_script_Apr17-2025.R" containing the following code:

```
1 #R.version.4.4.1.last.run.by.Coleen.Apr.15, 2025
2 -
3 #install.packages("tidyverse") #only run once, don't typically need to include in script
4 #repeat.for.any.other.packages.you.don't.have.installed.already-
5 -
6 library(tidyverse) #v2.0.0-
7 library(janitor) #v2.2.1, for clean_names()
8 library(scales) #v1.3.0, for scale_x_date() on graph-
9 -
10 #read.in.the.untidy.dataset:-
11 untidy_seminar_data <- read_csv("data/untidy_seminar_data.csv") |>-
12 #tidy.up.the.column.names-
13 janitor::clean_names()-
14 ¶
```

The right pane shows the "Environment" tab with a data frame named "untidy_seminar_data" containing 33 observations and 12 variables. The "Files" tab shows the project structure with folders for "data", "figures", "r_scripts", and the project itself.

COMMON “TIDIES”: RENAME

The screenshot shows the RStudio interface with a project titled "Tidy_seminar_Rproject". The left pane displays two tabs: "Tidy_seminar_script_Apr17-2025.R*" and "untidy_seminar_data". The "untidy_seminar_data" tab is active, showing a data frame with 14 rows and 8 columns. The columns are: tree, date_collected, bud_1_wax_mg, bud_1_fresh_weight_g, bud_1_dry_weight_mg, bud_2_wax_mg, bud_2_fr, and bud_2_fresh_weight_g. The data includes measurements for different trees collected on Feb7-2025 and Feb23-2025. The right pane shows the "Environment" tab with the "Global Environment" section containing the "untidy_seminar_data" object. The bottom pane shows the "Files" tab with a list of project files: "data", "figures", "r_scripts", and "Tidy_seminar_Rproject.Rproj". The status bar at the bottom indicates "R 4.4.1 · ~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject/".

	tree	date_collected	bud_1_wax_mg	bud_1_fresh_weight_g	bud_1_dry_weight_mg	bud_2_wax_mg	bud_2_fr
1	57	Feb7-2025	0.235	0.092704	6.142	0.194	0.115702
2	56	Feb7-2025	0.713	0.113864	5.759	0.420	0.088289
3	35	Feb7-2025	0.191	0.089647	6.070	0.149	0.100540
4	97	Feb7-2025	0.519	0.143839	6.029	0.338	0.084323
5	45	Feb7-2025	0.524	0.122457	5.759	0.652	0.091293
6	28	Feb7-2025	0.624	0.122933	5.817	0.592	0.112265
7	48	Feb7-2025	0.116	0.091300	5.943	0.216	0.130703
8	89	Feb7-2025	0.229	0.149313	6.271	0.303	0.111607
9	83	Feb7-2025	0.778	0.122321	5.750	0.781	0.085000
10	17	Feb7-2025	0.459	0.102999	5.972	0.359	0.149106
11	99	Feb7-2025	0.158	0.084239	5.875	0.160	0.103160
12	10	Feb23-2025	1.484	0.087159	6.585	1.129	0.131840
13	14	Feb23-2025	1.359	0.110853	6.118	1.083	0.120338
14	85	Feb23-2025	0.809	0.106154	6.493	1.473	0.118735

Showing 1 to 14 of 33 entries, 12 total columns

Console Background Jobs ×

R 4.4.1 · ~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject/ ↗

Rows: 33 Cols: 12

Environment History Connections

173 MiB Global Environment

untidy_seminar_data 33 obs. of 12 variables

Files Plots Packages Help View

data figures r_scripts Tidy_seminar_Rproject.Rproj

Name

Si: ..

25

more info on
mutating

COMMON “TIDIES”: MUTATING

#create a new column or modify an old column

```
>mutate(new_column = *modification*[existing_column])
```

#modification can be a formula (i.e. “3+”) or function

#can be used to change column type or set as a factor (with levels



COMMON “TIDIES”: REPLACING

#removing/replacing with NA

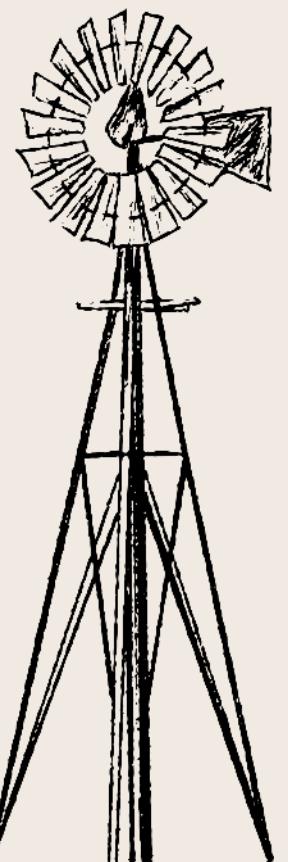
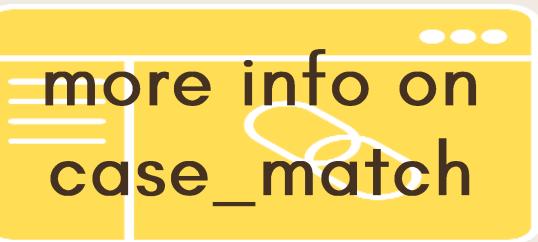
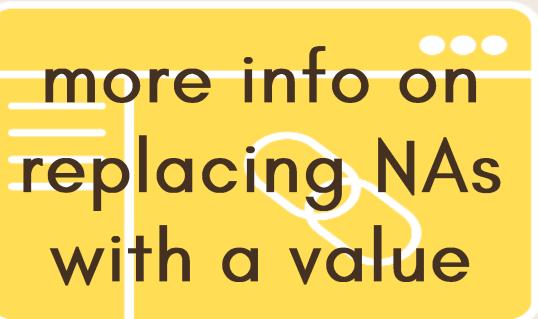
```
>na_if(value_or_string_to_replace)
```

#if it's a string, put it in “quotes”

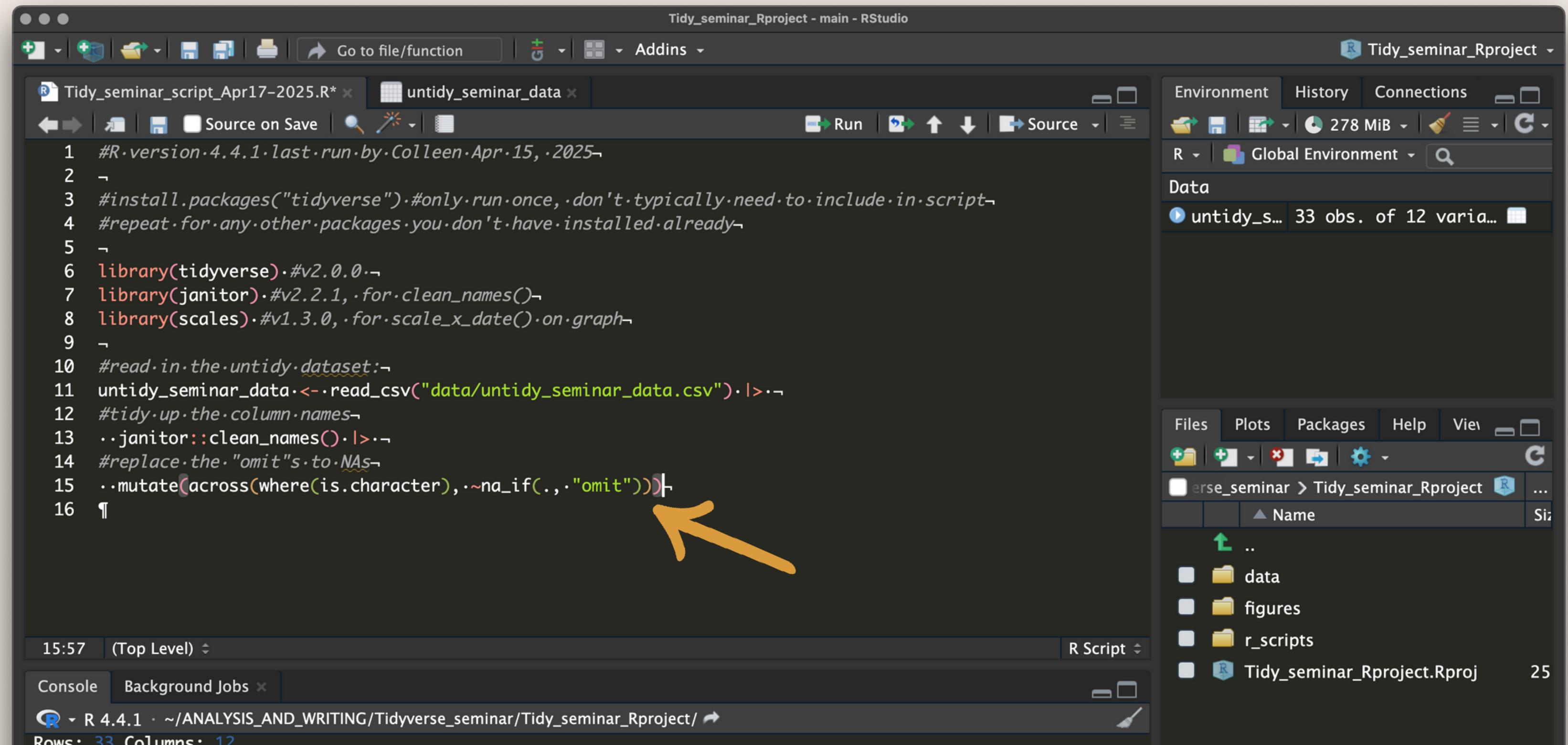
#replacing strings

```
>case_match(original_string ~ new_string,  
.default=column_name)
```

#don't forget the default, it fills in everything not replaced



COMMON “TIDIES”: REPLACING



The screenshot shows an RStudio interface with the following details:

- Title Bar:** Tidy_seminar_Rproject - main - RStudio
- File Tab:** Tidy_seminar_script_Apr17-2025.R*
- Code Editor:** The script contains R code for setting up the environment and reading a dataset.

```
1 #R.version.4.4.1.last.run.by.Coleen.Apr.15, 2025
2 -
3 #install.packages("tidyverse") #only run once, don't typically need to include in script
4 #repeat for any other packages you don't have installed already
5 -
6 library(tidyverse) #v2.0.0-
7 library(janitor) #v2.2.1, for clean_names()
8 library(scales) #v1.3.0, for scale_x_date() on graph
9 -
10 #read in the untidy dataset-
11 untidy_seminar_data <- read_csv("data/untidy_seminar_data.csv") |>-
12 #tidy up the column names-
13 janitor::clean_names() |>-
14 #replace the "omit"s to NAs-
15 mutate(across(where(is.character), ~na_if(., "omit")))
```

- Environment Tab:** Shows the Global Environment with a dataset named `untidy_seminar_data` containing 33 observations and 12 variables.
- Files Tab:** Shows the project structure: `Tidy_seminar > Tidy_seminar_Rproject`.
- Console Tab:** Shows the R version and current working directory.
- Page Number:** 25

A large orange arrow points from the bottom left towards the code editor area, highlighting the script's purpose.

COMMON “TIDIES”: REPLACING

The screenshot shows the RStudio interface with the project "Tidy_seminar_Rproject" open. In the center, a data frame titled "untidy_seminar_data" is displayed in a grid format. The columns are labeled: tree, date_collected, bud_1_wax_mg, bud_1_fresh_weight_g, bud_1_dry_weight_mg, bud_2_wax_mg, bud_2_fresh_weight_g, bud_2_dry_weight_mg, bud_3_wax_mg, bud_3_fresh_weight_g, bud_3_dry_weight_mg, and bud_4_wax_mg. The rows contain numerical values for each variable across 33 observations. An orange arrow points to the value "0.12503" in the "bud_1_fresh_weight_g" column of the 27th row. The right side of the screen shows the Environment, Data, Files, and Help panes.

tree	date_collected	bud_1_wax_mg	bud_1_fresh_weight_g	bud_1_dry_weight_mg	bud_2_wax_mg	bud_2_fresh_weight_g	bud_2_dry_weight_mg	bud_3_wax_mg	bud_3_fresh_weight_g	bud_3_dry_weight_mg	bud_4_wax_mg
19	6 Feb23-2025	0.763	0.095920	6.511	1.051	0.116277					
20	79 Feb23-2025	0.801	0.079822	6.474	0.870	0.083921					
21	96 Feb23-2025	1.152	0.088969	6.628	1.145	0.136308					
22	3 Feb23-2025	1.245	0.120130	6.682	1.259	0.114121					
23	43 Mar15-2025	1.611	0.177418	7.396	1.534	0.123918					
24	16 Mar15-2025	1.636	0.111890	6.940	1.542	0.104470					
25	41 Mar15-2025	1.384	0.093303	7.051	1.387	0.078832					
26	8 Mar15-2025	NA	NA	NA	1.550	0.134922					
27	36 Mar15-2025	1.564	0.12503	6.794	1.331	0.108223					
28	66 Mar15-2025	1.419	0.115823	6.859	1.520	0.152771					
29	70 Mar15-2025	1.752	0.115335	6.895	1.703	0.096307					
30	68 Mar15-2025	1.723	0.168143	7.216	1.496	0.150708					
31	82 Mar15-2025	1.651	0.116472	6.457	1.595	0.135611					
32	31 Mar15-2025	1.355	0.159025	6.802	1.587	0.122456					

Showing 18 to 31 of 33 entries, 12 total columns

Console Background Jobs ×

R 4.4.1 · ~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject/ ↵

Rows: 33 Columns: 12

Environment History Connections

278 MiB Global Environment

untidy_seminar.Rproj 33 obs. of 12 variables

Files Plots Packages Help View

Tidy_seminar_Rproject.Rproj

Name Size

..

data

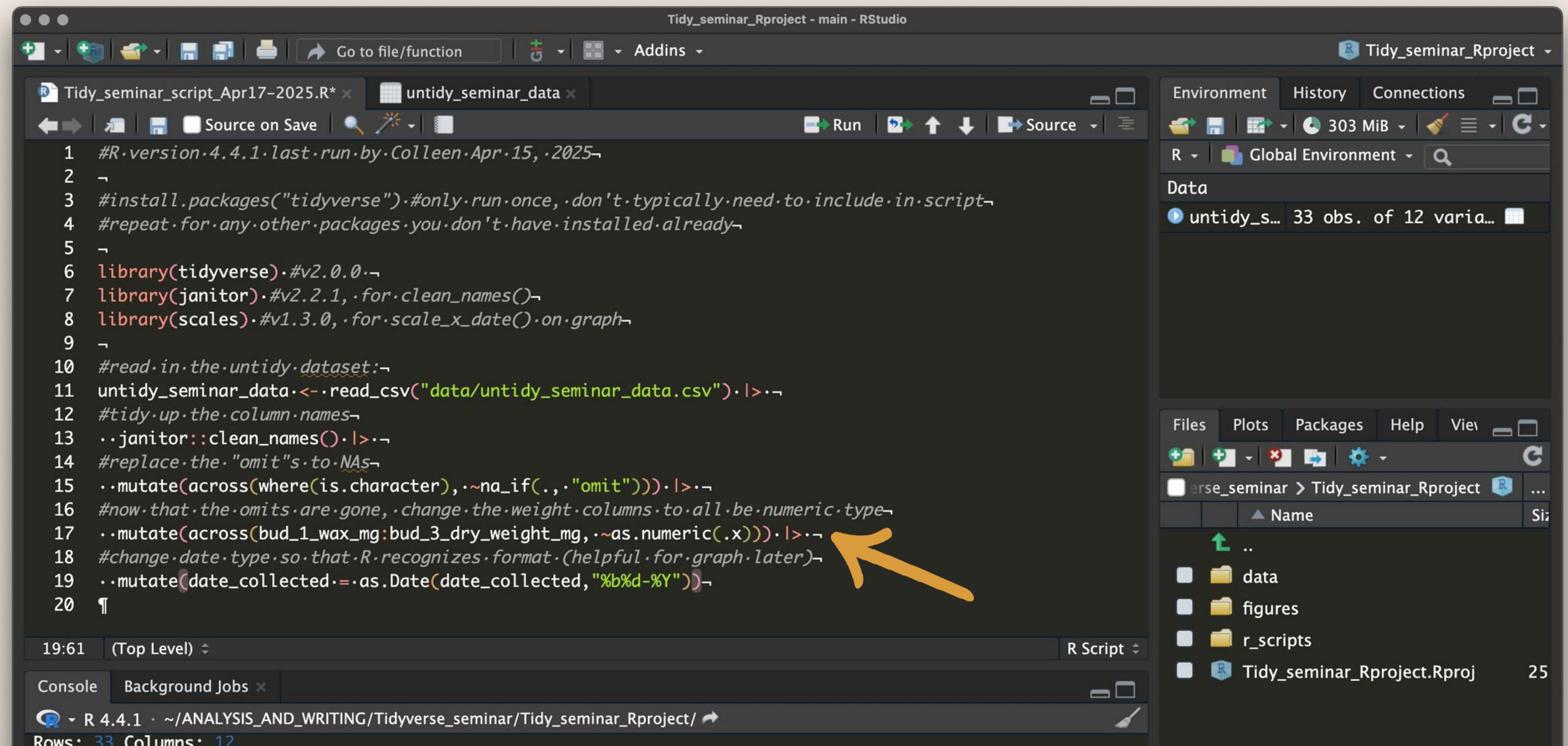
figures

r_scripts

Tidy_seminar_Rproject.Rproj

25

COMMON “TIDIES”: MUTATING

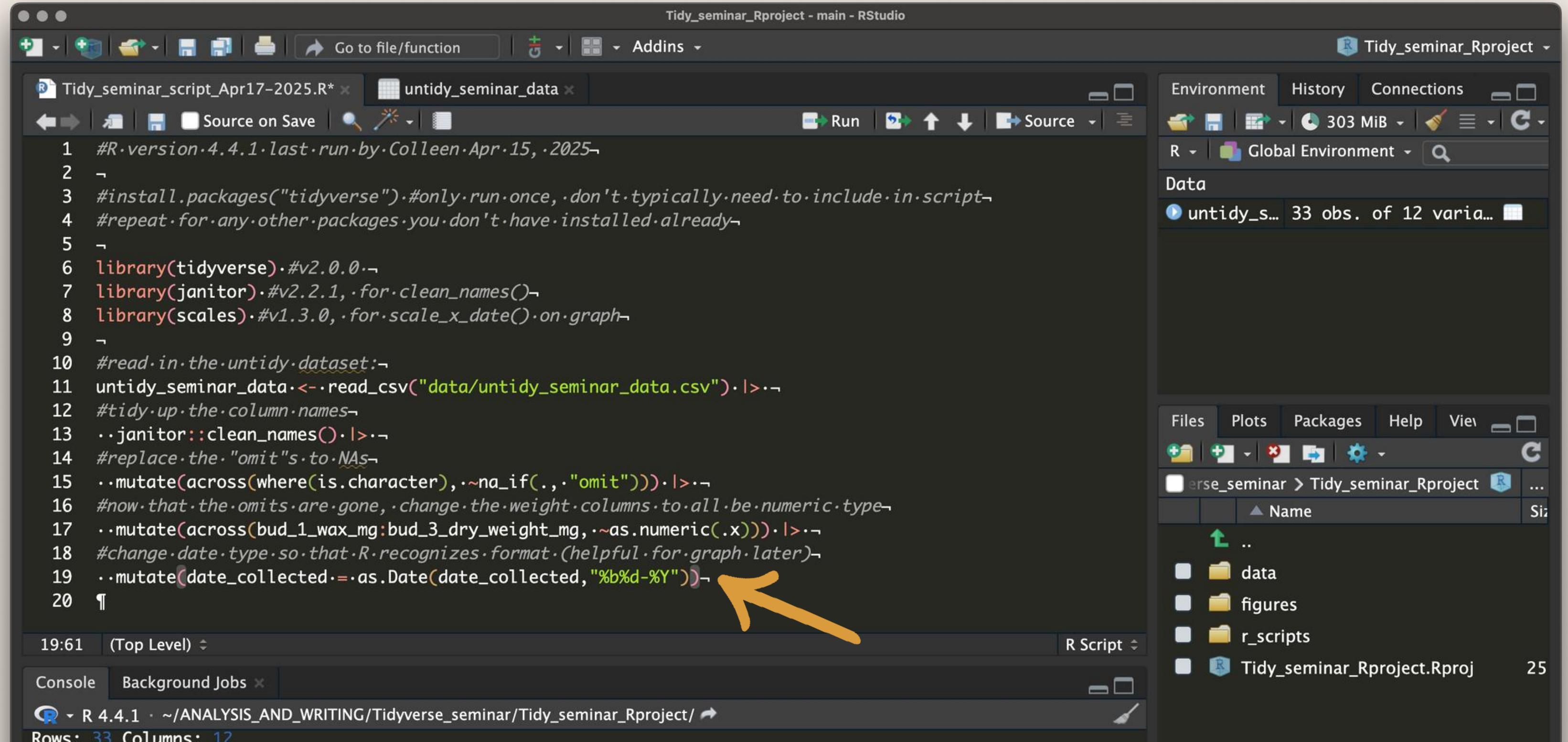


The screenshot shows an RStudio interface with the following details:

- Title Bar:** Tidy_seminar_Rproject - main - RStudio
- File Tab:** Tidy_seminar_script_Apr17-2025.R*
- Code Editor:** The script contains R code for data manipulation. An orange arrow points to the last line of code, which converts the 'date_collected' column to a Date object using the `as.Date` function with the format "%b%d-%Y".
- Environment Tab:** Shows the 'Global Environment' tab with a data frame named 'untidy_seminar_data' containing 33 observations and 12 variables.
- Files Tab:** Shows the project structure: Tidy_seminar > Tidy_seminar_Rproject. It includes folders for 'data', 'figures', 'r_scripts', and the project file 'Tidy_seminar_Rproject.Rproj'.
- Console Tab:** Shows the R version (4.4.1) and the current working directory (~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject).
- Status Bar:** Shows 'Rows: 33 Columns: 12'.

```
1 #R.version: 4.4.1.last.run.by:Colleen.Apr.15, 2025
2 -
3 #install.packages("tidyverse") #only run once, don't typically need to include in script
4 #repeat for any other packages you don't have installed already
5 -
6 library(tidyverse) #v2.0.0-
7 library(janitor) #v2.2.1, for clean_names()
8 library(scales) #v1.3.0, for scale_x_date() on graph
9 -
10 #read in the untidy dataset:
11 untidy_seminar_data <- read_csv("data/untidy_seminar_data.csv") |>-
12 #tidy up the column names
13 janitor::clean_names() |>-
14 #replace the "omit"s to NAs
15 mutate(across(where(is.character), ~na_if(., "omit")))|>-
16 #now that the omits are gone, change the weight columns to all be numeric type
17 mutate(across(bud_1_wax_mg:bud_3_dry_weight_mg, ~as.numeric(.x))|>-
18 #change date type so that R recognizes format (helpful for graph later)
19 mutate(date_collected = as.Date(date_collected, "%b%d-%Y"))|>-
20
```

COMMON “TIDIES”: MUTATING



The screenshot shows an RStudio interface with the following details:

- Title Bar:** Tidy_seminar_Rproject - main - RStudio
- File Tab:** Tidy_seminar_script_Apr17-2025.R*
- Code Editor:** The script contains R code for tidying seminar data. An orange arrow points to the line: `date_collected = as.Date(date_collected, "%b%d-%Y")`.
- Environment Tab:** Shows the Global Environment with a data frame named `untidy_seminar_data` containing 33 observations and 12 variables.
- Files Tab:** Shows the project structure: Tidy_seminar > Tidy_seminar_Rproject. It includes a `data` folder, a `figures` folder, an `r_scripts` folder, and the `Tidy_seminar_Rproject.Rproj` file.
- Console Tab:** Shows the R version (R 4.4.1) and the current working directory (~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject).
- Status Bar:** Shows the number of rows (33) and columns (12).

COMMON “TIDIES”: MUTATING

The screenshot shows the RStudio interface with a project titled "Tidy_seminar_Rproject". The main window displays a data frame named "untidy_seminar_data" with 33 rows and 12 columns. The columns are: tree, date_collected, bud_1_wax_mg, bud_1_fresh_weight_g, bud_1_dry_weight_mg, bud_2_wax_mg, bud_2_fresh_weight_g, bud_2_dry_weight_mg, bud_3_wax_mg, bud_3_fresh_weight_g, bud_3_dry_weight_mg, and bud_4_wax_mg. The "date_collected" column contains dates like "2025-02-07" and "2025-02-23". An orange arrow points to the first row of the "date_collected" column. The right sidebar shows the "Environment" tab with "Global Environment" selected, displaying the "untidy_seminar_data" object. The bottom left shows the "Console" tab with the R version "R 4.4.1" and the current working directory " ~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject/". The bottom right shows the page number "25".

tree	date_collected	bud_1_wax_mg	bud_1_fresh_weight_g	bud_1_dry_weight_mg	bud_2_wax_mg	bud_2_fresh_weight_g	bud_2_dry_weight_mg	bud_3_wax_mg	bud_3_fresh_weight_g	bud_3_dry_weight_mg	bud_4_wax_mg
1	57	2025-02-07	0.235	0.092704	6.142	0.194					
2	56	2025-02-07	0.713	0.113864	5.759	0.420					
3	35	2025-02-07	0.191	0.089647	6.070	0.149					
4	97	2025-02-07	0.519	0.143839	6.029	0.338					
5	45	2025-02-07	0.524	0.122457	5.759	0.652					
6	28	2025-02-07	0.624	0.122933	5.817	0.592					
7	48	2025-02-07	0.116	0.091300	5.943	0.216					
8	89	2025-02-07	0.229	0.149313	6.271	0.303					
9	83	2025-02-07	0.778	0.122321	5.750	0.781					
10	17	2025-02-07	0.459	0.102999	5.972	0.359					
11	99	2025-02-07	0.158	0.084239	5.875	0.160					
12	10	2025-02-23	1.484	0.087159	6.585	1.129					
13	14	2025-02-23	1.359	0.110853	6.118	1.083					
14	85	2025-02-23	0.809	0.106154	6.493	1.473					

Showing 1 to 14 of 33 entries, 12 total columns

Console Background Jobs x

R 4.4.1 · ~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject/ ↵

Rows: 33 Columns: 12

Environment History Connections

Global Environment

untidy_seminar_data 33 obs. of 12 variables

Files Plots Packages Help View

Tidy_seminar_Rproject.Rproj

25

more info on
pivoting

COMMON “TIDIES”: PIVOTING

`>pivot_longer()`

#a “tidy”

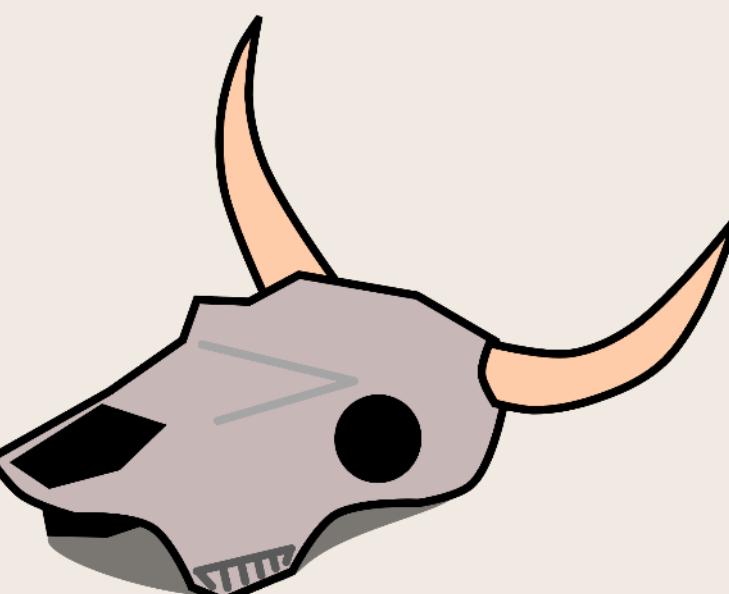
essential!

`>pivot_wider()`

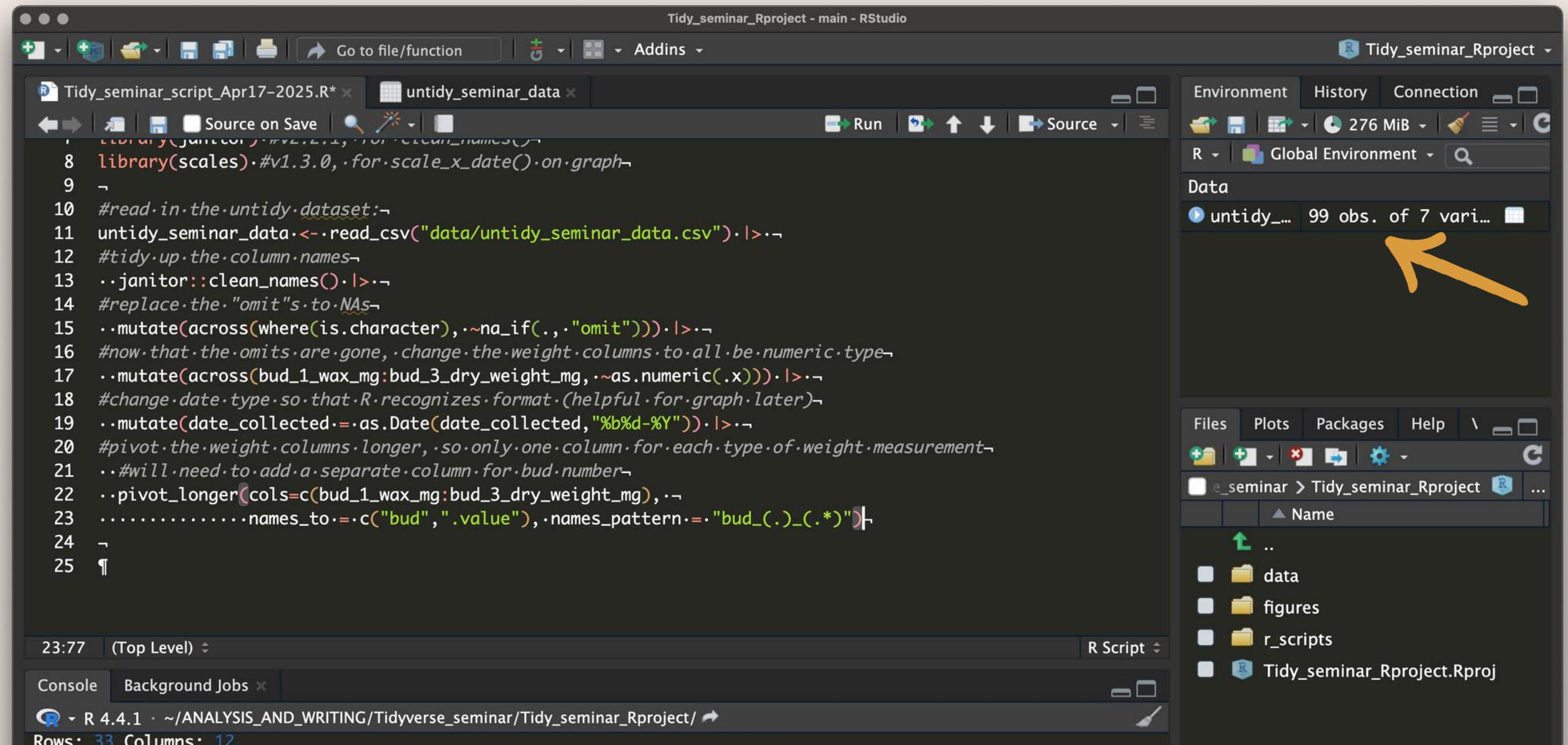
<https://github.com/gadenbuie/tidyexplain>

long		
id	key	val
1	x	a
2	x	b
1	y	c
2	y	d
1	z	e
2	z	f

wide			
id	x	y	z
1	a	c	e
2	b	d	f



COMMON “TIDIES”: PIVOTING



The screenshot shows an RStudio interface with the following details:

- Title Bar:** Tidy_seminar_Rproject - main - RStudio
- File Bar:** Tidy_seminar_script_Apr17-2025.R*
- Code Editor:** A script titled "untidy_seminar_data" containing R code for data manipulation. The code includes loading packages, reading a CSV file, cleaning names, replacing NAs, mutating columns, changing date types, and pivoting longer weight columns. A yellow arrow points from the bottom right towards the Global Environment pane.
- Environment Tab:** Shows the "untidy_seminar_data" dataset with 99 observations and 7 variables.
- Global Environment:** A list of objects including "data", "figures", "r_scripts", and "Tidy_seminar_Rproject".
- Bottom Status Bar:** R 4.4.1 · ~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject/

COMMON “TIDIES”: PIVOTING

The screenshot shows the RStudio interface with the following details:

- Title Bar:** Tidy_seminar_Rproject - main - RStudio
- File Tab:** Tidy_seminar_script_Apr17-2025.R*
- Data View:** untidy_seminar_data
- Table Content:** A data frame with 99 rows and 7 columns. The columns are: tree, date_collected, comments, bud, wax_mg, fresh_weight_g, and dry_weight_mg.
- Annotations:** Two orange arrows point from the text "date_collected" and "bud" in the title to the corresponding columns in the data view.
- Environment Tab:** Shows 276 MiB of memory usage.
- Data Tab:** Shows 99 obs. of 7 variables.
- Files Tab:** Shows project structure: data, figures, r_scripts, and Tidy_seminar_Rproject.Rproj.
- Console Tab:** R 4.4.1 - ~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject/

	tree	date_collected	comments	bud	wax_mg	fresh_weight_g	dry_weight_mg
1	57	2025-02-07	NA	1	0.235	0.092704	6.142
2	57	2025-02-07	NA	2	0.194	0.115702	5.812
3	57	2025-02-07	NA	3	0.156	0.088767	5.711
4	56	2025-02-07	NA	1	0.713	0.113864	5.759
5	56	2025-02-07	NA	2	0.420	0.088289	6.007
6	56	2025-02-07	NA	3	0.375	0.082331	5.832
7	35	2025-02-07	NA	1	0.191	0.089647	6.070
8	35	2025-02-07	NA	2	0.149	0.100540	5.868
9	35	2025-02-07	NA	3	0.219	0.118564	6.244
10	97	2025-02-07	NA	1	0.519	0.143839	6.029
11	97	2025-02-07	NA	2	0.338	0.084323	5.935
12	97	2025-02-07	NA	3	0.414	0.079606	6.083
13	45	2025-02-07	NA	1	0.524	0.122457	5.759
14	45	2025-02-07	NA	2	0.652	0.091293	5.727

Showing 1 to 14 of 99 entries, 7 total columns



COMMON “TIDIES”:

FILTER

#can filter for specific values or strings

```
>filter(column_name == "string_you_want_to_keep")
```

#or values you don't want

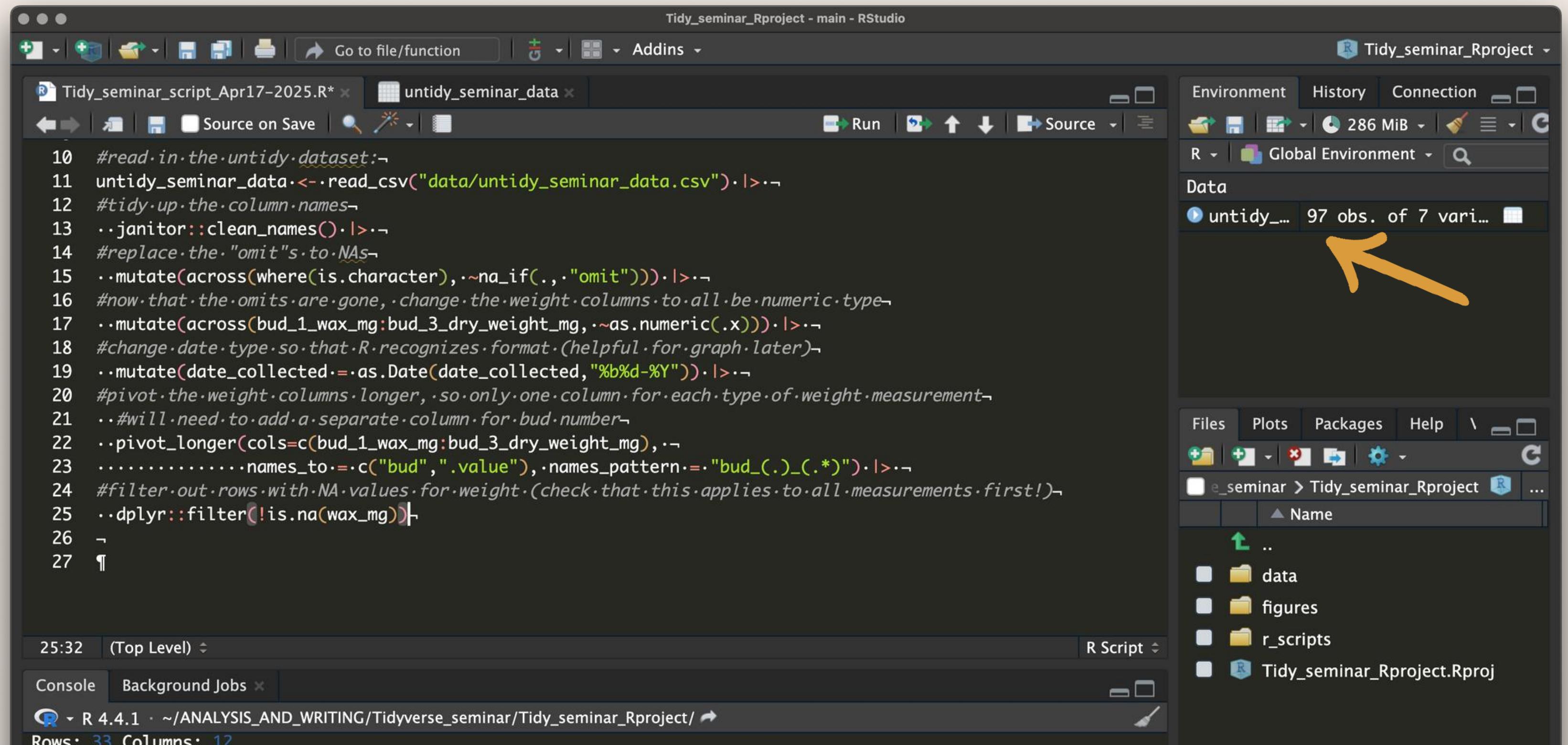
```
>filter(column_name != "string_you_HATE")
```

#multiple criteria:

```
>filter(column_name %in% c("string1", "string2"))
```



COMMON “TIDIES”: FILTER



The screenshot shows an RStudio interface with the following details:

- Title Bar:** Tidy_seminar_Rproject - main - RStudio
- File Bar:** Tidy_seminar_script_Apr17-2025.R*
- Code Editor:** A script containing R code for data tidying. Lines 10-27 are shown, starting with `#read.in.the.untidy.dataset:-` and ending with `#`.
- Global Environment Panel:** Shows a variable named `untidy_seminar_data` with 97 observations and 7 variables.
- Annotations:** A large orange arrow points from the text "Common ‘tidies’: FILTER" to the `filter` function call in the code editor.
- Bottom Status Bar:** Rows: 33 Columns: 12



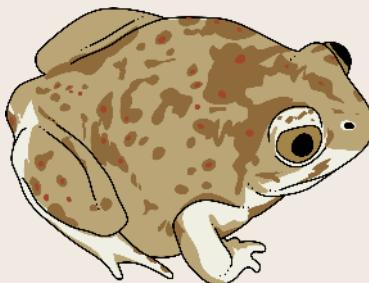
COMMON “TIDIES”: GROUP_BY

#works like “sort by” in excel - applies to lines following

>group_by(variable1, variable2)

#don't forget to ungroup() when you're done!

>ungroup()



COMMON “TIDIES”: GROUP_BY

The screenshot shows an RStudio interface with the following elements:

- Editor:** The main window displays an R script titled "Tidy_seminar_script_Apr17-2025.R". The code is written in dplyr syntax to manipulate a dataset. An orange arrow points from the bottom left towards the closing brace of the script.
- Global Environment:** The right-hand pane shows the "Global Environment" tab, which lists the "untidy_seminar_data" object. It is described as having 97 observations and 8 variables.
- File Explorer:** The bottom-right pane shows the project structure with folders for "data", "figures", "r_scripts", and the project itself.
- Console:** The bottom-left pane shows the R console output, indicating the script was run at 29:12 (Top Level) and the environment is R 4.4.1.

```
14 #replace the "omit"s to NAs
15 .~mutate(across(where(is.character), .~na_if(., "omit"))).|>..
16 #now that the omits are gone, change the weight columns to all be numeric type
17 .~mutate(across(bud_1_wax_mg:bud_3_dry_weight_mg, .~as.numeric(.x))).|>..
18 #change date type so that R recognizes format (helpful for graph later)
19 .~mutate(date_collected.=as.Date(date_collected, "%b%d-%Y")).|>..
20 #pivot the weight columns longer, so only one column for each type of weight measurement
21 .~#will need to add a separate column for bud number
22 .~pivot_longer(cols=c(bud_1_wax_mg:bud_3_dry_weight_mg), .~.
23 .....names_to.=c("bud",".value"), names_pattern.= "bud_(.)_(.*)").|>..
24 #filter out rows with NA values for weight (check that this applies to all measurements first!)
25 .~dplyr::filter(!is.na(wax_mg)).|>..
26 #create column with average weight across all buds for a tree
27 .~group_by(tree, date_collected).|>..
28 .~mutate(avg_wax_per_tree_mg.=mean(wax_mg)).|>..
29 .~ungroup()|
30 |
31 ¶
```

COMMON “TIDIES”: GROUP_BY

The screenshot shows the RStudio interface with a project titled "Tidy_seminar_Rproject". In the top-left, there are two tabs: "Tidy_seminar_script_Apr17-2025.R*" and "untidy_seminar_data". The main area displays a data frame with the following columns: tree, date_collected, comments, bud, wax_mg, fresh_weight_g, dry_weight_mg, and avg_wax_per_tree_mg. The data consists of 13 rows of measurements for trees 57, 56, 35, and 45 collected on 2025-02-07. A yellow arrow points from the "untidy_seminar_data" tab in the top-left to the "untidy_seminar_data" entry in the Environment pane on the right. The Environment pane also shows the Global Environment and Data sections, with "untidy_seminar_data" listed as having 97 observations and 8 variables. The bottom of the screen shows the R console and status bar.

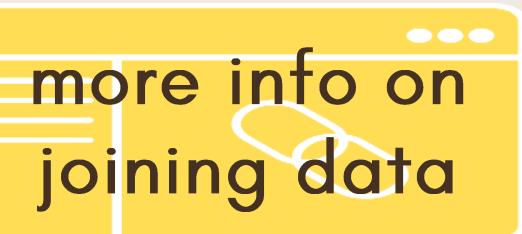
tree	date_collected	comments	bud	wax_mg	fresh_weight_g	dry_weight_mg	avg_wax_per_tree_mg
57	2025-02-07	NA	1	0.235	0.092704	6.142	0.1950000
57	2025-02-07	NA	2	0.194	0.115702	5.812	0.1950000
57	2025-02-07	NA	3	0.156	0.088767	5.711	0.1950000
56	2025-02-07	NA	1	0.713	0.113864	5.759	0.5026667
56	2025-02-07	NA	2	0.420	0.088289	6.007	0.5026667
56	2025-02-07	NA	3	0.375	0.082331	5.832	0.5026667
35	2025-02-07	NA	1	0.191	0.089647	6.070	0.1863333
35	2025-02-07	NA	2	0.149	0.100540	5.868	0.1863333
35	2025-02-07	NA	3	0.219	0.118564	6.244	0.1863333
97	2025-02-07	NA	1	0.519	0.143839	6.029	0.4236667
97	2025-02-07	NA	2	0.338	0.084323	5.935	0.4236667
97	2025-02-07	NA	3	0.414	0.079606	6.083	0.4236667
45	2025-02-07	NA	1	0.524	0.122457	5.759	0.6476667
45	2025-02-07	NA	2	0.652	0.091293	5.727	0.6476667

Showing 1 to 13 of 97 entries, 8 total columns

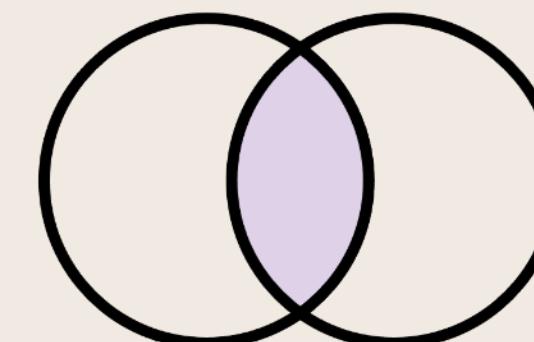
Console Background Jobs ×

R 4.4.1 · ~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject/ ↗

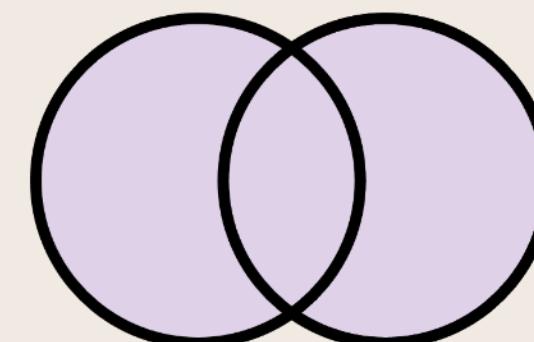
Rows: 33 Columns: 12



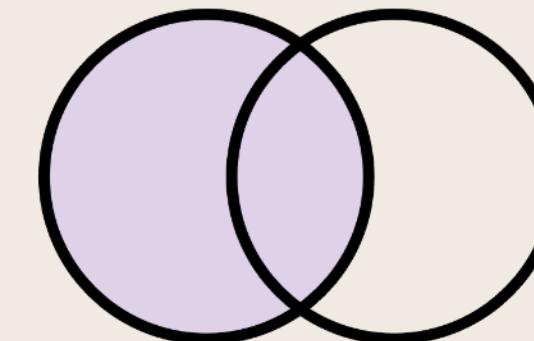
COMMON “TIDIES”: JOINING



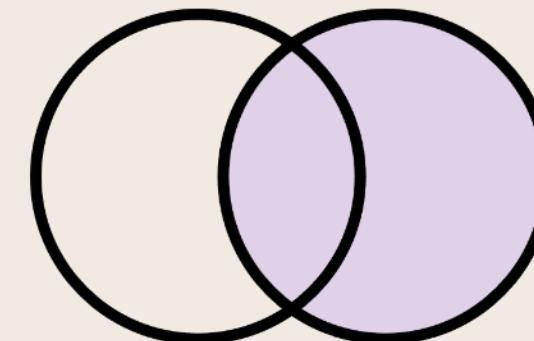
`inner_join(x, y)`



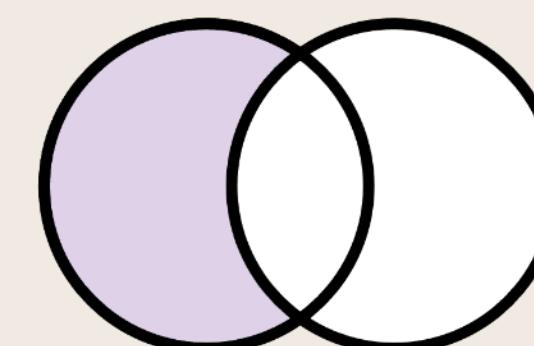
`full_join(x, y)`



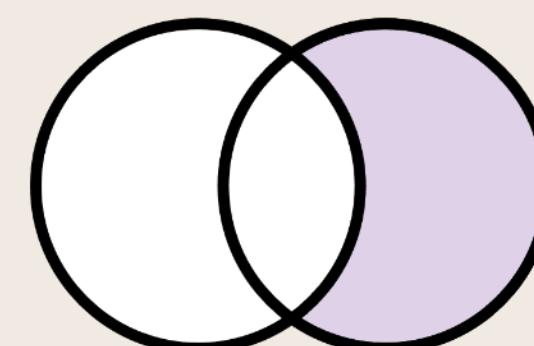
`left_join(x, y)`



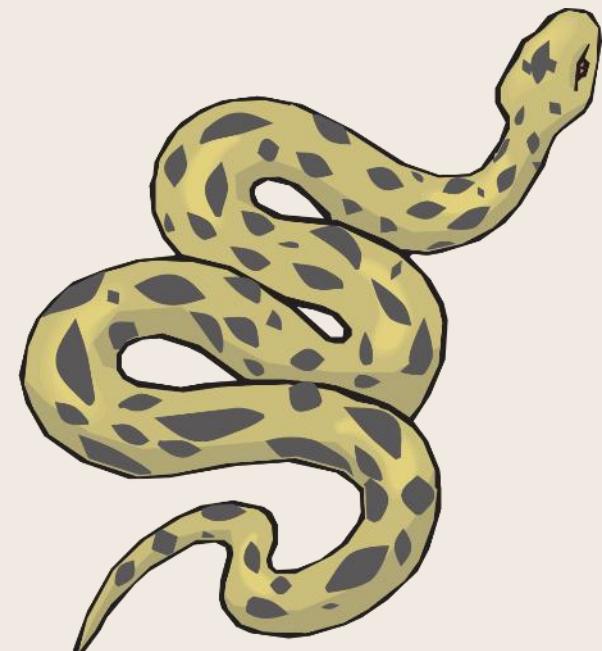
`right_join(x, y)`



`anti_join(x, y)`



`anti_join(y, x)`



COMMON “TIDIES”: JOINING

The screenshot shows an RStudio interface with the following details:

- Title Bar:** Tidy_seminar_Rproject - main - RStudio
- File Tab:** Tidy_seminar_script_Apr17-2025.R*
- Code Editor:** The script contains R code for data manipulation. It includes filtering rows with NA values, calculating average wax weight per tree, and joining datasets. The code uses the dplyr package.
- Environment Tab:** Shows three data frames: sample_(...), tidier_(...), and untidy_(...).
- Data View:** Displays the contents of the data frames.
- Files Tab:** Shows the project structure: e_seminar > Tidy_seminar_Rproject.
- Console Tab:** R 4.4.1 · ~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject/

```
39:1 (Top Level) R Script  
Console Background Jobs  
R 4.4.1 · ~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject/  
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40  
#tidy_seminar_data %>%  
#> #filter out rows with NA values for weight (check that this applies to all measurements first!)  
#> #dplyr::filter(!is.na(wax_mg))  
#> #create column with average weight across all buds for a tree  
#> #group_by(tree, date_collected)  
#> #mutate(avg_wax_per_tree_mg = mean(wax_mg))  
#> #ungroup()  
#>  
#read in sample metadata file to join sample codes  
sample_metadata <- read_csv("data/sample_metadata.csv")  
#>  
#janitor::clean_names()  
#>  
#create updated dataset and join sample metadata  
tidier_data <- untidy_seminar_data %>%  
#> left_join(sample_metadata, .  
#> .by = join_by(tree)) #if the join column names aren't identical, use columnA == columnB
```

COMMON “TIDIES”: JOINING

RStudio interface showing the "tidier_data" tab selected in the top navigation bar. The main area displays a data frame with 13 rows and 8 columns. An orange arrow points from the "Global Environment" section in the right sidebar to the "treatment" column header in the data frame.

bud	wax_mg	fresh_weight_g	dry_weight_mg	avg_wax_per_tree_mg	lims_code	time_point	treatment
1	0.235	0.092704	6.142	0.1950000	M187-01-01-BP	7 dpp	north exposure
2	0.194	0.115702	5.812	0.1950000	M187-01-01-BP	7 dpp	north exposure
3	0.156	0.088767	5.711	0.1950000	M187-01-01-BP	7 dpp	north exposure
1	0.713	0.113864	5.759	0.5026667	M187-01-02-BP	7 dpp	north exposure
2	0.420	0.088289	6.007	0.5026667	M187-01-02-BP	7 dpp	north exposure
3	0.375	0.082331	5.832	0.5026667	M187-01-02-BP	7 dpp	north exposure
1	0.191	0.089647	6.070	0.1863333	M187-01-03-BP	7 dpp	north exposure
2	0.149	0.100540	5.868	0.1863333	M187-01-03-BP	7 dpp	north exposure
3	0.219	0.118564	6.244	0.1863333	M187-01-03-BP	7 dpp	north exposure
1	0.519	0.143839	6.029	0.4236667	M187-01-04-BP	7 dpp	north exposure
2	0.338	0.084323	5.935	0.4236667	M187-01-04-BP	7 dpp	north exposure
3	0.414	0.079606	6.083	0.4236667	M187-01-04-BP	7 dpp	north exposure
1	0.524	0.122457	5.759	0.6476667	M187-01-05-BP	7 dpp	north exposure
2	0.652	0.091293	5.727	0.6476667	M187-01-05-BP	7 dpp	north exposure

Showing 1 to 13 of 97 entries, 11 total columns

Console Background Jobs R 4.4.1 ~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject/

Tidy_seminar_Rproject - main - RStudio

untidy_seminar_data tidier_data

Environment History Connection

276 MiB

Global Environment

sample_... 33 obs. of 4 vari...

tidier_... 97 obs. of 11 vari...

untidy_... 97 obs. of 8 vari...

Files Plots Packages Help

data figures r_scripts Tidy_seminar_Rproject.Rproj

Name ..

COMMON “TIDIES”: SEPARATE & UNITE



#splits a column of strings using regex (regular expression)

```
>separate(col2split, c("newcol1", "newcol2"), sep="-")
```

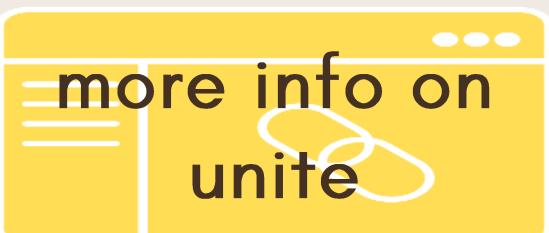
#sep argument looks for how to split up strings

#inverse is unite, which will join columns

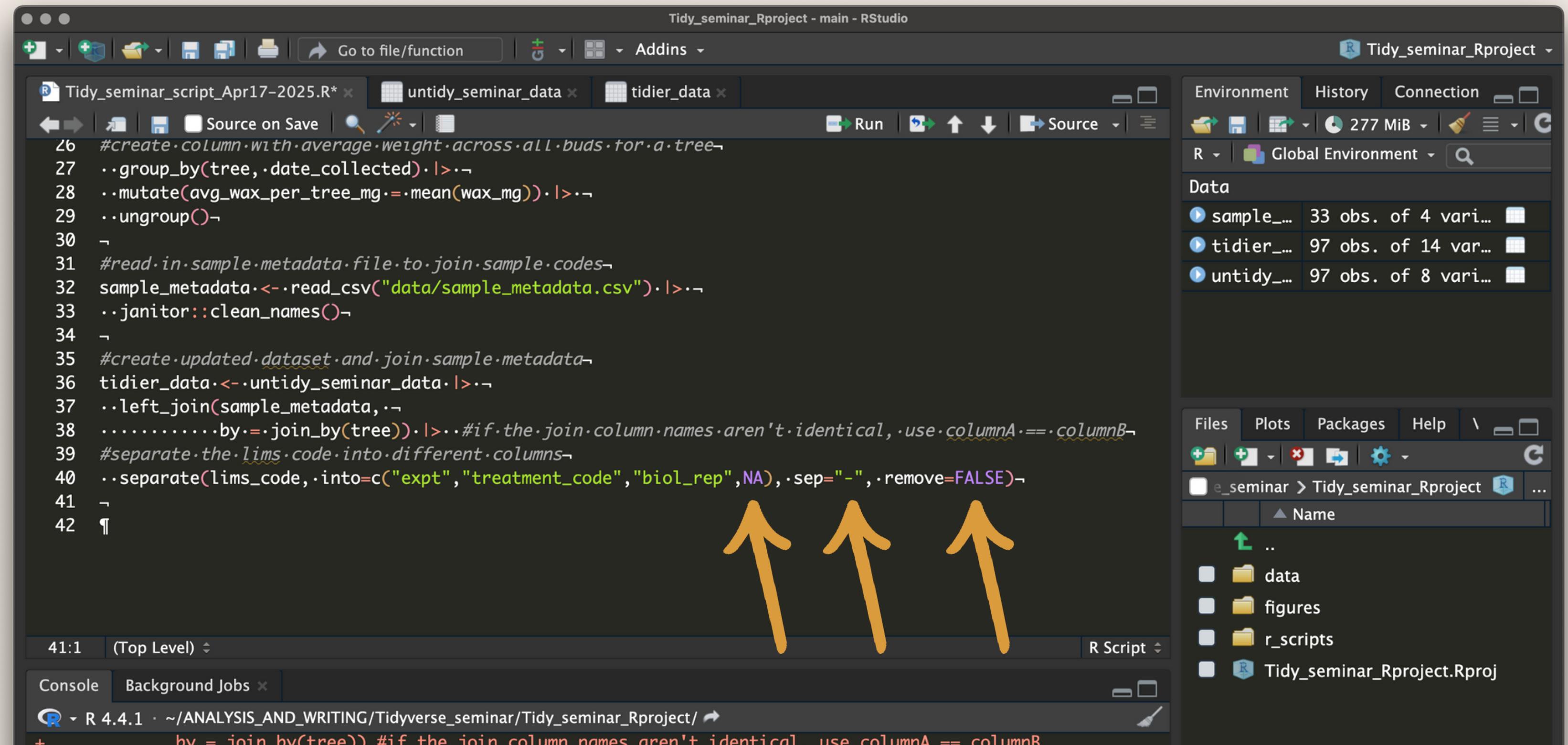
```
>unite("newcol", col1:col3, remove = FALSE)
```

#will unite columns 1, 2, and 3, but won't remove them

#can also have sep = "" argument



COMMON “TIDIES”: SEPARATE



The screenshot shows an RStudio interface with a script file open. The script contains R code demonstrating the use of the `separate` function from the `tidyverse` package. Three orange arrows point to the `separate` call at the bottom of the code block.

```
#create a column with average weight across all buds for a tree
..group_by(tree, .date_collected) |>-
  ..mutate(avg_wax_per_tree_mg = mean(wax_mg)) |>-
  ..ungroup() |>-
#read in sample metadata file to join sample codes
sample_metadata <- read_csv("data/sample_metadata.csv") |>-
  ..janitor::clean_names() |>-
#create updated dataset and join sample metadata
tidier_data <- untidy_seminar_data |>-
  ..left_join(sample_metadata, .by = join_by(tree)) |>-
    #if the join column names aren't identical, use columnA == columnB
#separate the lims code into different columns
..separate(lims_code, .into = c("expt", "treatment_code", "biol_rep", NA), .sep = "-", .remove = FALSE) |>-
  ..
```

The RStudio interface includes the following panels:

- Top Bar:** Tidy_seminar_Rproject - main - RStudio
- File Bar:** Tidy_seminar_script_Apr17-2025.R*, untidy_seminar_data, tidier_data
- Toolbar:** Go to file/function, Addins
- Code Editor:** Shows the R script with line numbers 26 to 42.
- Environment Panel:** Displays objects: sample..., tidier..., untidy... (33 obs. of 4 variables, 97 obs. of 14 variables, 97 obs. of 8 variables).
- Data View:** Shows the same three objects with their details.
- Files Panel:** Shows the project structure: Tidy_seminar > Tidy_seminar_Rproject.
- Console:** Shows the R version (R 4.4.1) and the current working directory (~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject).

COMMON “TIDIES”: SEPARATE

The screenshot shows the RStudio interface with a dark theme. The top bar displays the project name "Tidy_seminar_Rproject - main - RStudio". The left sidebar has tabs for "Tidy_seminar_script_Apr17-2025.R*", "untidy_seminar_data", and "tidier_data". The main area shows a data frame with 13 rows and 8 columns. The columns are: dry_weight_mg, avg_wax_per_tree_mg, lims_code, expt, treatment_code, biol_rep, time_point, and treatment. The data is as follows:

	dry_weight_mg	avg_wax_per_tree_mg	lims_code	expt	treatment_code	biol_rep	time_point	treatment
1	6.142	0.1950000	M187-01-01-BP	M187	01	01	7 dpp	north exposure
2	5.812	0.1950000	M187-01-01-BP	M187	01	01	7 dpp	north exposure
7	5.711	0.1950000	M187-01-01-BP	M187	01	01	7 dpp	north exposure
1	5.759	0.5026667	M187-01-02-BP	M187	01	02	7 dpp	north exposure
3	6.007	0.5026667	M187-01-02-BP	M187	01	02	7 dpp	north exposure
L	5.832	0.5026667	M187-01-02-BP	M187	01	02	7 dpp	north exposure
7	6.070	0.1863333	M187-01-03-BP	M187	01	03	7 dpp	north exposure
)	5.868	0.1863333	M187-01-03-BP	M187	01	03	7 dpp	north exposure
1	6.244	0.1863333	M187-01-03-BP	M187	01	03	7 dpp	north exposure
3	6.029	0.4236667	M187-01-04-BP	M187	01	04	7 dpp	north exposure
3	5.935	0.4236667	M187-01-04-BP	M187	01	04	7 dpp	north exposure
5	6.083	0.4236667	M187-01-04-BP	M187	01	04	7 dpp	north exposure
7	5.759	0.6476667	M187-01-05-BP	M187	01	05	7 dpp	north exposure
3	5.727	0.6476667	M187-01-05-BP	M187	01	05	7 dpp	north exposure

At the bottom, it says "Showing 1 to 13 of 97 entries, 14 total columns".

The right sidebar shows the "Environment" tab with three objects: "sample_..." (33 obs. of 4 vari...), "tidier_..." (97 obs. of 14 vari...), and "untidy_..." (97 obs. of 8 vari...). The "Global Environment" tab shows the same three objects.

The bottom navigation bar includes tabs for "Console", "Background Jobs", and "Plots". The console tab shows the R version "R 4.4.1" and the current working directory "~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject/".

COMMON “TIDIES”:

SELECT

#helpful for getting rid of unnecessary columns

```
>select(column1, column3, column4)
```

#the order you select is also the order your output will be

#can also use “starts_with” and “ends_with”

```
>select(starts_with("column") & ends_with("3"))
```



COMMON “TIDIES”:

SELECT

The screenshot shows the RStudio interface with the following details:

- Title Bar:** Tidy_seminar_Rproject - main - RStudio
- File Tab:** Tidy_seminar_script_Apr17-2025.R*
- Code Editor:** A script containing R code for data manipulation. The code uses the dplyr package to separate LIMS code into different columns, calculate percent water content, and select specific columns (wax_mg, fresh_weight_mg, fresh_weight_g, dry_weight_mg, percent_water_content, wax_per_mg_DW) from the tidier_data dataset.
- Environment Tab:** Shows three datasets: sample_... (33 obs. of 4 vars), tidier_... (97 obs. of 13 vars), and untidy_... (97 obs. of 8 vars).
- Files Tab:** Shows the project structure: e_seminar > Tidy_seminar_Rproject. It includes a data folder, figures folder, r_scripts folder, and the Tidy_seminar.Rproj file.
- Console Tab:** Shows the R version (R 4.4.1) and the current working directory (~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject).

```
59 #separate the LIMS code into different columns
60 ..separate(lims_code, .into=c("expt", "treatment_code", "biol_rep", NA), .sep="-", .remove=FALSE) |>..
61 #calculate percent water content, which can be compared to wax-- first need everything in same units
62 ..mutate(fresh_weight_mg = fresh_weight_g * 1000) |>..
63 #perc water content = weight of water / total weight
64 ..mutate(percent_water_content = (fresh_weight_mg - dry_weight_mg) / fresh_weight_mg) |>..
65 #calculate wax to dry weight ratio
66 ..mutate(wax_per_mg_DW = wax_mg / dry_weight_mg) |>..
67 #now take the average for each tree (tree == biol.rep, bud == tech.rep)
68 ..group_by(tree, .date_collected) |>..
69 ..mutate(avg_percent_water_per_tree = mean(percent_water_content)) |>..
70 ..mutate(avg_wax_per_mg_DW_per_tree = mean(wax_per_mg_DW)) |>..
71 ..ungroup() |>..
72 #remove unnecessary columns -- only want per tree avgs
73 #-c() will remove the columns you indicate, .c() will keep them. !! is another option
74 ..dplyr::select(-c(wax_mg, .fresh_weight_mg, .fresh_weight_g, ..
75 .....dry_weight_mg, .percent_water_content, .wax_per_mg_DW)) |>..
```

more info on
distinct

COMMON “TIDIES”: DISTINCT

#will remove rows that are identical in the variables you list
`>distinct(column1, column2, .keep_all=TRUE)`
#if `.keep_all=FALSE` (default), it will remove the other columns



COMMON “TIDIES”: DISTINCT

The screenshot shows the RStudio interface with the following details:

- Title Bar:** Tidy_seminar_Rproject - main - RStudio
- File Tab:** Tidy_seminar_script_Apr17-2025.R*
- Code Editor:** The script contains R code for tidying seminar data. It includes operations like separating columns, calculating percent water content, and calculating wax-to-dry weight ratios. The code uses the dplyr package.
- Environment Tab:** Shows the Global Environment with three data frames: sample_... (33 obs. of 4 variables), tidier_... (33 obs. of 12 variables), and untidy_... (97 obs. of 8 variables).
- Files Tab:** Shows the project structure: e_seminar > Tidy_seminar_Rproject. It includes folders for data, figures, r_scripts, and the project itself.
- Console Tab:** Shows the R version (R 4.4.1) and the current working directory (~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject).

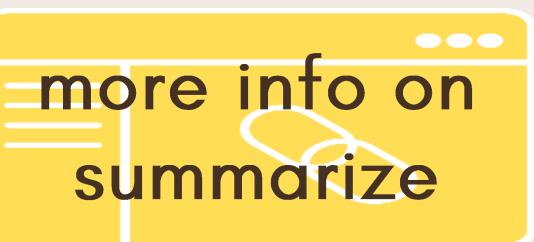
```
39 #separate the LIMS code into different columns
40 ..separate(lims_code,.into=c("expt","treatment_code","biol_rep",NA),.sep="-",.remove=FALSE).|>..
41 #calculate percent water content, which can be compared to wax-- first need everything in same units
42 ..mutate(fresh_weight_mg.=fresh_weight_g.*1000).|>..
43 #perc water content.=weight.of.water/total.weight
44 ..mutate(percent_water_content.= (fresh_weight_mg-dry_weight_mg)/fresh_weight_mg).|>..
45 #calculate wax.to.dry.weight.ratio
46 ..mutate(wax_per_mg_DW.=wax_mg/dry_weight_mg).|>..
47 #now take the average for each tree.(tree==biol.rep, .bud==tech.rep)-
48 ..group_by(tree,.date_collected).|>..
49 ..mutate(avg_percent_water_per_tree.=mean(percent_water_content)).|>..
50 ..mutate(avg_wax_per_mg_DW_per_tree.=mean(wax_per_mg_DW)).|>..
51 ..ungroup().|>..
52 #remove unnecessary columns--only want per.tree.avgs-
53 #-c() will remove the columns you indicate, .c() will keep them..!..is another option
54 ..dplyr::select(-c(wax_mg,.fresh_weight_mg,.fresh_weight_g,..
55 .....dry_weight_mg,.percent_water_content,.wax_per_mg_DW)).|>..
56 #since we only want one average value per tree, remove duplicate lines
57 ..distinct(across(!bud))#.need to exclude lines we know are not distinct
58 
```

COMMON “TIDIES”: DISTINCT

The screenshot shows the RStudio interface with the title bar "Tidy_seminar_Rproject - main - RStudio". In the top-left, there are tabs for "Tidy_seminar_script_Apr17-2025.R*" (active), "untidy_seminar_data", and "tidier_data". The main area displays a data frame with 13 rows and 9 columns. An orange arrow points to the second row's "tree" value, which is 56. The columns are: tree, date_collected, comments, avg_wax_per_tree_mg, lims_code, expt, treatment_code, biol_rep, and a unnamed column. The data is as follows:

	tree	date_collected	comments	avg_wax_per_tree_mg	lims_code	expt	treatment_code	biol_rep	
1	57	2025-02-07	NA	0.1950000	M187-01-01-BP	M187	01	01	
2	56	2025-02-07	NA	0.5026667	M187-01-02-BP	M187	01	02	
3	35	2025-02-07	NA	0.1863333	M187-01-03-BP	M187	01	03	
4	97	2025-02-07	NA	0.4236667	M187-01-04-BP	M187	01	04	
5	45	2025-02-07	NA	0.6476667	M187-01-05-BP	M187	01	05	
6	28	2025-02-07	NA	0.5093333	M187-02-01-BP	M187	02	01	
7	48	2025-02-07	NA	0.2450000	M187-02-02-BP	M187	02	02	
8	89	2025-02-07	NA	0.4000000	M187-02-03-BP	M187	02	03	
9	83	2025-02-07	NA	0.6080000	M187-02-04-BP	M187	02	04	
10	17	2025-02-07	NA	0.3600000	M187-02-05-BP	M187	02	05	
11	99	2025-02-07	NA	0.2910000	M187-02-06-BP	M187	02	06	
12	10	2025-02-23	NA	1.1516667	M187-03-01-BP	M187	03	01	
13	14	2025-02-23	NA	1.0686667	M187-03-02-BP	M187	03	02	
14	85	2025-02-23	NA	1.0276667	M187-03-03-BP	M187	03	03	

At the bottom left, it says "Showing 1 to 13 of 33 entries, 12 total columns". The right sidebar shows the "Environment" tab with objects: sample..., tidier..., and untidy... under the "Data" section. The "Files" tab lists "data", "figures", "r_scripts", and "Tidy_seminar.Rproj". The "Console" tab at the bottom shows the R session.



COMMON “TIDIES”: SUMMARIZE

#use with group_by to apply summary expressions

```
>group_by(col_sampleID) |>
  summarise ("col_avg" = mean(col_measurement))
```

#within mean, may want to specify na.rm = TRUE

#(excludes NAs from the calculation)



COMMON “TIDIES”: SUMMARIZE

The screenshot shows the RStudio interface with the following details:

- Title Bar:** Tidy_seminar_Rproject - main - RStudio
- File Tab:** Tidy_seminar_script_Apr17-2025.R*
- Code Editor:** The script contains R code for summarizing data using the dplyr package. The code includes mutate, ungroup, and summarize functions to calculate averages per tree and across treatments.
- Environment Tab:** Shows the Global Environment with four data frames: sample_, summarized_data, untidy_seminar_data, and tidier_data.
- Data View:** Displays the contents of the summarized_data frame.
- Files Tab:** Shows the project structure: e_seminar > Tidy_seminar_Rproject.
- Console Tab:** Shows the R version (R 4.4.1) and the current working directory (~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject).

```
49 ..mutate(avg_percent_water_per_tree := mean(percent_water_content)) |>-
50 ..mutate(avg_wax_per_mg_DW_per_tree := mean(wax_per_mg_DW)) |>-
51 ..ungroup() |>-
52 #remove.unnecessary.columns--only.want.per.tree.avgs-
53 #-c().will.remove.the.columns.you.indicate,.c().will.keep.them.!!is.another.option-
54 ..dplyr::select(-c(wax_mg, fresh_weight_mg, fresh_weight_g),-
55 .....dry_weight_mg, percent_water_content, wax_per_mg_DW)) |>-
56 #since.we.only.want.one.average.value.per.tree, remove.duplicate.lines-
57 ..distinct(across(!bud)) #need.to.exclude.lines.we.know.are.not.distinct-
58 -
59 #also.want.a.separate.summary.of.averages.by.time.point.&.treatment-
60 summarized_data <- tidier_data |>-
61 ..group_by(time_point, treatment) |>-
62 ..summarize(wax_per_mg_DW_treatment_avg := mean(avg_wax_per_mg_DW_per_tree),-
63 .....percent_water_treatment_avg := mean(avg_percent_water_per_tree),-
64 .....wax_mg_treatment_avg := mean(avg_wax_per_tree_mg)) |>-
65 ...
```

COMMON “TIDIES”: SUMMARIZE

The screenshot shows the RStudio interface with a project titled "Tidy_seminar_Rproject". The main area displays a data frame named "tidier_data" with the following structure:

	time_point	treatment	wax_per_mg_DW_treatment_avg	percent_water_treatment_avg	wax_mg_treatment_avg
1	23 dpp	north exposure	0.16938591	0.4290292	1.0585000
2	23 dpp	south exposure	0.16826726	0.4087619	1.0690000
3	43 dpp	north exposure	0.21713181	0.3943052	1.5154706
4	43 dpp	south exposure	0.23078479	0.4766418	1.5610000
5	7 dpp	north exposure	0.06623689	0.4015839	0.3910667
6	7 dpp	south exposure	0.06708883	0.4490694	0.4022222

The RStudio environment includes tabs for "Environment", "History", and "Connection". The "Files" panel shows a project structure with "data", "figures", "r_scripts", and the current project file. The "Console" tab at the bottom shows the command used to generate the summary.

```
R 4.4.1 · ~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject/ →
> #also want a separate summary of averages by time point & treatment
```

EXPORT YOUR DATA

The screenshot shows the RStudio interface with a project titled "Tidy_seminar_Rproject". The left pane displays an R script named "Tidy_seminar_script_Apr17-2025.R". The script contains code for tidying data using dplyr, specifically for calculating averages per tree across different treatments and time points. The right pane shows the "Environment" and "Data" panes, which list various data frames like "sample_...", "summarized_data", "tidier_data", and "untidy_seminar_data". The bottom pane shows the "Console" tab with the R command `summarize` being run.

```
#remove.unnecessary.columns--only.want.per.tree.avgs-
#-c() will.remove.the.columns.you.indicate,.c() will.keep.them..!..is.another.option-
..dplyr::select(-c(wax_mg,.fresh_weight_mg,.fresh_weight_g,..
.....dry_weight_mg,.percent_water_content,.wax_per_mg_DW))|>..
#since.we.only.want.one.average.per.tree,.remove.duplicate.lines-
..distinct(across(!bud))#need.to.exclude.lines.we.know.are.not.distinct-
-
#also.want.a.separate.summary.of.averages.by.time.point.&.treatment-
summarized_data<-tidier_data|>..
..group_by(time_point,.treatment)|>..
..summarize(wax_per_mg_DW_treatment_avg=mean(avg_wax_per_mg_DW_per_tree),-
.....percent_water_treatment_avg=mean(avg_percent_water_per_tree),-
.....wax_mg_treatment_avg=mean(avg_wax_per_tree_mg))-|
#export.datasets.as.csv--use.this.new.file.for.additional.analyses.so.don't.need.to.repeat.above.steps-
write_csv(tidier_data,"data/tidier_seminar_data.csv")-
```

69:1 (Top Level) R Script

Console Background Jobs

R 4.4.1 ~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject/

```
+ summarize(wax_per_mg_DW_treatment_avg = mean(avg_wax_per_mg_DW_per_tree),
+             percent_water_treatment_avg = mean(avg_percent_water_per_tree),
+             wax_mg_treatment_avg = mean(avg_wax_per_tree_mg))
`summarise()` has grouped output by 'time_point'. You can override using the `groups` argument.
> View(summarized_data)
```

Environment History Connection

252 MiB Global Environment

Data

Name	Type	Size
sample_...	33 obs. of 4 vari...	
summarized...	6 obs. of 5 varia...	
tidier_...	97 obs. of 13 vari...	
untidy_...	97 obs. of 8 vari...	

Files Plots Packages Help

e_seminar > Tidy_seminar_Rproject

Name

- ..
- data
- figures
- r_scripts
- Tidy_seminar_Rproject.Rproj
- tables



GT TABLES

The Parts of a gt Table



TABLE
HEADER

STUB
HEAD

STUB

TITLE			
SUBTITLE			
STUBHEAD LABEL	SPANNER COLUMN LABEL	COLUMN LABEL	COLUMN LABEL
ROW GROUP LABEL	COLUMN LABEL	COLUMN LABEL	COLUMN LABEL
ROW LABEL	Cell	Cell	Cell
ROW LABEL	Cell	Cell	Cell
SUMMARY LABEL	Summary Cell	Summary Cell	Summary Cell
FOOTNOTES			
SOURCE NOTES			

COLUMN
LABELS

TABLE
BODY

TABLE
FOOTER

GT TABLES

create headers

with 4 dashes (----)

The screenshot shows an RStudio interface with a script file open. The code generates a 'SUMMARY TABLE' using the gt library. A callout bubble points to the 'SUMMARY TABLE' section of the code with the text 'create headers with 4 dashes (----)'. Another callout bubble points to the resulting gt table in the RStudio preview pane with the text 'DATA ORGANIZATION SUMMARY TABLE'.

```
65 .....percent_water_treatment_avg..=mean(avg_percent_water_per_tree),-| DATA ORGANIZATION
66 .....wax_mg_treatment_avg..=mean(avg_wax_per_tree_mg))-| SUMMARY TABLE
67 -
68 #export.datasets.as.csv--use.this.new.file.for.additional.analyses.so.don't
69 write_csv(tidier_data,"data/tidier_seminar_data.csv")-
70 -
71 -
72 #####SUMMARY TABLE#####
73 library(gt)-
74 -
75 #data.is.piped.directly.into.table-
76 summarized_data |>-
77 #replace._.in.column.names.with.spaces.so.headers.look.nicer-
78 ..rename_with(~gsub("_",".",.,x,.fixed=TRUE),.ends_with("avg"))|>-
79 ..gt()-
80 ||
```

The resulting gt table is displayed in the preview pane:

treatment	wax per mg	perc DW	wa treatment	treatm avg
23 dpp				
north exposure	0.16938591	0.42902		
south exposure	0.16826726	0.40876		

Console output:

```
R 4.4.1 · ~/ANALYSIS_AND_WRITING/Tidyverse_seminar/Tidy_seminar_Rproject/
5 7 dpp      north exposure      0.0662      0.402      0.391
6 7 dpp      south exposure     0.0671      0.449      0.402
# i abbreviated name: `percent water treatment avg`
```

GT TABLES

Tidy_seminar_Rproject - main - RStudio

Go to file/function Addins Tidy_seminar_Rproject Connection

Viewer Zoom

65
66
67 -
68 #export
69 write_c
70 -
71 -
72 #####--
73 library
74 -
75 #data.n
76 summar
77 #repla
78 ..renam
79 ..gtO-
80 ||

t 4 vari... 5 vari... 13 vari... 8 vari...

Help port

mg perc
DW wa
ent wa
avg treatment

791 0.42902

80:1 # SUM

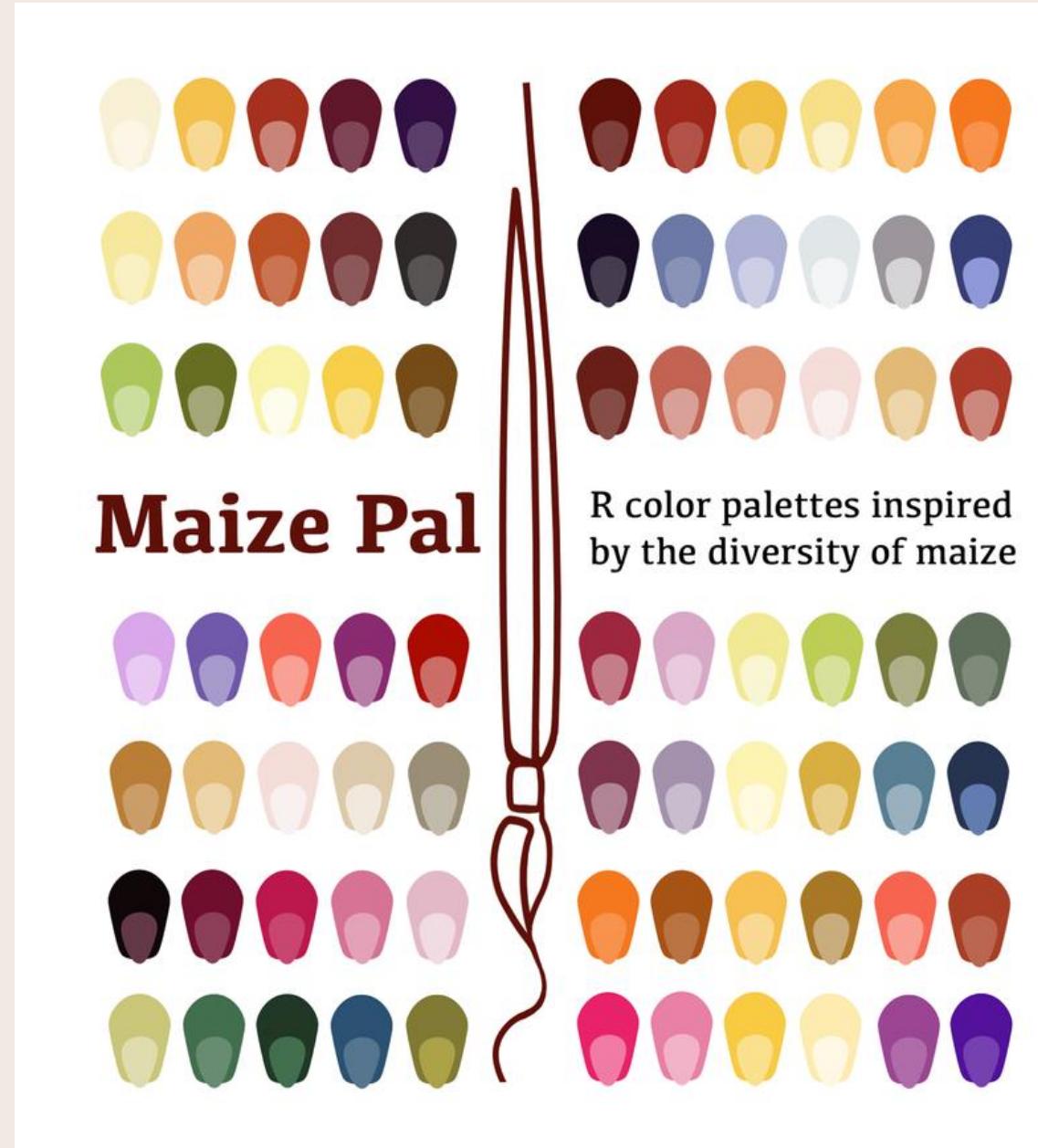
Console Background R 4.4.1

5 7 dpp north exposure 0.0662 0.402 0.391
6 7 dpp south exposure 0.0671 0.449 0.402
i abbreviated name: `percent water treatment avg`
> #data is piped directly into table
> summarized data >

treatment	wax per mg DW treatment avg	percent water treatment avg	wax mg treatment avg
23 dpp			
north exposure	0.16938591	0.4290292	1.0585000
south exposure	0.16826726	0.4087619	1.0690000
43 dpp			
north exposure	0.21713181	0.3943052	1.5154706
south exposure	0.23078479	0.4766418	1.5610000
7 dpp			
north exposure	0.06623689	0.4015839	0.3910667
south exposure	0.06708883	0.4490694	0.4022222

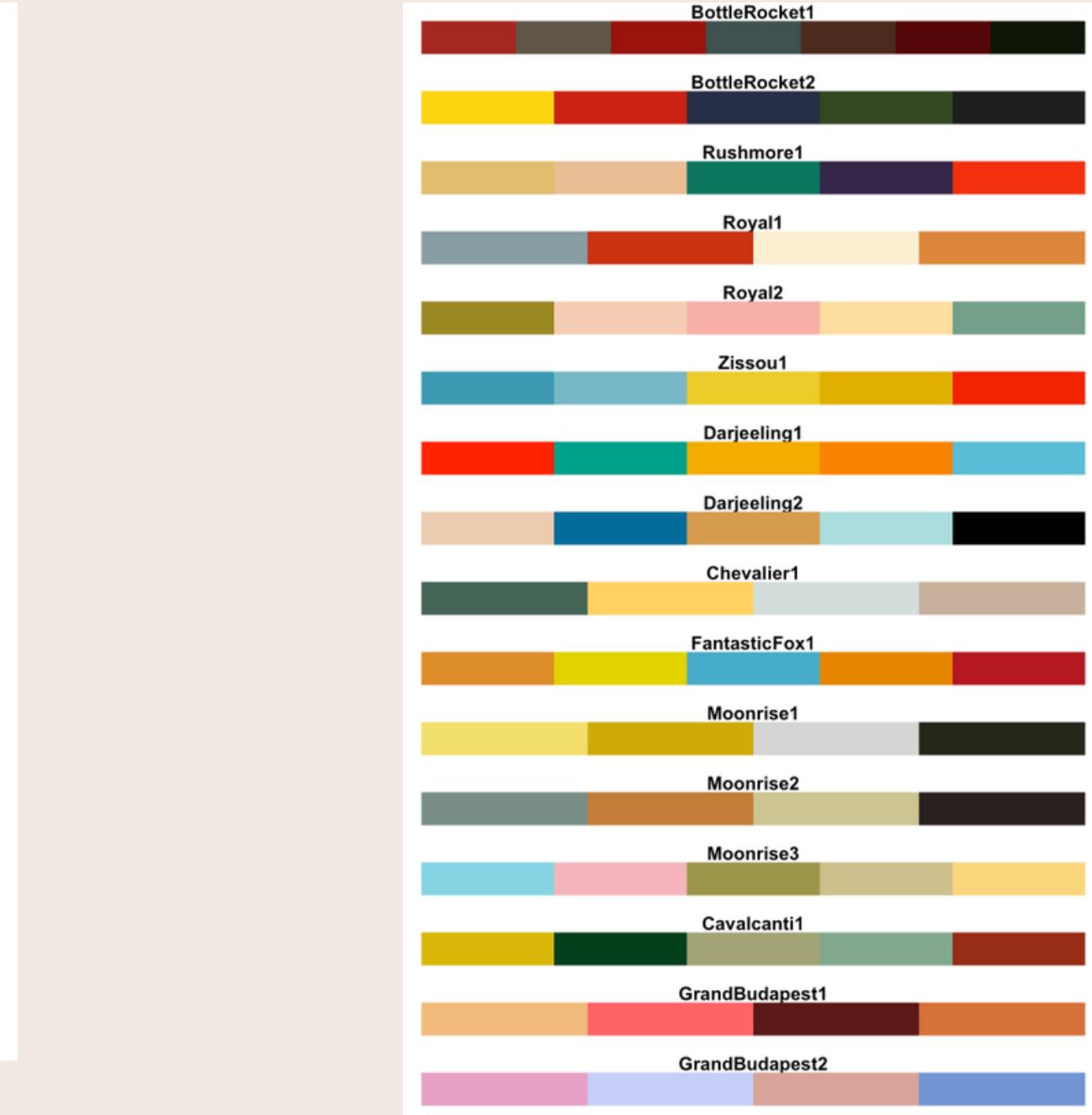
more info
on ggplot2

QUICK GRAPH IN GG PLOT 2



click to be
taken to link!

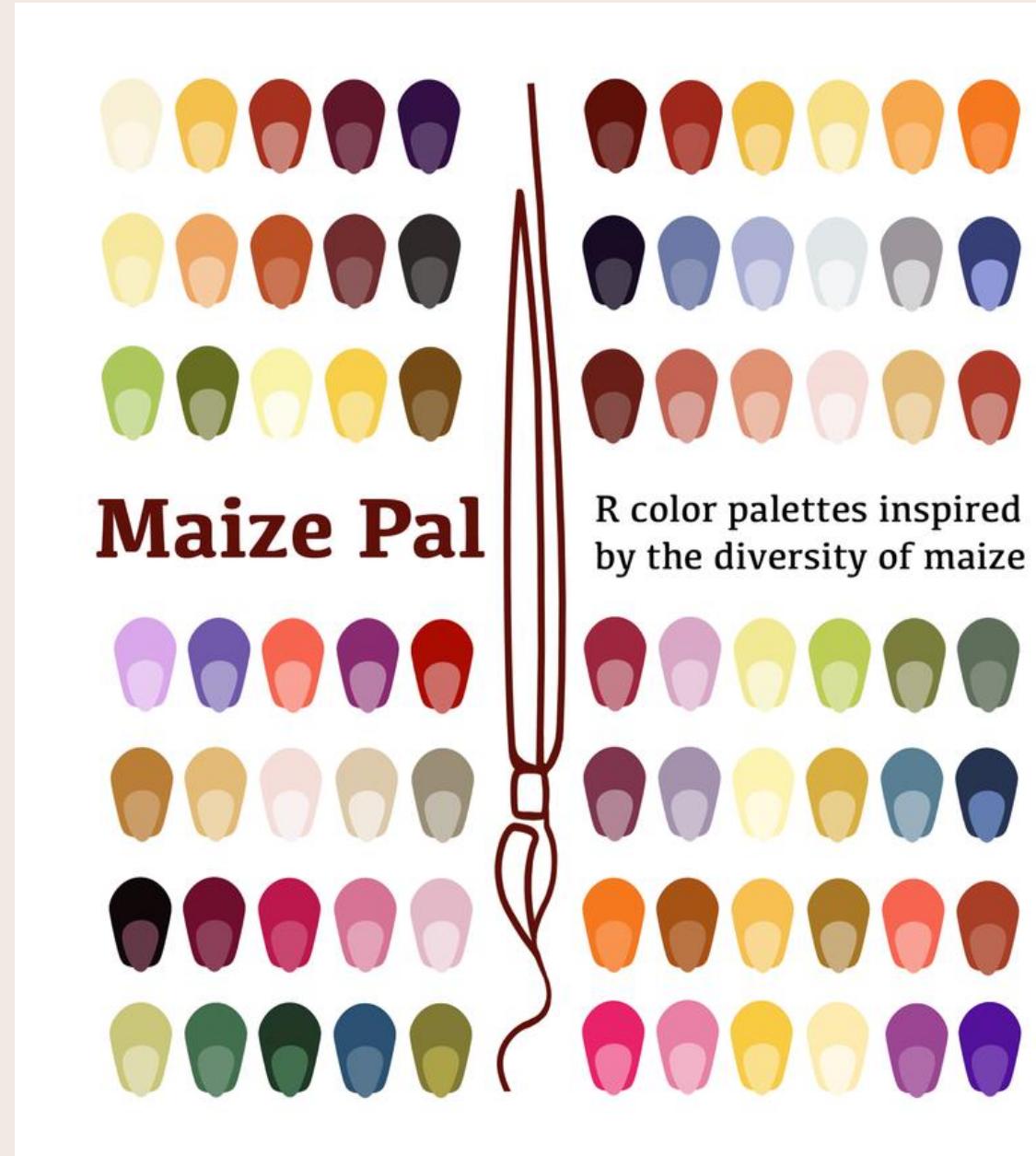
→ [MaizePal](#)



[wesanderson](#)



QUICK GRAPH IN GGPLOT2



MaizePal

GitHub - AndiKur4/MaizePal... +

github.com/AndiKur4/MaizePal?tab=readme-ov-file

Gmail Google Drive Google Calendar YouTube Netflix

README

colored pine cones with purple speckles, and still others that are pitch black when harvested and magenta when milled.

The variety of colors and shapes seems endless, and I could wax poetic about how wrong I was that corn is just yellow and green. Instead, I wrote an R package inspired by the colors and cobs I've seen. This collection of color palettes is not perfect nor complete, and I will continue to add to it as I come across more cobs I would like to document and share. In the meantime, I hope you find this package to be useful, colorful, and fun.

Installation

I have not yet released the package to CRAN, but intend to do so soon. For now, please download directly from github using:

```
devtools::install_github("AndiKur4/MaizePal")
```

Usage

```
library("MaizePal")  
# See all palettes  
names(maize_palettes)  
#> [1] "Anthocyanins1" "Anthocyanins2" "RubyGold"      "Sweetest"  
#> [5] "GlassGem"       "PodCorn"        "HighlandMAGIC" "MaizAzul"  
#> [9] "JimmyRed"        "FloweringTime"  "HopiBlue"       "Painted"  
#> [13] "MaizMorado"     "OaxacaGreen"
```

Palettes

Sweetest

```
maize_pal("Sweetest")
```

QUICK GRAPH IN GGPLOT2

Tidy_seminar_Rproject - main - RStudio

Tidy_seminar_script_Apr17-2025.R

```
84 ####---GRAPH---  
85 #make.a.boxplot.of.avg.wax.per.mg.DW.per.tree  
86 ~  
87 #load.MaizePal.package.for.custom.colors~  
88 #note.that.it.is.not.on.CRAN,.requires.install.using.the.devtools.package.(get.on.CRAN)~  
89 library(MaizePal).#v0.0.1..#devtools::install_github("andiKur4/MaizePal")~  
90 ~  
91 tidier_data. |>~  
92 #change.time.point.column.to.a.factor.with.levels--lets.you.specify.the.order~  
93 ..mutate(time_point.=factor(time_point,.levels.=c("7.dpp","23.dpp","43.dpp"))).|>~  
94 #note.that.you.can.use.pipes.going.in.to.ggplot,|but.it.uses.+.for.additional.arguments.(layers)~  
95 ..ggplot(aes(y=avg_wax_per_mg_DW_per_tree,.x=time_point,.fill=treatment))+-~  
96 ..geom_boxplot()+#adds.the.boxplot.layer~  
97 ..stat_boxplot(geom="errorbar",.position.=."dodge")+#adds.the.errorbars.above.and.below~  
98 ..theme_bw()+#removes.background.and.adds.border~  
99 ..theme(legend.position="top")+#move.the.legend.to.above.the.plot~  
100 ..scale_fill_manual(values.=maize_pal("GlassGem"))~  
101 ~  
102 #export.the.ggplot,.specify.file.type.and.dpi--it.will.save.the.most.recent.plot~  
103 ggsave("figures/tidier_data_plot.tiff",.dpi=600,.width.=5,.height.=5)~  
104 ¶
```

94:49 # GRAPH

Console Background Jobs

R 4.4.1 ~/ANALYSIS

```
> tidier_data |>  
+   mutate(time_point  
+   ggplot(aes(y=avg_...  
+   geom_boxplot()+#adds.the.boxplot.layer  
+   stat_boxplot(geom="errorbar",.position.=."dodge")+#adds.the.errorbars.above.and.below
```

Environment History Connections

Global Environment

Data

- sample_m... 33 obs. of 4 variab...
- summariz... 6 obs. of 5 variabl...
- tidier_d... 97 obs. of 13 varia...
- untidy_s... 97 obs. of 8 variab...

Files Plots Packages Help View

Zoom Export

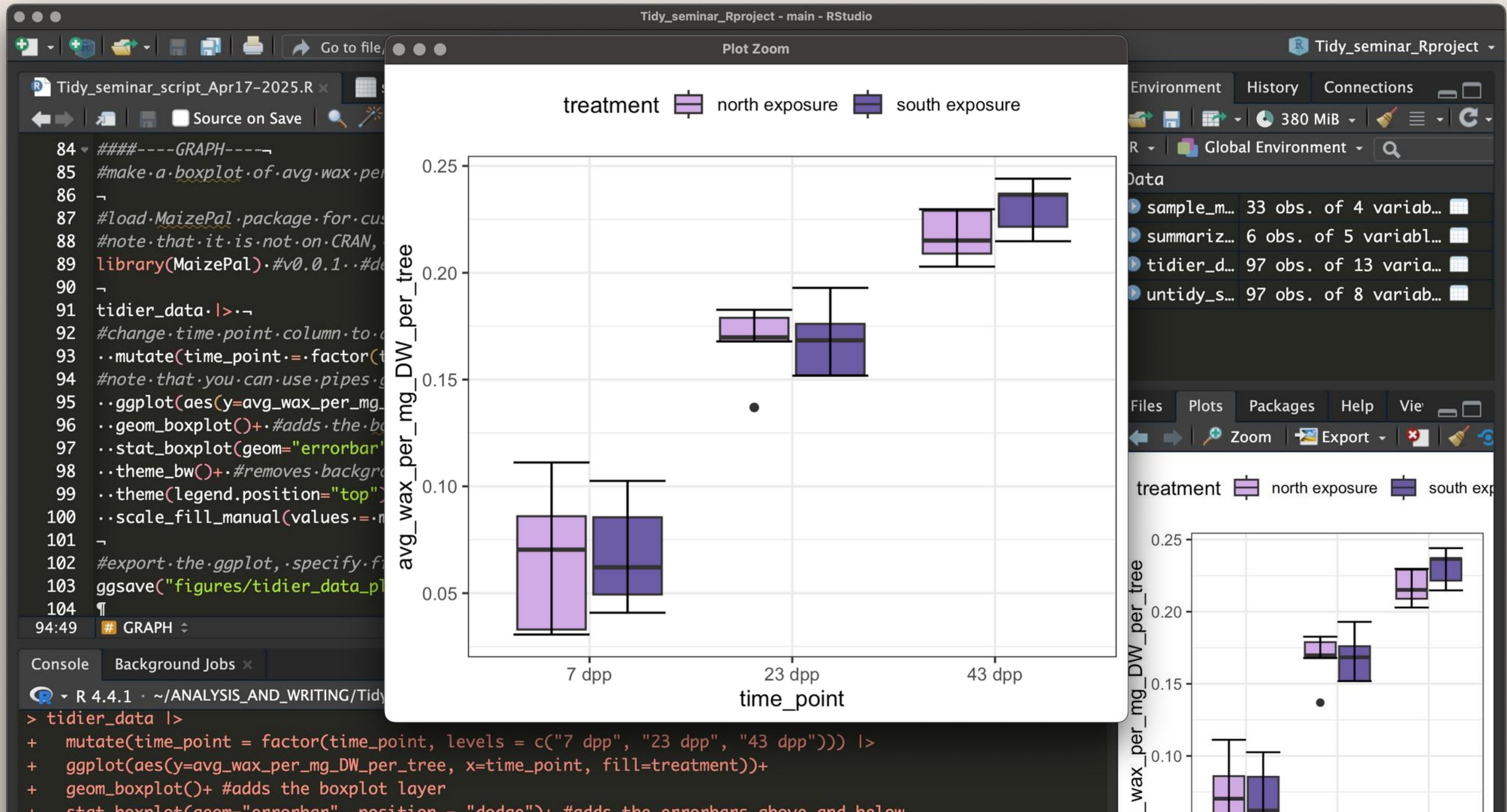
treatment

per_tree

north exposure

south exposure

QUICK GRAPHING IN GGPLOT2



more info on
citing packages

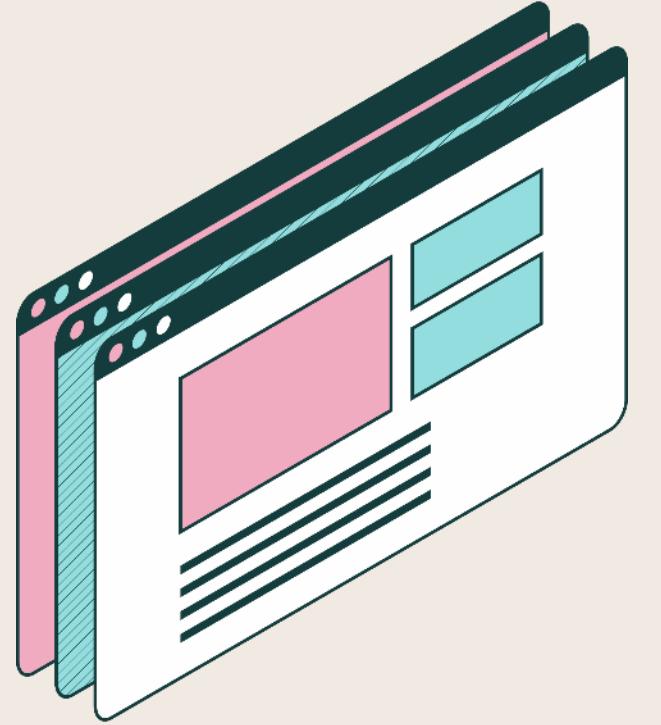
CITING PACKAGES

The screenshot shows the RStudio interface with the following components:

- Top Bar:** Titled "Tidy_seminar_Rproject - main - RStudio". Includes standard icons for file operations, a search bar labeled "Go to file/function", and a "Run" button.
- Left Panel:** A code editor window titled "Tidy_seminar_script_Apr17-2025.R". It contains R code for generating citations. Lines 103-119 are highlighted in yellow, indicating they are currently being run or have been run recently.
- Right Panel:** The "Environment" tab of the global environment pane, showing variables like "sample_m...", "summariz...", "tidier_d...", and "untidy_s...". The "Values" section shows "pkgs" as "chr [1:5] "tidyverse"".
- Bottom Panel:** The "Console" tab of the R Script pane, showing the command `write.bibtex` being run and its output: "Writing 5 Bibtex entries ... OK".

REMEMBER: THE POINT IS REPRODUCIBILITY & TRANSPARENCY

- Try NOT to modify your raw data before reading into R
- Never overwrite original data files/dataframes – use a new name
- if someone has your code (including packages and versions) and data files, they should be able to rerun everything you've done



LIKE THIS STUFF?



COME CHECK OUT THE
BIOINFORMATICS HELP CLUB!

ALL INFO AVAILABLE IN SLACK

(LET ME KNOW IF YOU NEED ACCESS)

THANKS Y'ALL!



INSTALLING R & R STUDIO

#FIRST

#Download and install base R

#SECOND

#Download and install RStudio



INSTALLING & LOADING PACKAGES

#Install:

```
>install.packages("tidyverse") #package name in quotes
```

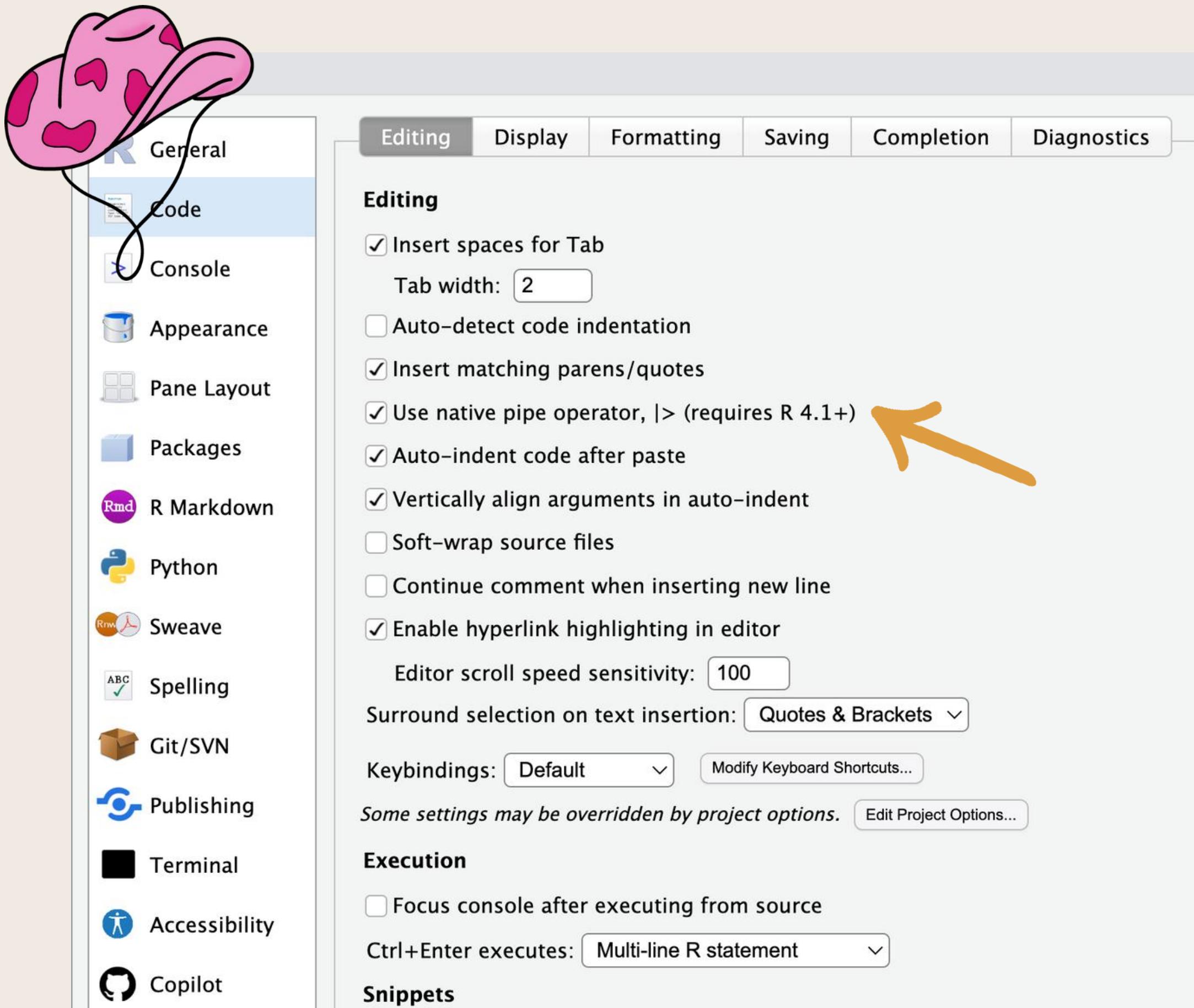
#only for packages available on CRAN

#other packages will usually include install code to use

#load the package:

```
>library(tidyverse) #no quotes
```

CUSTOMIZING R STUDIO



CUSTOMIZING R STUDIO

Options

General

Code

Console

Appearance

Pane Layout

Packages

R Markdown

Python

Sweave

Spelling

Git/SVN

Publishing

Terminal

Accessibility

Copilot

Editing Display Formatting Saving Completion Diagnostics

General

Highlight selected word

Highlight selected line

Show line numbers

Relative line numbers

Show margin

Margin column: 80

Show whitespace characters

Indentation guides: None

Blinking cursor

Allow scroll past end of document

Allow drag and drop of text

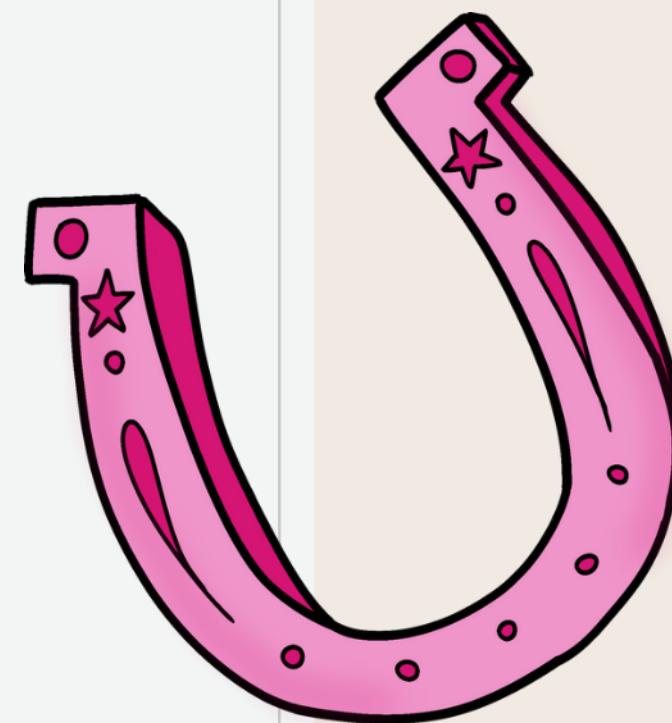
Fold style: Start and End

Syntax

Highlight R function calls

Enable preview of named and hexadecimal colors

Use rainbow parentheses



CUSTOMIZING R STUDIO



Options

R General

Code

> Console

Appearance

Pane Layout

Packages

Rmd R Markdown

Python

Rnw Sweave

ABC Spelling

Git/SVN

UBLISHING

Terminal

Accessibility

Copilot

RStudio theme: Modern

Zoom: 150%

Text rendering: (Default)

Editor font: Monaco

Editor font size: 10

Line height (%): 145

Help font size: 10

Editor theme: Solarized Dark
Solarized Light
SQL Server
Textmate (default)
Tomorrow
Tomorrow Night
Tomorrow Night 80s
Tomorrow Night Blue
Tomorrow Night Bright
Twilight
Vibrant Ink
Wombat
Xcode

```
# compute five-number summary
fivenum <- function(x) {

  # handle empty input
  n <- length(x)
  if (n == 0)
    return(rep.int(NA, 5))

  # compute quartile indices
  n5 <- 1
  n4 <- ((n + 3) %% 2) / 2
  n3 <- (n + 1) / 2
  n2 <- n + 1 - n4
  n1 <- n
  i <- c(n5, n4, n3, n2, n1)

  # compute quartile values
  x <- sort(x)
  xf <- x[floor(i)]
  xc <- x[ceiling(i)]
  0.5 * (xf + xc)

}
```

Add... Remove

OTHER GREAT RESOURCES

#UBC Library Research Commons Workshops

#and their free workshop resources

#including a whole section on Data Management

#Digital Research Alliance youtube

#and newsletter signup to get notified about free workshops

#UBC Advanced Research Computing Workshops

#UAlberta Research Bootcamp Workshop Recordings

OTHER GREAT RESOURCES

#Great Youtube Channels:

#Prof Andy Field - statistics, R & Rmd workflows

#Julia Silge - Tidy Tuesdays work-alongs, Tidy Models tutorials

#Posit PBC - (they own RStudio)

#SFU WestDRI - they co-host many workshops with DRAC