R-eproducible workflows

Brendan Palmer, University College Cork
Adam Kane, University College Dublin
Enrico Pirotta, Washington State University

BEES
School of Biological, Earth & Environmental Sciences at University College Cork, Ireland

CRF-C
HRB Clinical Research Facility Cork

UCD DUBLIN
University College Dublin
Ireland's Global University

WASHINGTON STATE UNIVERSITY

Data integrity

Goodbye point and click

Darren L Dahly
@statsepi

The number of published errors that are 100% due to poor/no training in "basic" data management and manipulation must be enormous.

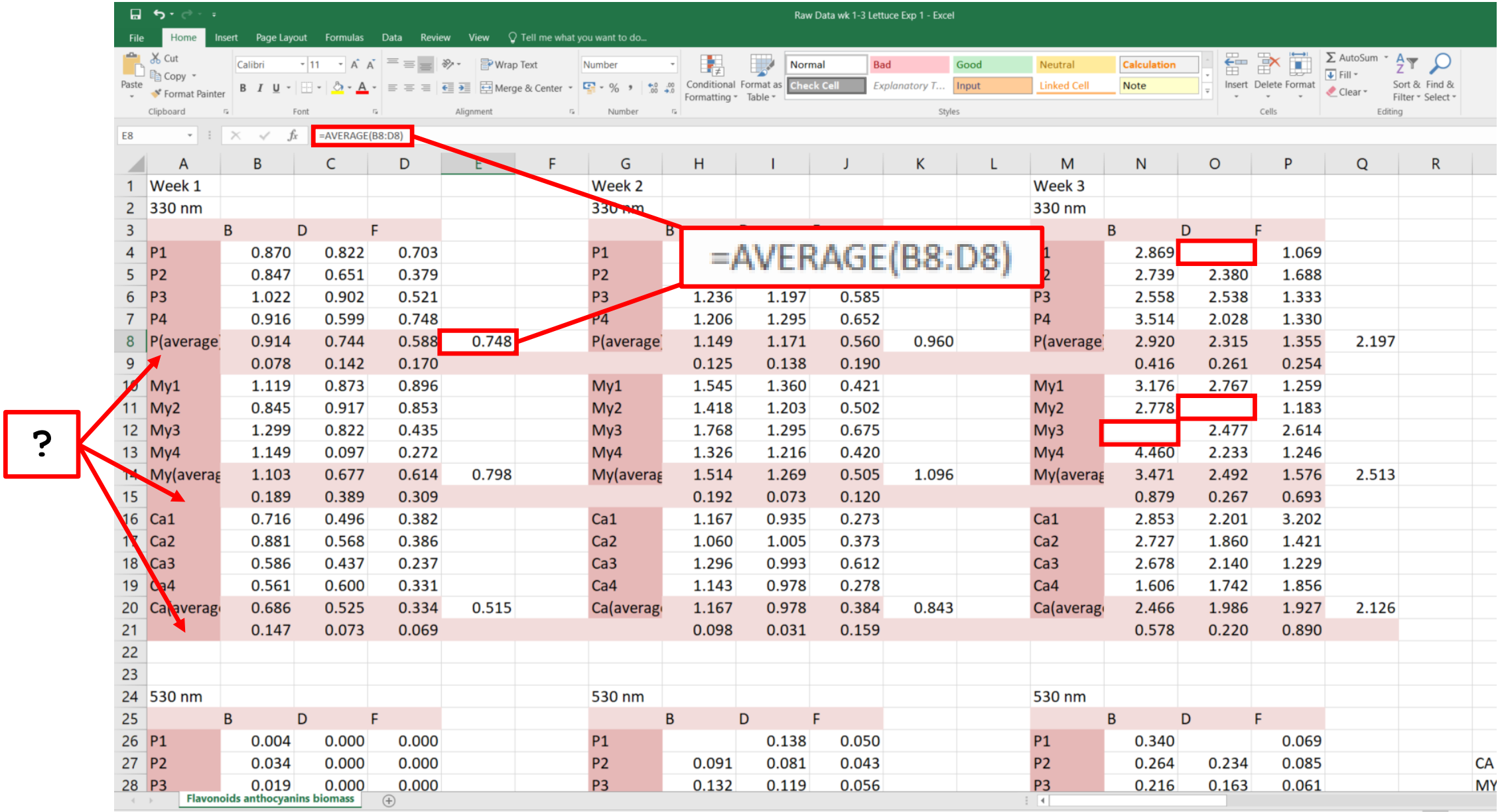
9:31 AM · Sep 17, 2019 · Twitter Web App

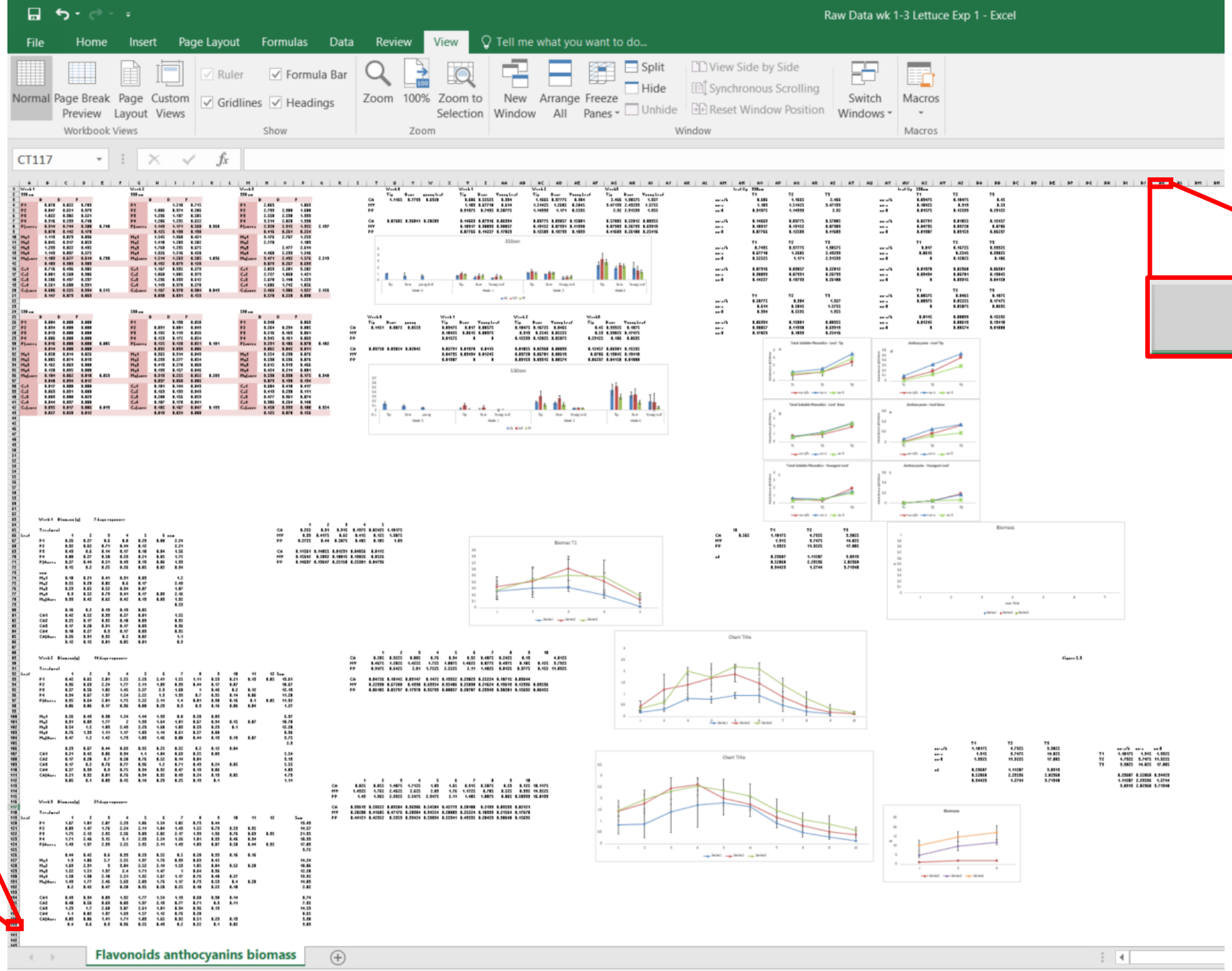
Our real life experiment

- UV light has potential to change the secondary metabolite composition (colour) of bronze/red lettuce

- Experimental setup:
  - 3 lettuce varieties
  - 3 UV filter conditions
  - 3 weeks duration

Real data comes with real problems




Raw Data wk 1-3 Lettuce Exp 1 - Excel
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Tell me what you want to do...
Normal
Page Break Preview
Page Layout
Custom Views
Workbook Views
Ruler
Gridlines
Formula Bar
Headings
Show
Zoom
100%
Zoom to Selection
Zoom
New Window
Arrange All
Freeze Panes
Split
Hide
Unhide
View Side by Side
Synchronous Scrolling
Reset Window Position
Window
Switch Windows
Macros
CT117
Flavonoids anthocyanins biomass

BK

139

Less stress, more success

	A	B	C	D	E	F	G	H	I	J	K	L
1	id	week_no	filter_nam	treatment	replicate_no	flavonoids	biomass	variety	date	investigator		
2	1	0	ptp	nofilter	1	1.061	0.39	cos	2019/04/01	Darren Dahly		
3	2	0	ptp	nofilter	2	1.1805	0.42	cos	2019/04/01	Darren Dahly		
4	3	0	ptp	nofilter	3	1.0345	0.62	cos	2019/04/01	Darren Dahly		
5	4	0	ptp	nofilter	4	1.094	0.63	cos	2019/04/01	Brendan Palmer		
6	1	0	my	nofilter	1	1.061	0.39	cos	2019/04/01	Brendan Palmer		
7	2	0	my	nofilter	2	1.1805	0.42	cos	2019/04/01	Brendan Palmer		
8	3	0	my	nofilter	3	1.0345	0.62	cos	2019/04/01	Brendan Palmer		
9	4	0	my	nofilter	4	1.094	0.63	cos	2019/04/01	Brendan Palmer		
10	1	0	ca	nofilter	1	1.061	0.39	cos	2019/04/01	Brendan Palmer		
11	2	0	ca	nofilter	2	1.1805	0.42	cos	2019/04/01	Brendan Palmer		
12	3	0	ca	nofilter	3	1.0345	0.62	cos	2019/04/01	Brendan Palmer		
13	4	0	ca	nofilter	4	1.094	0.63	cos	2019/04/01	Darren Dahly		
14	5	1	ptp	filter	1	0.87	0.76	cos	2019/04/08	Darren Dahly		
15	6	1	ptp	filter	2	0.847	0.95	cos	2019/04/08	Darren Dahly		
16	7	1	ptp	filter	3	1.022	0.95	cos	2019/04/08	Darren Dahly		
17	8	1	ptp	filter	4	0.916	0.95	cos	2019/04/08	Darren Dahly		
18	9	1	my	filter	1	1.119	1.55	cos	2019/04/08	Darren Dahly		
19	10	1	my	filter	2	0.845	3.16	cos	2019/04/08	Darren Dahly		
20	11	1	my	filter	3	1.299	4.9	cos	2019/04/08	Brendan Palmer		
21	12	1	my	filter	4	1.149	5.5	cos	2019/04/08	Brendan Palmer		
22	13	1	ca	filter	1	0.716	5.5	cos	2019/04/08	Brendan Palmer		
23	14	1	ca	filter	2	0.881	7.94	cos	2019/04/08	Brendan Palmer		
24	15	1	ca	filter	3	0.586	8.71	cos	2019/04/08	Brendan Palmer		
25	16	1	ca	filter	4	0.561	8.71	cos	2019/04/08	Brendan Palmer		
26	17	2	ptp	filter	1	0	14.45	cos	2019/04/15	Brendan Palmer		
27	18	2	ptp	filter	2	1.006	2.14	cos	2019/04/15	Brendan Palmer		
28	19	2	ptp	filter	3	1.236	1.86	cos	2019/04/15	Brendan Palmer		
29	20	2	ptp	filter	4	1.206	1.2	cos	2019/04/15	Brendan Palmer		
30	21	2	my	filter	1	1.545	2.45	cos	2019/04/15	Brendan Palmer		

data | dictionary | values

Less stress, more success

	A	B	C	D	E	F	G	H	I	J	K	L
1	id	week_no	filter_nam	treatment	replicate_no	flavonoids	biomass	variety	date	investigator		
2	1	0	ptp	nofilter	1	1.061	0.39	cos	2019/04/01	Darren Dahly		
3	2	0	ptp									
4	3	0	ptp									
5	4	0	ptp									
6	1	0	my									
7	2	0	my									
8	3	0	my									
9	4	0	my									
10	1	0	ca									
11	2	0	ca									
12	3	0	ca									
13	4	0	ca									
14	5	1	ptp									
15	6	1	ptp									
16	7	1	ptp									
17	8	1	ptp									
18	9	1	my									
19	10	1	my									
20	11	1	my									
21	12	1	my									
22	13	1	ca									
23	14	1	ca									
24	15	1	ca									
25	16	1	ca									
26	17	2	ptp									
27	18	2	ptp									
28	19	2	ptp									
29	20	2	ptp									
30	21	2	my									

data | dictionary

	A	B	C	D	E	F
1	field_name	data_type	data_format	example	standard_units	description
2	id	numeric	integer	23	NA	Unique identifier applied to each observation
3	week_no	numeric	integer	1	NA	Week number, 1 = 7 days exposure, 2 = 14 days exposure
4	filter_name	character	NA	my	NA	3 filter types; 'ptp' = polytunnel plastic blocks all UV light
5	treatment	character	NA	filter	NA	Presence or absence of a filter at the time of sampling
6	replicate_no	numeric	integer	1	NA	The number of replicates in each treatement
7	flavonoids	numeric	double	0.3421	parts per million (ppm)	Leaf disc taken from the tip of the most mature leaf at th
8	biomass	numeric	double		gram (g)	Above ground biomass on the day of harvest
9	variety	character	NA	cos	NA	3 commerical varieties of red lettuce used; 'cos' = Cos Di
10	date	date	YYYY/MM/DD	2019/06/28	ISO 8601	Experiment date
11	investigator	character	Firstname Lastname	Aoife Coffey	NA	Primary researcher who performed the experiment
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						

data | dictionary | values

Less stress, more success

	id	week_no	filter_nam	treatment	replicate_no	flavonoids	biomass	variety	date	investigator
2	1	0	ptp	nofilter	1	1.061	0.39	cos	2019/04/01	Darren Dahly
3	2	0	ptp							
4	3	0	ptp							
5	4	0	ptp							
6	1	0	my							
7	2	0	my							
8	3	0	my							
9	4	0	my							
10	1	0	ca							
11	2	0	ca							
12	3	0	ca							
13	4	0	ca							
14	5	1	ptp							
15	6	1	ptp							
16	7	1	ptp							
17	8	1	ptp							
18	9	1	my							
19	10	1	my							
20	11	1	my							
21	12	1	my							
22	13	1	ca							
23	14	1	ca							
24	15	1	ca							
25	16	1	ca							
26	17	2	ptp							
27	18	2	ptp							
28	19	2	ptp							
29	20	2	ptp							
30	21	2	my							

data | dictionary

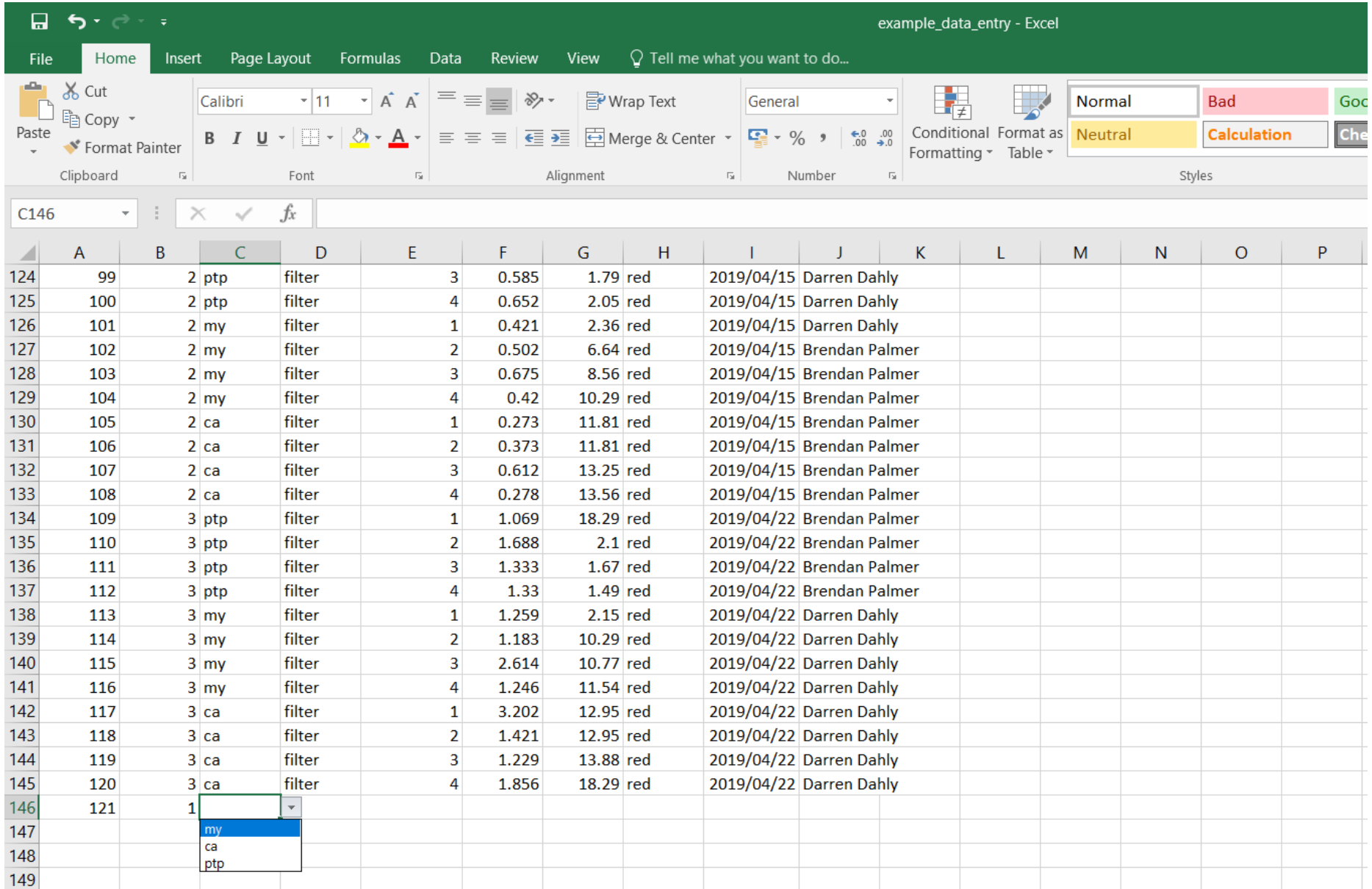
	field_name	data_type	data_format	example	standard_units	description
2	id	numeric	integer	23	NA	Unique identifier applied to each observation
3	week_no	numeric	integer			
4	filter_name	character	NA	n		
5	treatment	character	NA	fi		
6	replicate_no	numeric	integer			
7	flavonoids	numeric	double			
8	biomass	numeric	double			
9	variety	character	NA	c		
10	date	date	YYYY/MM/DD			
11	investigator	character	Firstname Lastname	A		

data | dictionary | values

	id	week_no	filter_name	treatment	replicate_no	flavonoids	biomass	variety	date	investigator
2		0	my	filter	1			cos		Brendan Palmer
3		1	ca	no_filter	2			sky		Darren Dahly
4		2	ptp		3			red		
5		3			4					

data | dictionary | values

Less stress, more success



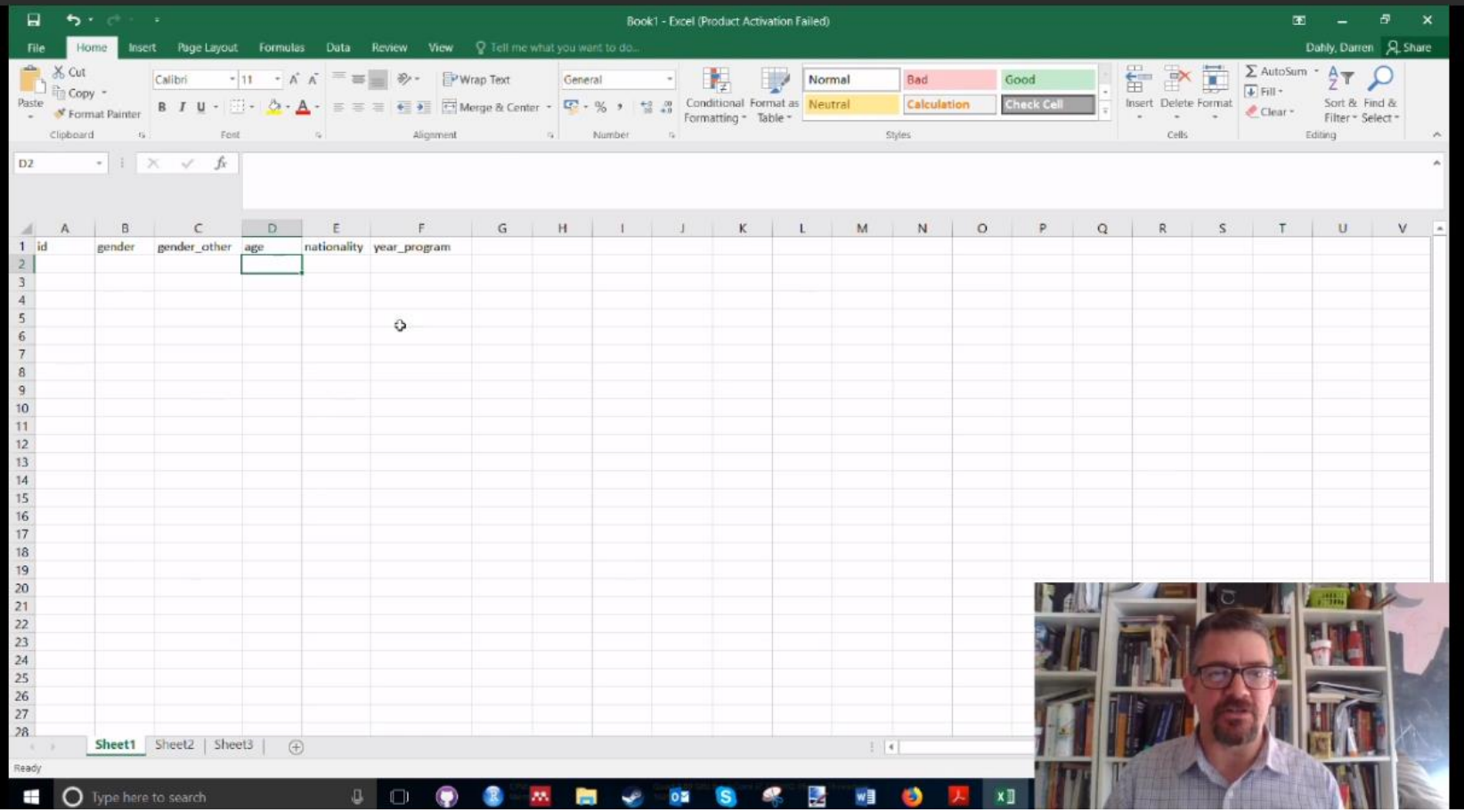
	A	B	C	D	E	F	G	H	I	J
124	99	2	ptp	filter	3	0.585	1.79	red	2019/04/15	Darren Dahly
125	100	2	ptp	filter	4	0.652	2.05	red	2019/04/15	Darren Dahly
126	101	2	my	filter	1	0.421	2.36	red	2019/04/15	Darren Dahly
127	102	2	my	filter	2	0.502	6.64	red	2019/04/15	Brendan Palmer
128	103	2	my	filter	3	0.675	8.56	red	2019/04/15	Brendan Palmer
129	104	2	my	filter	4	0.42	10.29	red	2019/04/15	Brendan Palmer
130	105	2	ca	filter	1	0.273	11.81	red	2019/04/15	Brendan Palmer
131	106	2	ca	filter	2	0.373	11.81	red	2019/04/15	Brendan Palmer
132	107	2	ca	filter	3	0.612	13.25	red	2019/04/15	Brendan Palmer
133	108	2	ca	filter	4	0.278	13.56	red	2019/04/15	Brendan Palmer
134	109	3	ptp	filter	1	1.069	18.29	red	2019/04/22	Brendan Palmer
135	110	3	ptp	filter	2	1.688	2.1	red	2019/04/22	Brendan Palmer
136	111	3	ptp	filter	3	1.333	1.67	red	2019/04/22	Brendan Palmer
137	112	3	ptp	filter	4	1.33	1.49	red	2019/04/22	Brendan Palmer
138	113	3	my	filter	1	1.259	2.15	red	2019/04/22	Darren Dahly
139	114	3	my	filter	2	1.183	10.29	red	2019/04/22	Darren Dahly
140	115	3	my	filter	3	2.614	10.77	red	2019/04/22	Darren Dahly
141	116	3	my	filter	4	1.246	11.54	red	2019/04/22	Darren Dahly
142	117	3	ca	filter	1	3.202	12.95	red	2019/04/22	Darren Dahly
143	118	3	ca	filter	2	1.421	12.95	red	2019/04/22	Darren Dahly
144	119	3	ca	filter	3	1.229	13.88	red	2019/04/22	Darren Dahly
145	120	3	ca	filter	4	1.856	18.29	red	2019/04/22	Darren Dahly
146	121	1								
147										
148										
149										

Step by step guide

https://www.youtube.com/watch?v=Ry2xjTBtNFE

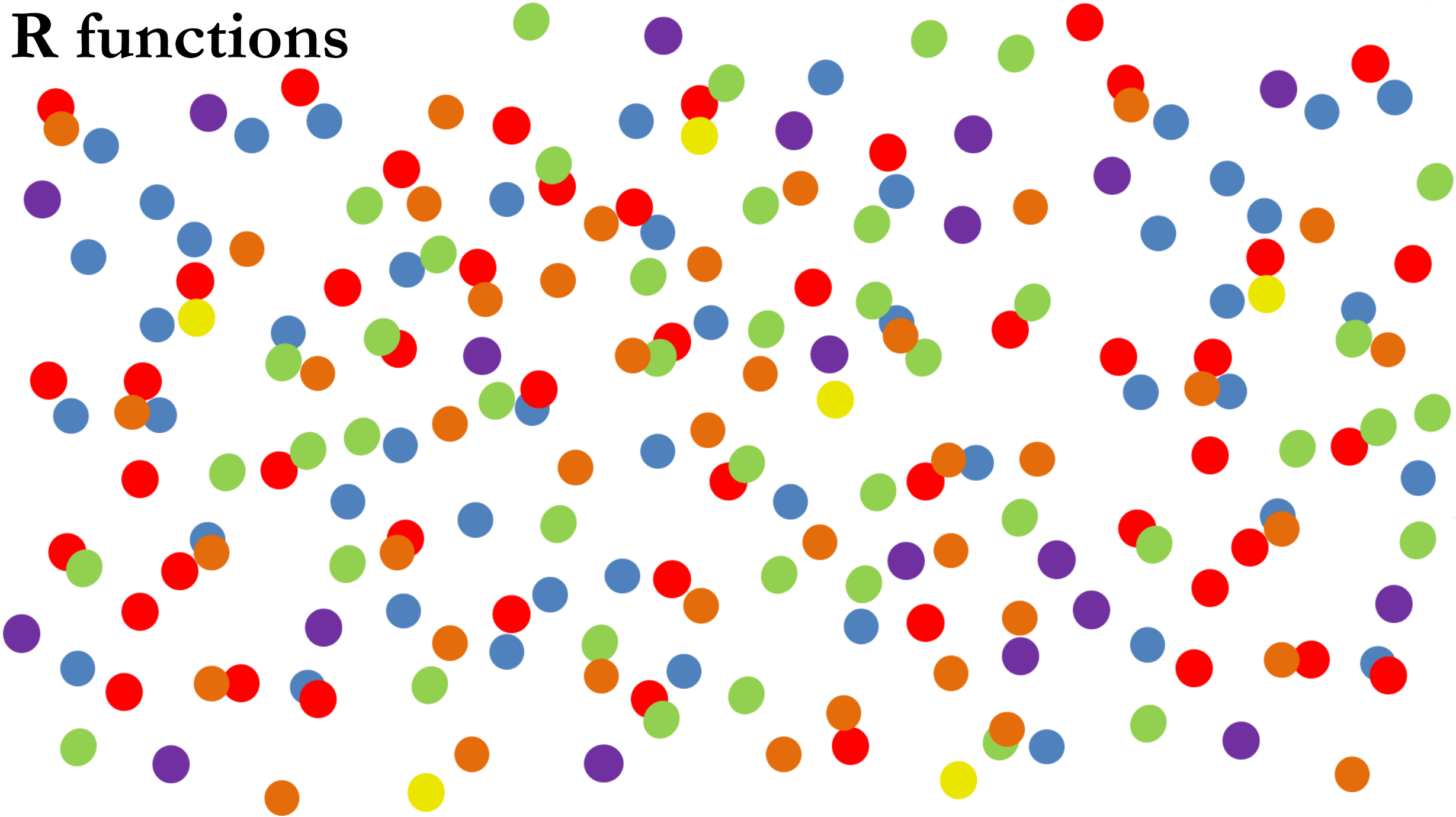
YouTube IE

Search

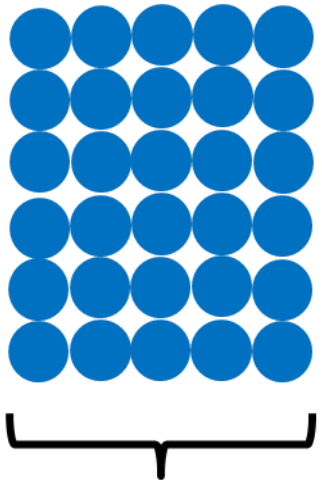


Tidyverse

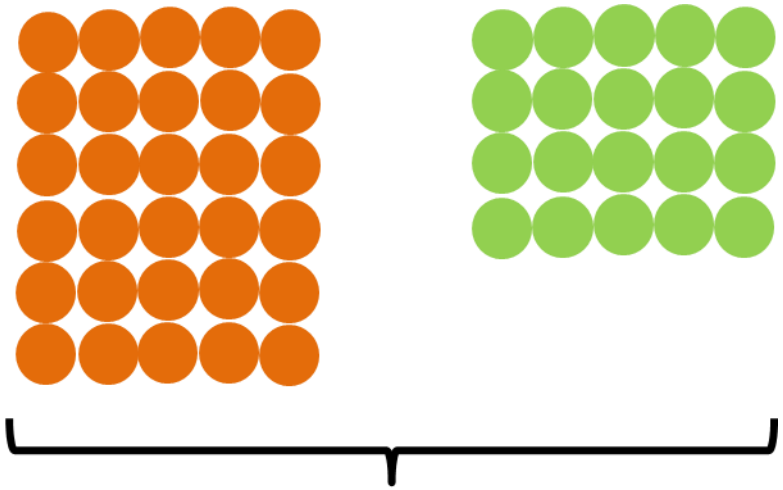
R functions



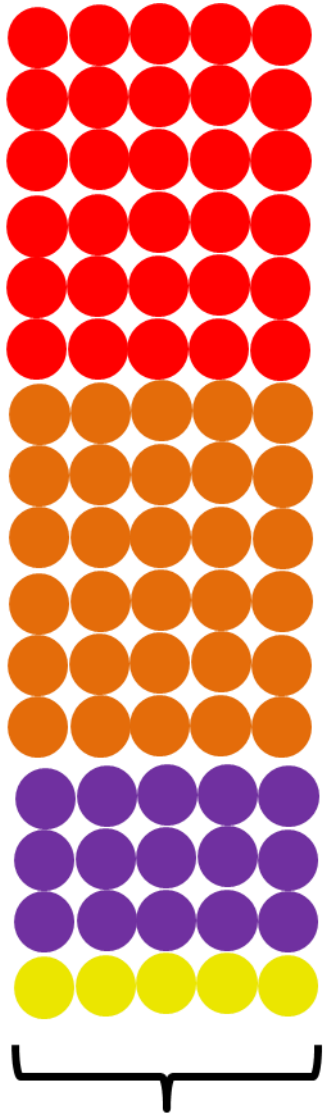
R packages



Base R:
Comes
pre-
loaded



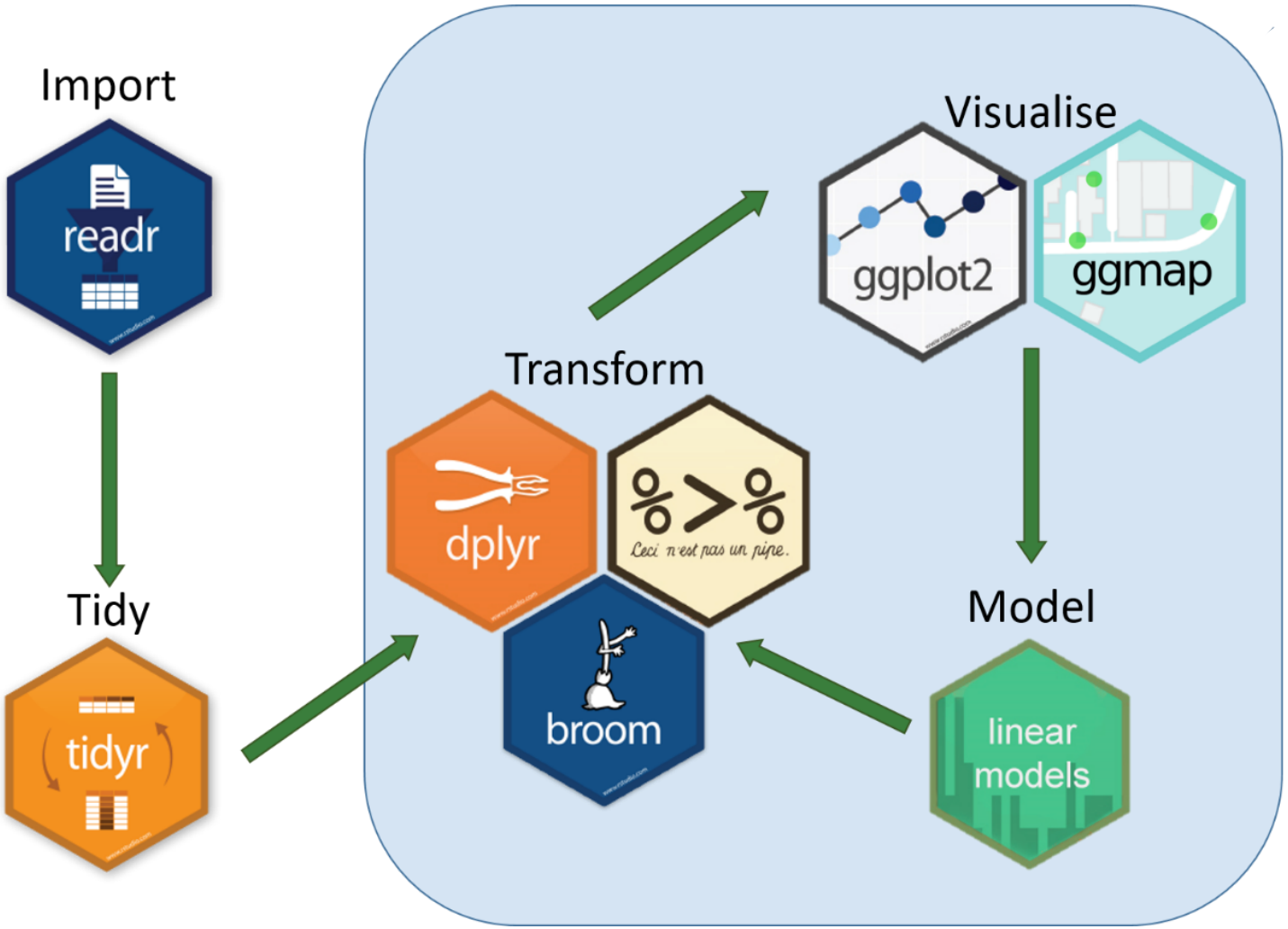
Other packages:
Install once
Update regularly
Load each session



core
tidyverse

Putting the pieces together

- Data analysis in a tidyverse nutshell



Line by line

~/R_Users_Workshop/8_weeks_Oct-Dec_17/Workshop_1/workshop_1_project - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | workshop_1_project — 8_weeks_Oct-Dec_17

ws1_script1_stepwise_Bauer_dataset_an... *

Source on Save | Run | Source

```
12
13  raw_gene_df <- read_delim("Brauer2008_DataSet1.tds", delim = "\t")
14
15  separated_gene_df <- separate(raw_gene_df, NAME,
16                                c("name", "BP", "MF", "systematic_name",
17                                  "number"),
18                                sep = "\\|\\|")
19
20  mutated_gene_df <- mutate_at(separated_gene_df,
21                               vars(name:systematic_name),
22                               funs(trimws)
23                               )
24
25  selected_gene_df <- select(mutated_gene_df, -number, -GID, -YORF, -GWEIGHT)
26
27  gathered_gene_df <- gather(selected_gene_df, sample, expression, G0.05:U0.3)
28
29  nearly_there_df <- separate(gathered_gene_df, sample,
30                              c("nutrient", "rate"), sep = 1, convert = TRUE)
31
32  nutrient_names <- c(G = "Glucose", L = "Leucine", P = "Phosphate",
33                      S = "Sulfate", N = "Ammonia", U = "Uracil")
34
35  cleaned_genes_df <- mutate(nearly_there_df,
36                             nutrient = plyr::revalue(nutrient, nutrient_names)
37                             ) %>%
38    filter(!is.na(expression), systematic_name != "")
```

15:1 | Section 1: Data import, tidying and transformation | R Script

Console Terminal

~/R_Users_Workshop/8_weeks_Oct-Dec_17/Workshop_1/workshop_1_project/

```
> raw_gene_df <- read_delim("Brauer2008_DataSet1.tds", delim = "\t")
Parsed with column specification:
cols(
  .default = col_double(),
  GID = col_character(),
  YORF = col_character(),
  NAME = col_character(),
  GWEIGHT = col_integer()
)
See spec(...) for full column specifications.
>
```

Environment History Connections

Import Dataset | Grid

Global Environment

Name	Type	Length	Size	Value
raw_gene_df	tbl_df	40	3.3 MB	5537 obs. of 40 variables

Files Plots Packages Help Viewer

New Folder | Delete | Rename | More

Home > R_Users_Workshop > 8_weeks_Oct-Dec_17 > Workshop_1 > workshop_1_project

Name	Size	Modified
..		
.RData	2.5 KB	Oct 2, 2017, 1:49 PM
.Rhistory	20.3 KB	Dec 6, 2017, 3:43 PM
Brauer2008_DataSet1.csv	1.6 MB	Sep 27, 2017, 11:32 PM
Brauer2008_DataSet1.tds	1.6 MB	Sep 28, 2017, 10:22 AM
house_completions.csv	4 KB	Sep 28, 2017, 1:35 PM
irish_population.csv	315 B	Aug 28, 2017, 4:21 PM
raw_house_completions.csv	16.2 KB	Aug 25, 2017, 3:45 PM
workshop_1.Rproj	217 B	Oct 18, 2018, 12:18 PM
ws1_script1_stepwise_Bauer_dataset_analysis.R	6.1 KB	Dec 5, 2017, 12:19 PM
ws1_script2_Bauer_dataset_analysis.R	2 KB	Dec 6, 2017, 2:33 PM
ws1_script3_house_completions.R	2.4 KB	Oct 2, 2017, 3:53 PM

Line by line

~/R_Users_Workshop/8_weeks_Oct-Dec_17/Workshop_1/workshop_1_project - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | workshop_1_project — 8_weeks_Oct-Dec_17

ws1_script1_stepwise_Bauer_dataset_an... *

Source on Save | Run | Source

```
12
13  raw_gene_df <- read_delim("Brauer2008_DataSet1.tds", delim = "\t")
14
15  separated_gene_df <- separate(raw_gene_df, NAME,
16                                c("name", "BP", "MF", "systematic_name",
17                                  "number"),
18                                sep = "\\|\\|")
19
20  mutated_gene_df <- mutate_at(separated_gene_df,
21                               vars(name:systematic_name),
22                               funs(trimws)
23                               )
24
25  selected_gene_df <- select(mutated_gene_df, -number, -GID, -YORF, -GWEIGHT)
26
27  gathered_gene_df <- gather(selected_gene_df, sample, expression, G0.05:U0.3)
28
29  nearly_there_df <- separate(gathered_gene_df, sample,
30                              c("nutrient", "rate"), sep = 1, convert = TRUE)
31
32  nutrient_names <- c(G = "Glucose", L = "Leucine", P = "Phosphate",
33                      S = "Sulfate", N = "Ammonia", U = "Uracil")
34
35  cleaned_genes_df <- mutate(nearly_there_df,
36                             nutrient = plyr::revalue(nutrient, nutrient_names)
37                             ) %>%
38    filter(!is.na(expression), systematic_name != "")
```

20:1 | Section 1: Data import, tidying and transformation | R Script

Console Terminal

~/R_Users_Workshop/8_weeks_Oct-Dec_17/Workshop_1/workshop_1_project/

```
> raw_gene_df <- read_delim("Brauer2008_DataSet1.tds", delim = "\t")
Parsed with column specification:
cols(
  .default = col_double(),
  GID = col_character(),
  YORF = col_character(),
  NAME = col_character(),
  GWEIGHT = col_integer()
)
See spec(...) for full column specifications.
> separated_gene_df <- separate(raw_gene_df, NAME,
+                               c("name", "BP", "MF", "systematic_name",
+                                 "number"),
+                               sep = "\\|\\|")
>
```

Environment History Connections

Import Dataset | Grid

Global Environment

Name	Type	Length	Size	Value
raw_gene_df	tbl_df	40	3.3 MB	5537 obs. of 40 variables
separated_gene...	tbl_df	44	3.6 MB	5537 obs. of 44 variables

Files Plots Packages Help Viewer

New Folder | Delete | Rename | More

Home > R_Users_Workshop > 8_weeks_Oct-Dec_17 > Workshop_1 > workshop_1_project

Name	Size	Modified
..		
.RData	2.5 KB	Oct 2, 2017, 1:49 PM
.Rhistory	20.3 KB	Dec 6, 2017, 3:43 PM
Brauer2008_DataSet1.csv	1.6 MB	Sep 27, 2017, 11:32 PM
Brauer2008_DataSet1.tds	1.6 MB	Sep 28, 2017, 10:22 AM
house_completions.csv	4 KB	Sep 28, 2017, 1:35 PM
irish_population.csv	315 B	Aug 28, 2017, 4:21 PM
raw_house_completions.csv	16.2 KB	Aug 25, 2017, 3:45 PM
workshop_1.Rproj	217 B	Oct 18, 2018, 12:18 PM
ws1_script1_stepwise_Bauer_dataset_analysis.R	6.1 KB	Dec 5, 2017, 12:19 PM
ws1_script2_Bauer_dataset_analysis.R	2 KB	Dec 6, 2017, 2:33 PM
ws1_script3_house_completions.R	2.4 KB	Oct 2, 2017, 3:53 PM

Line by line

~/R_Users_Workshop/8_weeks_Oct-Dec_17/Workshop_1/workshop_1_project - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | workshop_1_project — 8_weeks_Oct-Dec_17

ws1_script1_stepwise_Bauer_dataset_an... *

Source on Save | Run | Source

```
12
13  raw_gene_df <- read_delim("Brauer2008_DataSet1.tds", delim = "\t")
14
15  separated_gene_df <- separate(raw_gene_df, NAME,
16                                c("name", "BP", "MF", "systematic_name",
17                                  "number"),
18                                sep = "\\|\\|")
19
20  mutated_gene_df <- mutate_at(separated_gene_df,
21                               vars(name:systematic_name),
22                               funs(trimws)
23                               )
24
25  selected_gene_df <- select(mutated_gene_df, -number, -GID, -YORF, -GWEIGHT)
26
27  gathered_gene_df <- gather(selected_gene_df, sample, expression, G0.05:U0.3)
28
29  nearly_there_df <- separate(gathered_gene_df, sample,
30                              c("nutrient", "rate"), sep = 1, convert = TRUE)
31
32  nutrient_names <- c(G = "Glucose", L = "Leucine", P = "Phosphate",
33                      S = "Sulfate", N = "Ammonia", U = "Uracil")
34
35  cleaned_genes_df <- mutate(nearly_there_df,
36                             nutrient = plyr::revalue(nutrient, nutrient_names)
37                             ) %>%
38    filter(!is.na(expression), systematic_name != "")
```

27:1 | Section 1: Data import, tidying and transformation | R Script

Console Terminal

~/R_Users_Workshop/8_weeks_Oct-Dec_17/Workshop_1/workshop_1_project/

```
Parsed with column specification:
cols(
  .default = col_double(),
  GID = col_character(),
  YORF = col_character(),
  NAME = col_character(),
  GWEIGHT = col_integer()
)
See spec(...) for full column specifications.
> separated_gene_df <- separate(raw_gene_df, NAME,
+                               c("name", "BP", "MF", "systematic_name",
+                                 "number"),
+                               sep = "\\|\\|")
> mutated_gene_df <- mutate_at(separated_gene_df,
+                              vars(name:systematic_name),
+                              funs(trimws)
+                              )
> selected_gene_df <- select(mutated_gene_df, -number, -GID, -YORF, -GWEIGHT)
>
```

Environment History Connections

Import Dataset | Grid

Global Environment

Name	Type	Length	Size	Value
mutated_gene_df	tbl_df	44	3.5 MB	5537 obs. of 44 variables
raw_gene_df	tbl_df	40	3.3 MB	5537 obs. of 40 variables
selected_gene_df	tbl_df	40	2.4 MB	5537 obs. of 40 variables
separated_gene...	tbl_df	44	3.6 MB	5537 obs. of 44 variables

Files Plots Packages Help Viewer

New Folder | Delete | Rename | More

Home > R_Users_Workshop > 8_weeks_Oct-Dec_17 > Workshop_1 > workshop_1_project

Name	Size	Modified
..		
.RData	2.5 KB	Oct 2, 2017, 1:49 PM
.Rhistory	20.3 KB	Dec 6, 2017, 3:43 PM
Brauer2008_DataSet1.csv	1.6 MB	Sep 27, 2017, 11:32 PM
Brauer2008_DataSet1.tds	1.6 MB	Sep 28, 2017, 10:22 AM
house_completions.csv	4 KB	Sep 28, 2017, 1:35 PM
irish_population.csv	315 B	Aug 28, 2017, 4:21 PM
raw_house_completions.csv	16.2 KB	Aug 25, 2017, 3:45 PM
workshop_1.Rproj	217 B	Oct 18, 2018, 12:18 PM
ws1_script1_stepwise_Bauer_dataset_analysis.R	6.1 KB	Dec 5, 2017, 12:19 PM
ws1_script2_Bauer_dataset_analysis.R	2 KB	Dec 6, 2017, 2:33 PM
ws1_script3_house_completions.R	2.4 KB	Oct 2, 2017, 3:53 PM

Line by line

~/R_Users_Workshop/8_weeks_Oct-Dec_17/Workshop_1/workshop_1_project - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | workshop_1_project — 8_weeks_Oct-Dec_17

ws1_script1_stepwise_Bauer_dataset_an... *

Source on Save | Run | Source

```
12
13  raw_gene_df <- read_delim("Brauer2008_DataSet1.tds", delim = "\t")
14
15  separated_gene_df <- separate(raw_gene_df, NAME,
16                                c("name", "BP", "MF", "systematic_name",
17                                  "number"),
18                                sep = "\\|\\|")
19
20  mutated_gene_df <- mutate_at(separated_gene_df,
21                               vars(name:systematic_name),
22                               funs(trimws)
23                               )
24
25  selected_gene_df <- select(mutated_gene_df, -number, -GID, -YORF, -GWEIGHT)
26
27  gathered_gene_df <- gather(selected_gene_df, sample, expression, G0.05:U0.3)
28
29  nearly_there_df <- separate(gathered_gene_df, sample,
30                              c("nutrient", "rate"), sep = 1, convert = TRUE)
31
32  nutrient_names <- c(G = "Glucose", L = "Leucine", P = "Phosphate",
33                      S = "Sulfate", N = "Ammonia", U = "Uracil")
34
35  cleaned_genes_df <- mutate(nearly_there_df,
36                             nutrient = plyr::revalue(nutrient, nutrient_names)
37                             ) %>%
38    filter(!is.na(expression), systematic_name != "")
```

29:1 | Section 1: Data import, tidying and transformation | R Script

Console Terminal

~/R_Users_Workshop/8_weeks_Oct-Dec_17/Workshop_1/workshop_1_project/

```
cols(
  .default = col_double(),
  GID = col_character(),
  YORF = col_character(),
  NAME = col_character(),
  GWEIGHT = col_integer()
)
See spec(...) for full column specifications.
> separated_gene_df <- separate(raw_gene_df, NAME,
+                               c("name", "BP", "MF", "systematic_name",
+                                 "number"),
+                               sep = "\\|\\|")
> mutated_gene_df <- mutate_at(separated_gene_df,
+                              vars(name:systematic_name),
+                              funs(trimws)
+                              )
> selected_gene_df <- select(mutated_gene_df, -number, -GID, -YORF, -GWEIGHT)
> gathered_gene_df <- gather(selected_gene_df, sample, expression, G0.05:U0.3)
>
```

Environment History Connections

Import Dataset | Grid

Global Environment

Name	Type	Length	Size	Value
gathered_gene_df	tbl_df	6	9.8 MB	199332 obs. of 6 variables
mutated_gene_df	tbl_df	44	3.5 MB	5537 obs. of 44 variables
raw_gene_df	tbl_df	40	3.3 MB	5537 obs. of 40 variables
selected_gene_df	tbl_df	40	2.4 MB	5537 obs. of 40 variables
separated_gene...	tbl_df	44	3.6 MB	5537 obs. of 44 variables

Files Plots Packages Help Viewer

New Folder | Delete | Rename | More

Home > R_Users_Workshop > 8_weeks_Oct-Dec_17 > Workshop_1 > workshop_1_project

Name	Size	Modified
..		
.RData	2.5 KB	Oct 2, 2017, 1:49 PM
.Rhistory	20.3 KB	Dec 6, 2017, 3:43 PM
Brauer2008_DataSet1.csv	1.6 MB	Sep 27, 2017, 11:32 PM
Brauer2008_DataSet1.tds	1.6 MB	Sep 28, 2017, 10:22 AM
house_completions.csv	4 KB	Sep 28, 2017, 1:35 PM
irish_population.csv	315 B	Aug 28, 2017, 4:21 PM
raw_house_completions.csv	16.2 KB	Aug 25, 2017, 3:45 PM
workshop_1.Rproj	217 B	Oct 18, 2018, 12:18 PM
ws1_script1_stepwise_Bauer_dataset_analysis.R	6.1 KB	Dec 5, 2017, 12:19 PM
ws1_script2_Bauer_dataset_analysis.R	2 KB	Dec 6, 2017, 2:33 PM
ws1_script3_house_completions.R	2.4 KB	Oct 2, 2017, 3:53 PM

Line by line

~/R_Users_Workshop/8_weeks_Oct-Dec_17/Workshop_1/workshop_1_project - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | workshop_1_project — 8_weeks_Oct-Dec_17

ws1_script1_stepwise_Bauer_dataset_an... *

Source on Save | Run | Source

```
12
13  raw_gene_df <- read_delim("Brauer2008_DataSet1.tds", delim = "\t")
14
15  separated_gene_df <- separate(raw_gene_df, NAME,
16                                c("name", "BP", "MF", "systematic_name",
17                                  "number"),
18                                sep = "\\|\\|")
19
20  mutated_gene_df <- mutate_at(separated_gene_df,
21                               vars(name:systematic_name),
22                               funs(trimws)
23                               )
24
25  selected_gene_df <- select(mutated_gene_df, -number, -GID, -YORF, -GWEIGHT)
26
27  gathered_gene_df <- gather(selected_gene_df, sample, expression, G0.05:U0.3)
28
29  nearly_there_df <- separate(gathered_gene_df, sample,
30                              c("nutrient", "rate"), sep = 1, convert = TRUE)
31
32  nutrient_names <- c(G = "Glucose", L = "Leucine", P = "Phosphate",
33                      S = "Sulfate", N = "Ammonia", U = "Uracil")
34
35  cleaned_genes_df <- mutate(nearly_there_df,
36                             nutrient = plyr::revalue(nutrient, nutrient_names)
37                             ) %>%
38    filter(!is.na(expression), systematic_name != "")
```

32:1 | Section 1: Data import, tidying and transformation | R Script

Console Terminal

~/R_Users_Workshop/8_weeks_Oct-Dec_17/Workshop_1/workshop_1_project/

```
  GID = col_character(),
  YORF = col_character(),
  NAME = col_character(),
  GWEIGHT = col_integer()
)
See spec(...) for full column specifications.
> separated_gene_df <- separate(raw_gene_df, NAME,
+                               c("name", "BP", "MF", "systematic_name",
+                                 "number"),
+                               sep = "\\|\\|")
> mutated_gene_df <- mutate_at(separated_gene_df,
+                              vars(name:systematic_name),
+                              funs(trimws)
+                              )
> selected_gene_df <- select(mutated_gene_df, -number, -GID, -YORF, -GWEIGHT)
> gathered_gene_df <- gather(selected_gene_df, sample, expression, G0.05:U0.3)
> nearly_there_df <- separate(gathered_gene_df, sample,
+                             c("nutrient", "rate"), sep = 1, convert = TRUE)
> |
```

Environment History Connections

Import Dataset | Grid

Global Environment

Name	Type	Length	Size	Value
gathered_gene_df	tbl_df	6	9.8 MB	199332 obs. of 6 variables
mutated_gene_df	tbl_df	44	3.5 MB	5537 obs. of 44 variables
nearly_there_df	tbl_df	7	11.3 MB	199332 obs. of 7 variables
raw_gene_df	tbl_df	40	3.3 MB	5537 obs. of 40 variables
selected_gene_df	tbl_df	40	2.4 MB	5537 obs. of 40 variables
separated_gene...	tbl_df	44	3.6 MB	5537 obs. of 44 variables

Files Plots Packages Help Viewer

New Folder | Delete | Rename | More

Home > R_Users_Workshop > 8_weeks_Oct-Dec_17 > Workshop_1 > workshop_1_project

Name	Size	Modified
..		
.RData	2.5 KB	Oct 2, 2017, 1:49 PM
.Rhistory	20.3 KB	Dec 6, 2017, 3:43 PM
Brauer2008_DataSet1.csv	1.6 MB	Sep 27, 2017, 11:32 PM
Brauer2008_DataSet1.tds	1.6 MB	Sep 28, 2017, 10:22 AM
house_completions.csv	4 KB	Sep 28, 2017, 1:35 PM
irish_population.csv	315 B	Aug 28, 2017, 4:21 PM
raw_house_completions.csv	16.2 KB	Aug 25, 2017, 3:45 PM
workshop_1.Rproj	217 B	Oct 18, 2018, 12:18 PM
ws1_script1_stepwise_Bauer_dataset_analysis.R	6.1 KB	Dec 5, 2017, 12:19 PM
ws1_script2_Bauer_dataset_analysis.R	2 KB	Dec 6, 2017, 2:33 PM
ws1_script3_house_completions.R	2.4 KB	Oct 2, 2017, 3:53 PM

Line by line

~/R_Users_Workshop/8_weeks_Oct-Dec_17/Workshop_1/workshop_1_project - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | workshop_1_project — 8_weeks_Oct-Dec_17

ws1_script1_stepwise_Bauer_dataset_an... *

Source on Save | Run | Source

```
12
13  raw_gene_df <- read_delim("Brauer2008_DataSet1.tds", delim = "\t")
14
15  separated_gene_df <- separate(raw_gene_df, NAME,
16                                c("name", "BP", "MF", "systematic_name",
17                                  "number"),
18                                sep = "\\|\\|")
19
20  mutated_gene_df <- mutate_at(separated_gene_df,
21                               vars(name:systematic_name),
22                               funs(trimws)
23                               )
24
25  selected_gene_df <- select(mutated_gene_df, -number, -GID, -YORF, -GWEIGHT)
26
27  gathered_gene_df <- gather(selected_gene_df, sample, expression, G0.05:U0.3)
28
29  nearly_there_df <- separate(gathered_gene_df, sample,
30                              c("nutrient", "rate"), sep = 1, convert = TRUE)
31
32  nutrient_names <- c(G = "Glucose", L = "Leucine", P = "Phosphate",
33                      S = "Sulfate", N = "Ammonia", U = "Uracil")
34
35  cleaned_genes_df <- mutate(nearly_there_df,
36                             nutrient = plyr::revalue(nutrient, nutrient_names)
37                             ) %>%
38    filter(!is.na(expression), systematic_name != "")
```

35:1 | Section 1: Data import, tidying and transformation | R Script

Console Terminal

~/R_Users_Workshop/8_weeks_Oct-Dec_17/Workshop_1/workshop_1_project/

```
  NAME = col_character(),
  GWEIGHT = col_integer()
)
See spec(...) for full column specifications.
> separated_gene_df <- separate(raw_gene_df, NAME,
+                               c("name", "BP", "MF", "systematic_name",
+                                 "number"),
+                               sep = "\\|\\|")
> mutated_gene_df <- mutate_at(separated_gene_df,
+                              vars(name:systematic_name),
+                              funs(trimws)
+                              )
> selected_gene_df <- select(mutated_gene_df, -number, -GID, -YORF, -GWEIGHT)
> gathered_gene_df <- gather(selected_gene_df, sample, expression, G0.05:U0.3)
> nearly_there_df <- separate(gathered_gene_df, sample,
+                             c("nutrient", "rate"), sep = 1, convert = TRUE)
> nutrient_names <- c(G = "Glucose", L = "Leucine", P = "Phosphate",
+                     S = "Sulfate", N = "Ammonia", U = "Uracil")
> |
```

Environment | History | Connections

Import Dataset | Grid

Global Environment

Name	Type	Length	Size	Value
gathered_gene_df	tbl_df	6	9.8 MB	199332 obs. of 6 variables
mutated_gene_df	tbl_df	44	3.5 MB	5537 obs. of 44 variables
nearly_there_df	tbl_df	7	11.3 MB	199332 obs. of 7 variables
nutrient_names	character	6	984 B	Named chr [1:6] "Glucose" "...
raw_gene_df	tbl_df	40	3.3 MB	5537 obs. of 40 variables
selected_gene_df	tbl_df	40	2.4 MB	5537 obs. of 40 variables
separated_gene...	tbl_df	44	3.6 MB	5537 obs. of 44 variables

Files | Plots | Packages | Help | Viewer

New Folder | Delete | Rename | More

Home > R_Users_Workshop > 8_weeks_Oct-Dec_17 > Workshop_1 > workshop_1_project

Name	Size	Modified
..		
.RData	2.5 KB	Oct 2, 2017, 1:49 PM
.Rhistory	20.3 KB	Dec 6, 2017, 3:43 PM
Brauer2008_DataSet1.csv	1.6 MB	Sep 27, 2017, 11:32 PM
Brauer2008_DataSet1.tds	1.6 MB	Sep 28, 2017, 10:22 AM
house_completions.csv	4 KB	Sep 28, 2017, 1:35 PM
irish_population.csv	315 B	Aug 28, 2017, 4:21 PM
raw_house_completions.csv	16.2 KB	Aug 25, 2017, 3:45 PM
workshop_1.Rproj	217 B	Oct 18, 2018, 12:18 PM
ws1_script1_stepwise_Bauer_dataset_analysis.R	6.1 KB	Dec 5, 2017, 12:19 PM
ws1_script2_Bauer_dataset_analysis.R	2 KB	Dec 6, 2017, 2:33 PM
ws1_script3_house_completions.R	2.4 KB	Oct 2, 2017, 3:53 PM

Line by line

~/R_Users_Workshop/8_weeks_Oct-Dec_17/Workshop_1/workshop_1_project - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | workshop_1_project — 8_weeks_Oct-Dec_17

ws1_script1_stepwise_Bauer_dataset_an... *

Source on Save | Run | Source

```
15 separated_gene_df <- separate(raw_gene_df, NAME,
16                               c("name", "BP", "MF", "systematic_name",
17                                 "number"),
18                               sep = "\\|\\|")
19
20 mutated_gene_df <- mutate_at(separated_gene_df,
21                              vars(name:systematic_name),
22                              funs(trimws)
23                              )
24
25 selected_gene_df <- select(mutated_gene_df, -number, -GID, -YORF, -GWEIGHT)
26
27 gathered_gene_df <- gather(selected_gene_df, sample, expression, G0.05:U0.3)
28
29 nearly_there_df <- separate(gathered_gene_df, sample,
30                              c("nutrient", "rate"), sep = 1, convert = TRUE)
31
32 nutrient_names <- c(G = "Glucose", L = "Leucine", P = "Phosphate",
33                     S = "Sulfate", N = "Ammonia", U = "Uracil")
34
35 cleaned_genes_df <- mutate(nearly_there_df,
36                            nutrient = plyr::revalue(nutrient, nutrient_names)
37                            ) %>%
38   filter(!is.na(expression), systematic_name != "")
39
40
41
```

44:1 | Section 1: Data import, tidying and transformation | R Script

Console Terminal

~/R_Users_Workshop/8_weeks_Oct-Dec_17/Workshop_1/workshop_1_project/

```
> separated_gene_df <- separate(raw_gene_df, NAME,
+                               c("name", "BP", "MF", "systematic_name",
+                                 "number"),
+                               sep = "\\|\\|")
> mutated_gene_df <- mutate_at(separated_gene_df,
+                              vars(name:systematic_name),
+                              funs(trimws)
+                              )
> selected_gene_df <- select(mutated_gene_df, -number, -GID, -YORF, -GWEIGHT)
> gathered_gene_df <- gather(selected_gene_df, sample, expression, G0.05:U0.3)
> nearly_there_df <- separate(gathered_gene_df, sample,
+                             c("nutrient", "rate"), sep = 1, convert = TRUE)
> nutrient_names <- c(G = "Glucose", L = "Leucine", P = "Phosphate",
+                     S = "Sulfate", N = "Ammonia", U = "Uracil")
> cleaned_genes_df <- mutate(nearly_there_df,
+                            nutrient = plyr::revalue(nutrient, nutrient_names)
+                            ) %>%
+   filter(!is.na(expression), systematic_name != "")
>
```

Environment History Connections

Import Dataset | Grid

Global Environment

Name	Type	Length	Size	Value
cleaned_genes_df	tbl_df	7	11.3 MB	198430 obs. of 7 variables
gathered_gene_df	tbl_df	6	9.8 MB	199332 obs. of 6 variables
mutated_gene_df	tbl_df	44	3.5 MB	5537 obs. of 44 variables
nearly_there_df	tbl_df	7	11.3 MB	199332 obs. of 7 variables
nutrient_names	character	6	984 B	Named chr [1:6] "Glucose" "...
raw_gene_df	tbl_df	40	3.3 MB	5537 obs. of 40 variables
selected_gene_df	tbl_df	40	2.4 MB	5537 obs. of 40 variables
separated_gene...	tbl_df	44	3.6 MB	5537 obs. of 44 variables

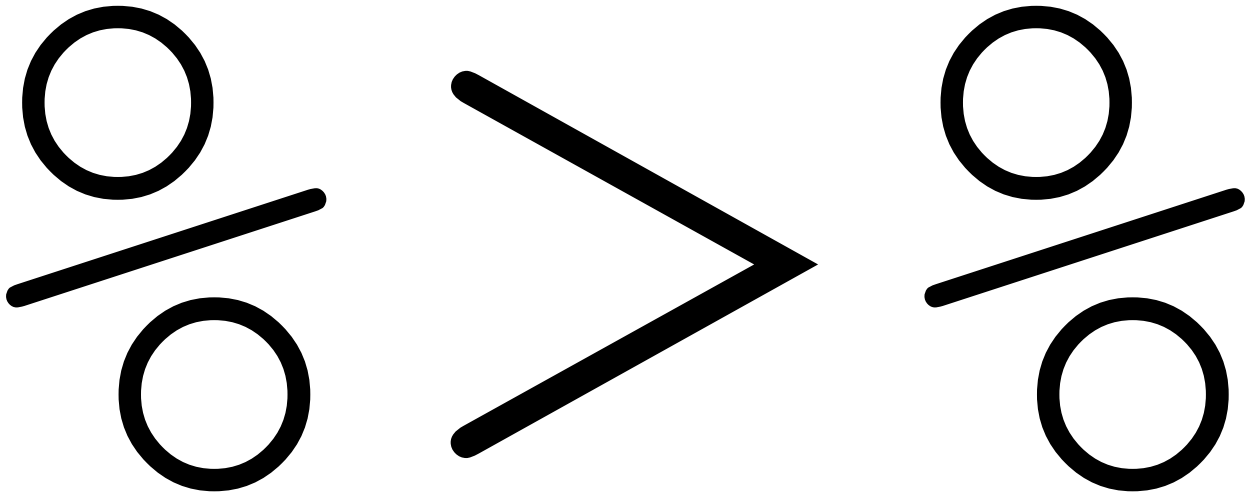
Files Plots Packages Help Viewer

New Folder | Delete | Rename | More

Home > R_Users_Workshop > 8_weeks_Oct-Dec_17 > Workshop_1 > workshop_1_project

Name	Size	Modified
..		
.RData	2.5 KB	Oct 2, 2017, 1:49 PM
.Rhistory	20.3 KB	Dec 6, 2017, 3:43 PM
Brauer2008_DataSet1.csv	1.6 MB	Sep 27, 2017, 11:32 PM
Brauer2008_DataSet1.tds	1.6 MB	Sep 28, 2017, 10:22 AM
house_completions.csv	4 KB	Sep 28, 2017, 1:35 PM
irish_population.csv	315 B	Aug 28, 2017, 4:21 PM
raw_house_completions.csv	16.2 KB	Aug 25, 2017, 3:45 PM
workshop_1.Rproj	217 B	Oct 18, 2018, 12:18 PM
ws1_script1_stepwise_Bauer_dataset_analysis.R	6.1 KB	Dec 5, 2017, 12:19 PM
ws1_script2_Bauer_dataset_analysis.R	2 KB	Dec 6, 2017, 2:33 PM
ws1_script3_house_completions.R	2.4 KB	Oct 2, 2017, 3:53 PM

Putting the pieces together



Tidyverse code structure

new_object - assign the output to a new object
<- - the assign operator
input_data - data to be manipulated
%>% - the magrittr/pipe operator
function - the function you are calling on
data_ - elements of the input data to use
arguments_ - how you want to apply the function

Piped

~/R_Users_Workshop/8_weeks_Oct-Dec_17/Workshop_1/workshop_1_project - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | workshop_1_project — 8_weeks_Oct-Dec_17

ws1_script1_stepwise_Bauer_dataset_an... | ws1_script2_Bauer_dataset_analysis.R*

Source on Save | Run | Source

```
1  nutrient_names <- c(G = "Glucose", L = "Leucine", P = "Phosphate",
2                      S = "Sulfate", N = "Ammonia", U = "Uracil"
3                      )
4
5  cleaned_genes_df <- read_delim("Brauer2008_DataSet1.tds", delim = "\t"
6                                 ) %>%
7
8    separate(NAME, c("name", "BP", "MF", "systematic_name", "number"), sep = "\\|\\|"
9             )  %>%
10
11   mutate_at(vars(name:systematic_name), funs(trimws)
12             ) %>%
13
14   select(-number, -GID, -YORF, -GWEIGHT
15          ) %>%
16
17   gather(sample, expression, G0.05:U0.3
18          ) %>%
19
20   separate(sample, c("nutrient", "rate"), sep = 1, convert = TRUE
21            ) %>%
22
23   mutate(nutrient = plyr::revalue(nutrient, nutrient_names)
24          ) %>%
25
26   filter(!is.na(expression), systematic_name != ""
27          )
```

9:18 (Top Level) R Script

Console | Terminal

~/R_Users_Workshop/8_weeks_Oct-Dec_17/Workshop_1/workshop_1_project/

```
+
+   separate(sample, c("nutrient", "rate"), sep = 1, convert = TRUE
+            ) %>%
+
+   mutate(nutrient = plyr::revalue(nutrient, nutrient_names)
+          ) %>%
+
+   filter(!is.na(expression), systematic_name != ""
+          )
Parsed with column specification:
cols(
  .default = col_double(),
  GID = col_character(),
  YORF = col_character(),
  NAME = col_character(),
  GWEIGHT = col_integer()
)
See spec(...) for full column specifications.
>
```

Environment | History | Connections

Import Dataset | Grid

Global Environment

Name	Type	Length	Size	Value
cleaned_genes_df	tbl_df	7	11.3 MB	198430 obs. of 7 variables
nutrient_names	character	6	984 B	Named chr [1:6] "Glucose" "Le...

Files | Plots | Packages | Help | Viewer

New Folder | Delete | Rename | More

Home > R_Users_Workshop > 8_weeks_Oct-Dec_17 > Workshop_1 > workshop_1_project

Name	Size	Modified
..		
.RData	2.5 KB	Oct 2, 2017, 1:49 PM
.Rhistory	20.3 KB	Dec 6, 2017, 3:43 PM
Brauer2008_DataSet1.csv	1.6 MB	Sep 27, 2017, 11:32 PM
Brauer2008_DataSet1.tds	1.6 MB	Sep 28, 2017, 10:22 AM
house_completions.csv	4 KB	Sep 28, 2017, 1:35 PM
irish_population.csv	315 B	Aug 28, 2017, 4:21 PM
raw_house_completions.csv	16.2 KB	Aug 25, 2017, 3:45 PM
workshop_1.Rproj	217 B	Oct 18, 2018, 12:18 PM
ws1_script1_stepwise_Bauer_dataset_analysis.R	6.1 KB	Dec 5, 2017, 12:19 PM
ws1_script2_Bauer_dataset_analysis.R	2 KB	Dec 6, 2017, 2:33 PM
ws1_script3_house_completions.R	2.4 KB	Oct 2, 2017, 3:53 PM

R Markdown

R Markdown

- R Markdown combines the code you wrote, the output produced and you own comments

- You can view it as a digital lab notebook, where you are both recording what you're doing, and what you were thinking while you were doing it!

- R Markdown outputs can take many forms
    - Word documents, PDFs, slideshows etc.

R Markdown

YAML header

```
---
title: "Diamond sizes"
date: 2016-08-25
output: html_document
---
```

Chunks of code

````
```{r setup, include = FALSE}
library(ggplot2)
library(dplyr)
smaller <- diamonds %>%
filter(carat <= 2.5)
```
````

Plain text with data outputs from R code

````
We have data about `r nrow(diamonds)`
diamonds. Only
`r nrow(diamonds) - nrow(smaller)` are
larger than
2.5 carats. The distribution of the
remainder is shown below:
````

Chunks of code

````
```{r, echo = FALSE}
smaller %>%
ggplot(aes(carat)) +
geom_freqpoly(binwidth = 0.01)
```
````

R Markdown

~/Open_Science/Digital_Badge/RCR - master - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

lettuce_report.Rmd*

Knit | Insert | Run

```
---
title: "This is a reproducible document"
author: "Dr. Brendan Palmer"
date: "18th June 2019"
output:
  word_document:
    fig_height: 4
    fig_width: 6
---
# This is the beginning of the project

Our initial reports might be restricted to lab meetings etc. We can use `R Markdown` to show the code we are using, so that the meetings are not just a demonstration of the results, but also an examination of the `code` used to obtain them.

## Data overview
```{r packages and setup, include = FALSE}

knitr::opts_chunk$set(echo = FALSE, message = FALSE, warning = FALSE)

# Load your packages here
library(tidyverse)
library(knitr)
```

The plot below is call from the ggplot object entitled `report_plot` created in the script `03_final_analysis.R`.

```{r Plots from script, echo = FALSE}

source("scripts/03_final_analysis.R")

# The location of the Rmd file dictates whether the path to other files is intact
```

This is a reproducible document

Dr. Brendan Palmer

18th June 2019

This is the beginning of the project

Our initial reports might be restricted to lab meetings etc. We can use R Markdown to show the code we are using, so that the meetings are not just a demonstration of the results, but also an examination of the code used to obtain them.

Data overview

The plot below is call from the ggplot object entitled report_plot created in the script 03_final_analysis.R.

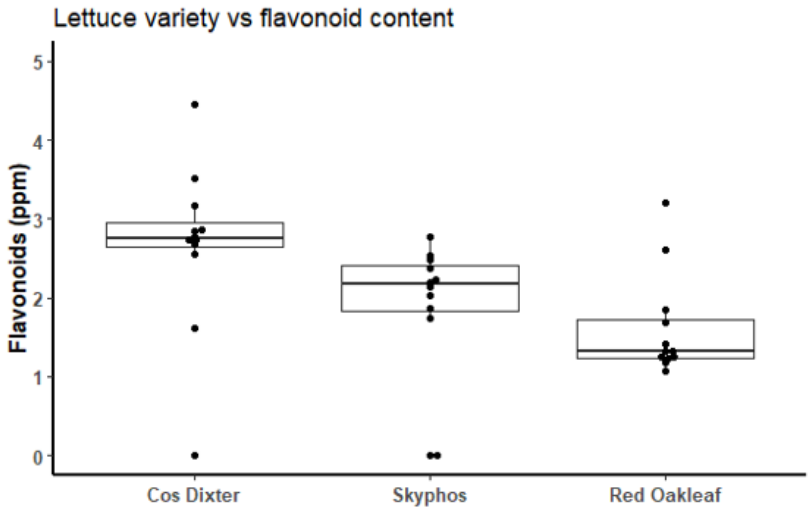
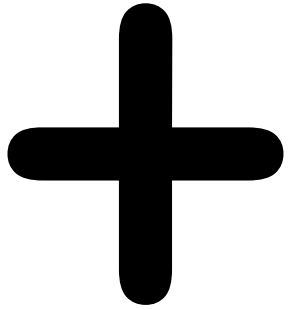


Fig. 1. Flavonoid content of three lettuce varieties under three experimental conditions.

Or we can also recreate the code within the R Markdown document as seen below.
```

The moral of the story…….

You can go from this

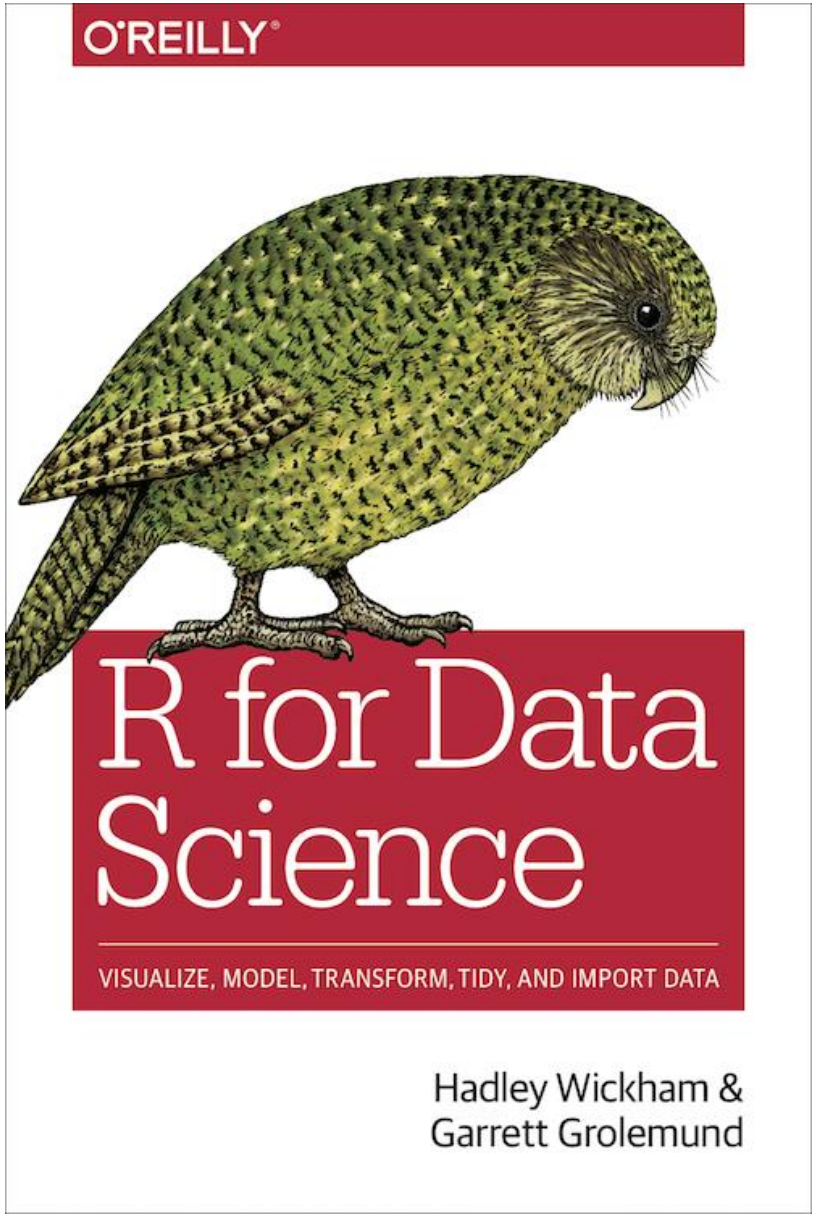


To this!!

Completely ordinary

Master Builder!

You could write a book on that!!



R for Data Science webpage

R projects

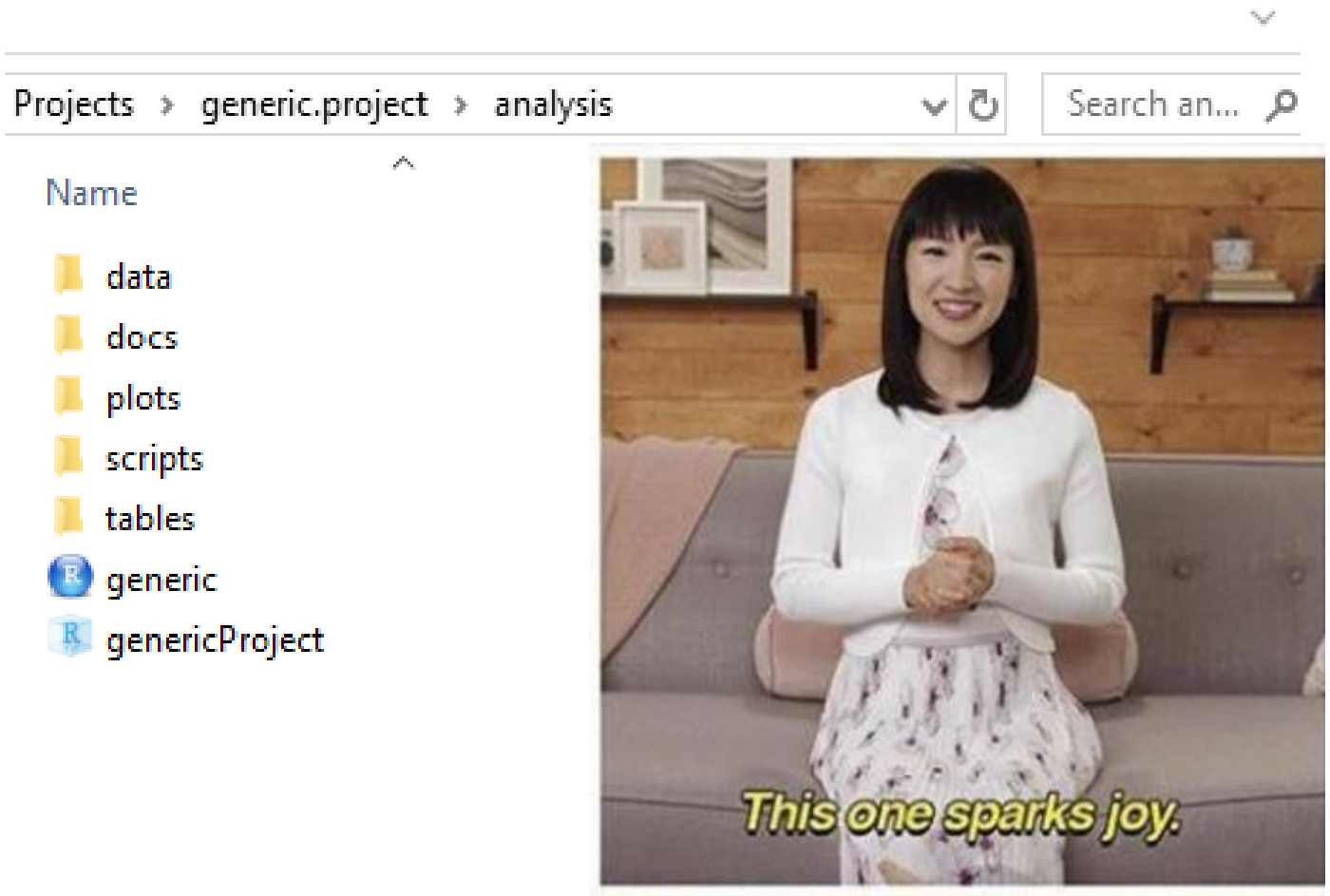
Still haven't found what I'm looking for

- Help your future-self

B_Palmer_Medicine_Files › 4a Project › Pyrosequencing_analysis › Pyrosequencing_Paper › Draft_Paper_incl_Figs › Submission › JVI_Resubmission › JVI_resubmission_files › Final Final version

Name	Date modified
Cover_letter_B_A_Palmer_Sept_2014	10/09/2014 17:05
Fig_1_Sept_14	11/09/2014 10:31
Fig_1_Sept_14	10/09/2014 23:07
Fig_2_Sept_14	11/09/2014 10:31
Fig_2_Sept_14	10/09/2014 23:07
Fig_3_Sept_14	11/09/2014 10:31
Fig_3_Sept_14	10/09/2014 23:07
Fig_4_Sept_14	11/09/2014 10:31
Fig_4_Sept_14	10/09/2014 23:07
Fig_5_Sept_14	11/09/2014 10:33
Fig_5_Sept_14	10/09/2014 23:07
HCV_UDPS_B_A_Palmer_Sept_14	17/09/2014 12:21
Response_to_Reviewer_Sept_14	10/09/2014 22:42
Supplementry_Figure_B_A_Palmer_Sept_14	29/08/2014 13:21
Supplementry_Figure_B_A_Palmer_Sept_14	10/09/2014 22:31
Tables_B_A_Palmer_Sept_2014	10/09/2014 22:09

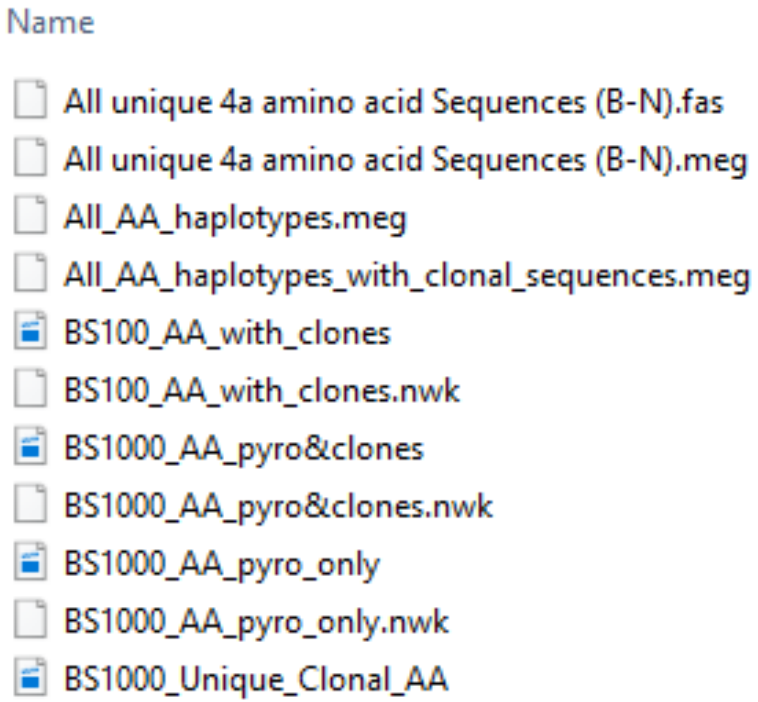
Define a generic project structure



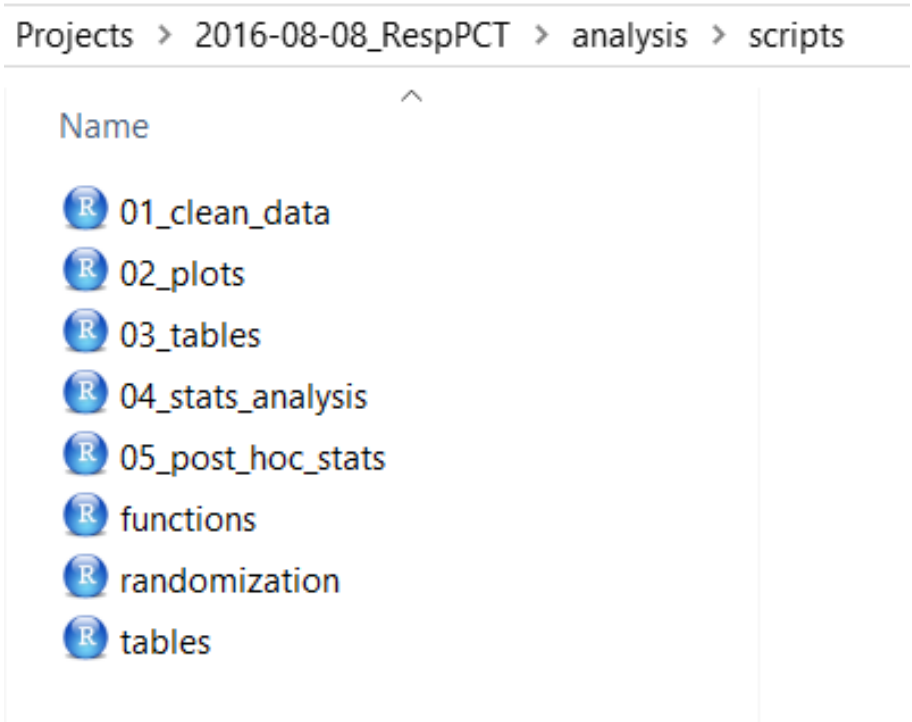
Give your files and folders informative names

- **Make your file names:**
  1. **Machine readable**
  2. **Human readable**
  3. **Work with default ordering**

NO



Yes



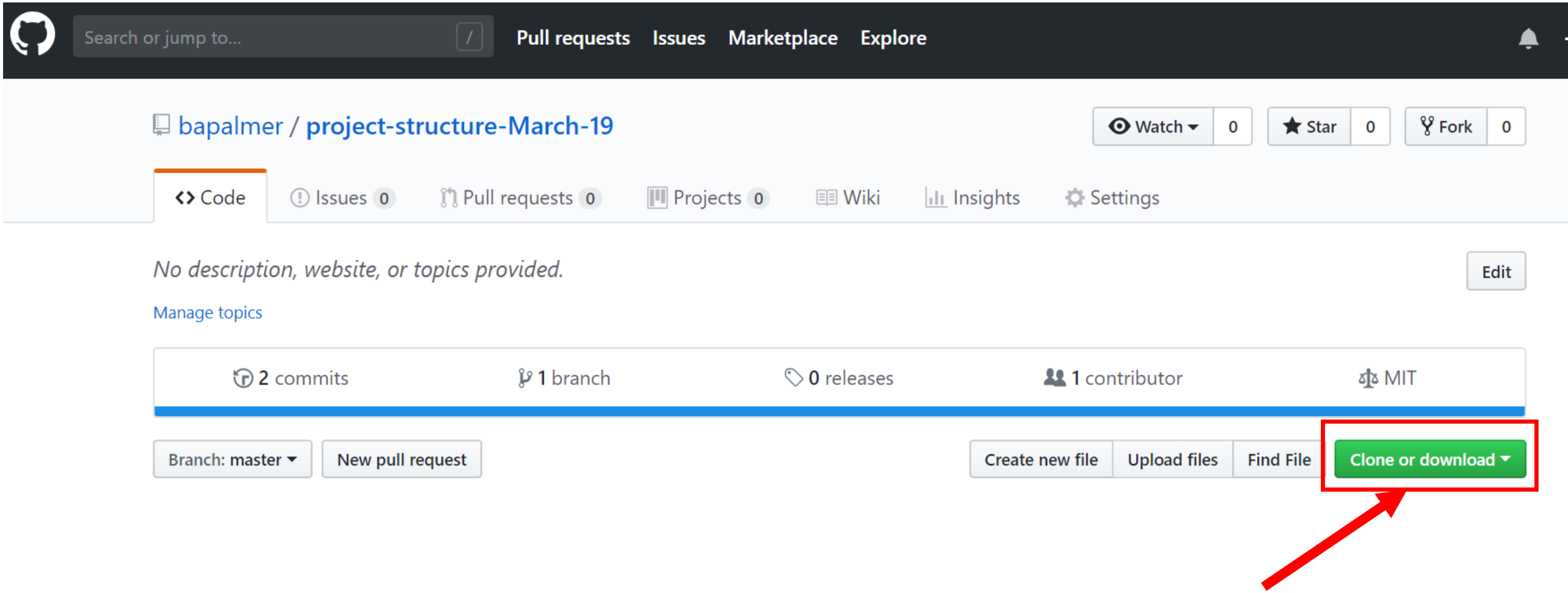
Scripted workflows

- The R scripts you generate should be human readable
    - Annotate the code
    - Break up the scripts into dedicated tasks
    - Interlink with other within project scripts

```
1  # Data ----
2  # Eight tibbles returned from the 01_data_import_and_tidying_master_file.R
3  # 1. fgf23_data => FGF23 readings from study centres 01-03
4  # 2. food_level_data => Food diary entries
5  # 3. grouped_data => Dialysis and nondialysis diary entries by component
6  # 4. k_data => Serum potassium
7  # 5. master_data_clean => all the clean master file data if required
8  # 6. p_data => Serum phosphate
9  # 7. pth_data => Parathyroid hormone readings
10 # 8. pulses_nuts_data
11
12 source("scripts/01_data_import_and_tidying_master_file.R")
```

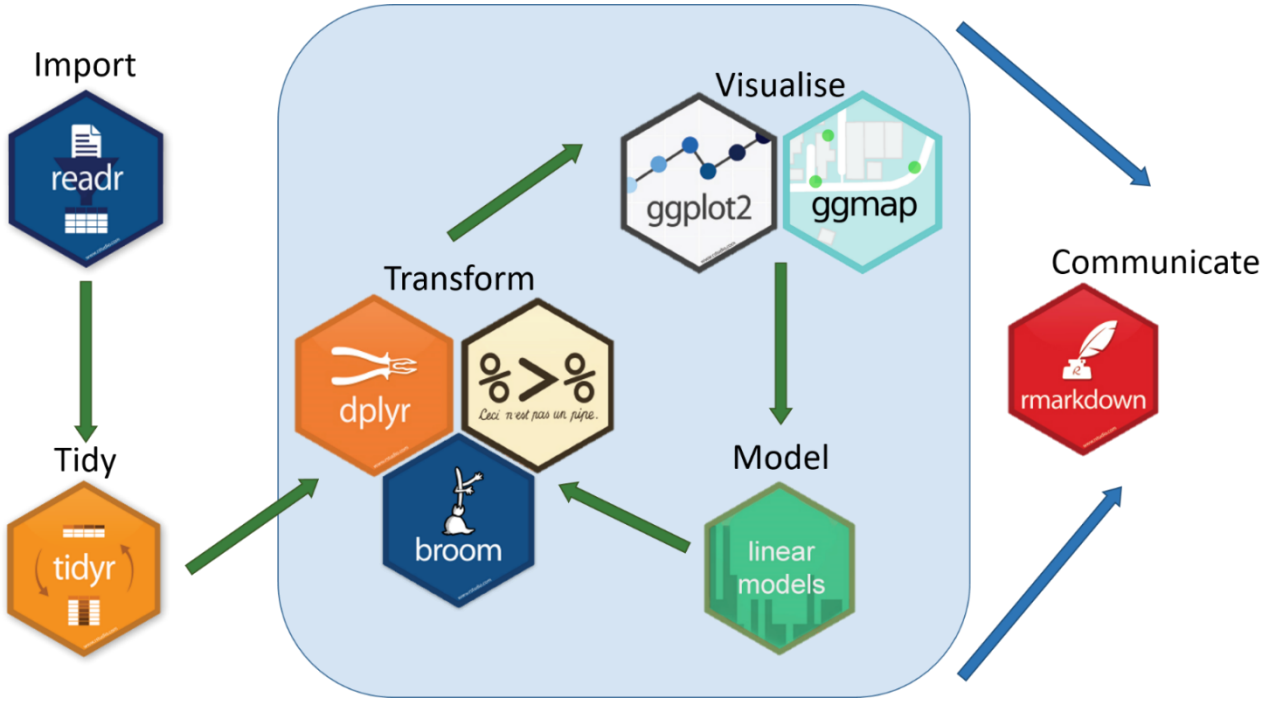
R projects

- Here's one I made earlier……

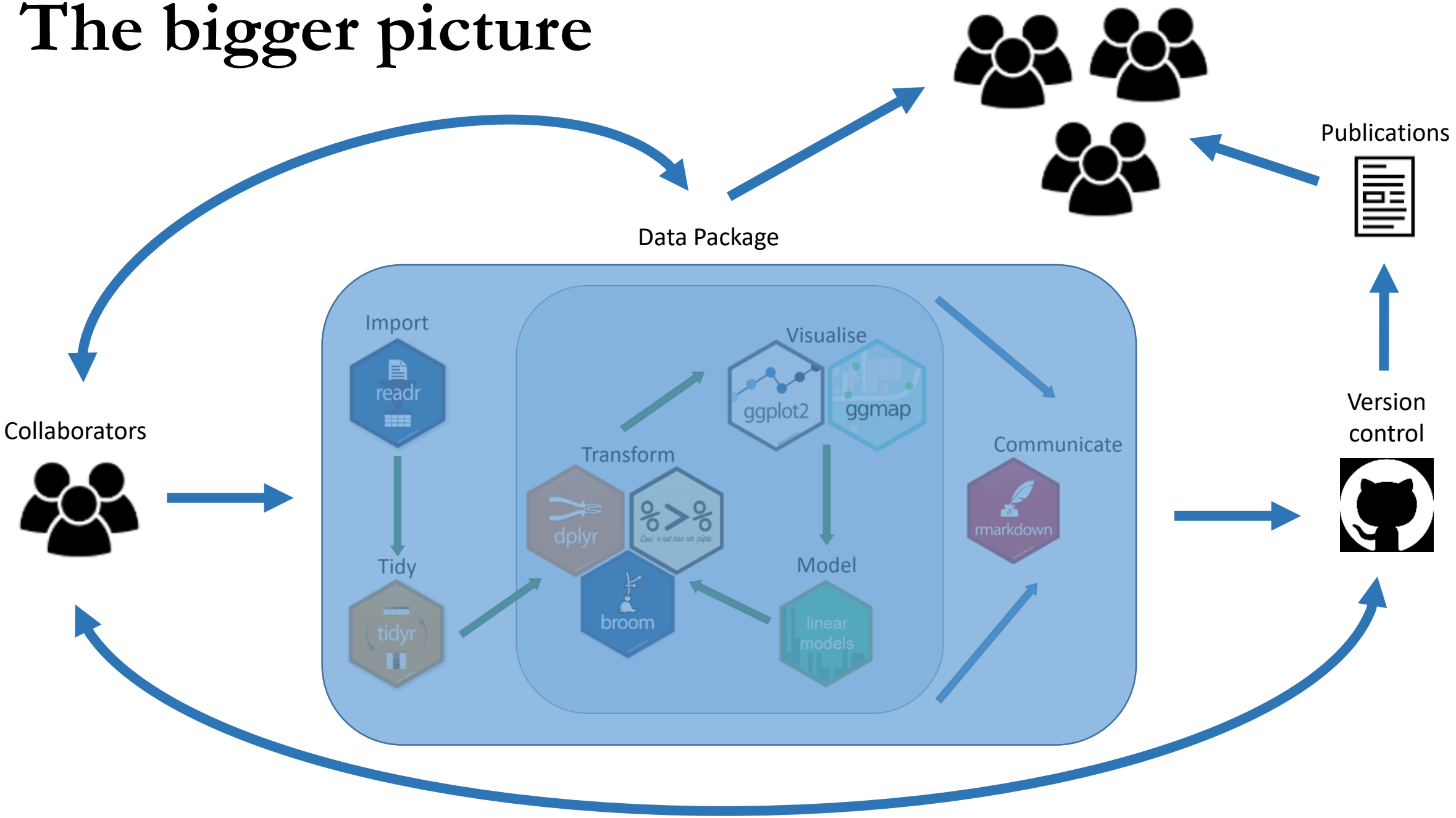


Packaging

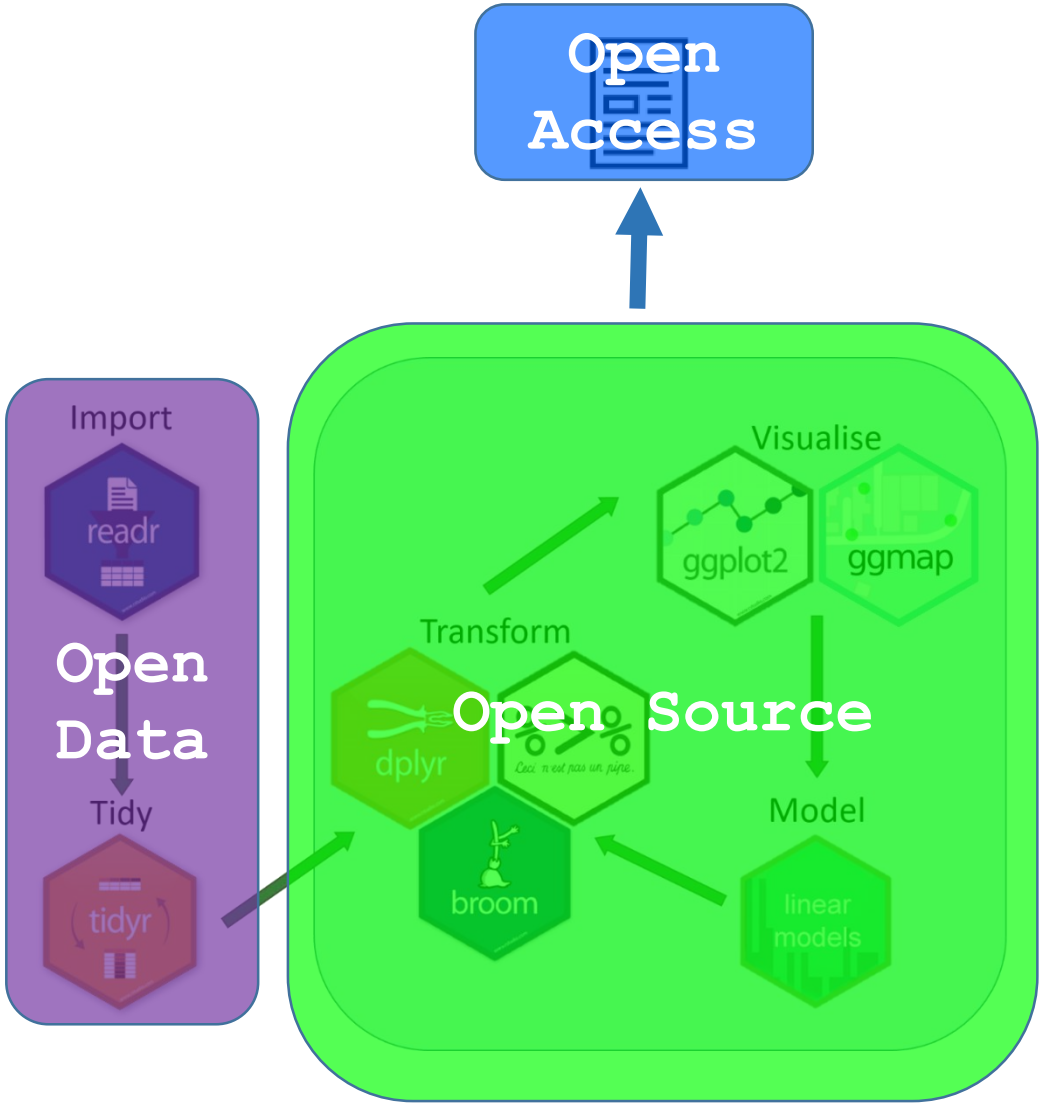
Putting the pieces together using R



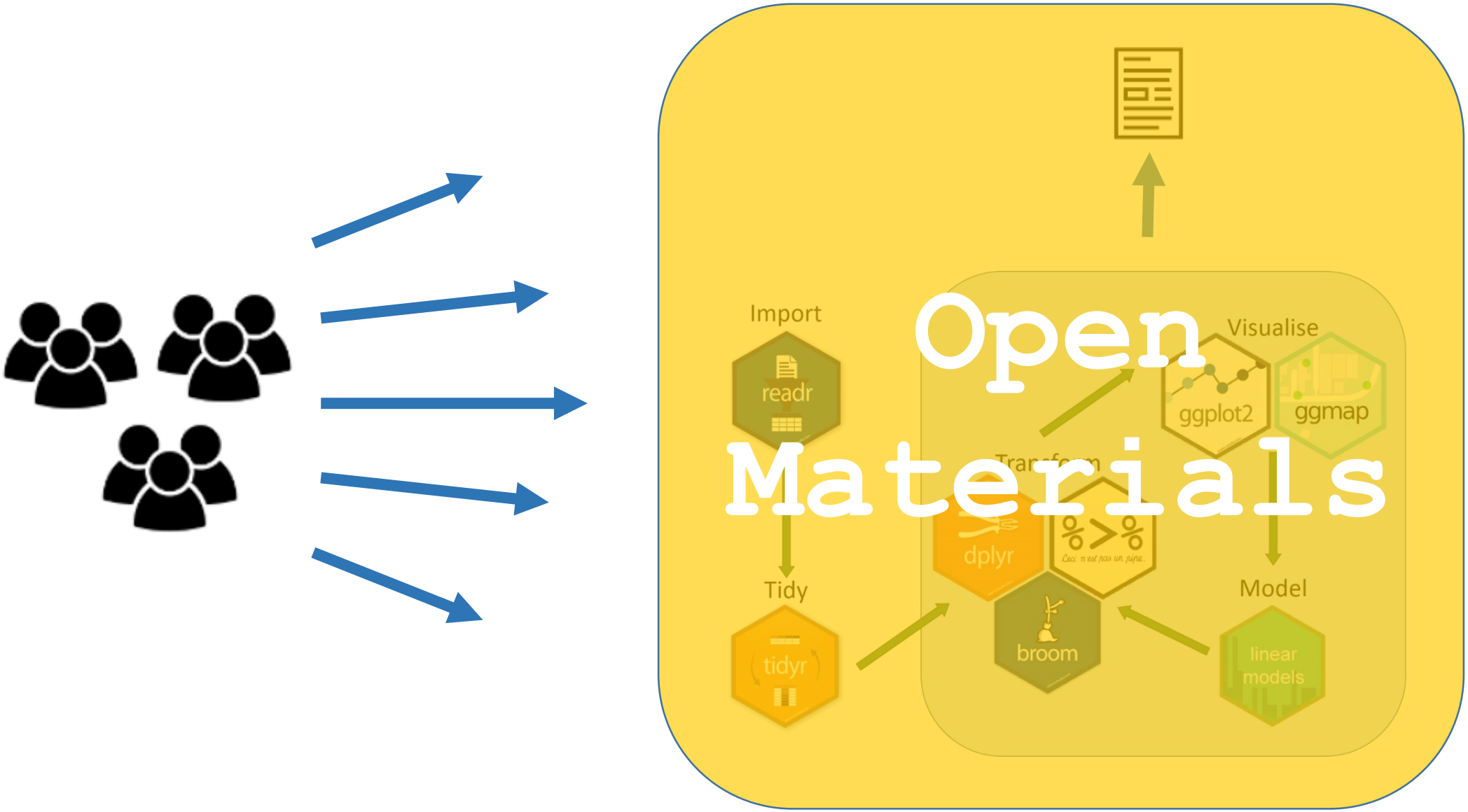
The bigger picture



The ‘Open Science’ picture



The ‘Open Science’ picture



Project Packaging

Turn a Git repo into a collection of interactive notebooks

Have a repository full of Jupyter notebooks? With Binder, open those notebooks in an executable environment, making your code immediately reproducible by anyone, anywhere.

Build and launch a repository

GitHub repository name or URL

GitHub repository name or URL GitHub

Git branch, tag, or commit

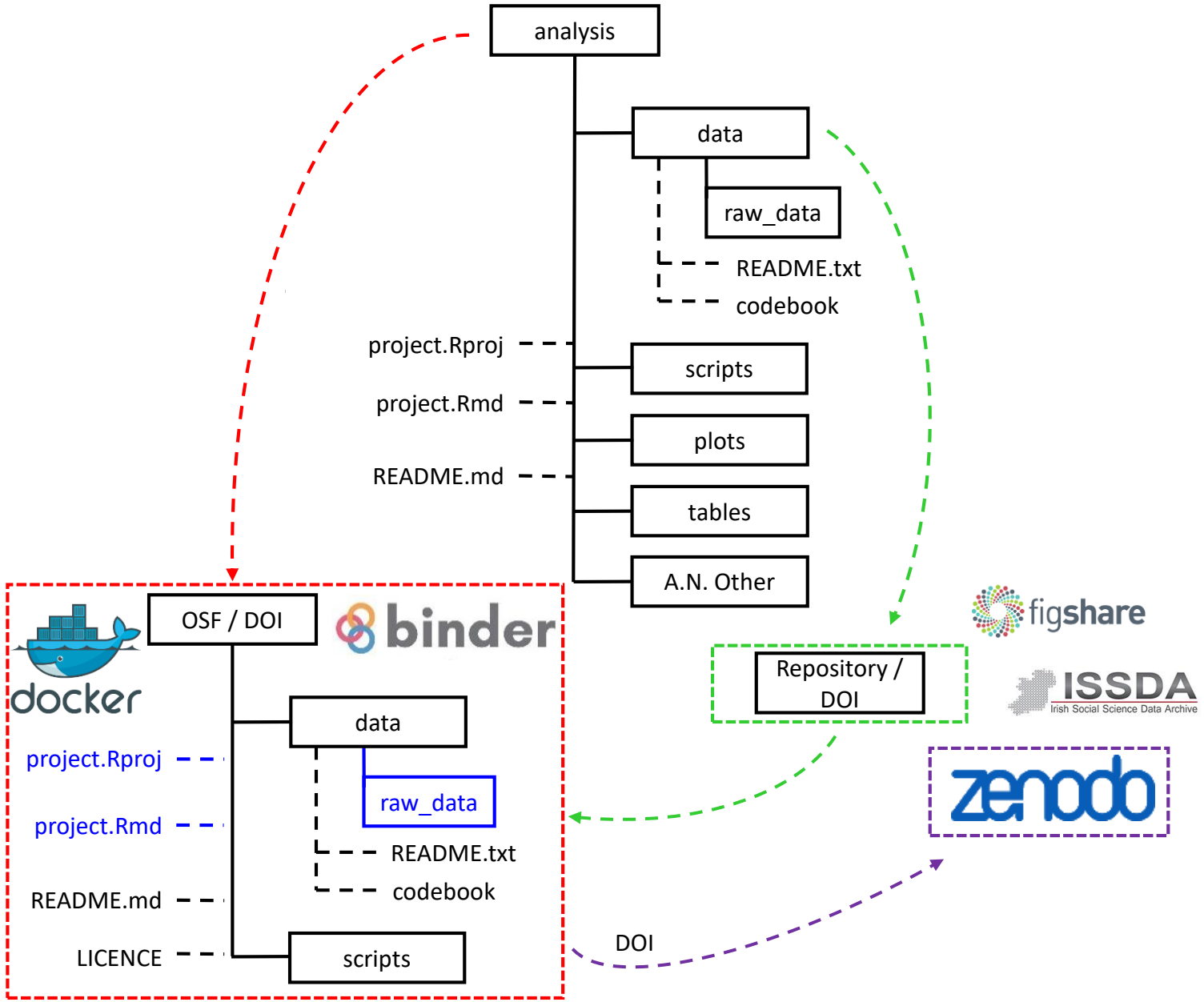
Git branch, tag, or commit

Path to a notebook file (optional)

Path to a notebook file (optional) File

launch

What does this allow us to do?



The butterfly has started flapping its wings

Plan S
Making full and immediate Open Access a reality

Why Plan S | 10 Principles | Funders & support | Implementation | About | Contact

"After 1 January 2020 scientific publications on the results from research funded by public grants provided by national and European research councils and funding bodies, must be published in compliant Open Access Journals or on compliant Open Access Platforms."

HRB Health Research Board

Funding | Data collections & evidence | Publications | Success stories | News | About

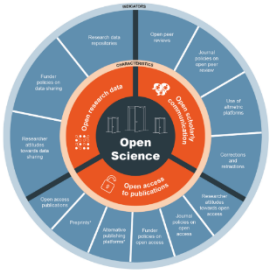
Home > Funding > Policies and principles > Open Research

- Funding schemes
- EU funding support
- Manage a grant
- Funding awarded
- Evaluation
- GDPR guidance for researchers
- Policies and principles
  - EU legislation
  - Gender
  - Good research practice
  - Open Research

Open Research

The HRB is committed to ensuring that its funded research is open, accessible and usable, so it can have the greatest possible impact.

There is a fundamental shift across Europe towards making research more transparent, collaborative, accessible and efficient. This Open Science movement is a strategic priority for the European Commission in research and innovation policy and an EU high-level Expert Group, the Open Science Policy Platform (OSPP 2016-2018) has been established to consider key implementation areas.



EUROPEAN COMMISSION
Directorate-General for Research & Innovation

H2020 Programme

Guidelines on

FAIR Data Management in Horizon 2020

Science Foundation Ireland sfi For what's next

Funding | Engagement | Events | Research News | SFI Research Centres

Science Foundation Ireland joins DORA

14th February 2019, Dublin – Science Foundation Ireland has become a signatory to the San Francisco Declaration of Research Assessment (DORA), making a formal commitment to assessing the quality and impact of research through means other than journal impact factors.

Putting the final pieces into place

Make Your Code Citable Using GitHub and Zenodo: A How-to Guide

By Open Science MOOC on July 24, 2018

Try it out yourself

Search or jump to... / Pull requests Issues Marketplace Explore

Set status

Brendan Palmer

bapalmer

Edit profile

Overview Repositories 14 Projects 0 Stars 1 Followers 12 Following 10

Pinned Order updated. Customize your pins

RSS_Belfast_2019

Data FAIRification using R/RStudio workflows

R

R-A_Hitchhikers_Guide_to_Reproducible_Research

A 3-day R course given in University College Cork that encompasses various elements off reproducible research facilitated through RStudio projects, the R tidyverse language and reporting using R Ma...

HTML ★ 2

RCR

Section of the UCC Reproducible Conduct of Research digital badge dedicated to exposing researchers to reproducible research practices.

HTML

lunchtime_sessions

Short 1 hour introductions to R-related topics such as creating R projects, using GitHub through RStudio and more

HTML ★ 1

meetup

Cork (Ireland) R-Users Group