

**TEACHING
REPRODUCIBLE DATA
ANALYSIS IN R**

SCHOOL OF PSYCHOLOGY TEACHING TEAM

It's not just about changing what you teach...
it's about building a community

It's not just about R, it's about:

- building confidence and independence
- enabling more efficient data analysis workflows
- instilling values of reproducibility and transparency

WHY, WHAT, AND HOW

- our backstory
- what flavor of R to teach
- how to teach it

WHY?

AutoSave Off MovieClip_1 - Excel

File Home Insert Page Layout Formulas Data Review View Tableau Help FOXIT PDF Tell me what you want to do Share

Clipboard Font Alignment Number Styles Cells Editing

Clipboard: Cut, Copy, Paste, Format Painter
 Font: Calibri, 11, Bold, Italic, Underline, Text Color, Background Color
 Alignment: Wrap Text, Merge & Center
 Number: General, Percentage, Decimals
 Styles: Conditional Formatting, Format as Table, Cell Styles
 Cells: Insert, Delete, Format
 Editing: AutoSum, Fill, Clear, Sort & Filter, Find & Select

Formula Bar: A1, DUAL

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	DUAL	DUAL	DUAL	DUAL	FIX	FIX	FIX	FIX	FREE	FREE	DUAL	DUAL	FIX	FIX	DUAL	DUAL	FREE	FREE	FIX	FIX	FIX
2	AKW26	AKW26	AML05	AML05	AMS16	AMS16	ATM12	ATM12	AWA11	AWA11	BPS04	BPS04	CAN07	CAN07	CDN08	CDN08	CGE15	CGE15	DON17	DON17	DVE12
3	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X
4		34	96	-1	37	4	143	-117	213	-168	211	-5	40	45	-7	-98	48	-95	62	-24	24
5		32	-100	3	-48	-3	-47	-28	-31	-151	19	-28	49	16	-22	-57	-10	-84	-270	-28	-79
6		9	-70	27	-31	-11	10	-45	-14	-266	137	-2	-7	15	-15	-61	-169	-29	-146	-27	0
7		0	-57	31	-20	-27	-78	-67	-41	-93	174	-10	-38	19	-6	-36	-25	11	-82	-11	2
8		6	-43	3	-4	-2	8	-66	84	228	254	16	15	29	-20	-44	59	4	-86	-31	-90
9		5	-7	-1	-3	-1	-3	86	93	-92	173	8	-31	43	-22	205	19	-2	-101	-32	-88
10		21	-101	5	-5	-17	-71	-118	120	-146	154	12	7	42	-9	-61	-33	-5	-54	-16	-100
11		4	-11	13	-17	-5	9	-121	134	-118	91	1	15	41	-3	-95	41	21	-80	-15	-32
12		13	-18	-22	-24	-14	-10	-120	132	-95	109	-1	15	41	5	-117	-45	54	-112	-16	-29
13		22	-97	2	3	-19	-17	-118	136	-103	147	-12	-41	57	53	109	-52	14	-78	-17	-1
14		9	-25	3	1	-26	-71	-133	124	-244	133	2	4	62	38	72	-108	18	-44	-25	-8
15		4	-26	6	-3	-57	-92	-127	117	-148	83	-4	-28	59	29	-45	86	14	-78	-22	-47
16		4	-22	3	4	-12	-15	-146	8	-175	121	-5	15	50	27	-35	-60	4	-93	-6	24
17		-3	-26	-3	6	-81	5	-147	121	-260	109	23	-45	52	34	36	-49	-141	13	-10	9
18		9	-116	-8	8	43	11	-165	119	-182	130	-33	-46	50	29	37	-124	-55	36	-21	11
19		7	-22	-6	3	-31	-97	-413	125	70	136	2	23	50	29	-16	24	152	47	-15	12
20		-1	-158	-11	7	-9	-115	-168	135	125	141	-41	-53	47	22	-50	-67	64	-24	-18	15
21		7	-68	-9	21	-24	-28	-166	117	68	131	14	-27	57	16	-49	68	22	-118	-17	24
22		27	-239	-6	16	-31	-25	-174	124	26	108	-3	6	51	-13	-54	81	25	-117	-19	25
23		2	-63	-5	15	-31	-13	-169	137	251	113	-44	-51	119	38	17	6	25	-115	-30	23

Ready

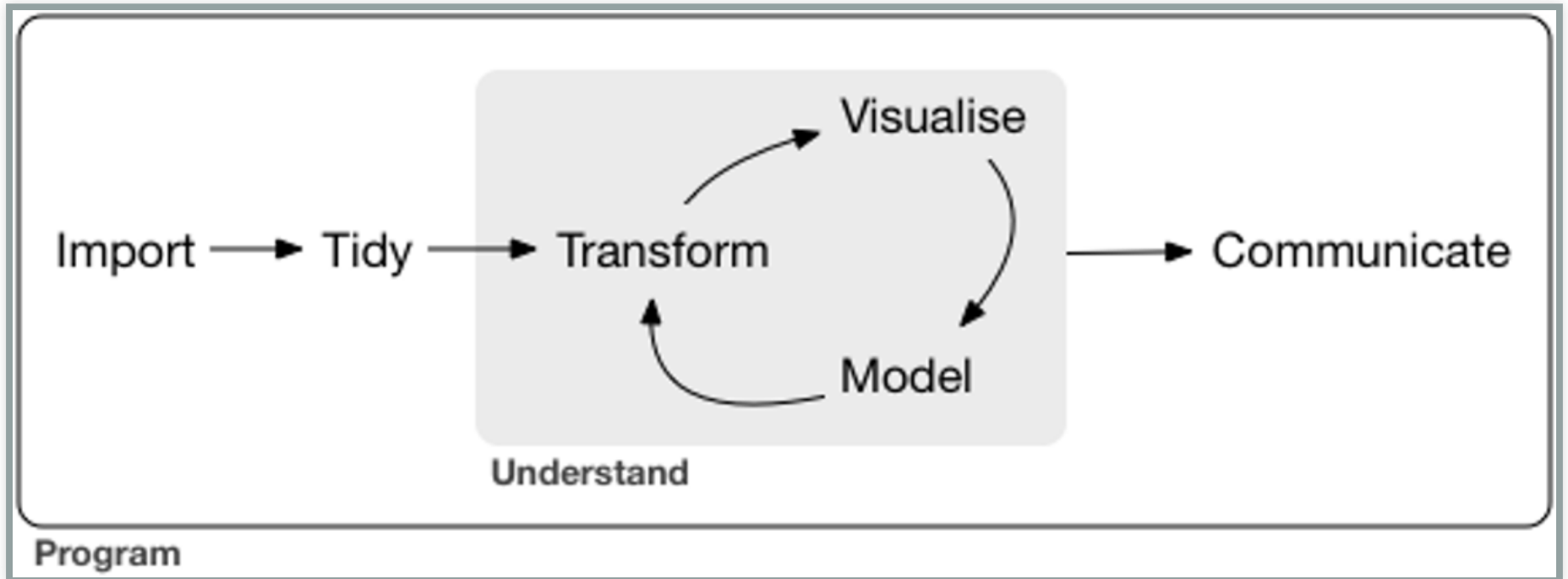
“I want to have the data file so there are only 4 columns
 - PARTICIPANT CODE, CONDITION, X, Y.”

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2	Id	Gender	Age	Participate	I often notice small	usually concentrat	I find it easy to do	If there is an interr	I find it easy to rea	I know how to tell if	When I'm reading a	I like to collect
3	16	Male	20	1	Slightly Disagree	Definitely Agree	Slightly Disagree	Definitely Disagree	Slightly Agree	Slightly Agree	Slightly Agree	Definitely Disa
4	17	Male	40	1	Definitely Agree	Slightly Agree	Slightly Agree	Definitely Agree	Definitely Agree	Definitely Agree	Slightly Agree	Slightly Disagr
5	18	Male	33	1	Definitely Agree	Definitely Agree	Slightly Agree	Definitely Agree	Definitely Agree	Definitely Agree	Slightly Agree	Definitely Agree
6	19	Male	18	1	Definitely Agree	Definitely Agree	Definitely Agree	Slightly Agree	Definitely Agree	Definitely Agree	Slightly Disagree	Slightly Disagr
7	20	Male	24	1	Definitely Disagree	Slightly Disagree	Definitely Agree	Slightly Agree	Slightly Agree	Slightly Agree	Slightly Disagree	Slightly Disagr
8	21	Female	42	1	Slightly Disagree	Slightly Disagree	Definitely Agree	Slightly Agree	Slightly Disagree	Slightly Disagree	Definitely Disagree	Definitely Disa
9	22	Female	19	1	Slightly Agree	Definitely Agree	Slightly Disagree	Slightly Disagree	Definitely Disagree	Slightly Agree	Definitely Disagree	Slightly Agree
10	28	Female	49	1	Slightly Disagree	Slightly Disagree	Slightly Agree	Slightly Agree	Slightly Agree	Slightly Agree	Slightly Disagree	Slightly Disagr
11	29	Female	18	1	Slightly Agree	Slightly Disagree	Definitely Agree	Definitely Agree	Definitely Agree	Definitely Agree	Slightly Disagree	Slightly Agree
12	31	Male	18	1	Slightly Agree	Slightly Disagree	Slightly Agree	Slightly Agree	Slightly Agree	Slightly Agree	Slightly Disagree	Definitely Disa
13	34	Female	32	1	Definitely Agree	Slightly Agree	Definitely Agree	Definitely Agree	Slightly Agree	Slightly Agree	Slightly Disagree	Definitely Disa
14	35	Female	18	1	Slightly Agree	Slightly Disagree	Definitely Agree	Definitely Agree	Definitely Agree	Definitely Agree	Slightly Agree	Definitely Agree
15	39	Female	20	1	Slightly Agree	Slightly Agree	Definitely Agree	Slightly Disagree	Definitely Agree	Definitely Agree	Slightly Disagree	Definitely Disa
16	40	Male	20	1	Slightly Agree	Definitely Agree	Slightly Disagree	Definitely Agree	Definitely Agree	Slightly Agree	Slightly Agree	Definitely Agree

Id	AQ	Gender	Age	Participate
(int)	(dbl)	(chr)	(int)	(int)
52	9	Male	25	0
55	8	Female	23	1
86	8	Male	77	1
99	8	Male	25	0
46	7	Female	21	1
74	7	Female	48	1
22	6	Female	19	1
40	5	Male	39	1
45	5	Female	58	1
51	5	Female	20	1

http://talklab.psy.gla.ac.uk/r_training/scoring_the_AQ/

DATA COMPREHENSION



Grolemund & Wickham, [R for Data Science](#)

WHAT?

WHAT FLAVOR OF R?

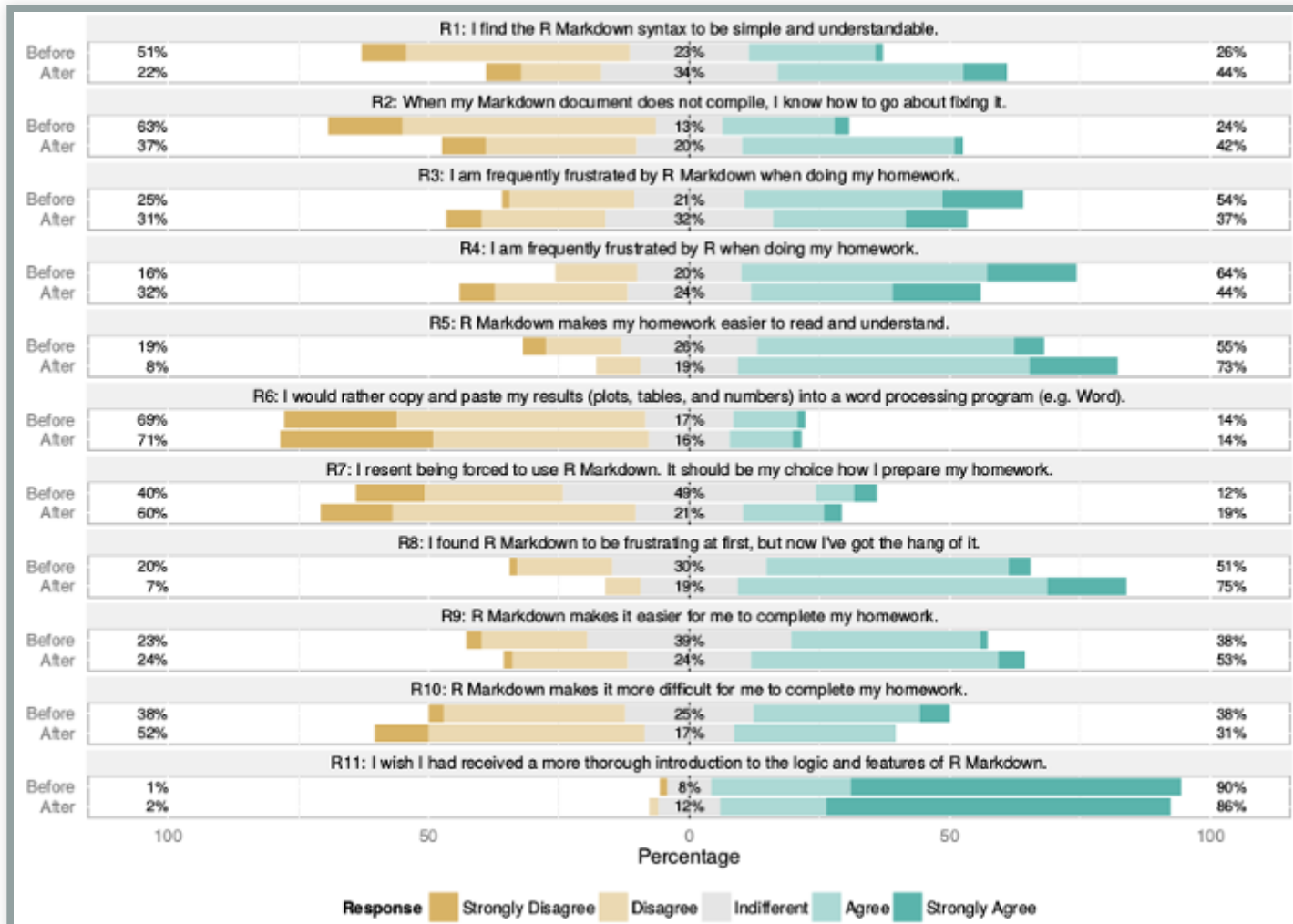


REPRODUCIBLE REPORTS IN RMARKDOWN

The image displays a side-by-side comparison of R Markdown source code and its rendered output. On the left, the RStudio editor shows the source file 'Untitled.Rmd' with the following code:

```
1- ---
2- title: "Untitled"
3- author: "Garrett"
4- date: "July 10, 2014"
5- output: html_document
6- runtime: shiny
7- ---
8
9- This R Markdown document is made interactive using Shiny. Unlike the more
10- traditional workflow of creating static reports, you can now create
11- documents that allow your readers to change the assumptions underlying
12- your analysis and see the results immediately.
13
14- To learn more, see \[Interactive Documents\](http://rmarkdown.rstudio.com/authoring_shiny.html).
15
16- ## Inputs and Outputs
17
18- You can embed Shiny inputs and outputs in your document. Outputs are
19- automatically updated whenever inputs change. This demonstrates how a
20- standard R plot can be made interactive by wrapping it in the Shiny
21- 'renderPlot' function. The 'selectInput' and 'sliderInput' functions
22- create the input widgets used to drive the plot.
23
24- ```{r, echo=FALSE}
25- inputPanel(
26-   selectInput("n_breaks", label = "Number of bins:",
27-             choices = c(10, 20, 35, 50), selected = 20),
28-   sliderInput("bw_adjust", label = "Bandwidth adjustment:",
29-             min = 0.2, max = 2, value = 1, step = 0.2)
30- )
31- renderPlot({
```

On the right, the rendered report is shown in a browser window titled 'Untitled'. It features a header with the author's name 'Garrett' and the date 'July 10, 2014'. The main text explains the interactive nature of the document and provides a link to 'Interactive Documents'. Below this, a section titled 'Inputs and Outputs' describes the Shiny components. The report includes two interactive input widgets: a dropdown menu for 'Number of bins' (set to 20) and a slider for 'Bandwidth adjustment' (set to 1). At the bottom, a histogram titled 'Geyser eruption duration' is displayed, showing the density distribution of eruption durations with a blue kernel density estimate curve overlaid.



Baumer, Cetinkaya-Rundel, Bray, Loi, & Horton (2014)

TIDYVERSE FIRST!

- **Tidy Data**
 - Rows = observations
 - Columns = variables
 - Table = observation unit
- **Tidy Tools**
 - tidy input -> tidy output
- **Visualization with ggplot2**

<http://varianceexplained.org/r/teach-tidyverse/>

THE “WICKHAM SIX” DPLYR VERBS

Six verbs cover 90% of data tidying - Hadley Wickham

<code>select()</code>	choose columns
<code>filter()</code>	choose rows
<code>mutate()</code>	create new columns
<code>arrange()</code>	sort the rows
<code>group_by()</code>	establish groups
<code>summarise()</code>	summarise data/groups

- also: combining data sources, restructuring data

THE PROGRAMME

Y1	R/RStudio/RMarkdown, data import, tidying, viz, probability & descriptive stats
Y2	sampling distributions, GLM, correlation, regression, t-test, data simulation
Y3	mixed-model ANOVA, multilevel regression
Y4	advanced topics (factor analysis, psychometrics, etc)

HOW?

MAKING THE TRANSITION

- *Ongoing*: stats journal club
- R/RStudio training sessions
- Phase in gradually
 - translate descriptive/inferential stats materials into R
 - introduce data wrangling labs at intro level
- Support staff and students using slack.com messaging
 - separate workspaces for staff and students

HOW DO WE GET STAFF TO BUY IN TO USING R?

SOLUTION: EMBED R INTO TEACHING

- generating academic web pages on github
- using R in marking
- tracking student engagement with Moodle logs
- make exams with the exams package
- generate self-guided web exercises with RMarkdown and webex
- semi-automated assessment/feedback on RMarkdown-based assignments with assessr

WHY? *reproducibility, efficiency, competence, confidence*

WHAT? *data wrangling, R + RStudio + RMarkdown +
tidyverse*

HOW? *gradually, embed R in everyday practice*

WHEN? *ASAP!*

WHERE? *EVERYWHERE!*

TODAY'S SCHEDULE

09:00 - 09:30	Coffee and chat (58 Hillhead Street)
09:30 - 10:00	Introduction and philosophy
10:00 - 10:30	Our approach
10:30 - 10:45	Coffee and chat
10:45 - 11:15	Practicalities
11:15 - 11:45	Staff skill development
11:45 - 12:15	Student engagement
12:15 - 13:00	Lunch (move to Boyd Orr)
13:00 - 14:00	Lab demo
14:00 - 14:30	Coffee and questions
14:30 - 15:00	Assessment with assessr
15:00 - 15:30	Web exercises with webex
15:30 - 16:30	Wine and discussion