

# Digital badge

- Reproducible research

---

Brendan Palmer,

Clinical Research Facility - Cork &  
School of Public Health



## 1,500 scientists lift the lid on reproducibility

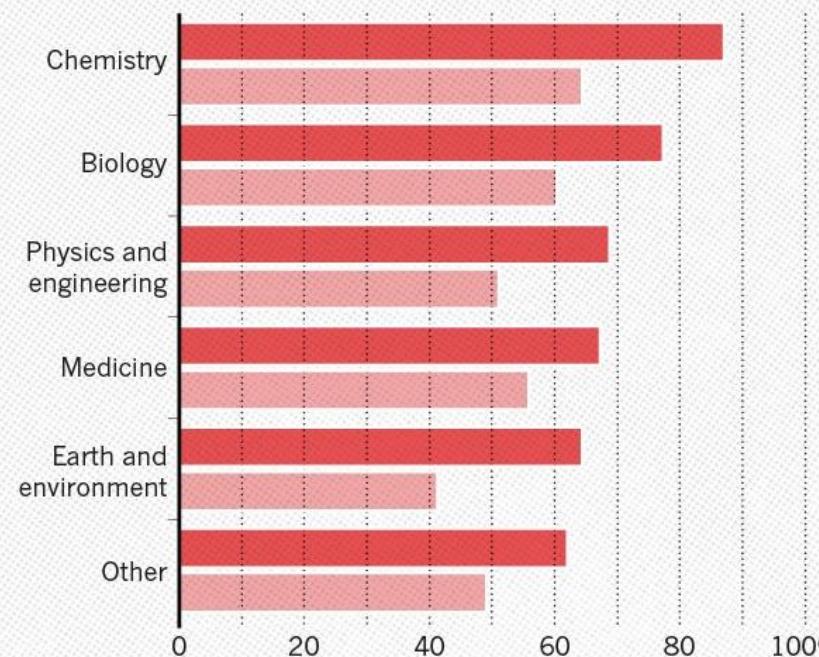
Survey sheds light on the 'crisis' rocking research.

Monya Baker

### HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.

● Someone else's    ● My own



# Reproducible or replicable

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

# Reproducible or replicable

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

## COMMENTARY

# Scientists behaving badly

To protect the integrity of science, we must look beyond falsification, fabrication and plagiarism, to a wider range of questionable research practices, argue **Brian C. Martinson, Melissa S. Anderson and Raymond de Vries**.

**Table 1 | Percentage of scientists who say that they engaged in the behaviour listed within the previous three years (*n* = 3,247)**

Top ten behaviours	All	Mid-career	Early-career
1. Falsifying or 'cooking' research data	0.3	0.2	0.5
2. Ignoring major aspects of human-subject requirements	0.3	0.3	0.4
3. Not properly disclosing involvement in firms whose products are based on one's own research	0.3	0.4	0.3
4. Relationships with students, research subjects or clients that may be interpreted as questionable	1.4	1.3	1.4
5. Using another's ideas without obtaining permission or giving due credit	1.4	1.7	1.0
6. Unauthorized use of confidential information in connection with one's own research	1.7	2.4	0.8 ***
7. Failing to present data that contradict one's own previous research	6.0	6.5	5.3
8. Circumventing certain minor aspects of human-subject requirements	7.6	9.0	6.0 **
9. Overlooking others' use of flawed data or questionable interpretation of data	12.5	12.2	12.8
10. Changing the design, methodology or results of a study in response to pressure from a funding source	15.5	20.6	9.5 ***
<b>Other behaviours</b>			
11. Publishing the same data or results in two or more publications	4.7	5.9	3.4 **
12. Inappropriately assigning authorship credit	10.0	12.3	7.4 ***
13. Withholding details of methodology or results in papers or proposals	10.8	12.4	8.9 **
14. Using inadequate or inappropriate research designs	13.5	14.6	12.2
15. Dropping observations or data points from analyses based on a gut feeling that they were inaccurate	15.3	14.3	16.5
16. Inadequate record keeping related to research projects	27.5	27.7	27.3

# Some cautionary tales

The Atlantic

Popular

Latest

Sections ▾

SCIENCE

## A Waste of 1,000 Research Papers

Decades of early research on the genetics of depression were built on nonexistent foundations. How did that happen?

ED YONG MAY 17, 2019

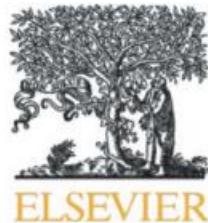


[Am J Psychiatry. 2019 May 1;176\(5\):376-387. doi: 10.1176/appi.ajp.2018.18070881. Epub 2019 Mar 8.](#)

**No Support for Historical Candidate Gene or Candidate Gene-by-Interaction Hypotheses for Major Depression Across Multiple Large Samples.**

[Border R<sup>1</sup>, Johnson EC<sup>1</sup>, Evans LM<sup>1</sup>, Smolen A<sup>1</sup>, Berley N<sup>1</sup>, Sullivan PF<sup>1</sup>, Keller MC<sup>1</sup>.](#)

# 241 shades of grey



Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: [www.elsevier.com/locate/ynim](http://www.elsevier.com/locate/ynim)



## Full Length Articles

### The secret lives of experiments: Methods reporting in the fMRI literature

Joshua Carp

*University of Michigan, Department of Psychology, 530 Church Street, Ann Arbor, MI, 48109, USA*

---

#### ARTICLE INFO

*Article history:*  
Accepted 3 July 2012  
Available online 10 July 2012

*Keywords:*  
fMRI  
Methods reporting  
Reproducibility  
Experimental design  
Analysis methods  
Statistical power

---

#### ABSTRACT

Replication of research findings is critical to the progress of scientific understanding. Accordingly, most scientific journals require authors to report experimental procedures in sufficient detail for independent researchers to replicate their work. To what extent do research reports in the functional neuroimaging literature live up to this standard? The present study evaluated methods reporting and methodological choices across 241 recent fMRI articles. Many studies did not report critical methodological details with regard to experimental design, data acquisition, and analysis. Further, many studies were underpowered to detect any but the largest statistical effects. Finally, data collection and analysis methods were highly flexible across studies, with nearly as many unique analysis pipelines as there were studies in the sample. Because the rate of false positive results is thought to increase with the flexibility of experimental designs, the field of functional neuroimaging may be particularly vulnerable to false positives. In sum, the present study documented significant gaps in methods reporting among fMRI studies. Improved methodological descriptions in research reports would yield significant benefits for the field.

# Who benefits most from reproducibility?



**Casey Greene**  
@GreeneScientist

Follow

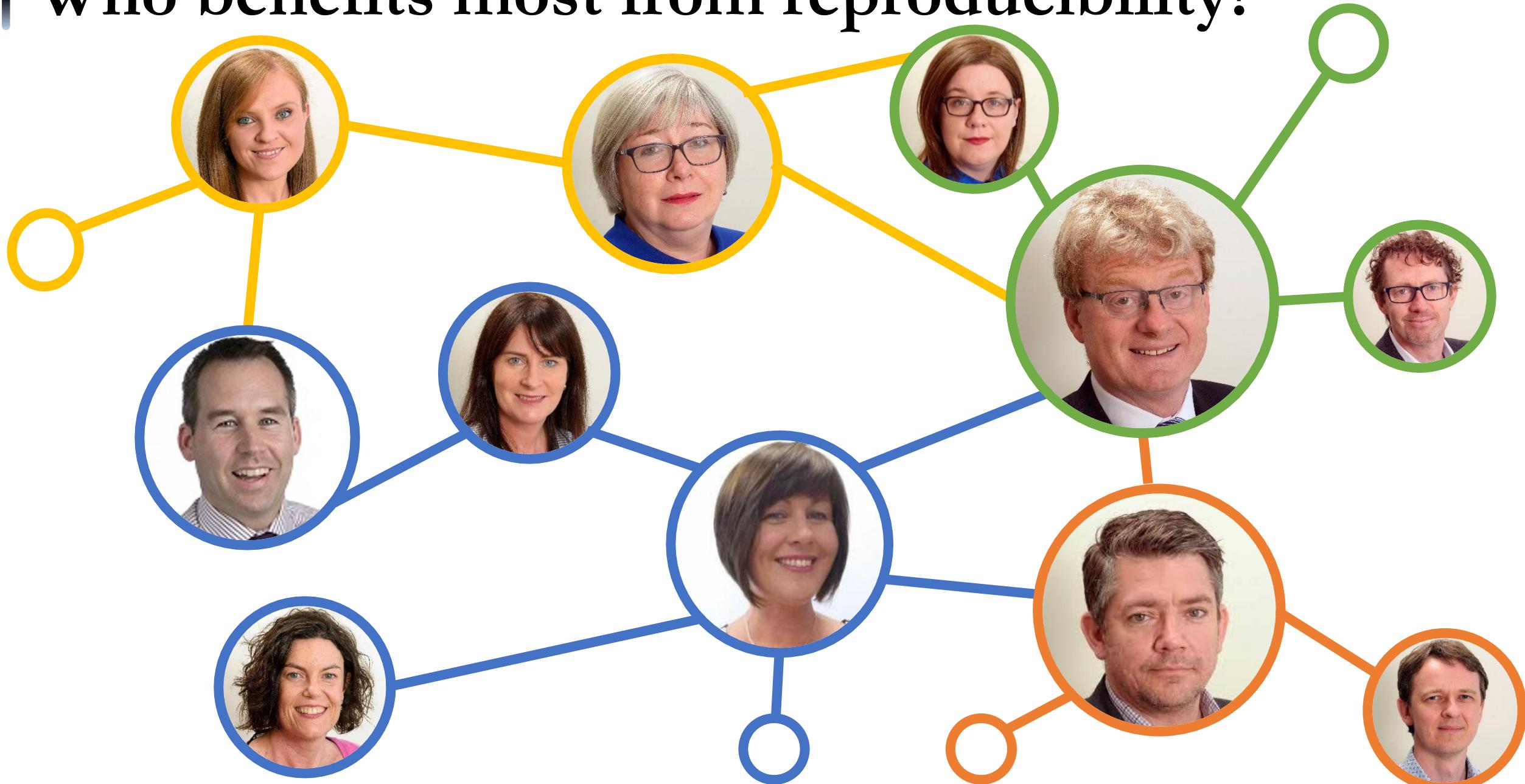


Reproducibility is important because the you  
of 3 months ago is terrible at answering  
email! - [@tracykteal](#) at [#2016dssummit](#)

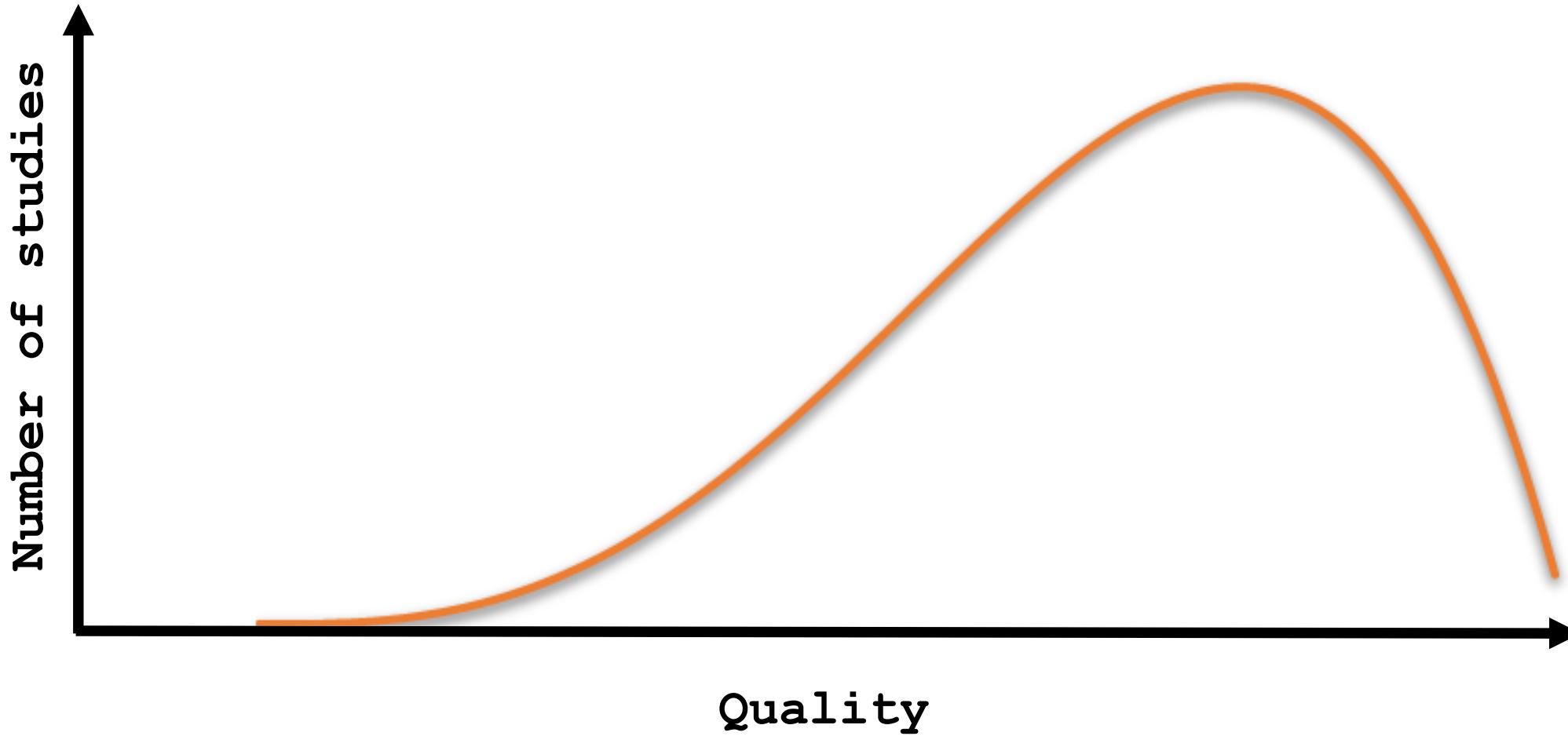
1:17 PM - 26 Oct 2016 from [Manhattan, NY](#)



# Who benefits most from reproducibility?



# Today



# Past failings



"In short, peer review misses all the hard stuff, and a worrying amount of the easy stuff"

James Heathers,  
Northwestern University

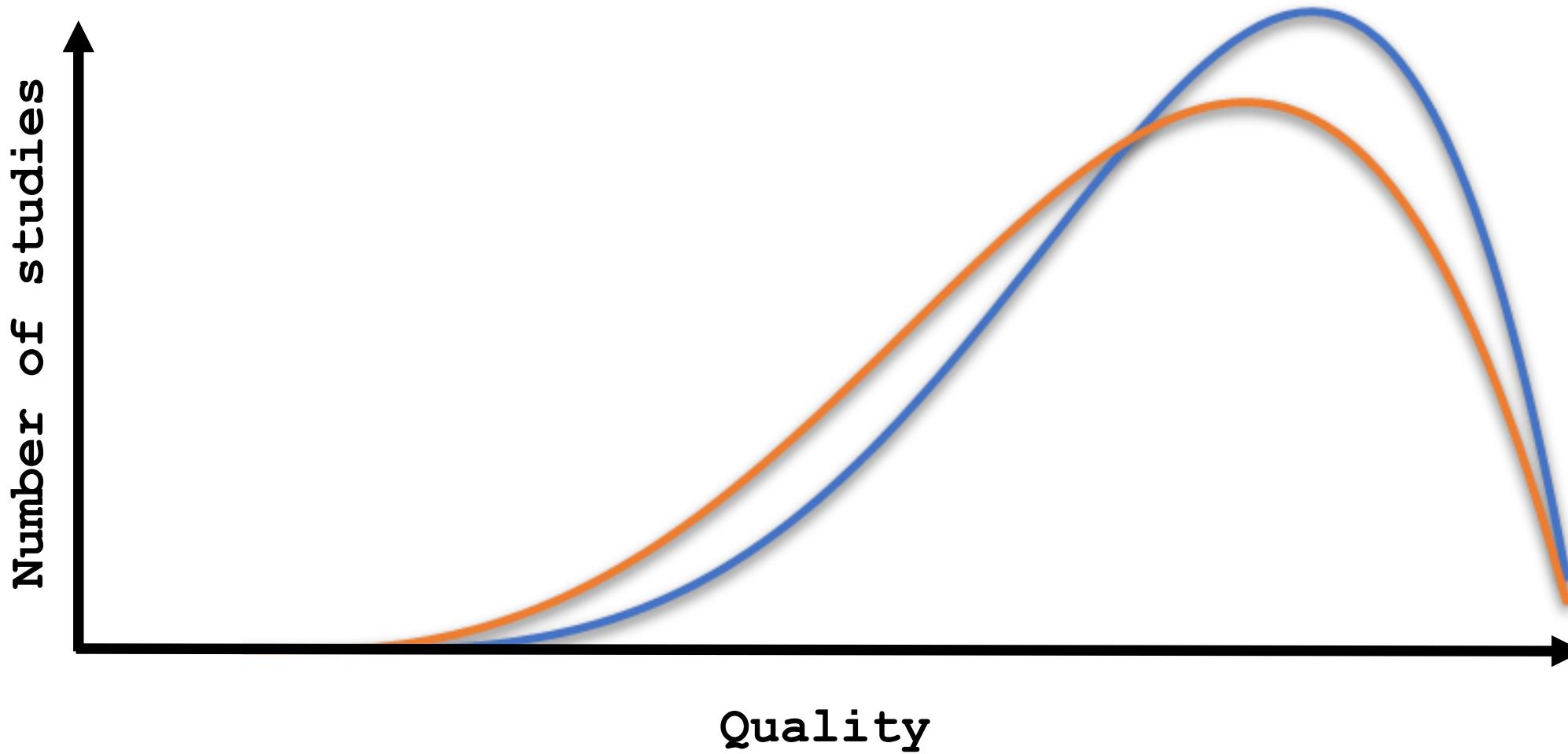
#datathugs



## Brian Wansink: The grad student who never said no

"Every day we would scratch our heads, ask "Why," and come up with another way to reanalyze the data with yet another set of plausible hypotheses. Eventually we started discovering solutions"

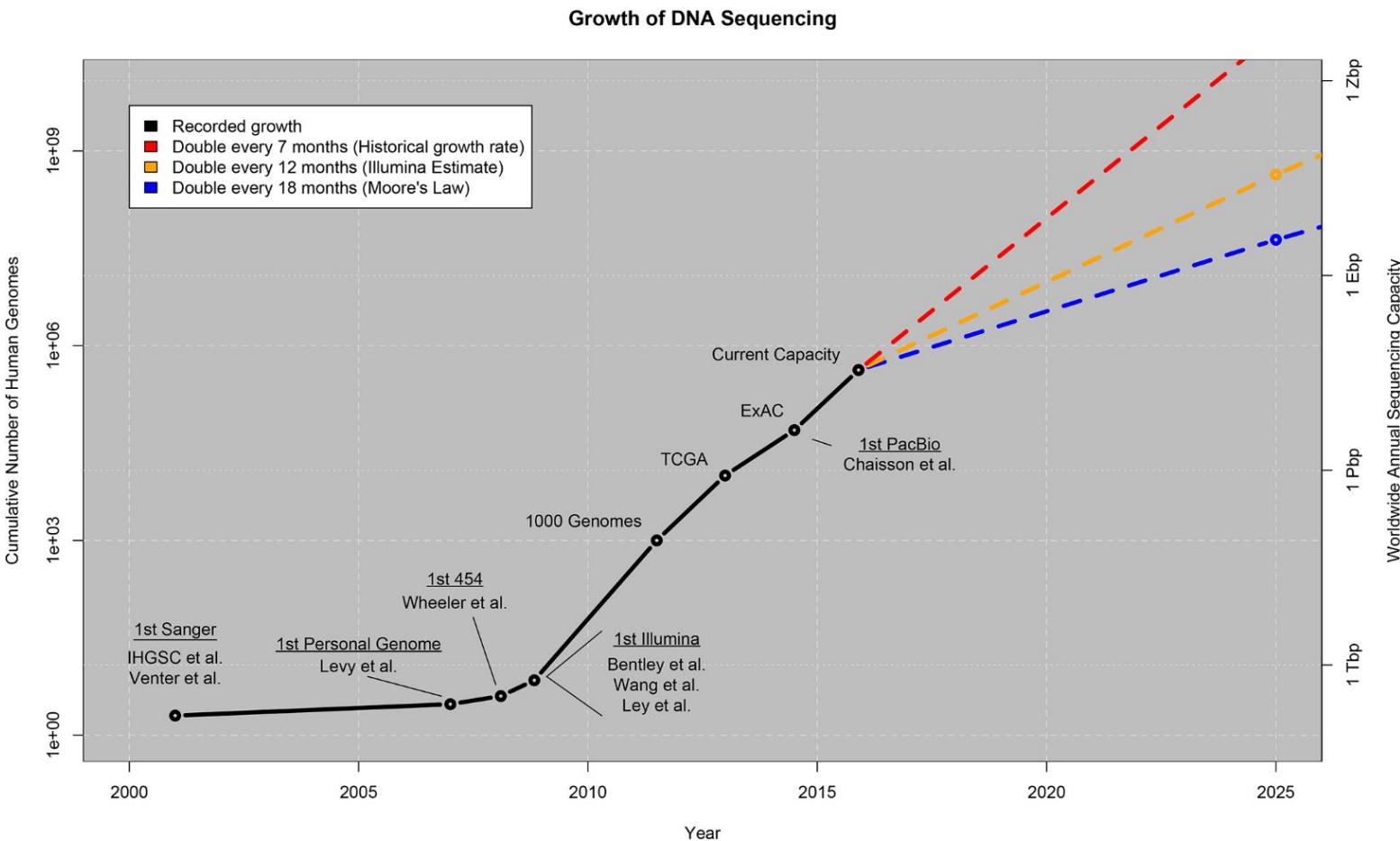
# Tomorrow



# Where to begin...



# The challenge



Follow

Congratulations to Dr Katie Bouman!  
This is the woman who created the algorithm  
to crunch the 5 petabytes of data from 500  
kg of hard drives from 8 radio telescopes to  
make the first image of the #EHTBlackHole  
#BlackHole



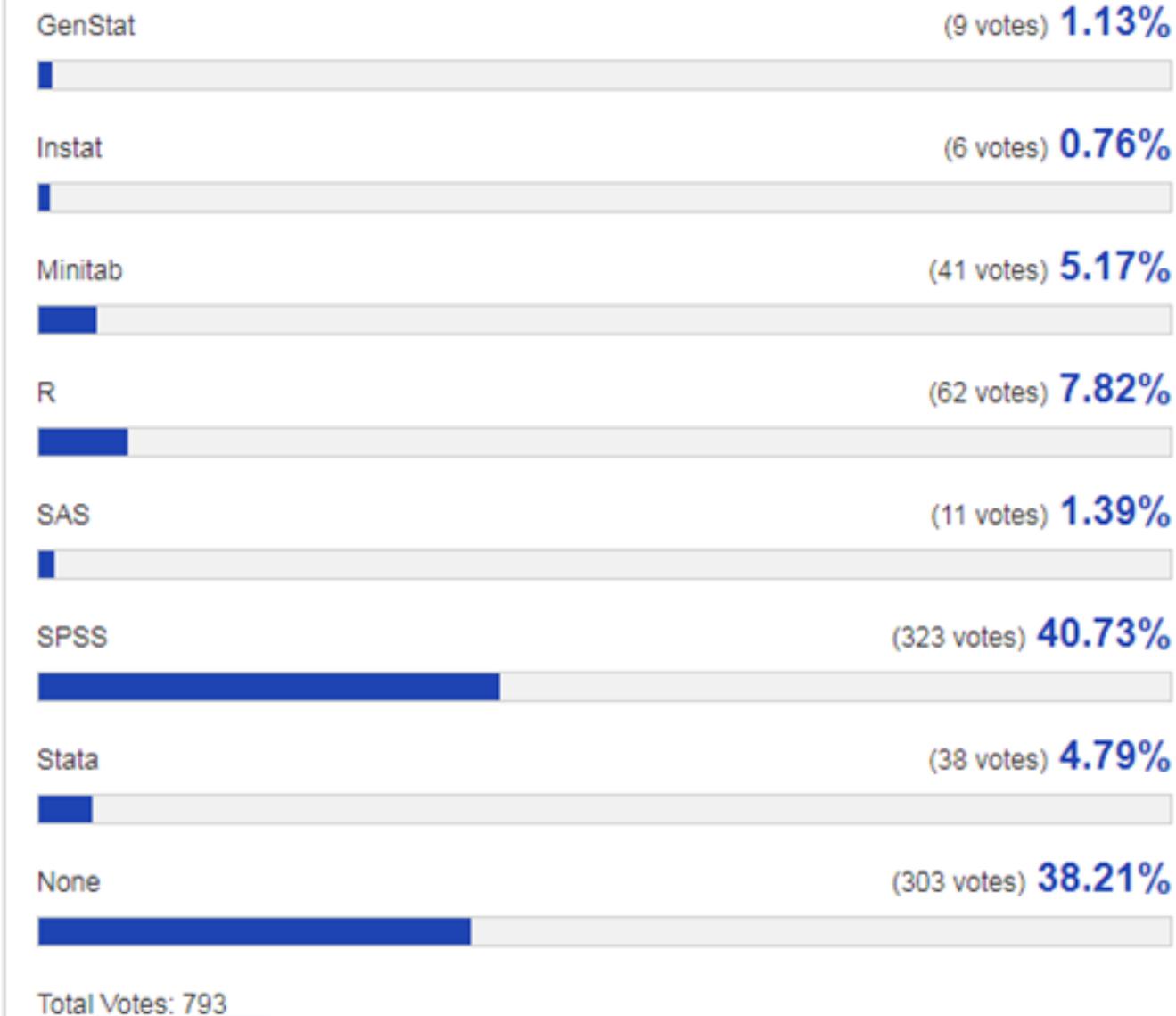
2:55 PM - 10 Apr 2019

# Fundamental problem



I'm not in the office at the moment. Send any work to be translated

# Where are we all coming from?



# This is a big problem

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
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																			
41																			
42																			
43																			
44																			
45																			

## Enron North America - West Gas

November 9, 2001

### ENA - West Gas Contacts

#### Houston Office

Barry Tycholiz	(713) 853-1587
Kim Ward	(713) 853-0685
Stephanie Miller	(713) 853-1688
Philip Polsky	(713) 853-5181

#### Regional Offices

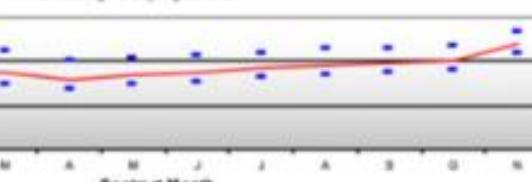
Mark Whitt	(303) 575-6473	Denver
Paul Lucci	(303) 575-6474	Denver
Tyrell Harrison	(303) 575-6478	Denver
Dave Fuller	(503) 464-3732	Portland

### Forward Prices (US\$/MMBtu)

#### NYMEX

NETTLE	A
Cash	
ROM	
Dec-01	2.960 0.090
Dec-01 to Mar-02	3.088 0.083
Apr-02 to Oct-02	3.166 0.084
Nov-02 to Mar-03	3.651 0.090
One Year Strip*	3.165 0.084

Forward NYMEX Strip  
with trailing 10-day highs/lows



#### IF NWPL Rocky Mountains

Fixed Price	Basis	Fixed Price	Basis
BID	OFFER	BID	OFFER
1.890	1.910	2.060	2.080
2.395	2.415	(0.565)	(0.545)
2.594	2.614	(0.494)	(0.474)
2.581	2.601	(0.585)	(0.565)
3.356	3.376	(0.295)	(0.275)
2.634	2.654	(0.520)	(0.510)

#### IF EL Paso San Juan

Fixed Price	Basis	Fixed Price	Basis
BID	OFFER	BID	OFFER
2.450	2.470	2.350	2.370
2.560	2.580	(0.400)	(0.380)
2.743	2.763	(0.345)	(0.325)
2.801	2.821	(0.365)	(0.345)
3.421	3.441	(0.230)	(0.210)
2.817	2.837	(0.347)	(0.327)

#### AECO / NIT

Fixed Price	Basis	Fixed Price	Basis
BID	OFFER	BID	OFFER
2.376	2.396	2.480	2.500
2.398	2.418	2.460	2.480
2.552	2.572	(0.408)	(0.388)
2.616	2.636	(0.472)	(0.452)
2.681	2.681	(0.505)	(0.485)
3.216	3.236	(0.435)	(0.415)
2.676	2.696	(0.488)	(0.468)

#### IF NWPL Canadian Border (Sumas)

Fixed Price	Basis	Fixed Price	Basis
BID	OFFER	BID	OFFER
2.800	2.820	(0.160)	(0.140)
2.892	2.912	(0.196)	(0.176)
2.796	2.816	(0.370)	(0.350)
3.706	3.726	0.055	0.075
2.880	2.900	(0.285)	(0.265)

#### IF PEPL TX-OK

Fixed Price	Basis	Fixed Price	Basis
BID	OFFER	BID	OFFER
2.530	2.550	2.530	2.550
2.828	2.848	(0.133)	(0.113)
2.958	2.978	(0.130)	(0.110)
3.046	3.066	(0.120)	(0.100)
3.531	3.551	(0.120)	(0.100)
3.041	3.061	(0.123)	(0.103)

# 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	mv	filter	1	1.545	2.45	cos	2019/04/15	Brendan Palmer		

data

dictionary

values



# 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 week duration

# Real data comes with real problems

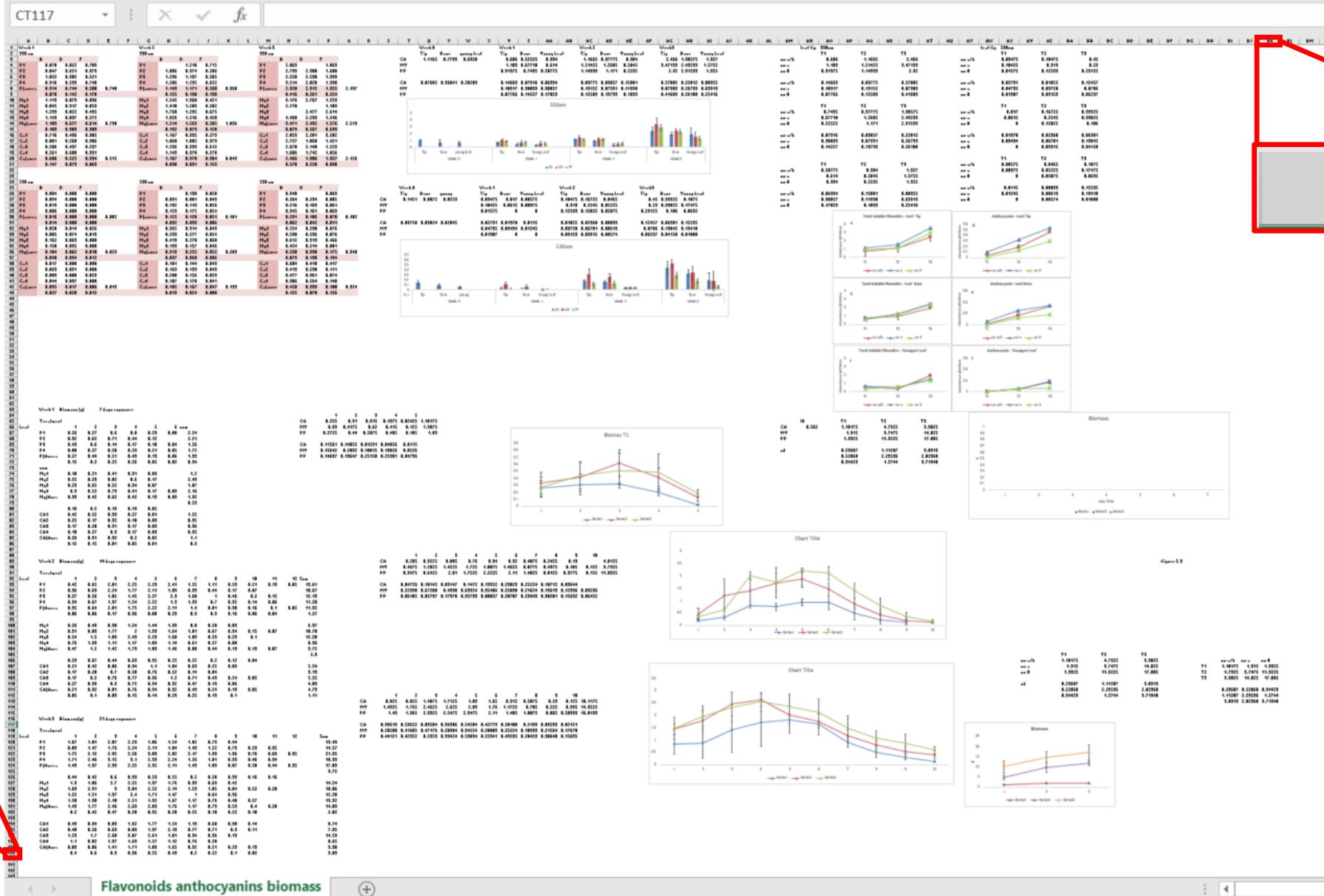
Raw Data wk 1-3 Lettuce Exp 1 - Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Week 1						Week 2						Week 3					
2	330 nm						330 nm						330 nm					
3		B	D	F				B	D	F				B	D	F		
4	P1	0.870	0.822	0.703			P1						1	2.869		1.069		
5	P2	0.847	0.651	0.379			P2						2	2.739	2.380	1.688		
6	P3	1.022	0.902	0.521			P3	1.236	1.197	0.585			P3	2.558	2.538	1.333		
7	P4	0.916	0.599	0.748			P4	1.206	1.295	0.652			P4	3.514	2.028	1.330		
8	P(average)	0.914	0.744	0.588	0.748		P(average)	1.149	1.171	0.560	0.960		P(average)	2.920	2.315	1.355	2.197	
9		0.078	0.142	0.170				0.125	0.138	0.190				0.416	0.261	0.254		
10	My1	1.119	0.873	0.896			My1	1.545	1.360	0.421			My1	3.176	2.767	1.259		
11	My2	0.845	0.917	0.853			My2	1.418	1.203	0.502			My2	2.778		1.183		
12	My3	1.299	0.822	0.435			My3	1.768	1.295	0.675			My3		2.477	2.614		
13	My4	1.149	0.097	0.272			My4	1.326	1.216	0.420			My4	4.460	2.233	1.246		
14	My(average)	1.103	0.677	0.614	0.798		My(average)	1.514	1.269	0.505	1.096		My(average)	3.471	2.492	1.576	2.513	
15		0.189	0.389	0.309				0.192	0.073	0.120				0.879	0.267	0.693		
16	Ca1	0.716	0.496	0.382			Ca1	1.167	0.935	0.273			Ca1	2.853	2.201	3.202		
17	Ca2	0.881	0.568	0.386			Ca2	1.060	1.005	0.373			Ca2	2.727	1.860	1.421		
18	Ca3	0.586	0.437	0.237			Ca3	1.296	0.993	0.612			Ca3	2.678	2.140	1.229		
19	Ca4	0.561	0.600	0.331			Ca4	1.143	0.978	0.278			Ca4	1.606	1.742	1.856		
20	Ca(average)	0.686	0.525	0.334	0.515		Ca(average)	1.167	0.978	0.384	0.843		Ca(average)	2.466	1.986	1.927	2.126	
21		0.147	0.073	0.069				0.098	0.031	0.159				0.578	0.220	0.890		
22																		
23																		
24	530 nm						530 nm						530 nm					
25		B	D	F				B	D	F				B	D	F		
26	P1	0.004	0.000	0.000			P1		0.138	0.050				P1	0.340		0.069	
27	P2	0.034	0.000	0.000			P2		0.091	0.081	0.043			P2	0.264	0.234	0.085	CA
28	P3	0.019	0.000	0.000			P3		0.132	0.119	0.056			P3	0.216	0.163	0.061	MY

File Home Insert Page Layout Formulas Data Review View Tell me what you want to do...

Normal Page Break Preview Custom Layout Views Workbook Views

Ruler Formula Bar Gridlines Headings Zoom 100% Zoom to Selection Window New Arrange Freeze All Panes Hide Synchronous Scrolling Reset Window Position Window Switch Windows Macros Macros



# Take small steps to big changes

THE AMERICAN STATISTICIAN  
2018, VOL. 72, NO. 1, 2–10  
<https://doi.org/10.1080/00031305.2017.1375989>



OPEN ACCESS



## Data Organization in Spreadsheets

Karl W. Broman<sup>a</sup> and Kara H. Woo<sup>b</sup>

<sup>a</sup>Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison, Madison, WI; <sup>b</sup>Information School, University of Washington, Seattle, WA

### ABSTRACT

Spreadsheets are widely used software tools for data entry, storage, analysis, and visualization. Focusing on the data entry and storage aspects, this article offers practical recommendations for organizing spreadsheet data to reduce errors and ease later analyses. The basic principles are: be consistent, write dates like YYYY-MM-DD, do not leave any cells empty, put just one thing in a cell, organize the data as a single rectangle (with subjects as rows and variables as columns, and with a single header row), create a data dictionary, do not include calculations in the raw data files, do not use font color or highlighting as data, choose good names for things, make backups, use data validation to avoid data entry errors, and save the data in plain text files.

### ARTICLE HISTORY

Received June 2017  
Revised August 2017

### KEYWORDS

Data management; Data organization; Microsoft Excel; Spreadsheets

# Less stress, more success

	A	B	C	D	E	F	G	H	I	J	K	L
1	id	week_no	filter_name	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		A	B	C	D	E			
4	3	0	ptp		1	field_name	data_type	data_format	example	standard_units	description	
5	4	0	ptp		2	id	numeric	integer	23	NA	Unique identifier applied to each observation	
6	1	0	my		3	week_no	numeric	integer	1	NA	Week number, 1 = 7 days exposure, 2 = 14 days exposure	
7	2	0	my		4	filter_name	character	NA	my	NA	3 filter types; 'ptp' = polytunnel plastic blocks all UV light	
8	3	0	my		5	treatment	character	NA	filter	NA	Presence or absence of a filter at the time of sampling	
9	4	0	my		6	replicate_no	numeric	integer	1	NA	The number of replicates in each treatment	
10	1	0	ca		7	flavonoids	numeric	double	0.3421	parts per million (ppm)	Leaf disc taken from the tip of the most mature leaf at th	
11	2	0	ca		8	biomass	numeric	double		gram (g)	Above ground biomass on the day of harvest	
12	3	0	ca		9	variety	character	NA	cos	NA	3 commerical varieties of red lettuce used; 'cos' = Cos Di	
13	4	0	ca		10	date	date	YYYY/MM/DD	2019/06/28	ISO 8601	Experiment date	
14	5	1	ptp		11	investigator	character	Firstname Lastname	Aoife Coffey	NA	Primary researcher who performed the experiment	
15	6	1	ptp		12							
16	7	1	ptp		13							
17	8	1	ptp		14							
18	9	1	my		15							
19	10	1	my		16							
20	11	1	my		17							
21	12	1	my		18							
22	13	1	ca		19							
23	14	1	ca		20							
24	15	1	ca		21							
25	16	1	ca		22							
26	17	2	ptp		23							
27	18	2	ptp		24							
28	19	2	ptp		25							
29	20	2	ptp		26							
30	21	2	mv		27							
			dictionary		28							
					29							
					30							

# Less stress, more success

1	A	B	C	D	E	F	G	H	I	J	K	L
2	1	0	ptp	nofilter	1	1.061	0.39	cos	2019/04/01	Darren Dahly		
3	2	0	ptp	A	B	C	D	E				
4	3	0	ptp	1	field_name	data_type	data_format	example	standard_units	description		
5	4	0	ptp	2	id	numeric	integer	23	NA	Unique identifier applied to each observation		
6	1	0	my	3	week_no	numeric	integer					
7	2	0	my	4	filter_name	character	NA					
8	3	0	my	5	treatment	character	NA					
9	4	0	my	6	replicate_no	numeric	integer					
10	1	0	ca	7	flavonoids	numeric	double					
11	2	0	ca	8	biomass	numeric	double					
12	3	0	ca	9	variety	character	NA					
13	4	0	ca	10	date	date	YYYY/MM/DD					
14	5	1	ptp	11	investigator	character	Firstname Lastname					
15	6	1	ptp	12								
16	7	1	ptp	13								
17	8	1	ptp	14								
18	9	1	my	15								
19	10	1	my	16								
20	11	1	my	17								
21	12	1	my	18								
22	13	1	ca	19								
23	14	1	ca	20								
24	15	1	ca	21								
25	16	1	ca	22								
26	17	2	ptp	23								
27	18	2	ptp	24								
28	19	2	ptp	25								
29	20	2	ptp	26								
30	21	2	mv	27								
		data	dictionary	28								
				29								
				30								

The screenshot shows a data entry interface with two tabs: 'data' and 'dictionary'. The 'data' tab displays a grid of experimental data. The 'dictionary' tab provides a detailed schema for each column, including field\_name, data\_type, data\_format, example, standard\_units, and description. A tooltip for 'id' indicates it is a unique identifier applied to each observation.

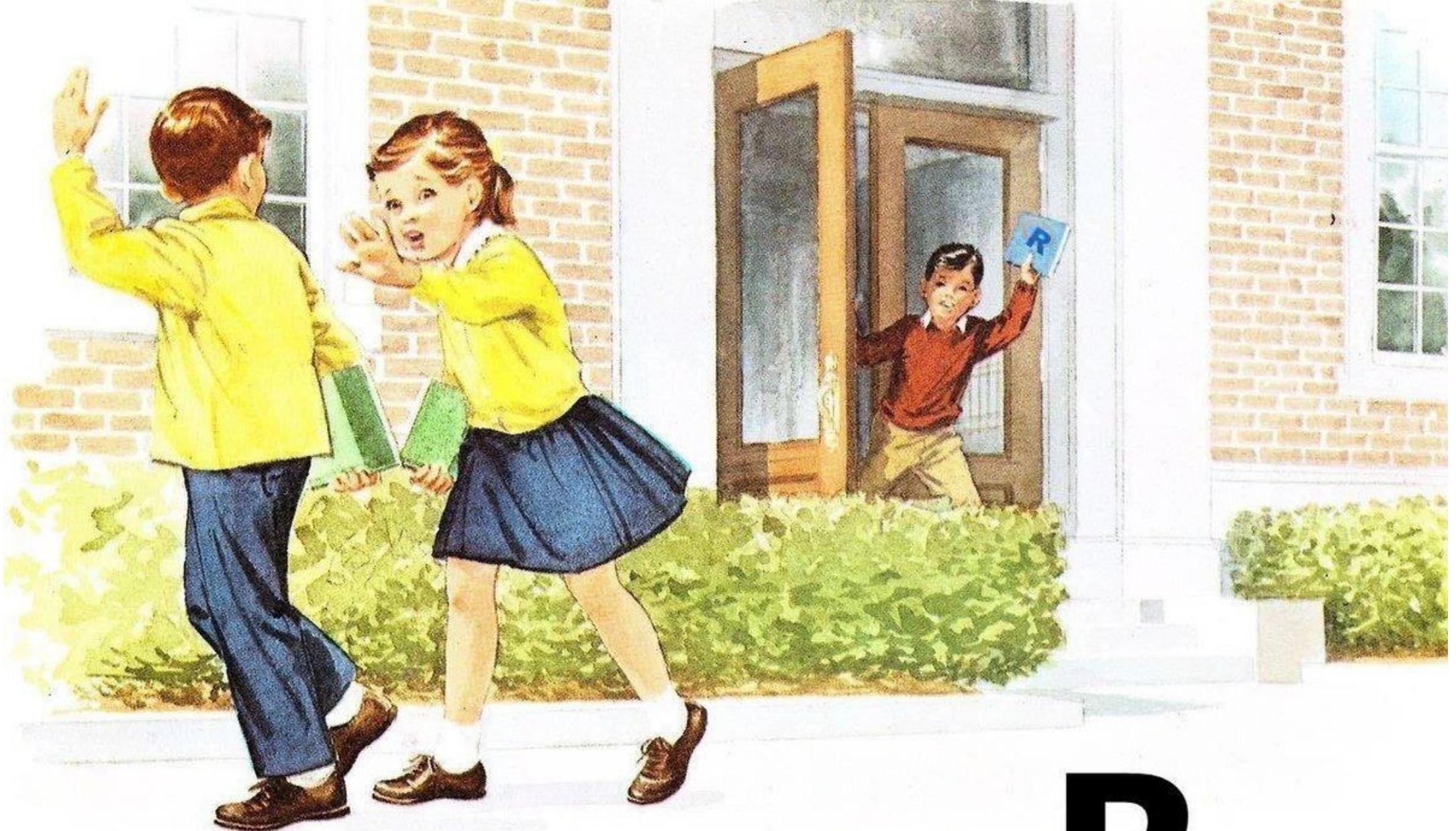
Below the tabs, there are navigation buttons for the data grid: back, forward, data, dictionary, values, and a plus sign.

# Less stress, more success

# Putting the pieces together

- A: Define a project structure
- B: Set a naming convention
- C: Use scripted workflows
- D: Digital notebooks

Reproducible  
research



Run, or he's going to tell us about  
again!

R

# Still haven't found what I'm looking for

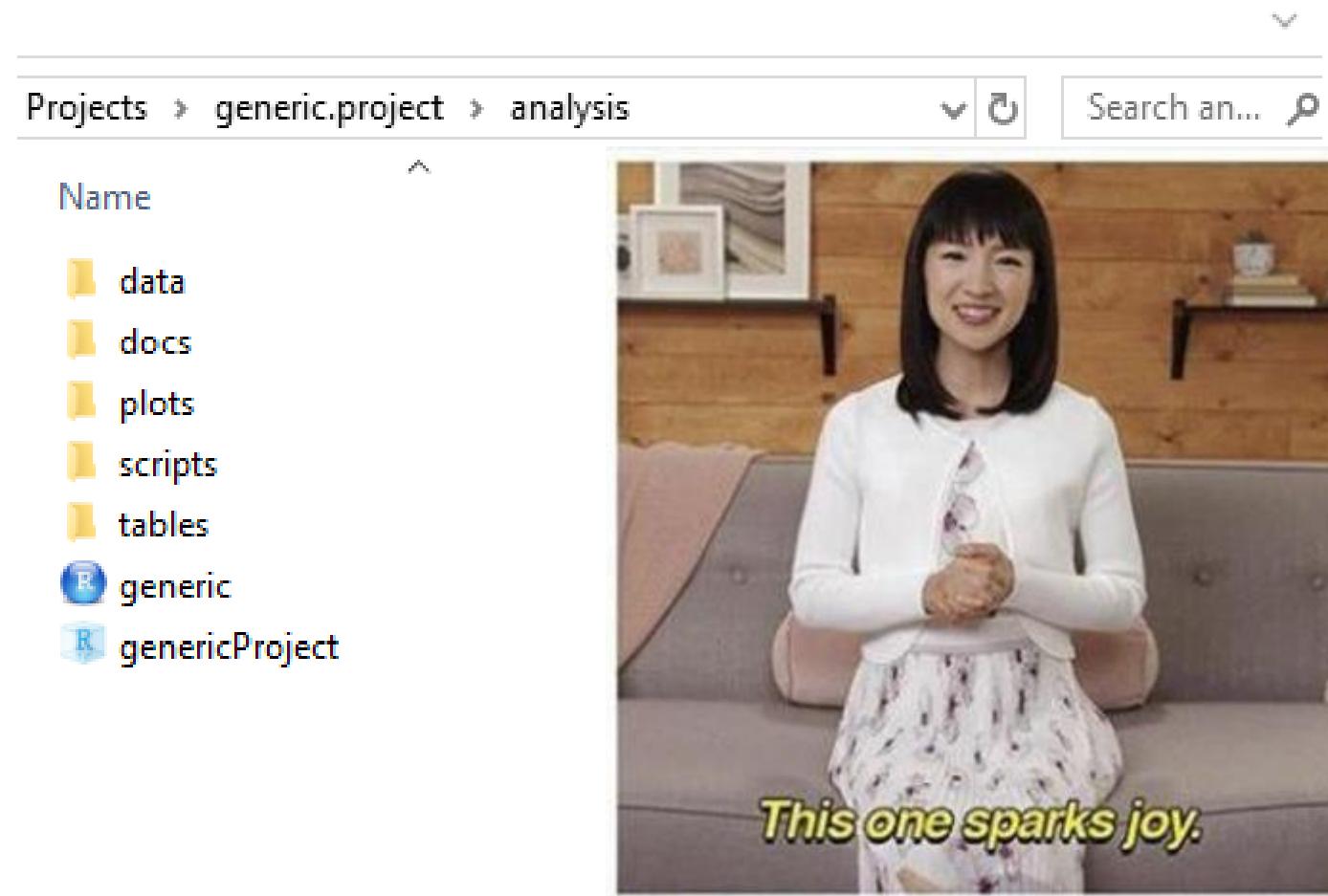
- Help your future-self

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
Supplementary_Figure_B_A_Palmer_Sept_14	29/08/2014 13:21
Supplementary_Figure_B_A_Palmer_Sept_14	10/09/2014 22:31
Tables_B_A_Palmer_Sept_2014	10/09/2014 22:09



# A: Define a generic project structure



# B: Give your files and folders informative names

This PC > Documents > Projects > **2016-08-08\_RespPCT** > analysis > data

Name	Date modified
raw_data	21/01/2019 21:06
2018-11-06_abx	06/11/2018 13:10
2018-11-06_monitoring	06/11/2018 13:09
2018-11-06_pct	06/11/2018 13:08
2018-11-06_pt_info	06/11/2018 13:07

# Outline a file naming convention

## Machine readable:

- Inherent order
- Avoid spaces
- Avoid punctuation
- Remove case-sensitivity

## Human readable:

- Contains info on content
- Avoid spaces
- Avoid punctuation
- Remove case sensitivity

## Metadata:

Separate with underscores ("\_")

- Avoid punctuation
- Remove case-sensitivity

01\_marshall-data.r  
02\_pre-dea-filtering.r  
03\_dea-with-limma-voom.r  
04\_explore-dea-results.r  
90\_limma-model-term-name-fiasco.r  
helper01\_load-counts.r  
helper02\_load-exp-des.r  
helper03\_load-focus-statinf.r  
helper04\_extract-and-tidy.r

# Everything in its right place

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

**NO**

Name
All unique 4a amino acid Sequences (B-N).fas
All unique 4a amino acid Sequences (B-N).meg
All_AA_haplotypes.meg
All_AA_haplotypes_with_clonal_sequences.meg
BS100_AA_with_clones
BS100_AA_with_clones.nwk
BS1000_AA_pyro&clones
BS1000_AA_pyro&clones.nwk
BS1000_AA_pyro_only
BS1000_AA_pyro_only.nwk
BS1000_Uncle_Clonal_AA

**Yes**

Projects > 2016-08-08\_RespPCT > analysis > scripts

Name
R 01_clean_data
R 02_plots
R 03_tables
R 04_stats_analysis
R 05_post_hoc_stats
R functions
R randomization
R tables

# C: Joined up thinking

- The R scripts should also be human readable
  - Annotate the code
  - Break up the scripts into dedicated tasks
  - Interlink to other 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")
```

# Work from the raw data ALWAYS!!



**Tom Webb** @tomjwebb · 16 Jan 2015

If you could tell a new PhD student one thing to help make their data more useful/shareable, what would it be?

27

11

7



**Dr Gavin Simpson**

@ucfagls

Follow

Replying to @tomjwebb

@tomjwebb don't, not even with a barge pole, not for one second, touch or otherwise edit the raw data files. Do any manipulations in script

7:15 AM - 16 Jan 2015

# D: R Markdown

- R Markdown combines the code you wrote, the output produced and your 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.

# D: R Markdown

R ~/Open\_Science/Digital\_Badge/RCR - master - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

lettuce\_report.Rmd\* Go to file/function Addins

```
1 ---  
2 title: "This is a reproducible document"  
3 author: "Dr. Brendan Palmer"  
4 date: "18th June 2019"  
5 output:  
6   word_document:  
7     fig_height: 4  
8     fig_width: 6  
9 ---  
10 # This is the beginning of the project  
11  
12 our initial reports might be restricted to lab meetings etc. We can use `R  
13 Markdown` to show the code we are using, so that the meetings are not just a  
14 demonstration of the results, but also an examination of the `code` used to obtain  
15 them.  
16  
17 knitr::opts_chunk$set(echo = FALSE, message = FALSE, warning = FALSE)  
18  
19 # Load your packages here  
20 library(tidyverse)  
21 library(knitr)  
22  
23  
24 The plot below is call from the ggplot object entitled `report_plot` created in  
25 the script `03_final_analysis.R`.  
26  
27 {r Plots from script, echo = FALSE}  
28 source("scripts/03_final_analysis.R")  
29  
30 # 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.

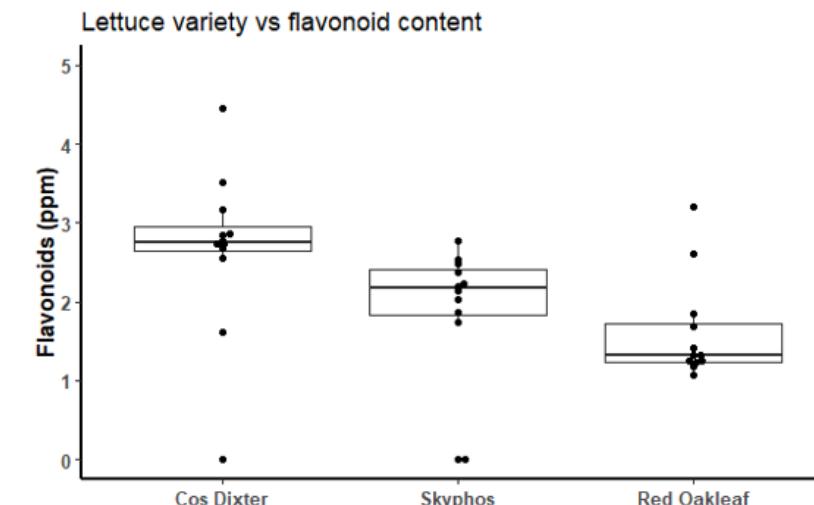
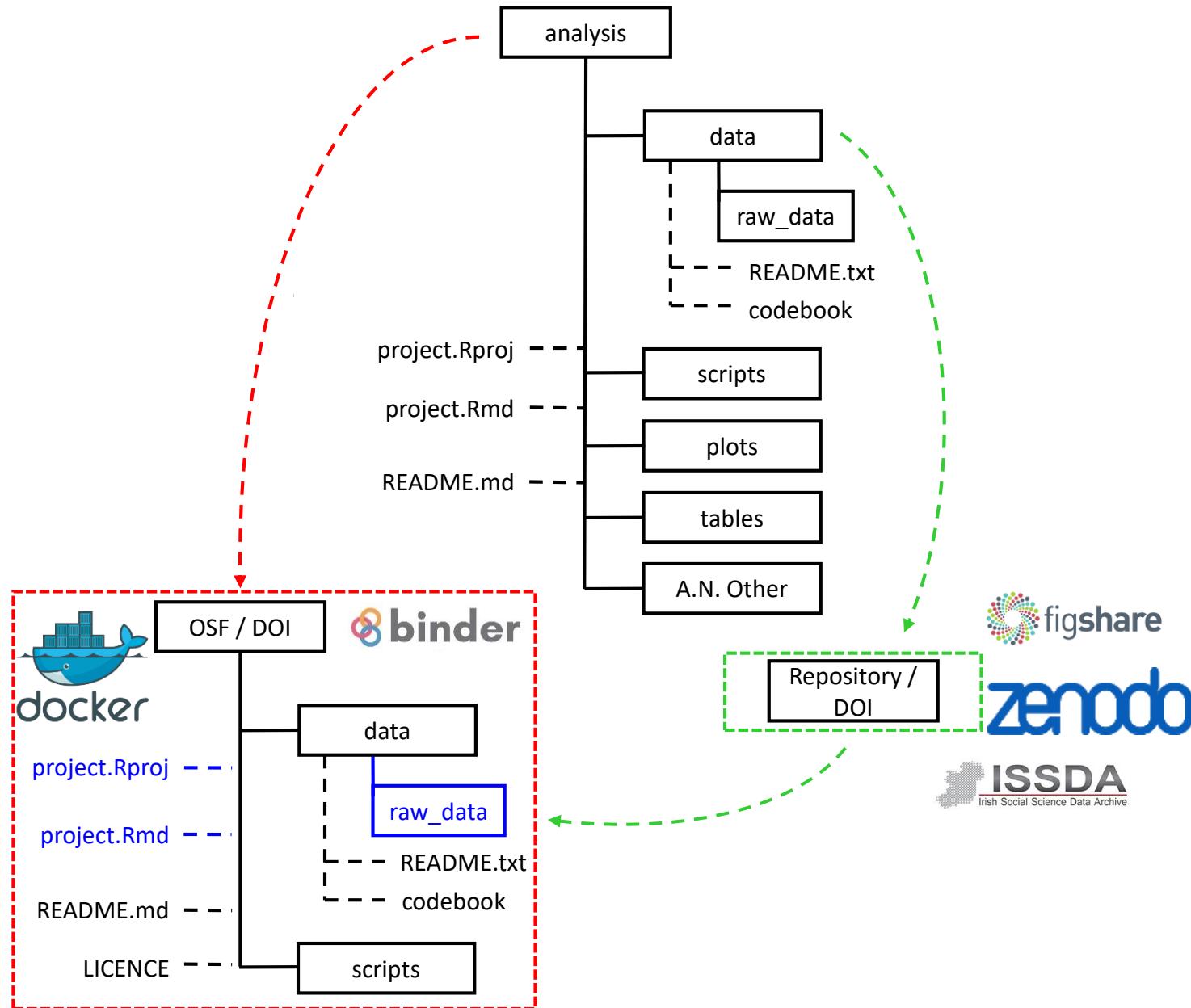


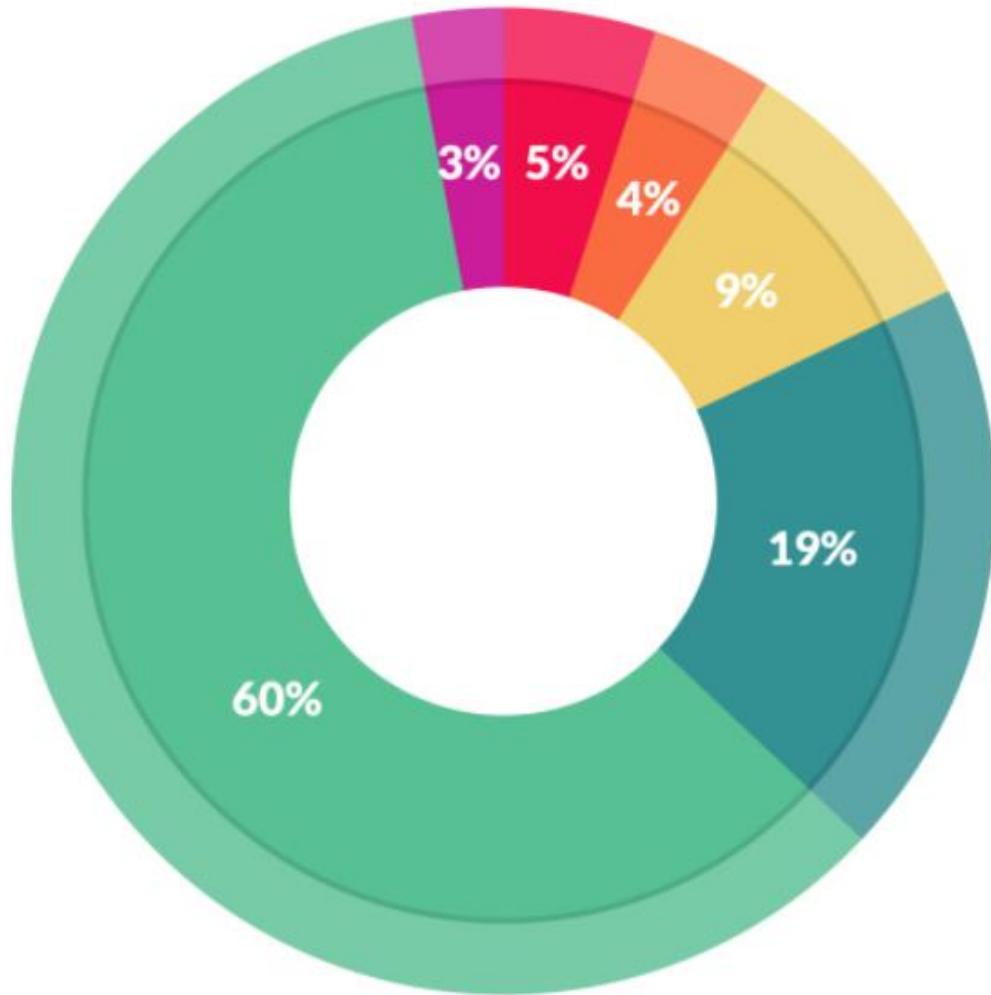
Fig. 1. Flavonoid content of three lettuce varieties under three experimental conditions.

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

# What does this allow us to do?



# Resources are being wasted by not doing this



What data scientists spend the most time doing

- *Building training sets: 3%*
- *Cleaning and organizing data: 60%*
- *Collecting data sets; 19%*
- *Mining data for patterns: 9%*
- *Refining algorithms: 4%*
- *Other: 5%*

# RStudio Cloud



[Log In](#) [Sign Up](#)

## Welcome to RStudio Cloud alpha

Do, share, teach and learn data science with R.

[Get Started](#)

If you already have an RStudio shinyapps.io account, you can log in using your existing credentials.

# GitHub



Search or jump to...

Pull requests Issues Marketplace Explore



Set status

**Brendan Palmer**  
bapalmer

[Edit profile](#)

Twitter: @B\_A\_Palmer  
[🔗 https://crfcsdau.github.io/about/](https://crfcsdau.github.io/about/)

Organizations



Overview    **Repositories 14**    Projects 0    Stars 0    Followers 10    Following 10

Find a repository...

Type: All ▾

Language: All ▾

New

## R-A\_Hitchhikers\_Guide\_to\_Reproducible\_Research

Star

A 3-day R course given in University College Cork that encompasses RStudio projects, an introduction to the R tidyverse language, project workflows and reporting using R Markdown.

MIT License    Updated 6 days ago

## lunchtime\_sessions

Star

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

HTML    1    MIT License    Updated 12 days ago

## RCR

Star

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

HTML    MIT License    Updated 23 days ago

# The butterfly has started flapping its wings



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."



EUROPEAN COMMISSION  
Directorate-General for Research & Innovation

## H2020 Programme

Guidelines on  
FAIR Data Management in Horizon 2020

Home > Funding > Policies and principles > **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. The 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.

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



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.

# Install the Chrome plugin PubPeer

NCBI Resources ▾ How To ▾ Sign in to NCBI

PubMed ▾ Is the Power Threshold of 0.8 Applicable to Surgical Science? Search

PubMed.gov US National Library of Medicine National Institutes of Health Create RSS Create alert Advanced Help

Format: Abstract ▾ Send to ▾

See 1 citation found by title matching your search:

J Surg Res. 2019 Apr 26;241:235-239. doi: 10.1016/j.jss.2019.03.062. [Epub ahead of print]

23 comments on PubPeer (by: Andrew D. Althouse, Thom Baguley, Guillaume A. Rousselet, Timothy Feeney, Paul M Brown, Frank E. Harrell, David Nunan, Samantha R. Seals, Raj Mehta, Yevgeniy Feyman, Ionomidotis Irregularis, Andrew Gelman, Aleksi Reito, Daniel E. Leisman, Pavlos Msaouel, Ryan Miller, Maarten Van Smeden, Zad Rafi Chow)

**Is the Power Threshold of 0.8 Applicable to Surgical Science?-Empowering the Underpowered Study.**

Bababekov YJ<sup>1</sup>, Hung YC<sup>2</sup>, Hsu YT<sup>2</sup>, Udelsman BV<sup>2</sup>, Mueller JL<sup>2</sup>, Lin HY<sup>2</sup>, Stapleton SM<sup>2</sup>, Chang DC<sup>2</sup>.

Author information

Abstract

**BACKGROUND:** Many articles in the surgical literature were faulted for committing type 2 error, or concluding no difference when the study was "underpowered". However, it is unknown if the current power standard of 0.8 is reasonable in surgical science.

Full text links ELSEVIER FULL-TEXT ARTICLE

Save items

Similar articles

Review Interventions to Prevent Falls in Community-L Agency for Healthcare Research...]

Review Is There Truly "No Significant Difference"? Underl J Bone Joint Surg Am. 2015]

Review Randomized controlled trials and neurosurgery: the ideal fit or : [J Neurosurg. 2016]

Review Low-Dose Aspirin for the Prevention of Morbidity anc Agency for Healthcare Research...]

# Further reading



Sam Westwood

@westwoodsam1

Following



I am embarking on my own [#PaperPerDayChallenge](#) where I read at least one paper, well, per day for a whole year. To kick start, [nature.com/articles/43573...](http://nature.com/articles/43573...) inspired by [@ukrepro](#) Reproducibility Workshop [@CumberlandLodge](#) and a talk by [@MarcusMunafo](#)



## Scientists behaving badly

In a questionnaire-based survey of US biomedical researchers, respondents admitted to a range of dubious practices. Transgressions included failing to present data [nature.com](http://nature.com)

# Twitter



**UK Reproducibility Network**

@ukrepro

UK Reproducibility Network: a peer-led consortium to investigate factors which contribute to robust research, provide training, and disseminate best practice.



**Malcolm Macleod #FBPE**

@Macломaclee Follows you

clinical neurologist, stroke trialist, and interested in improving the quality of laboratory research



**Open Science MOOC**

@OpenScienceMOOC Follows you

A community designed for students and researchers to help make 'Open' the default setting for the future of research. Slack: [osmooc.herokuapp.com](https://osmooc.herokuapp.com)

⌚ Everywhere



**Brian Nosek**

@BrianNosek

Executive Director @ Center for Open Science, Professor @ University of Virginia, and co-Founder of Project Implicit



**Kate Button**

@ButtonKate Follows you

Academic. Psychologist. Cognitive mechanisms of depression & anxiety. Meta-science & scientific rigour. Sporadic Twitterer.



**Darren L Dahly**

@statsep1 Follows you

Principal Statistician, Epidemiologist, Sr Lecturer | @HRBIreland Clinical Research Facility @CRF\_CORK | Cork #Rstats Users Group [meetup.com/Cork-Ireland-R...](https://meetup.com/Cork-Ireland-R-/)



**Dorothy Bishop**

@deevybee

Professor of developmental neuropsychology. Blog on [deevybee.blogspot.com](http://deevybee.blogspot.com) Main focus #devlangdis, see: [youtube.com/radld](https://youtube.com/radld)



**Elisabeth Bik**

@MicrobiomDigest

Science consultant, PhD. Harbers-Bik LLC. Microbiome, research integrity & misconduct. Ex @Stanford. MicrobiomeDigest/Bik's Picks. Dutch/USA. My views.



**Retraction Watch**

@RetractionWatch

Tracking retractions as a window into the scientific process. Sign up for our daily newsletter: [eepurl.com/bNR1Un](http://eepurl.com/bNR1Un) Tips? team@retractionwatch.com



**Jenny Bryan**

@JennyBryan

Software engineer @rstudio, humane #rstats, adjunct prof @UBC where I created @STAT545, part of @ropensci