

effective computing for research reproducibility

Laura Fortunato

University of Oxford

Santa Fe Institute



@anthrolog



<http://www.santafe.edu/~fortunato/>

preliminaries

disclaimer

- ▶ I am not a statistician
- ▶ I am not a data scientist
- ▶ I am not a computer scientist

preliminaries

disclaimer

- ▶ I am not a statistician
- ▶ I am not a data scientist
- ▶ I am not a computer scientist
- ▶ I am a *computational researcher*

preliminaries

disclaimer

- ▶ I am not a statistician
- ▶ I am not a data scientist
- ▶ I am not a computer scientist
- ▶ I am a *computational researcher*

broader context

“open science” < “open research” < “open scholarship”

Reproducible Research Oxford



@RR_Oxford



<http://rroxford.github.io/>



Reproducible Research Oxford



@RR_Oxford



<http://rroxford.github.io/>



The Carpentries



@thecarpentries



<http://carpentries.org/>



Reproducible Research Oxford

 @RR_Oxford
 <http://rroxford.github.io/>



The Carpentries

 @thecarpentries
 <http://carpentries.org/>



UK Reproducibility Network

 @ukrepro
 <http://ukrn.org/>



Reproducible Research Oxford

 @RR_Oxford
 <http://rroxford.github.io/>



The Carpentries

 @thecarpentries
 <http://carpentries.org/>



UK Reproducibility Network

 @ukrepro
 <http://ukrn.org/>



Project TIER

 @Project_TIER
 <https://www.projecttier.org/>



Reproducible Research Oxford

 @RR_Oxford
 <http://rroxford.github.io/>



The Carpentries

 @thecarpentries
 <http://carpentries.org/>



UK Reproducibility Network

 @ukrepro
 <http://ukrn.org/>



Project TIER

 @Project_TIER
 <https://www.projecttier.org/>



Software Freedom Conservancy

 @conservancy
 <https://sfconservancy.org/>



research reproducibility

the problem, loosely defined

results can only be trusted if they can be *re-derived*
by the original researchers and/or by others working independently

research reproducibility

the problem, loosely defined

results can only be trusted if they can be *re-derived*
by the original researchers and/or by others working independently

trust issues

research reproducibility

the problem, loosely defined

results can only be trusted if they can be *re-derived*
by the original researchers and/or by others working independently

trust issues

- ▶ can I trust the result?

research reproducibility

the problem, loosely defined

results can only be trusted if they can be *re-derived*
by the original researchers and/or by others working independently

trust issues

- ▶ can I trust the result?
- ▶ can I trust the researcher?

research reproducibility

the problem, loosely defined

results can only be trusted if they can be *re-derived*
by the original researchers and/or by others working independently

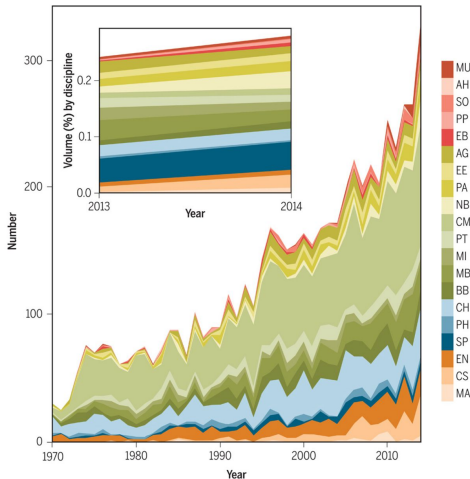
trust issues

- ▶ can I trust the result?
- ▶ can I trust the researcher?
- ▶ can I trust the process?

reproducibility crisis

Goodman et al. (2016) *Science Translational Medicine* 8(341): 341ps12

What does research reproducibility mean?



title or abstract including
one or more of:

“research reproducibility”
“reproducibility of research”
“reproducibility of results”
“results reproducibility”
“reproducibility of study”
“study reproducibility”
“reproducible research”
“reproducible finding”
“reproducible result”

“research on research”

Ioannidis et al. (2015) *PLOS Biology* 13(10): e1002264

Meta-research: evaluation and improvement of research methods and practices

Meta-research area	Specific interests (nonexhaustive list)
Methods: “performing research”—study design, methods, statistics, research synthesis, collaboration, and ethics	Biases and questionable practices in conducting research, methods to reduce such biases, meta-analysis, research synthesis, integration of evidence, crossdesign synthesis, collaborative team science and consortia, research integrity and ethics
Reporting: “communicating research”—reporting standards, study registration, disclosing conflicts of interest, information to patients, public, and policy-makers	Biases and questionable practices in reporting, explaining, disseminating and popularizing research, conflicts of interest disclosure and management, study registration and other bias-prevention measures, and methods to monitor and reduce such issues
Reproducibility: “verifying research”—sharing data and methods, repeatability, replicability, reproducibility, and self-correction	Obstacles to sharing data and methods, replication studies, replicability and reproducibility of published research, methods to improve them, effectiveness of correction and self-correction of the literature, and methods to improve them
Evaluation: “evaluating research”—prepublication peer review, postpublication peer review, research funding criteria, and other means of evaluating scientific quality	Effectiveness, costs, and benefits of old and new approaches to peer review and other science assessment methods, and methods to improve them
Incentives: “rewarding research”: promotion criteria, rewards, and penalties in research evaluation for individuals, teams, and institutions	Accuracy, effectiveness, costs, and benefits of old and new approaches to ranking and evaluating the performance, quality, value of research, individuals, teams, and institutions

xkcd: trouble for science

<https://xkcd.com/1574/>

MANY COMMERCIAL ANTIBODY-BASED
IMMUNOASSAYS ARE UNRELIABLE

ANTIBODY-BASED IMMUNOASSAYS ARE THE MAINSTAY OF BIOLOGICAL RESEARCH, BUT A NEW STUDY SUGGESTS THAT MANY COMMERCIAL ASSAYS ARE UNRELIABLE. THE STUDY, WHICH WAS CONDUCTED BY A TEAM OF RESEARCHERS AT THE UNIVERSITY OF CALIFORNIA, SAN DIEGO, FOUND THAT ONLY 37% OF THE ASSAYS THEY TESTED WERE ABLE TO DETECT THE TARGET PROTEIN AT THE LEVELS CLAIMED BY THE MANUFACTURERS. THE RESEARCHERS WERE CONCERNED THAT THIS MIGHT BE A PROBLEM FOR OTHER RESEARCHERS WHO RELY ON THESE ASSAYS TO MAKE IMPORTANT SCIENTIFIC DECISIONS.

PROBLEMS WITH THE p -VALUE AS AN
INDICATOR OF SIGNIFICANCE

THE p -VALUE IS A COMMONLY USED STATISTICAL TOOL TO DETERMINE IF A RESULT IS SIGNIFICANT. HOWEVER, IT HAS BEEN CRITICIZED FOR BEING MISUSED AND MISUNDERSTOOD. MANY RESEARCHERS HAVE FOUND THAT THE p -VALUE CAN BE EASILY MANIPULATED TO MAKE A RESULT APPEAR SIGNIFICANT WHEN IT IS NOT. THIS HAS LEAD TO A LOSS OF TRUST IN SCIENTIFIC RESEARCH AND A CALL FOR ALTERNATIVE METHODS OF ANALYSIS.

OVERFEEDING OF LABORATORY RODENTS
COMPROMISES ANIMAL MODELS

LABORATORY RODENTS ARE USED IN A WIDE RANGE OF SCIENTIFIC RESEARCH, BUT A NEW STUDY SUGGESTS THAT OVERFEEDING THEM CAN COMPROMISE THE RESULTS. THE STUDY, WHICH WAS CONDUCTED BY A TEAM OF RESEARCHERS AT THE UNIVERSITY OF CALIFORNIA, SAN DIEGO, FOUND THAT OVERFEEDING RODENTS LEADS TO CHANGES IN METABOLISM AND OTHER PHYSIOLOGICAL FUNCTIONS THAT CAN AFFECT THE OUTCOME OF EXPERIMENTS. THE RESEARCHERS WERE CONCERNED THAT THIS MIGHT BE A PROBLEM FOR OTHER RESEARCHERS WHO RELY ON THESE ANIMAL MODELS TO MAKE IMPORTANT SCIENTIFIC DECISIONS.

REPLICATION STUDY FAILS TO REPRODUCE
MANY PUBLISHED RESULTS

A REPLICATION STUDY OF 100 PUBLISHED RESULTS FROM A VARIETY OF SCIENTIFIC JOURNALS FOUND THAT ONLY 37% OF THE RESULTS WERE ABLE TO BE REPRODUCED. THIS IS A CONCERNING FINDING THAT SUGGESTS THAT MANY PUBLISHED RESULTS MAY BE UNRELIABLE. THE REPLICATION STUDY WAS CONDUCTED BY A TEAM OF RESEARCHERS AT THE UNIVERSITY OF CALIFORNIA, SAN DIEGO, AND THE RESULTS WERE PUBLISHED IN THE JOURNAL OF SCIENTIFIC DATA.

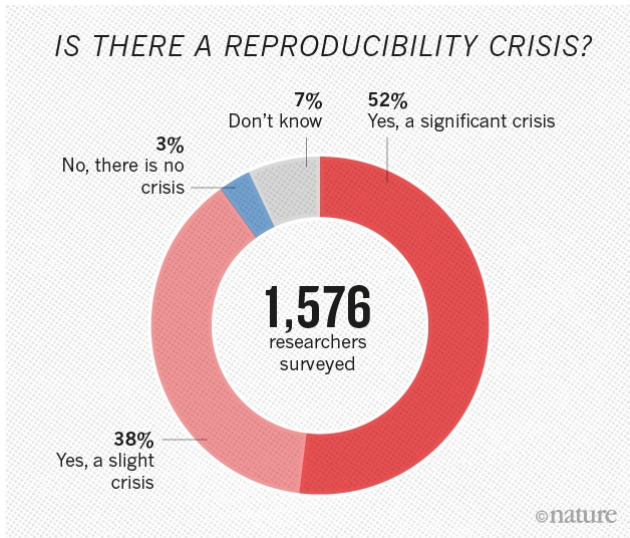
CONTROLLED TRIALS SHOW BUNSEN
BURNERS MAKE THINGS COLDER

A NEW STUDY HAS SHOWN THAT BUNSEN BURNERS CAN MAKE THINGS COLDER. THE STUDY, WHICH WAS CONDUCTED BY A TEAM OF RESEARCHERS AT THE UNIVERSITY OF CALIFORNIA, SAN DIEGO, FOUND THAT THE HEAT FROM A BUNSEN BURNER CAN CAUSE A DROPPING TEMPERATURE IN THE SURROUNDING AREA. THIS IS BECAUSE THE BURNER IS CONCENTRATING THE HEAT IN ONE PLACE, WHICH CAUSES THE AIR TO RISE AND CARRY THE HEAT AWAY. THE RESEARCHERS WERE CONCERNED THAT THIS MIGHT BE A PROBLEM FOR OTHER RESEARCHERS WHO RELY ON BUNSEN BURNERS TO MAKE IMPORTANT SCIENTIFIC DECISIONS.

perceptions among scientists

Baker (2016) *Nature* 533(7604): 452–454

1,500 scientists lift the lid on reproducibility



perceptions among the public

The Economist, issue October 19, 2013



crisis vs. “crisis narrative”

PNAS Sackler Colloquium on Improving the Reproducibility of Scientific Research
published March 13, 2018 by the NAS — National Academy of Sciences USA

Crisis or self-correction: Rethinking media narratives about the well-being of science

Kathleen Hall Jamieson

Opinion: Is science really facing a reproducibility crisis, and do we need it to?

Daniele Fanelli

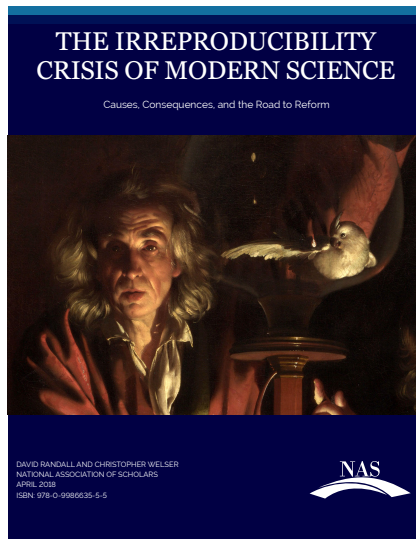
Scientific progress despite irreproducibility: A seeming paradox

Richard M. Shiffrin, Katy Börner, and Stephen M. Stigler

credibility crisis

NAS Irreproducibility Report

published April 17, 2018 by the NAS — National Association of Scholars



it's complicated. . .

Peng (2011) *Science* 334(6060): 1226–1227

Reproducible research in computational science

reproducibility spectrum

it's complicated. . .

Peng (2011) *Science* 334(6060): 1226–1227

Reproducible research in computational science

reproducibility spectrum

reproducibility *identical* results with same data

it's complicated. . .

Peng (2011) *Science* 334(6060): 1226–1227

Reproducible research in computational science

reproducibility spectrum

reproducibility *identical* results with same data

replicability *consistent* results with new data

it's complicated. . .

Goodman et al. (2016) *Science Translational Medicine* 8(341): 341ps12

What does research reproducibility mean?

a conceptual framework

it's complicated. . .

Goodman et al. (2016) *Science Translational Medicine* 8(341): 341ps12

What does research reproducibility mean?

a conceptual framework

1. methods reproducibility:
enough detail about study procedures and data

it's complicated. . .

Goodman et al. (2016) *Science Translational Medicine* 8(341): 341ps12

What does research reproducibility mean?

a conceptual framework

1. methods reproducibility:
enough detail about study procedures and data
2. results reproducibility:
same results from closely matched independent study

it's complicated. . .

Goodman et al. (2016) *Science Translational Medicine* 8(341): 341ps12

What does research reproducibility mean?

a conceptual framework

1. methods reproducibility:
enough detail about study procedures and data
2. results reproducibility:
same results from closely matched independent study
3. inferential reproducibility:
qualitatively similar conclusions from reanalysis or replication

it's complicated. . .

Stark (2018) *Nature* 557: 613

Before reproducibility must come preproducibility

“preproducibility”

it's complicated. . .

Stark (2018) *Nature* 557: 613

Before reproducibility must come preproducibility

“preproducibility”

a “scientific recipe” for others to repeat the experiment or analysis

it's complicated. . .

Stark (2018) *Nature* 557: 613

Before reproducibility must come preproducibility

“preproducibility”

a “scientific recipe” for others to repeat the experiment or analysis

*“An experiment or analysis is preproducible
if it has been described in adequate detail
for others to undertake it.*

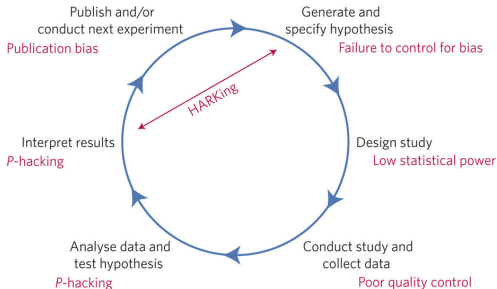
*Preproducibility is a prerequisite for reproducibility,
and the idea makes sense across disciplines.”*

points of view

Munafò et al. (2017) *Nature Human Behaviour* 1: 0021

A manifesto for reproducible science

e.g. the view from psychology...



hypothetico-deductive model
of the scientific method
and **potential threats**

points of view

Redish et al. (2018) *PNAS* 115(20): 5042–5046

Reproducibility failures are essential to scientific inquiry

... vs. the view from maths, computer science

- ▶ e.g. no statistical problems, no issues with experimental design
- ▶ failures to replicate essential to integration of conflicting observations and ideas into coherent theory

points of view

Redish et al. (2018) *PNAS* 115(20): 5042–5046

Reproducibility failures are essential to scientific inquiry

... vs. the view from maths, computer science

- ▶ e.g. no statistical problems, no issues with experimental design
- ▶ failures to replicate essential to integration of conflicting observations and ideas into coherent theory

“The discovery that an experiment does not replicate is not a lack of success but an opportunity.

... A failure to reproduce is only the first step in scientific inquiry. In many ways, how science responds to these failures is what determines whether it succeeds.”

research reproducibility

the problem, loosely defined

results can only be trusted if they can be *re-derived*
by the original researchers and/or by others working independently

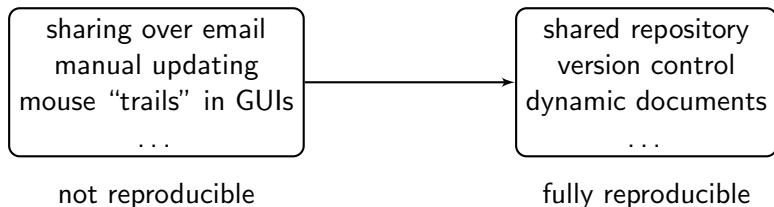
research reproducibility

the problem, loosely defined

results can only be trusted if they can be *re-derived*
by the original researchers and/or by others working independently

(part of) the solution

change in approach to computing in research



e.g. data management malfunction

Brown et al. (2005) *Nature* 438: 1148–1150

Dance reveals symmetry especially in young men



e.g. data management malfunction

Brown et al. (2005) *Nature* 438: 1148–1150

Dance reveals symmetry especially in young men

Reich (2013) *Nature* 497: 170–171

Symmetry study deemed a fraud



“... Brown says that it is unclear which data set is the original because many versions exist.”

e.g. data analysis malfunction

Reinhart and Rogoff (2010)

American Economic Review 100(2): 573–578

Growth in a time of debt

e.g. data analysis malfunction

	B	C	I	J	K	L	M
2			Real GDP growth				
3			(Index/GDP)				
4	Country	Coverage	30 or less	30 to 60	60 to 90	90 or above	30 or less
26			1.6	0.3	1.3	1.8	0.8
27	Minimum		5.4	4.9	10.2	3.6	13.8
28	Maximum						
29							
30	US	1946-2009	n.a.	3.4	3.3	-2.0	n.a.
31	UK	1946-2009	n.a.	2.4	2.5	2.4	n.a.
32	Sweden	1946-2009	3.6	2.9	2.7	n.a.	6.3
33	Spain	1946-2009	1.5	3.4	4.2	n.a.	9.8
34	Portugal	1952-2009	4.8	2.5	0.3	n.a.	7.9
35	New Zealand	1948-2009	2.5	2.9	3.9	-7.9	7.8
36	Netherlands	1956-2009	4.1	2.7	3.1	n.a.	6.4
37	Norway	1947-2009	3.4	5.1	n.a.	n.a.	5.4
38	Japan	1946-2009	7.0	4.8	1.0	0.7	7.0
39	Italy	1951-2009	5.4	2.1	1.8	1.8	5.6
40	Ireland	1948-2009	4.4	4.5	4.0	2.4	2.8
41	Greece	1970-2009	4.0	0.5	2.7	2.6	13.5
42	Germany	1946-2009	3.9	0.9	n.a.	n.a.	3.2
43	France	1949-2009	4.9	2.7	3.0	n.a.	5.2
44	Finland	1946-2009	3.8	2.4	5.5	n.a.	7.8
45	Denmark	1950-2009	3.5	1.7	2.4	n.a.	5.6
46	Canada	1951-2009	1.9	3.6	4.1	n.a.	2.2
47	Belgium	1947-2009	n.a.	4.2	3.1	2.6	n.a.
48	Austria	1948-2009	5.2	3.3	-0.8	n.a.	5.7
49	Australia	1951-2009	3.2	4.9	4.0	n.a.	5.8
50							
51			4.1	2.8	2.8	DAVERAGE	1.01148

Reinhart and Rogoff (2010)

American Economic Review 100(2): 573–578

Growth in a time of debt

Herndon et al. (2014)

Cambridge Journal of Economics 38(2): 257–279

Does high public debt consistently stifle economic growth?

A critique of Reinhart and Rogoff

e.g. software malfunction

[https://support.microsoft.com/en-us/help/214326/](https://support.microsoft.com/en-us/help/214326/excel-incorrectly-assumes-that-the-year-1900-is-a-leap-year)

[excel-incorrectly-assumes-that-the-year-1900-is-a-leap-year](https://support.microsoft.com/en-us/help/214326/excel-incorrectly-assumes-that-the-year-1900-is-a-leap-year)

The screenshot shows a web browser window with the URL <https://support.microsoft.com/en-us/help/214326/excel-incorrectly-assumes-that-the-year-1900-is-a-leap-year>. The page features the Microsoft logo and navigation links for Office, Windows, Surface, Xbox, Deals, Support, and More. A blue header bar contains the text "Microsoft Support". The main content area has the title "Excel incorrectly assumes that the year 1900 is a leap year" in large, bold black font. To the right of the title are links for "Email" and "Print". Below the title, the text reads: "Applies to: Microsoft Excel for Mac 2011, Excel for Mac for Office 365, Microsoft Office Excel 2003, Microsoft Office Excel 2007, Excel 2010, Excel 2013, Excel 2016, [Less](#)". A vertical "Site feedback" button is on the right. The "Symptoms" section follows, with the text: "Microsoft Excel incorrectly assumes that the year 1900 is a leap year. This article explains why the year 1900 is treated as a leap year, and outlines the behaviors that may occur if this specific issue is corrected." At the bottom, there is a "More Information" section and a blue button labeled "Try the Virtual Agent" with a small avatar icon.

Excel incorrectly assumes that the year 1900 is a leap year - Mozilla Firefox

Excel incorrectly assum... X

<https://support.microsoft.com/en-us/help/214326/excel-incorrectly-assumes-that-the-year-1900-is-a-leap-year>

Microsoft Office Windows Surface Xbox Deals Support More

Microsoft Support

Excel incorrectly assumes that the year 1900 is a leap year

Email Print

Applies to: Microsoft Excel for Mac 2011, Excel for Mac for Office 365, Microsoft Office Excel 2003, Microsoft Office Excel 2007, Excel 2010, Excel 2013, Excel 2016, [Less](#)

Symptoms

Microsoft Excel incorrectly assumes that the year 1900 is a leap year. This article explains why the year 1900 is treated as a leap year, and outlines the behaviors that may occur if this specific issue is corrected.

More Information

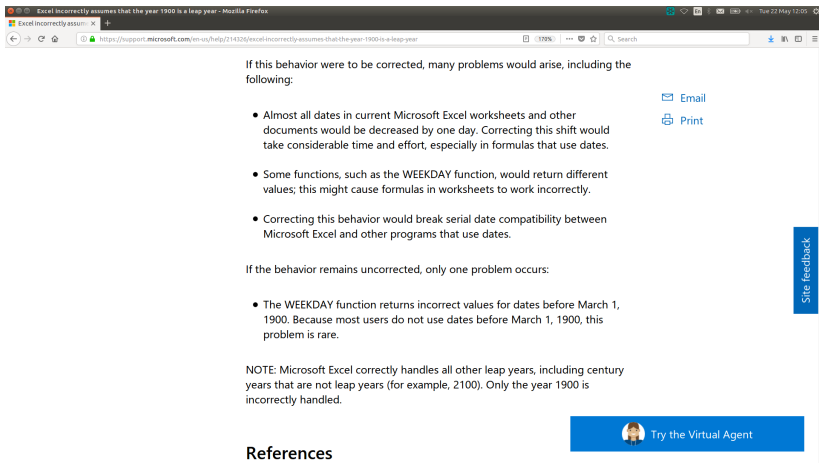
Try the Virtual Agent

Site feedback

e.g. software malfunction

[https://support.microsoft.com/en-us/help/214326/](https://support.microsoft.com/en-us/help/214326/excel-incorrectly-assumes-that-the-year-1900-is-a-leap-year)

[excel-incorrectly-assumes-that-the-year-1900-is-a-leap-year](https://support.microsoft.com/en-us/help/214326/excel-incorrectly-assumes-that-the-year-1900-is-a-leap-year)



The screenshot shows a web browser window with the title 'Excel incorrectly assumes that the year 1900 is a leap year - Mozilla Firefox'. The address bar shows the URL 'https://support.microsoft.com/en-us/help/214326/excel-incorrectly-assumes-that-the-year-1900-is-a-leap-year'. The page content includes a paragraph about the consequences of correcting the 1900 leap year assumption, followed by a bulleted list of three points. To the right of the list are links for 'Email' and 'Print'. Below the list is another paragraph about the WEEKDAY function. At the bottom, a 'NOTE' explains that Excel correctly handles other leap years. A 'References' section is at the bottom left, and a 'Try the Virtual Agent' button is at the bottom right. A vertical 'Site feedback' button is on the far right edge.

If this behavior were to be corrected, many problems would arise, including the following:

- Almost all dates in current Microsoft Excel worksheets and other documents would be decreased by one day. Correcting this shift would take considerable time and effort, especially in formulas that use dates.
- Some functions, such as the WEEKDAY function, would return different values; this might cause formulas in worksheets to work incorrectly.
- Correcting this behavior would break serial date compatibility between Microsoft Excel and other programs that use dates.

If the behavior remains uncorrected, only one problem occurs:

- The WEEKDAY function returns incorrect values for dates before March 1, 1900. Because most users do not use dates before March 1, 1900, this problem is rare.

NOTE: Microsoft Excel correctly handles all other leap years, including century years that are not leap years (for example, 2100). Only the year 1900 is incorrectly handled.

References

Try the Virtual Agent

Site feedback

a slogan

Buckheit and Donoho (1995) WaveLab and reproducible research
in Antoniadis and Oppenheim (eds.) *Wavelets and statistics*, Springer

*An article about computational science
in a scientific publication is **not** the scholarship itself,
it is merely **advertising** of the scholarship.
The actual scholarship is the complete
software development environment and the complete
set of instructions which generated the figures.*

another slogan

Stark (2018) *Nature* 557: 613

Before reproducibility must come preproducibility

*Science should be 'show me',
not 'trust me';
it should be 'help me if you can',
not 'catch me if you can'.*

Reproducible Research Oxford

 @RR_Oxford
 <http://rroxford.github.io/>



The Carpentries

 @thecarpentries
 <http://carpentries.org/>



UK Reproducibility Network

 @ukrepro
 <http://ukrn.org/>



Project TIER

 @Project_TIER
 <https://www.projecttier.org/>



Software Freedom Conservancy

 @conservancy
 <https://sfconservancy.org/>

