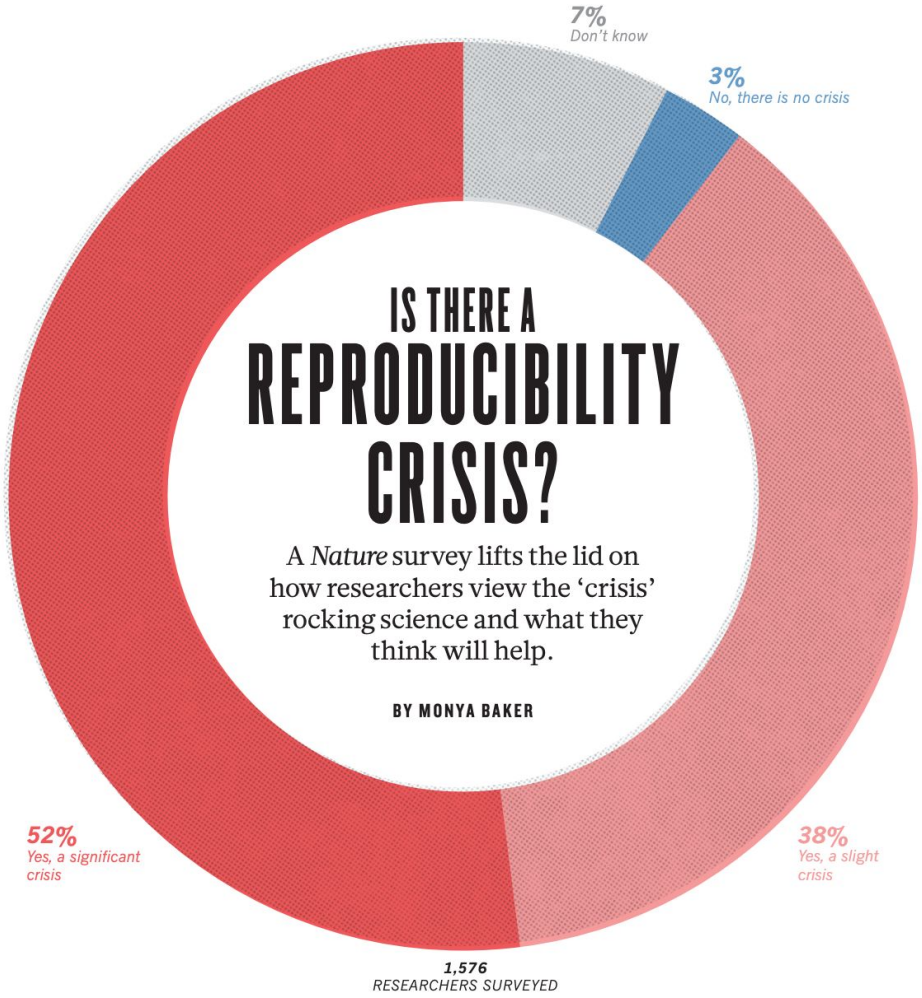
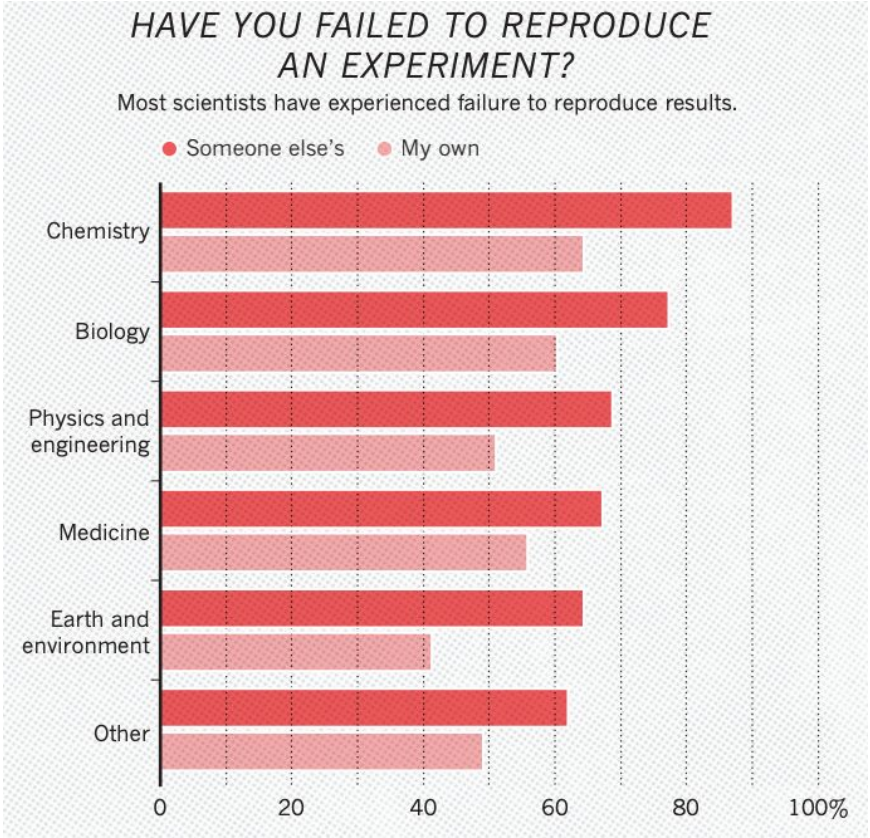
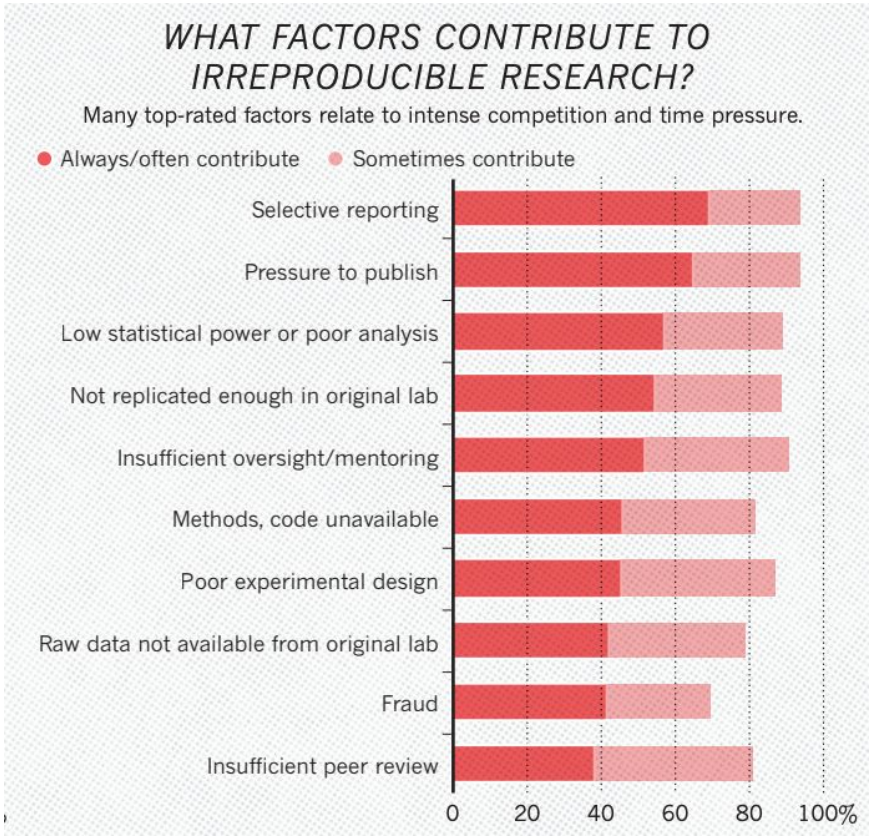
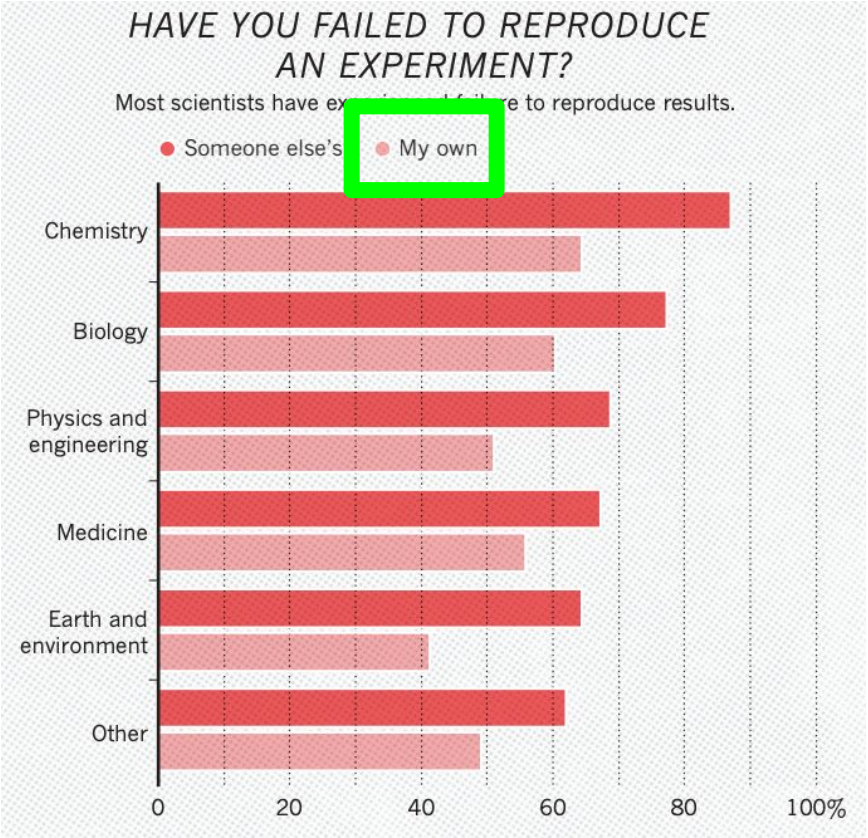
Code Review

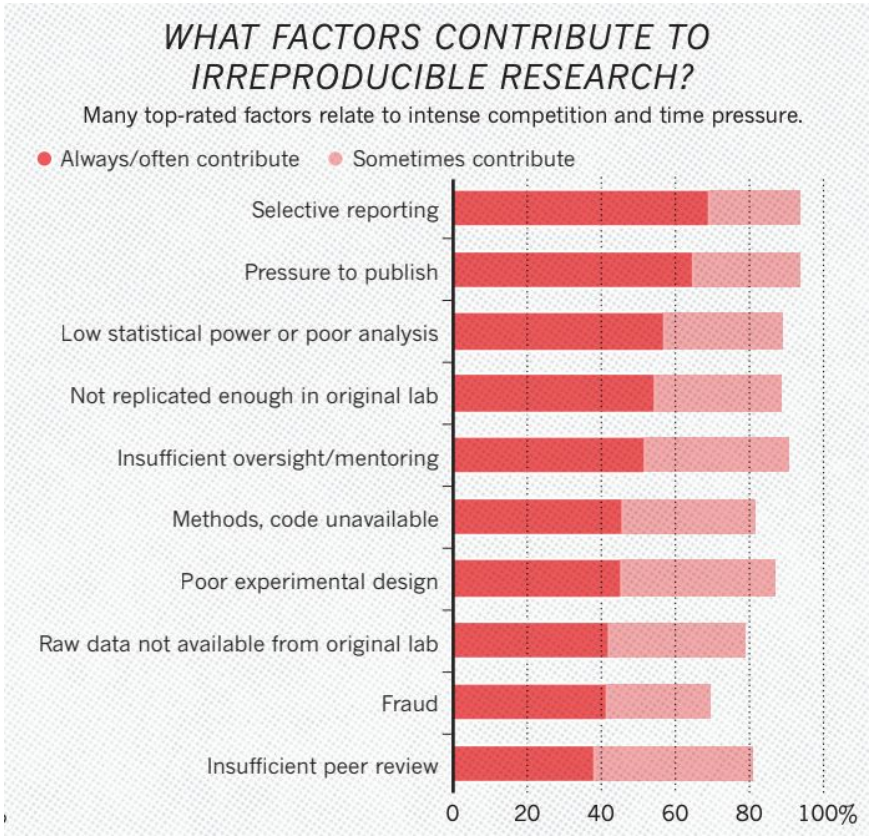
Software Engineering for Scientists

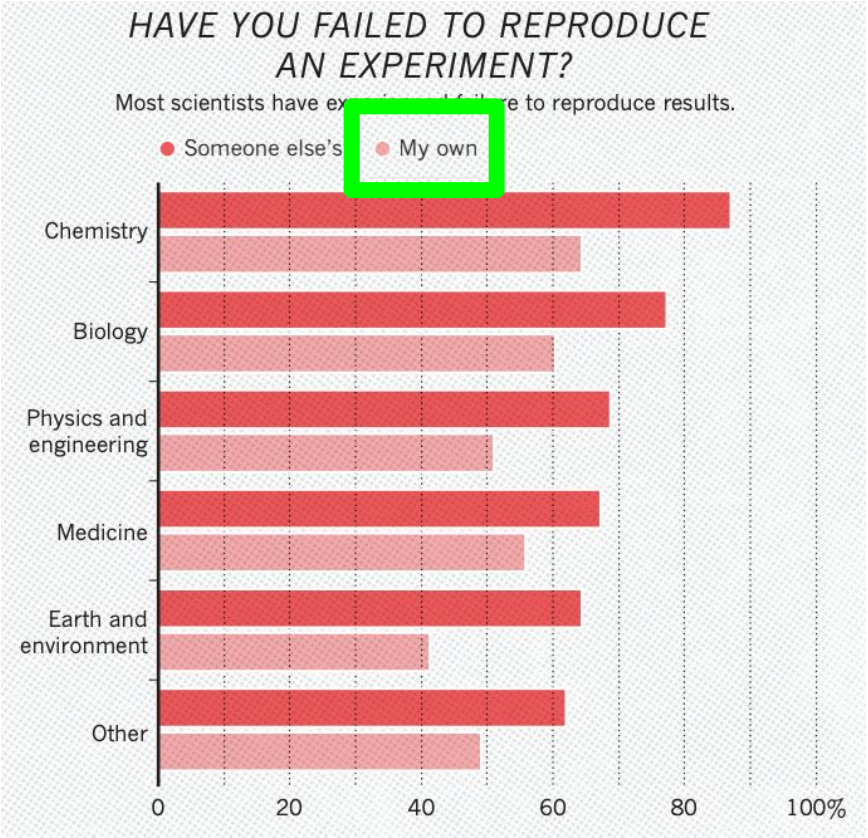

7%
Don't know
3%
No, there is no crisis
IS THERE A
REPRODUCIBILITY
CRISIS?
A *Nature* survey lifts the lid on how researchers view the 'crisis' rocking science and what they think will help.
BY MONYA BAKER
52%
Yes, a significant crisis
38%
Yes, a slight crisis
1,576
RESEARCHERS SURVEYED

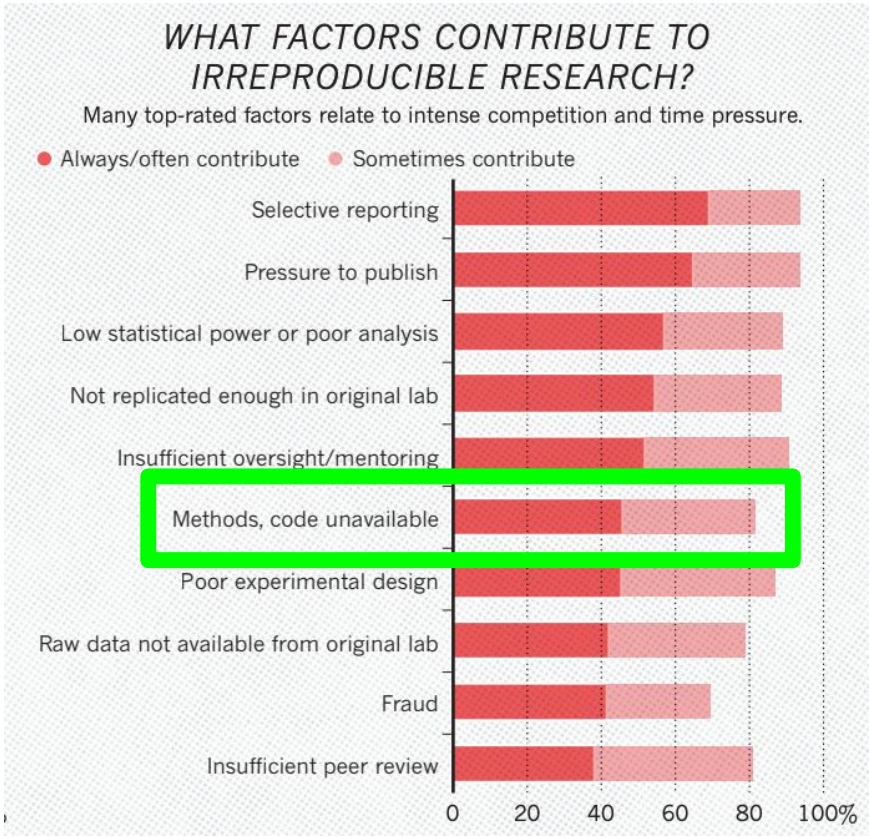

HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?
Most scientists have experienced failure to reproduce results.
Someone else's
My own
Chemistry
Biology
Physics and engineering
Medicine
Earth and environment
Other
0
20
40
60
80
100%


WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?
Many top-rated factors relate to intense competition and time pressure.
Always/often contribute
Sometimes contribute
Selective reporting
Pressure to publish
Low statistical power or poor analysis
Not replicated enough in original lab
Insufficient oversight/mentoring
Methods, code unavailable
Poor experimental design
Raw data not available from original lab
Fraud
Insufficient peer review
0
20
40
60
80
100%

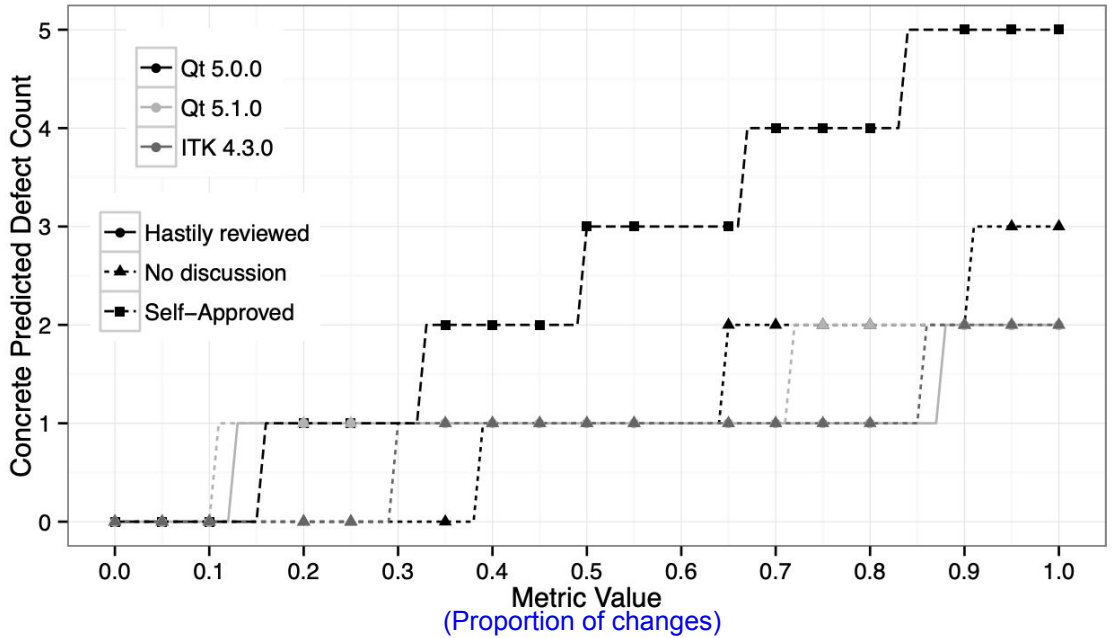

HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?
Most scientists have ex to reproduce results.
Someone else's
My own
Chemistry
Biology
Physics and engineering
Medicine
Earth and environment
Other
0
20
40
60
80
100%


WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?
Many top-rated factors relate to intense competition and time pressure.
Always/often contribute
Sometimes contribute
Selective reporting
Pressure to publish
Low statistical power or poor analysis
Not replicated enough in original lab
Insufficient oversight/mentoring
Methods, code unavailable
Poor experimental design
Raw data not available from original lab
Fraud
Insufficient peer review
0
20
40
60
80
100%


HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?
Most scientists have ex to reproduce results.
Someone else's
My own
Chemistry
Biology
Physics and engineering
Medicine
Earth and environment
Other
0
20
40
60
80
100%


WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?
Many top-rated factors relate to intense competition and time pressure.
Always/often contribute
Sometimes contribute
Selective reporting
Pressure to publish
Low statistical power or poor analysis
Not replicated enough in original lab
Insufficient oversight/mentoring
Methods, code unavailable
Poor experimental design
Raw data not available from original lab
Fraud
Insufficient peer review
0
20
40
60
80
100%

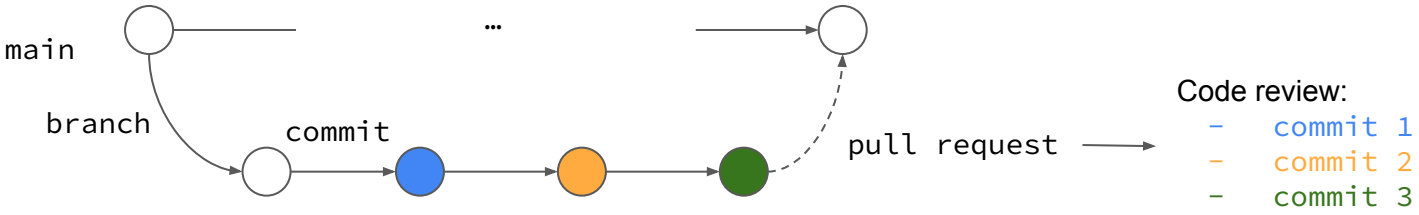
Discussion help us write better/more reproducible code



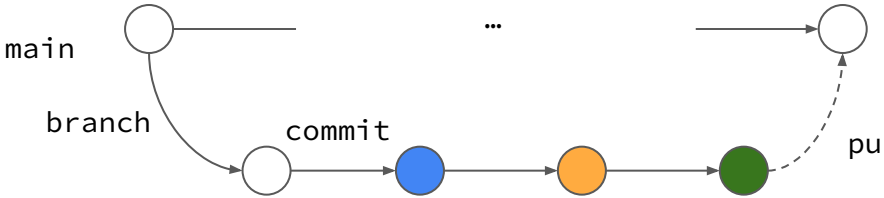
Lack of participation in code review has a negative impact on software quality. Reviews without discussion are associated with higher post-release defect counts, suggesting that the amount of discussion generated during review should be considered when making integration decisions.

McIntosh, 2014

Industry code review

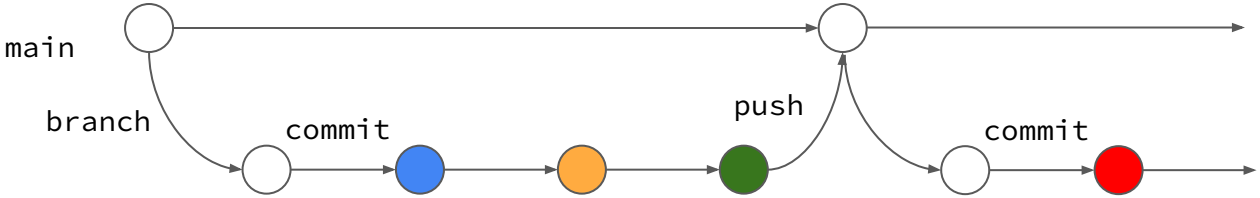


Industry code review



Code review:
- commit 1
- commit 2
- commit 3

Academic code review



Paper / Meeting Deadline

Code review:
- commit 1
- commit 2
- commit 3
- commit 4
- commit 5
- commit 6
- commit 7
- commit 8

Goals of code review

1. Improve the code
2. Improve the coder
3. Improve the reviewer

Pillars of good code

1. Safe from bugs
2. Easy to understand
3. Ready for change

What to look for

Don’t Repeat Yourself
Avoid magic numbers
No global variables

Comments where needed
One purpose for each function/variable
Return results, don’t print them

Fail fast
Use good names
Easy to test

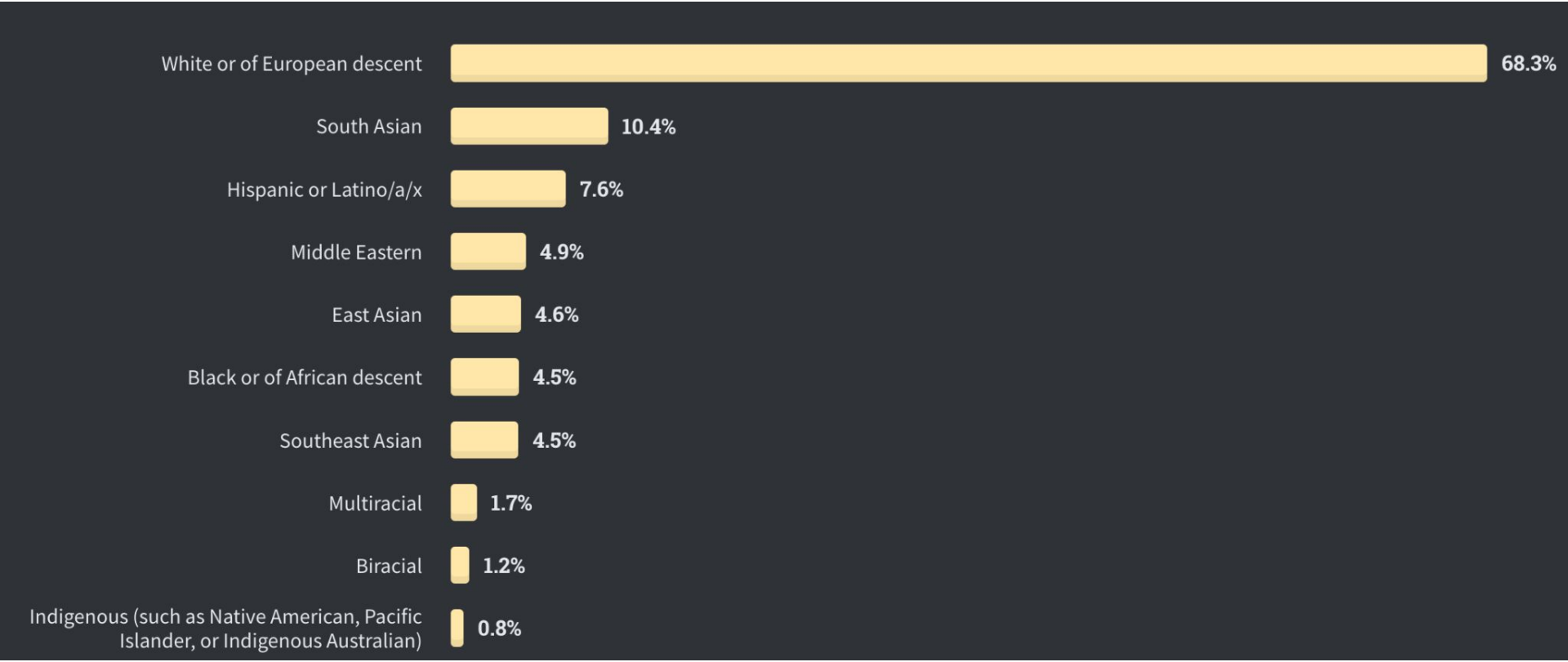
https://ocw.mit.edu/ans7870/6/6.005/s16/classes/04-code-review/

Great! We are all experts on code review

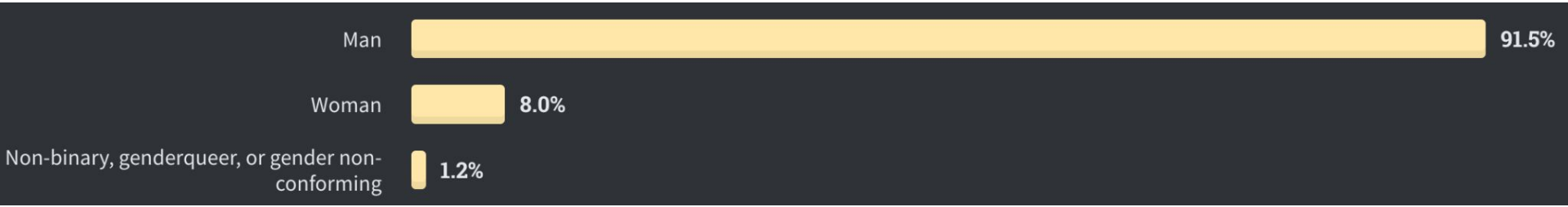
Great! We are all experts on code review

Or are we?

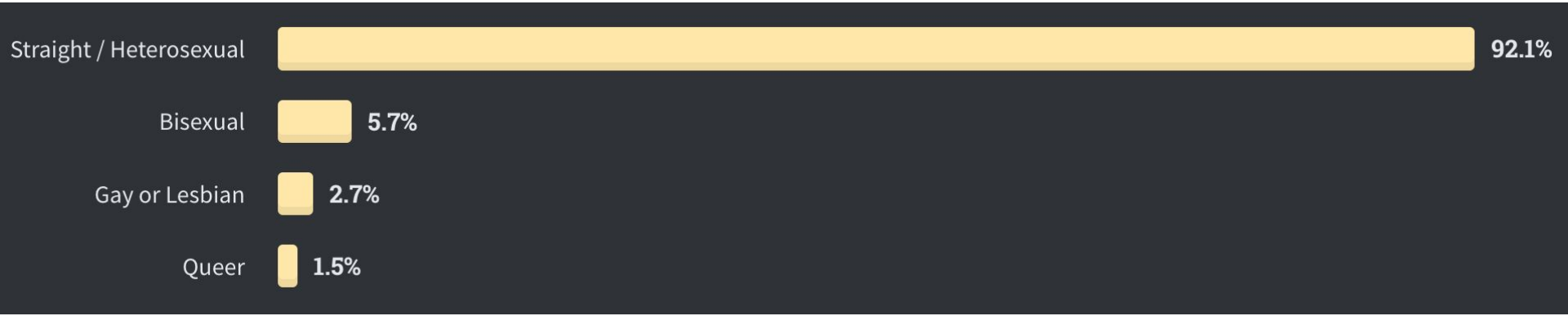
Computer Science is notorious for its equality and inclusivity problems



https://insights.stackoverflow.com/survey/2020



https://insights.stackoverflow.com/survey/2020



https://insights.stackoverflow.com/survey/2020

Storing sex (gender) in database

I want to store a users gender in a database with as little (size/performance) cost as possible.

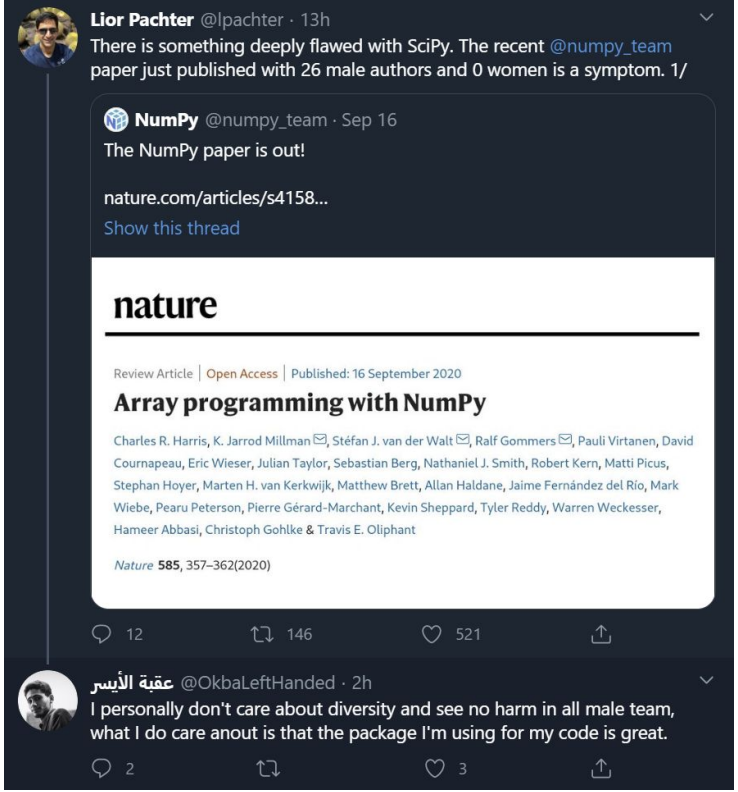
So far, 3 scenarios come to mind

1. **Int** - *aligned with Enum in code (1 = Male, 2 = Female, 3 = Hopefully no need for this?)*
2. **char(1)** - *Store **m** or **f***
3. **Bit** *(boolean) - and call the column isMale (sorry ladies :p)?*

No sexist offense intended with option 3 :-)

Option 3 is your best bet, but not all DB engines have a "bit" type. If you don't have a bit, then TinyINT would be your best bet.

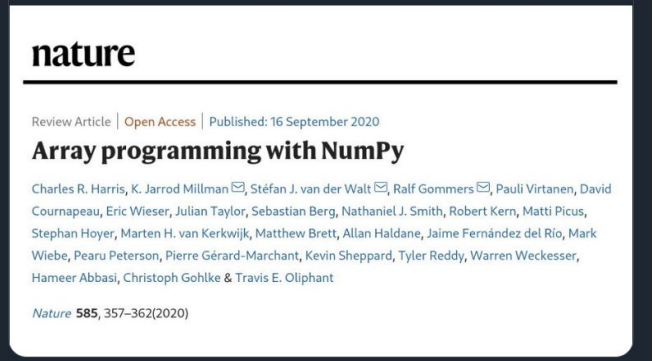

Lior Pachter @lpachter · 13h
There is something deeply flawed with SciPy. The recent @numpy_team paper just published with 26 male authors and 0 women is a symptom. 1/
NumPy @numpy_team · Sep 16
The NumPy paper is out!
nature.com/articles/s4158...
Show this thread
nature
Review Article | Open Access | Published: 16 September 2020
Array programming with NumPy
Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke & Travis E. Oliphant
Nature 585, 357–362(2020)
12
146
521


Lior Pachter @lpachter · 13h
There is something deeply flawed with SciPy. The recent @numpy_team paper just published with 26 male authors and 0 women is a symptom. 1/
NumPy @numpy_team · Sep 16
The NumPy paper is out!
nature.com/articles/s4158...
Show this thread
nature
Review Article | Open Access | Published: 16 September 2020
Array programming with NumPy
Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke & Travis E. Oliphant
Nature 585, 357–362(2020)
12
146
521
عقبة الأيسر @OkbaLeftHanded · 2h
I personally don't care about diversity and see no harm in all male team, what I do care anout is that the package I'm using for my code is great.
2
3

Lior Pachter @lpachter · 13h

There is something deeply flawed with SciPy. The recent @numpy_team paper just published with 26 male authors and 0 women is a symptom. 1/

> **NumPy** @numpy_team · Sep 16
>
> The NumPy paper is out!
>
> nature.com/articles/s4158...
>
> Show this thread



nature

Review Article | Open Access | Published: 16 September 2020

Array programming with NumPy

Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke & Travis E. Oliphant

Nature **585**, 357–362(2020)

12 146 521

عقبة الأيسر @OkbaLeftHanded · 2h

I personally don't care about diversity and see no harm in all male team, what I do care anout is that the package I'm using for my code is great.

2 3

Zbib of Quraish زبيب القرشي @zbiiib · 2h

Replying to @lpachter and @numpy_team

How the fuck a community driven open source project is supposed to include minorities? Drag them and force them to work on an open source project by force?
There's a big fucking button that says contribute in their website.
Stop the witch hunt and the virtue signalling. Get a nap

3


vegetable minister (abolish all police)
@no_reply
DataCamp quietly published a "community" post today, in which they admit an executive sexually assaulted ("uninvited physical contact") an employee.
datacamp.com/community/blog... #Python #rstats #datascience #MeToo
A note to our community
datacamp.com
3:56 PM · Apr 4, 2019
340
263 people are Tweeting about this


Sharla Gelfand
@sharlagelfand
these snake jokes write themselves right??
Anaconda @anacondainc
Anaconda is proud to partner with @DataCamp to bring interactive #datascience training to our customers. Read the full press release here: bit.ly/3edjOuE
We're proud to partner with DataCamp!
DataCamp
ANACONDA.
6:48 PM · Jul 1, 2020

Inclusive Code Review

https://microsoft.github.io/code-with-engineering-playbook/code-reviews/inclusion-in-code-review/
https://builtin.com/software-engineering-perspectives/code-review-etiquette

Beginning a code review with the right mindset is crucial to both giving and receiving feedback.
Getting critical feedback can be difficult even when the feedback is helpful and well intentioned.

For the Author

- **Be respectful of the reviewers time**
- Write code that is easy to read and review
- Describe your motivations and how your code achieves its goals
- Ask for targeted feedback
- Respond clearly to questions asked by the reviewers
- Acknowledge the reviewer's effort

For the the Reviewer

- **Assume positive intent from the author.**
- Write clear and elaborate comments.
- Be clear about what is an opinion, personal preference, best practice, and fact. Personal preferences and opinions should be avoided.
- If you do not understand the code properly, communicate with the author to get a basic understanding of their work.
- **Be suggestive and not prescriptive.** A reviewer should suggest changes and not prescribe changes, let the author decide if they really want to accept the changes proposed.
- Comments don't all need to be critical
- Avoid nitpicking

Culture and Code Reviews

Culture and communication style of a particular geography also influences how people interact. Assuming positive intent of the author and reviewer is a good start to start analyzing quality of code reviews.

Impostor Syndrome

Anyone can have valuable insights.

A fresh new pair of eyes are always welcome.

Study the review until you have clearly understood it, check the corner cases and look for ways to improve it.

If something is not clear, a simple specific question should be asked.

If you learned something, you can always compliment the author.