Knowledge and solutions
for a changing world

Be boundless

Advancing data-intensive
discovery in all fields

Reproducibility with Dave

David A. C. Beck

Chemical Engineering & eScience Institute

Facets of reproducibility

- Technical reproducibility
  - Run code with same data and get same result
- Algorithmic reproducibility
  - Re-implement algorithm according to description
  - Run with same data
  - Get same result (or is a very similar result acceptable?)
- Study replicability
  - Run code or run re-implementation
  - Using statistically similar but different data
  - Get similar result

Terminology

- Reviewable research
  - The descriptions of the research methods can be independently assessed and the results judged credible.
- Replicable research
  - Tools are made available that would allow one to duplicate the results of the research.
- Confirmable research
  - The main conclusions of the research can be attained independently without the use of physical materials provided by the author.

Stodden *et al.* Setting the default to reproducible. ICERM report, 2015

Terminology

- Auditable research
  - Sufficient records (including data and software) have been archived so that the research can be defended later if necessary or differences between independent confirmations resolved. The archive might be private, as with traditional laboratory notebooks.
- Reproducible research
  - Well-documented and materials and data that are available that would allow one to (a) fully audit the procedure, (b) replicate and also independently reproduce the results of the research, and (c) extend the results or apply the method to new problems.

Stodden *et al.* Setting the default to reproducible. ICERM report, 2015

Epic fail parade

Epic fail

- In 2011, Bayer (pharmaceuticals) tried to replicate 67 important papers
  - Oncology
  - Women's health
  - Cardiovascular medicine

Only about 21% were reproducible

Begley, C. G.; Ellis, L. M. (2012). "Drug development: Raise standards for preclinical cancer research". *Nature* **483** (7391): 531–533.

Epic fail, part 2

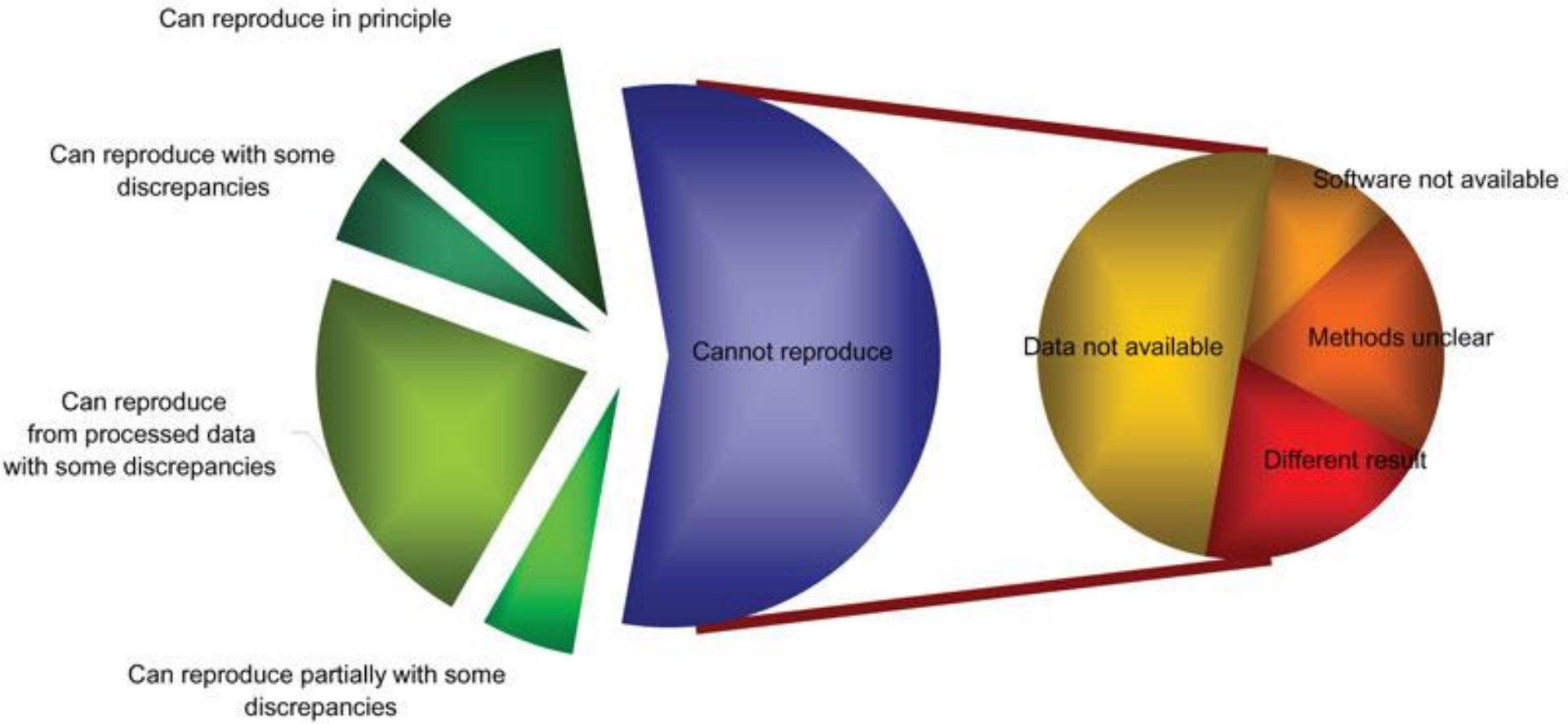
- In 2012, Amgen published a report in Nature
  - Examined 53 landmark studies in cancer

6 of 53 (11%) were reproducible

Begley, C. G.; Ellis, L. M. (2012). "Drug development: Raise standards for preclinical cancer research". *Nature* **483** (7391): 531–533.

Epic fail, part 3

Attempt to reproduce 18 tables and figures papers published in Nature Genetics using microarrays



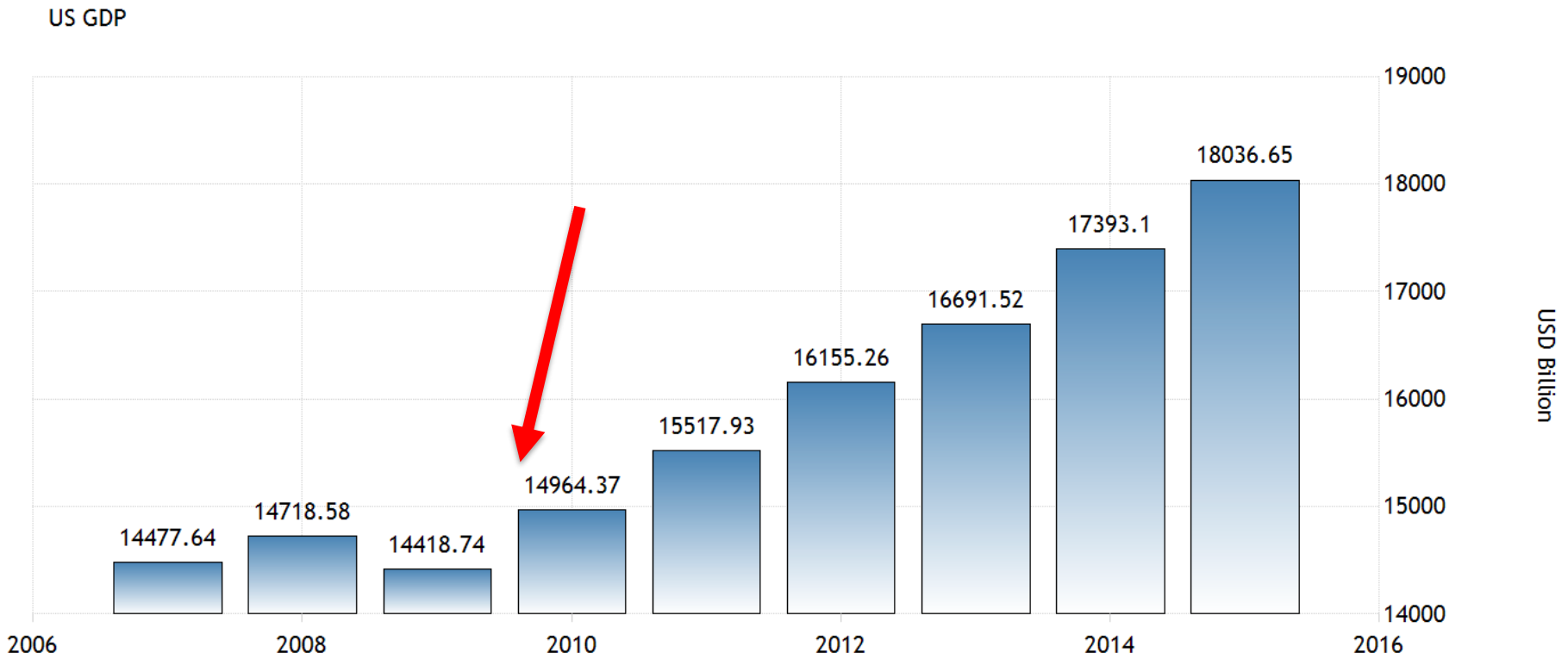
Ionnidis, P. et al. *Repeatability of published microarray gene expression analyses. Nat Gen , 41:2, Feb 2009*

Epic fails in medicine

- What are the repercussions of irreproducible results in medicine?
  - Biotech companies
  - Government
  - People?

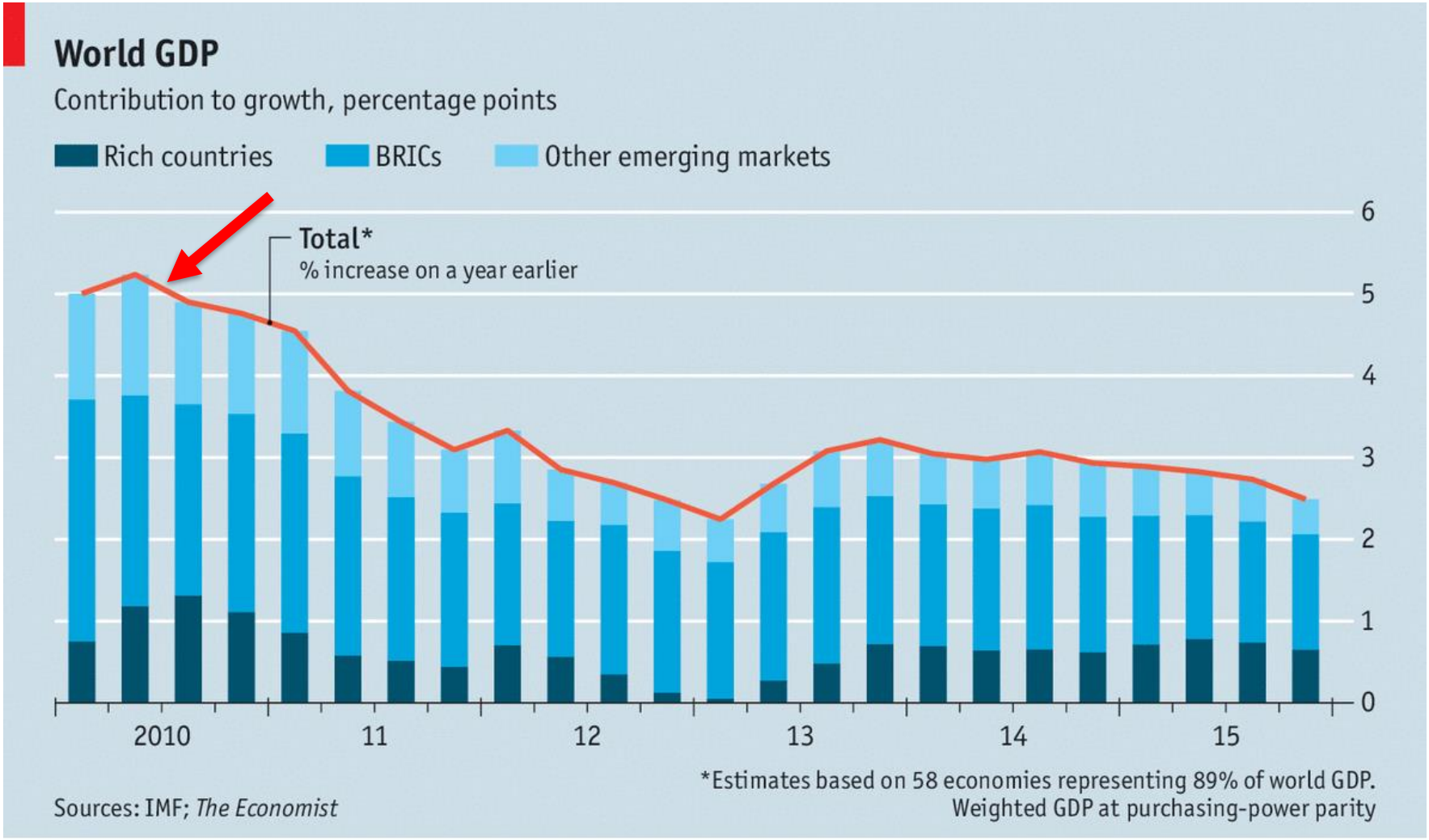
Epic fail, global impact

- Grab your way-back hat and put it on!



SOURCE: WWW.TRADINGECONOMICS.COM | WORLD BANK

Epic fail, global impact



Epic fail, global impact

- 2010 paper by Reinhart & Rogoff “Growth in a Time of Debt”
  - **…high debt/GDP levels (90 percent and above) are associated with notably lower growth outcomes.**
  - **Debt to GDP ratios over 90% have real GDP growth of -0.1%**
  - **Seldom do countries “grow” their way out of debts.**

Reinhart, Carmen M., and Kenneth S. Rogoff. 2010. "Growth in a Time of Debt." *American Economic Review*, 100(2): 573-78.

Epic fail, global impact

- Paper was widely cited by
  - Political parties
  - Governments
  - International lending agencies
- To show that **austerity** was the solution to the global recession
- Even part of the 2012 US presidential election!

Reinhart, Carmen M., and Kenneth S. Rogoff. 2010. "Growth in a Time of Debt." *American Economic Review*, 100(2): 573-78.

Epic fail, global impact

- UMass Amherst Graduate student Thomas Herndon
  - Tried to reproduce the results of the paper for a class: **couldn't**
  - Requested the 'code' for the computations from R&R: got an Excel spreadsheet
  - Found multiple errors

Reinhart, Carmen M., and Kenneth S. Rogoff. 2010. "Growth in a Time of Debt." *American Economic Review*, 100(2): 573-78.
Thomas Herndon, Michael Ash & Robert Pollin, Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff

Epic fail, global impact

- UMass Amherst Graduate student Thomas Herndon
  - Found multiple errors

Coding errors, selective exclusion of available data, and unconventional weighting of summary statistics lead to serious errors that inaccurately represent the relationship between public debt and GDP growth.

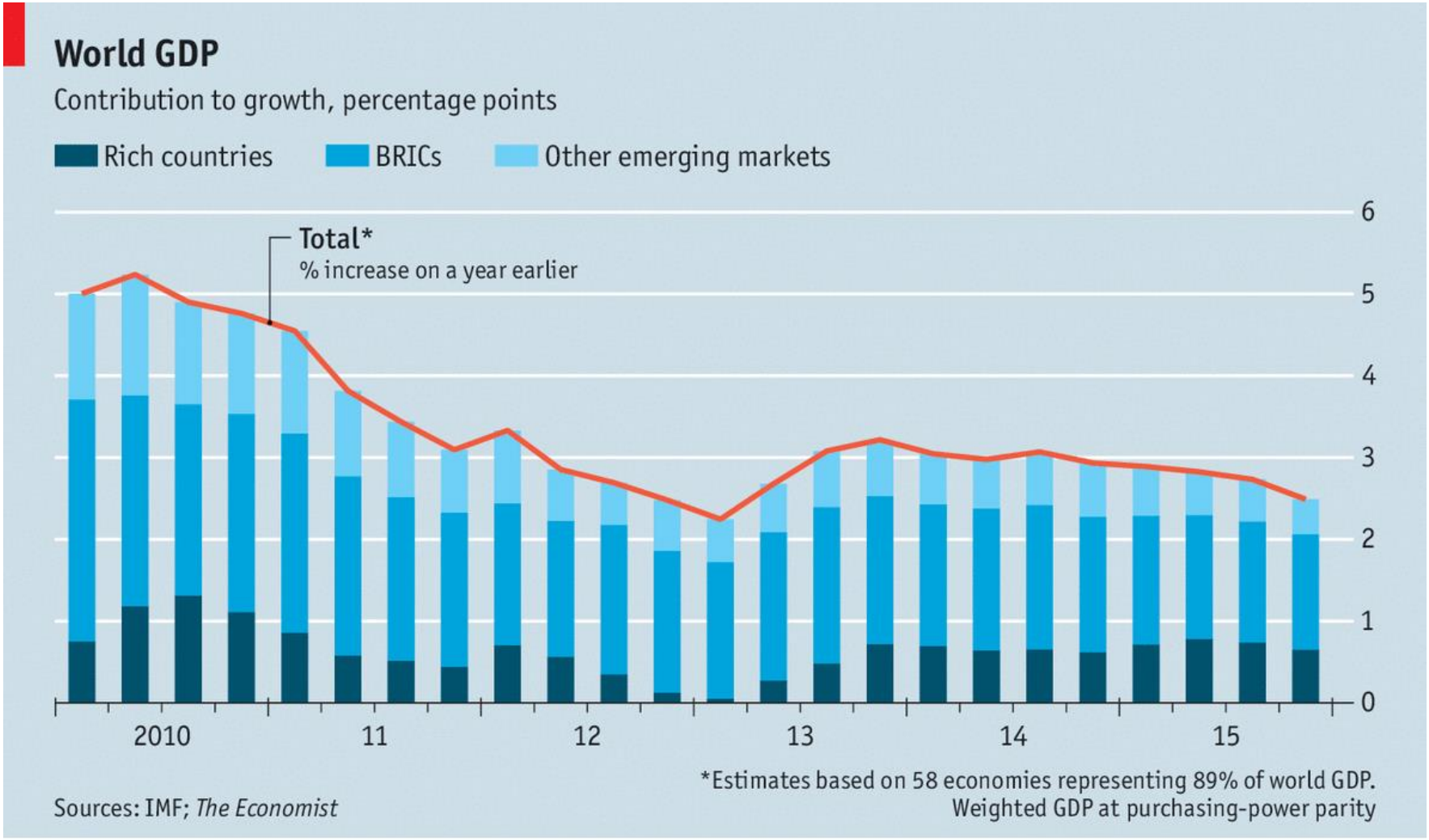
Reinhart, Carmen M., and Kenneth S. Rogoff. 2010. "Growth in a Time of Debt." *American Economic Review*, 100(2): 573-78.
Thomas Herndon, Michael Ash & Robert Pollin, Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff

Epic fail, global impact

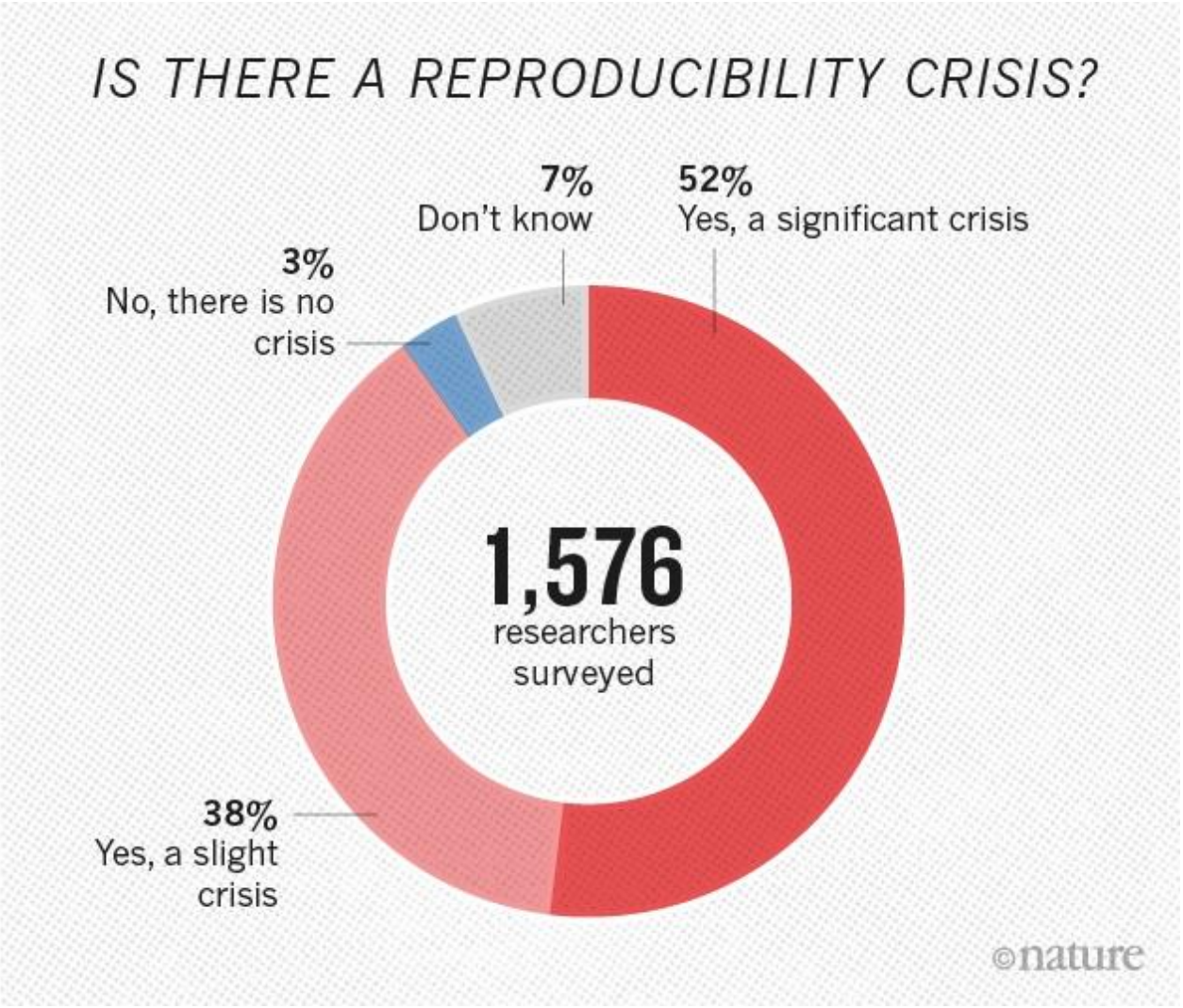
- Herndon fixed the errors and reexamined claims
- Original claims
  - Debt to GDP ratios over 90% have real GDP growth of **-0.1%**
  - In a recession: Austerity good, spending bad
- Modified claims
  - Debt to GDP ratios over 90% have real GDP growth of **2.2%**
  - In a recession: Spending good

Reinhart, Carmen M., and Kenneth S. Rogoff. 2010. "Growth in a Time of Debt." *American Economic Review*, 100(2): 573-78.
Thomas Herndon, Michael Ash & Robert Pollin, Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff

Epic fail, global impact



Science in crisis?



Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **533**, 452-454 (2016).

Why do we care?

"Non-reproducible single occurrences are of no significance to science."

– Karl Popper

- Faith in science and the scientific method
  - General population
  - Government budgeting perspective

Popper, K. R. 1959. The logic of scientific discovery. Hutchinson, London, United Kingdom.

Why to avoid epic failure

- To achieve greater scientific validity and integrity by making it easier to verify published results.
- To provide the evidence needed to evaluate the reliability of scientific claims.
- To increase productivity of current and future researchers on funded projects.
- To increase the impact of the research performed, software developed, and papers published.
- To help promote data and code as first class research products. (Get a job or get promoted!)
- To increase access to and usability of research products by other researchers.

Why to avoid epic failure

Publishing retractions doesn't contribute to your citation count

Why is this happening?

- Poor working knowledge of statistics
  - Underpowered experimental design
  - Misuse of statistical tests
  - Misinterpretation of statistical tests

  - Journal are getting better at recognizing these issues and requiring statistical review

How to avoid epic failure

- Use version control (lik
  - Detailed record of your
  - Bug tracking and identi
  - Treat code as a first class research product

Overheard at a previous GeoHackWeek:
My code was working yesterday but today it isn't. [unhappy face] I didn't commit the script when it was working.

- Write tests
  - Formal unit tests are best
    - For more info look at `pytest` http://doc.pytest.org/en/latest/
  - Example code that demonstrates use cases
    - Example data, ideally small
    - With easy to identify expected outcomes

How to avoid epic failure

- Documentation (Documentation! Documentation!)
  - Comments in code
    - Accurate description
    - Updated when code changes
    - Inaccurate comment is a bug
  - Accurate description of approach and functions
    - Users know what to expect
    - Users know how to properly use the tool
    - Users can find/meet/install dependencies (include version #s)

How to avoid epic failure

- Reproducible computing environments
  - Computing environment matters!
    - System libraries change
    - Math libraries change
    - E.g. square root of 4 is 2, but a computer knows it as
      - 1.9999999999999999999
      - 2.0000000000000000001
    - These differences compound and can yield different results!
  - Catalog the precise work environment
    - Operating system dependencies
    - Python dependencies
    - External dependencies, e.g. GEE version

How to avoid epic failure

- Reproducible computing environments
  - Use `docker` to create a static image of environment
  - Create an AWS machine image (AMI)

- Data!
  - Open data deposited in a third party site
  - Data metadata is clear, e.g. units of measurement
  - Clear description of provenance
  - DOI
  - Treat data as a first class research product

Further reading

- UW Reproducibility Working Group Guidelines
  - http://uwescience.github.io/reproducible/
- Ten Simple Rules for the Care and Feeding of Scientific Data
  - http://www.ploscompbiol.org/article/info:doi/10.1371/journal.pcbi.1003542
- Literate Programming Tools
  - http://www.literateprogramming.com/tools.html
- Detailed guide to using vagrant for science here:
  - http://hplgit.github.io/vagrantbox/doc/pub/._vagrant_box001.html
- Thoughts on metadata annotation & standards
  - https://github.com/mozillascience/code-research-object/issues/2#issuecomment-35610035

Discussion time

- Questions to seed discussion
  - What has been your previous experience with
    - other peoples code and data?
    - your previous code and data?
    - Do you have tips to offer others?
  - What specific actions has your team taken this week to ensure your code is reproducible?
  - What specific actions can your team take going forward?