Knowledge and solutions for a changing world

Be boundless

Advancing data-intensive discovery in all fields

Software Engineering for Data Scientists

Reproducible computations

David A. C. Beck (dacb)
Chemical Engineering & eScience Institute

Overview

- Terminology
- What are we guarding against?
- What are the tools we can use for defense?

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 software provided by the author.

Terminology

- Auditable research
  - The main conclusions of the research can be attained independently without the use of software provided by the author.
- Reproducible research
  - Well-documented and code and data that are available that would allow one to (a) fully audit the computational procedure, (b) replicate and also independently reproduce the results of the research, and (c) extend the results or apply the method to new problems.

Reproducibility

- Can an experimental result be reproduced?
- Reproducibility comes in different flavors
  - Same data, same analyses (Reproducible)
  - Similar data, same analyses (Replicability)
  - Same data, similar analyses (Robustness)
  - Others?

  - Today I'll use **Reproducibility** to cover all of these

Epic fail Schadenfreude parade*

*a feeling of joy that comes from seeing or hearing about another person's troubles or failures. – Wikipedia

I share this with you so you can be an advocate for these ideas in your team.

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

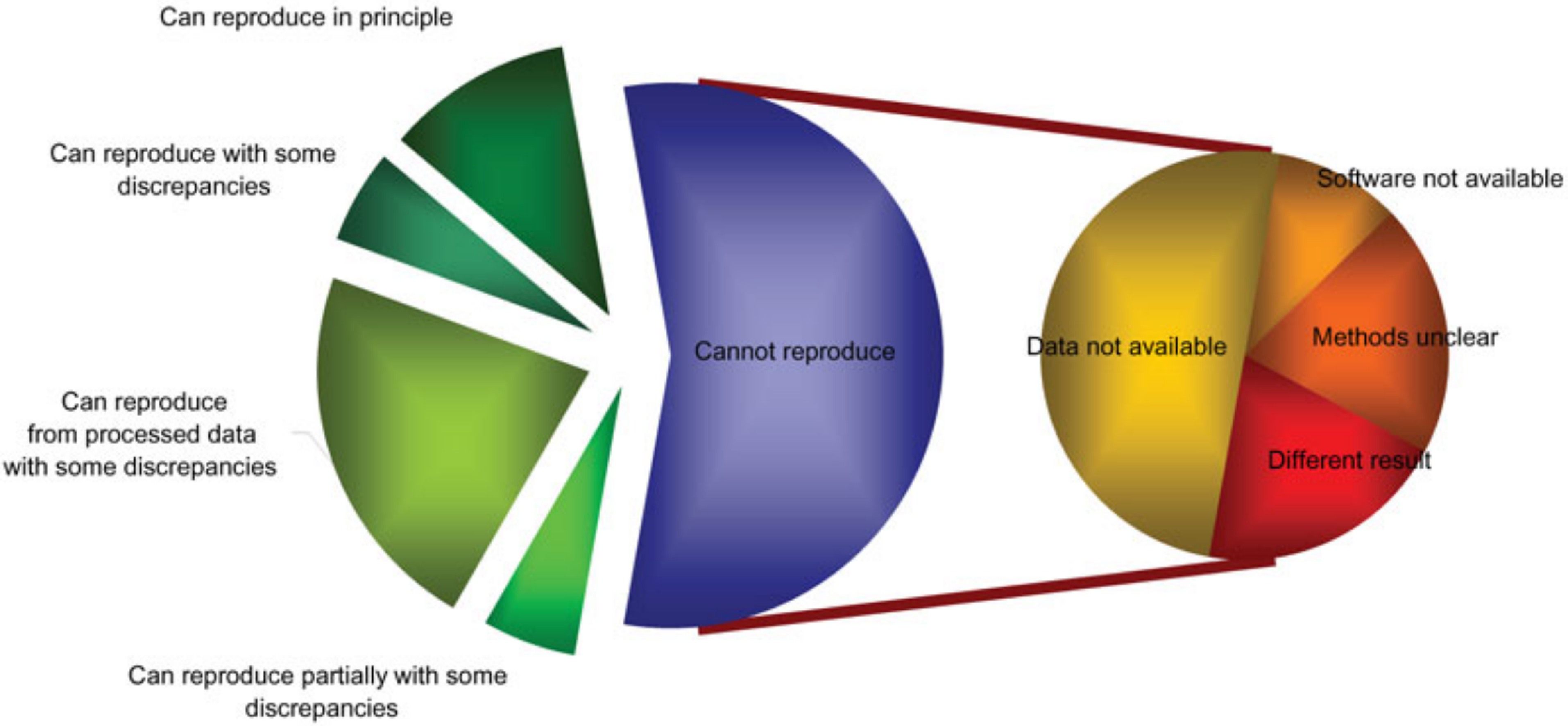
- In 2012, Amgen published a report in Nature
  - Examined 53 landmark studies in cancer

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



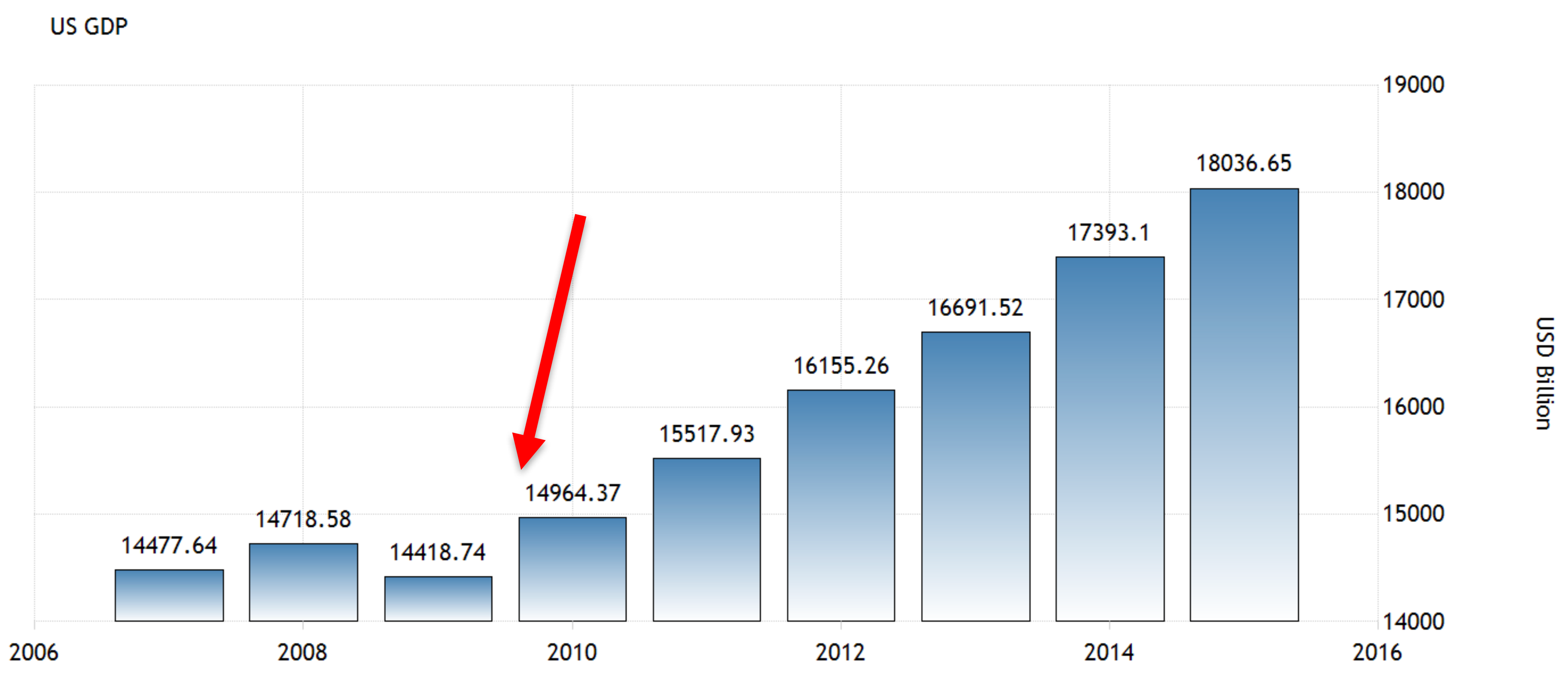
Ionnidis, P. et al. *Repeatability of published microarray gene expression analyses. Nat Gen , 41:2, Feb 2009*

Epic fails in medicine

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

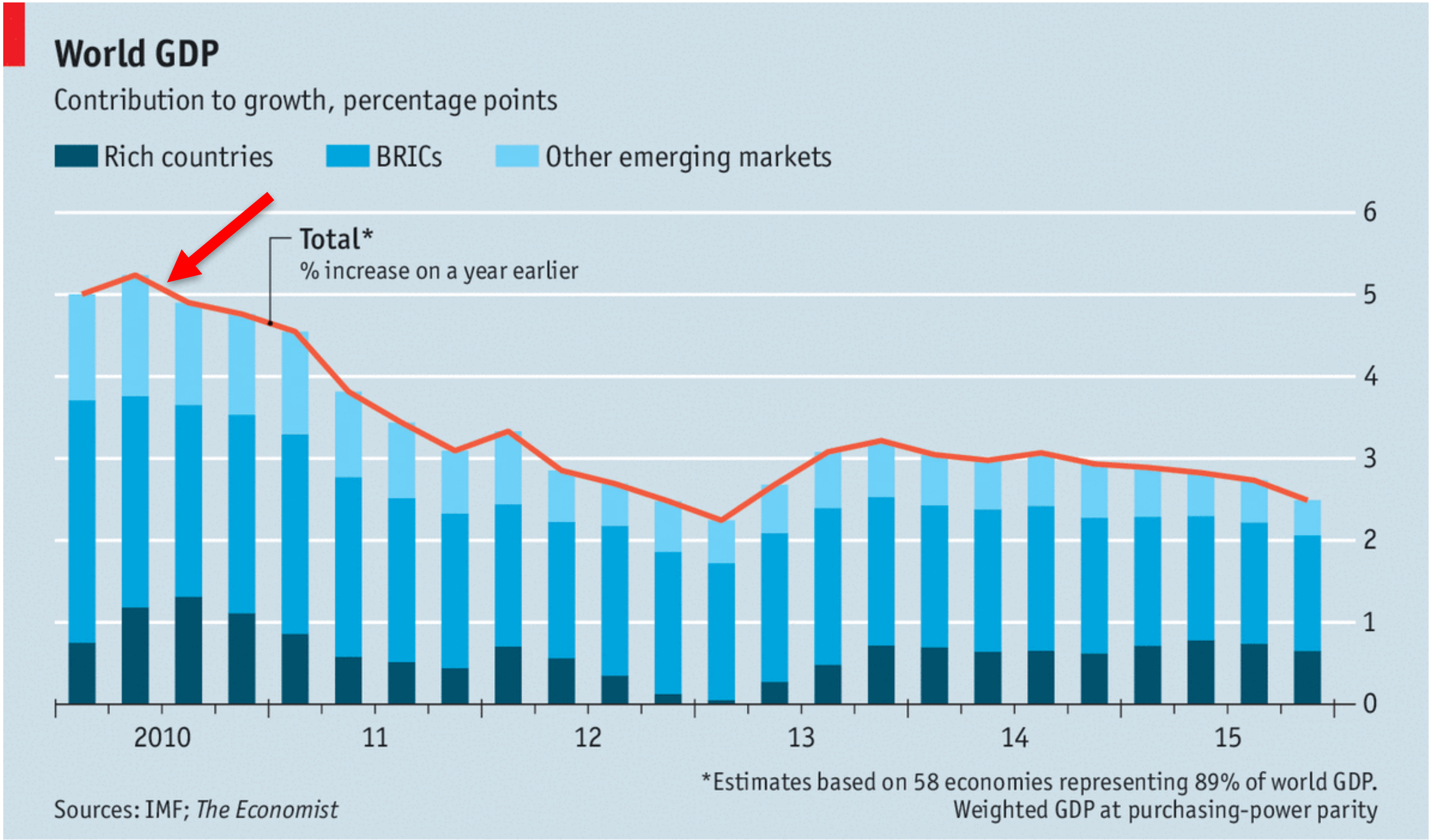
Epic fail, global impact

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



SOURCE: WWW.TRADINGECONOMICS.COM | WORLD BANK

Epic fail, global impact



World GDP

Contribution to growth, percentage points

*Estimates based on 58 economies representing 89% of world GDP. Weighted GDP at purchasing-power parity

Sources: IMF; *The Economist*

Economist.com

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 read 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.

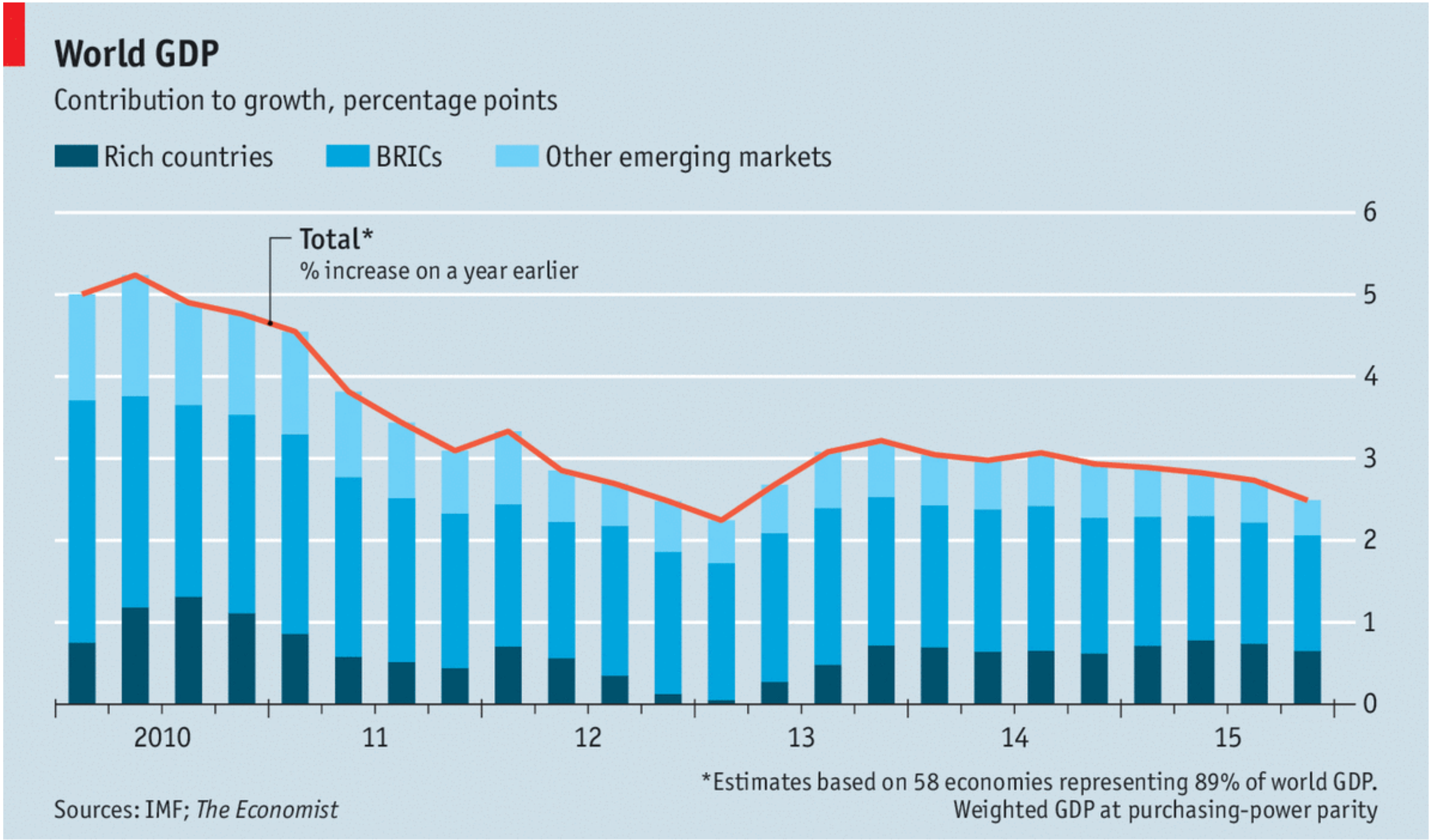
Reinhart, Carmen M., and Kenneth S. Rogoff. 2010. "Growth in a Time of Debt." *American Economic Review*, 100(2): 573-78.
Thomas Herndon, Michael Ash & Robert Pollin, Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff

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


World GDP

Contribution to growth, percentage points

*Estimates based on 58 economies representing 89% of world GDP. Weighted GDP at purchasing-power parity

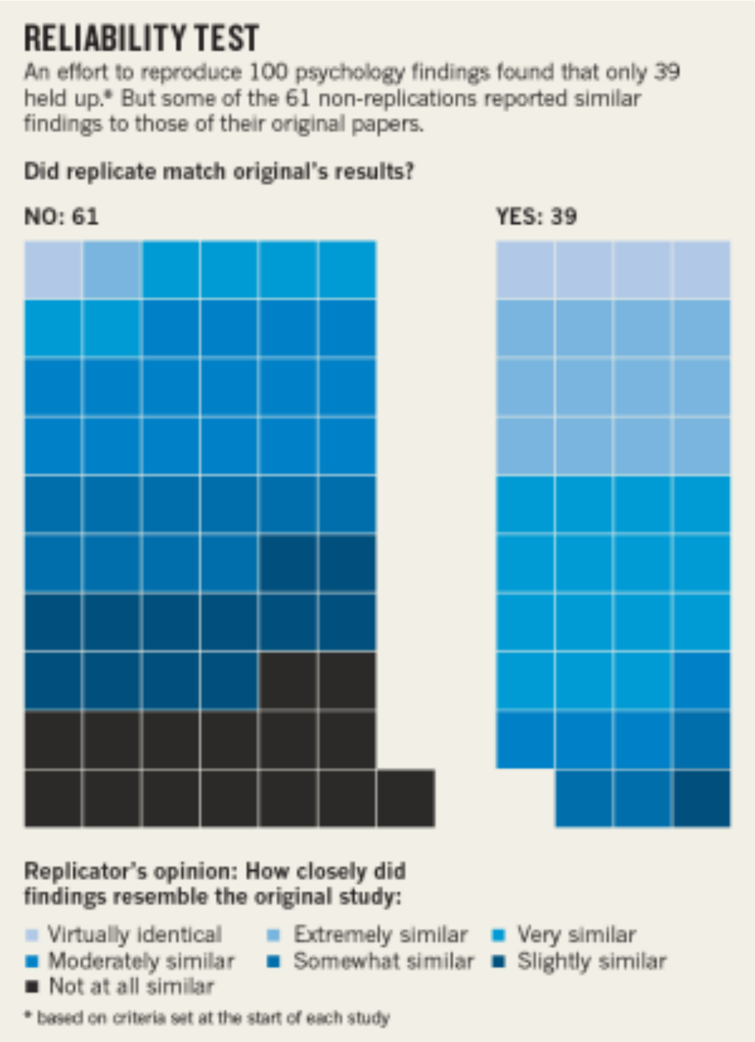
Sources: IMF; *The Economist*

Economist.com

Epic fail, global impact

- What effect did the incorrect R&R paper have?

Epic failure, part 4



RELIABILITY TEST

An effort to reproduce 100 psychology findings found that only 39 held up.* But some of the 61 non-replications reported similar findings to those of their original papers.

Did replicate match original's results?

Replicator's opinion: How closely did findings resemble the original study:

* based on criteria set at the start of each study

http://www.nature.com/news/over-half-of-psychology-studies-fail-reproducibility-test-1.18248

Why is this happening?

- Social factors, e.g.
  - Fraud, misconduct
  - Pressure to publish

Important but not Data Science related.
WE ARE WORKING ON THESE!

- *p*-hacking
- Poor experimental design
  - Small effect size
  - Small sample size
- **Data not available or disclosed**
- **Software not available or other software issue**

p-hacking

- Do a study to test some hypothesis
  - E.g. an apple a day keeps the Dr. away
- Use a *p*-value of 0.05
  - i.e. 5% chance of seeing a difference at least as big as we have, by chance alone
- Perform 1000s of statistical tests
- What happens?

Some significant results by chance alone

[1] Simmons, J.P., N.D. Nelson, and U. Simonsohn. 2011. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* 22(11):1359-1366.

Data

- Data availability
  - – Deposit data in a repository

- Versioning
  - – Explicit versioning in file names
  - – Record date retrieved in file names
  - – Compute and record md5sum of files (demo)

- Different data or new data

Data

- Why would more data change the result?
  - Use a *p*-value of 0.05
    - i.e. 5% chance of seeing a difference at least as big as we have, by chance alone
  - Was the effect size small?
  - Report more than the *p*-value
    - Sample size
    - Effect size

Why is this happening?

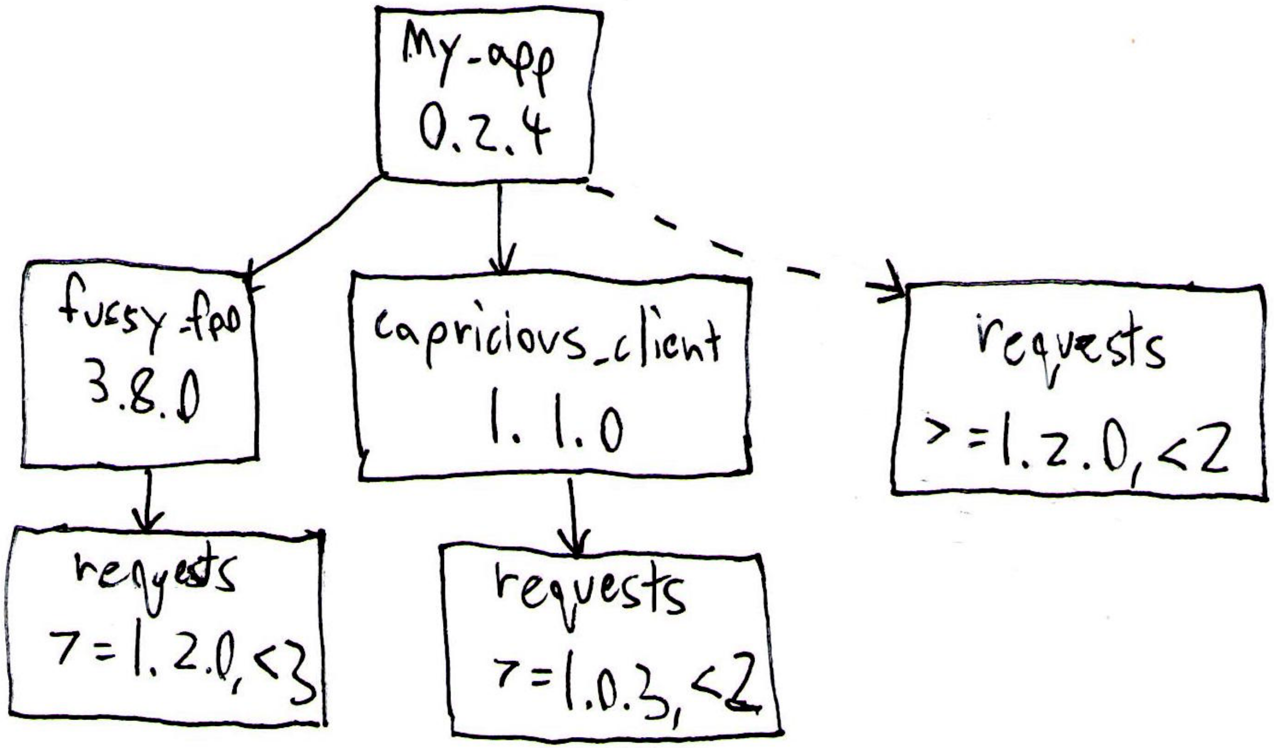
- Software not available or other software issue
  - Software or source code private
    - ’Open’ software is reviewable
    - ’Open’ software is replicable
    - ’Open’ software is producible
    - ’Open’ software is auditable
    - Code reviews for quality, purpose and intent
      - You learn stuff from reviewing other people’s code
      - People help you find stuff in your own code
      - Remember that compliment sandwiches taste great!

Why is this happening?

- Software not available or other software issue
  - Dependency hell

C …)

on A)



Why is this happening?

- Software not available or other software issue
  - Dependency hell
    - Virtual environments (for python)
      - conda create –n <name> <options>
      - Where options are things like what python version to use
      - https://conda.io/docs/using/envs.html
      - After creating, switch to it with ‘source activate <name>’
        - Do a `which python`
      - Switch out with ‘source deactivate <name>’
      - Use conda or pip to install packages
      - Export environment: `conda env export > environment.yml`
      - Create an environment: `conda env create -f environment.yml`
      - List environments: `conda env list`
      - Play time!

- Sof ssue
  - D

Why is this happening?

Example trace command:

```
% echo ‘print(“hello world”)’ > test.py
% reprozip trace python test.py
```

- Reprozip (http://reprozip.readthedocs.io/en/1.0.x/reprozip.html)
  - Captures all the necessary components in a single, distributable package
  - Install: ’pip install reprozip’
  - Trace:
    - » First time: ’reprozip trace **<command line>**’
    - » Subsequent: ’reprozip trace --continue <command line>’
  - Package: ‘reprozip pack package.rpz’
  - Share the .rpz

Why is this happening?

- Software not available or other software issue
  - Dependency hell
    - Reprozip
      - Info: `reprounzip info <package>.rpz`
      - Unpack the rpz:
        - Pick a compatible unpacker

`reprounzip <unpacker> setup <package>.rpz <path>`

`reprounzip <unpacker> run <path>`

`reprounzip <unpacker> run <path> --cmdline <new-command-line>`

- ALWAYS cleanup with

`reprounzip <unpacker> destroy  <path>`

Why is this happening?

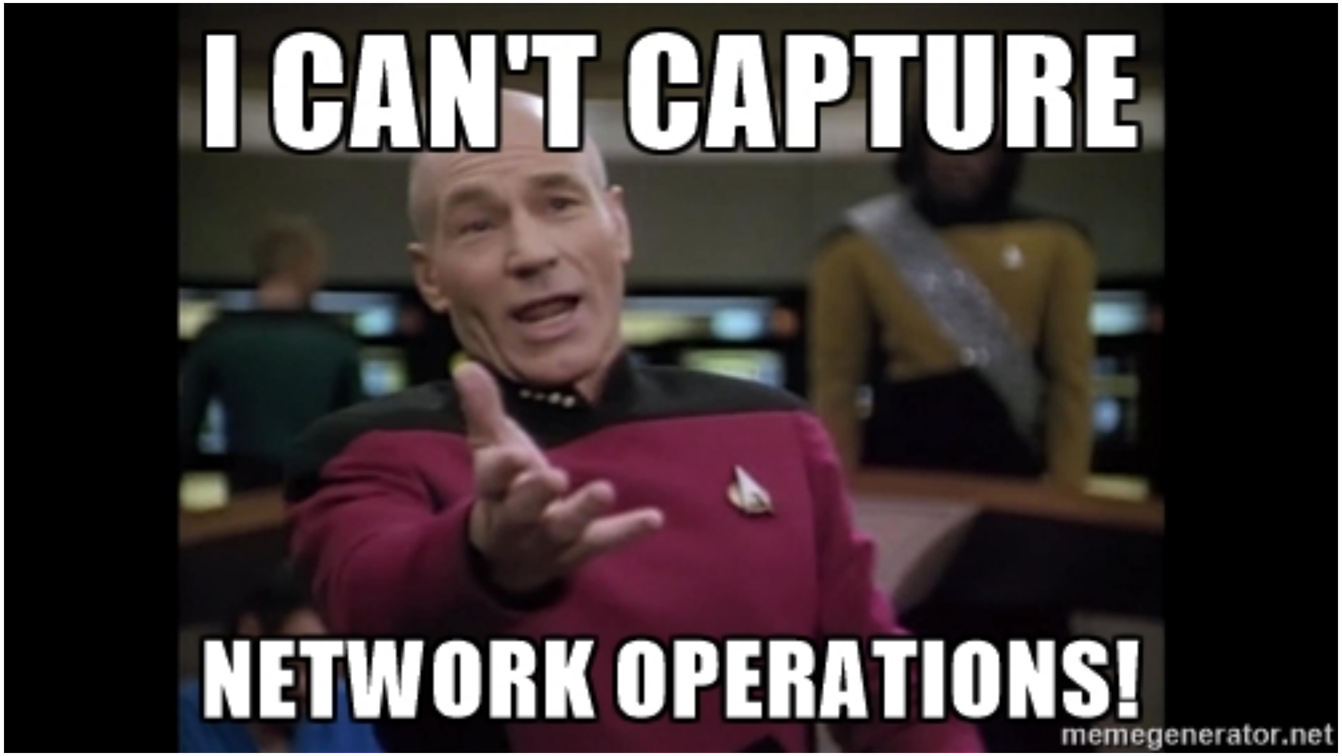
- Software not available or other software issue
  - Dependency hell
    - Reprozip
      - Variety of unpackers:
        - » Directory – puts everything in a directory, failure prone
        - » Chroot– creates a copy of the OS and 'changes the root'
        - » Installpkgs – tries to install packages into the OS environment (linux only)
        - » Vagrant – makes a virtual machine (requires Vagrant and Virtual Box)
        - » Docker – makes a docker container (light weight virtual machine)
      - Play time! (Try the directory unpacker, did it work?)

Why is this happening?

- Software not available or other software issue
  - Dependency hell
    - Manual creation of a virtual
      - Virtual Box
      - Amazon machine image
      - Docker image

Why is this happening?

- Software not available or other software issue
  - Dependency hell



Why is this happening?

- Software not available or other software issue
  - Software version
    - Don’t know or can’t remember what version was used
  - Version control
    - Use `tags` to label versions
      - The bad: tags can be moved
    - Use commit hashes to identify versions
      - Commit ID cannot be moved
    - Use Zenodo to get a DOI (GitHub)

Software

- Software not available or other software issue
  - Testing
    - Unit tests prevent again regression bugs
    - Use continuous integration
      - Automatically runs unit tests
      - Computes coverage
      - Let’s you know if you broke ‘stuff’

Software

- Software not available or other software issue
  - – Lack of documentation

Software

- Software not available or other software issue
  - Lack of documentation
  - You change postings someone can run your code
  - Best reason:
    - A year later you can run your own code!

Software

- Software not available or other software issue
  - Lack of documentation
    - How to run the code
    - Dependencies (as you understand hell)
    - Provide an example notebook
    - Example dataset

Software

- Questions?