Reproducible Research

Mark van der Loo

https://github.com/markvanderloo/UFPEL2019

The reproducibility crisis

An alarming number of scientific papers contain Excel errors

By Christopher Ingraham August 26

(Jewel Samad/AFP/Getty Images)

A surprisingly high number of scientific papers in the errors introduced by Microsoft Excel, according to an in the journal Genome Biology.

Many scientific studies can't be replicated. That's a problem.

By Joel Achenbach August 27, 2015

Brian Nosek, co-founder of The Center for Open Science, in his Charlottesville, Va. offices. (Bill O'Leary/The Washington Post)

This post has been updated.

Maverick researchers have long argued that much of what gets published in elite scientific journals is fundamentally squishy — that the results tell a great story but

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive

News & Comment > News > 2016 > September > Article

NATURE | NEWS

Dutch agency launches first grants programme dedicated to replication

Three-year pilot devotes €3 million to verifying other studies.

Monya Baker

20 July 2016 | Updated: 20 July 2016

Rights & Permissions

The Netherlands has launched what researchers say is the world's first national fund dedicated to replication studies: a pot of €3 million (US$3.3 million) over the next 3 years for Dutch scientists to test whether they can reproduce important research results in social and medical sciences.

The pilot programme was announced on 19 July by the Netherlands Organisation for Scientific Research (NWO), the country's largest research-funding agency. It marks a tiny fraction of the agency's €700-million annual budget, but is an important step, says Brian Nosek, executive director of the Center for Open Science in Charlottesville, Virginia. "If my calculations are correct, this is an increase of infinity per cent of federal funding dedicated to replication studies," he says.

Nosek has led an effort to replicate work from 100 psychology publications, which relied on what he estimates was $4 million in donations of time and resources from participating researchers. Even limited funds for replication can make innovative research more efficient, because it helps researchers

Homework (if you want to know more)

Read the wikipedia article on ‘replication crisis’

ASTM definitions (in engineering)

Replicability

A measurement is *replicated* if it is performed in the same laboratory, by the same operator, using the same apparatus, within a short time span.

Reproducibility

A measurement may be *reproduced* in a different laboratory, by a different operator, using a different apparatus after an undetermined time span.

- These terms define how precision and accuracy generalize when relaxing certain conditions.

In official statistics

UN principles governing international statistical activities

"[...] methods and procedures employed in the production of international statistics are chosen to meet professional **scientific standards** and are made transparent for the users."

European Statistics Code of Practice

"Based on **scientific principles and methods**, the European Statistical System will offer and continuously improve a programme of harmonised European statistics"

- No exlicit reference to reproducibility, nor defined precisely, but implied reproducibility by reference to the scientific method.

Reproducibility and the scientific method

Scientific objectivity

Reproducibility is an aspect of what is called *process objectivity* of science. It adheres to the idea that scientific statements are objective in the sense that they are produced in a way that "neither depend on contingent social and ethical values, nor on the individual bias of a scientist"[1][2].

Notes

- Statistical inference quantifies the measure of reproducibility of statistical (sampling) experiments.
- Reproducible data analyses is a very basic requirement underlying reproducibility which in practice turns out to be non-trivial.

[1] Stanford encyclopedia of philosophy

[2] Definition of concepts is not part of this process

Practical reasons for reproducibility

- Accountability, transparancy
- Efficiency
- Knowledge sharing
- Process quality
  - Usability (understandable)
  - Maintainability (testable, analyzable)

Reproducibility of statistical experiments

Statistical inference

Given the result of my randomized experiment, how do I expect this result to vary if I repeat the experiment?

Examples

- *p*-value: estimated probability that I will see a different result when the experiment is repeated.
- 95%-confidence interval: estimated interval where I expect to see the result when the experiment is repeated.

Note

p values, confidence intervals, etc express an estimate of the amount of evidence gatherd by an experiment for a given result. It does *not* say whether the result is important.

Statistical inference

1. **Estimate a population parameter, based on a randomized sample**

2. **Analytical route**

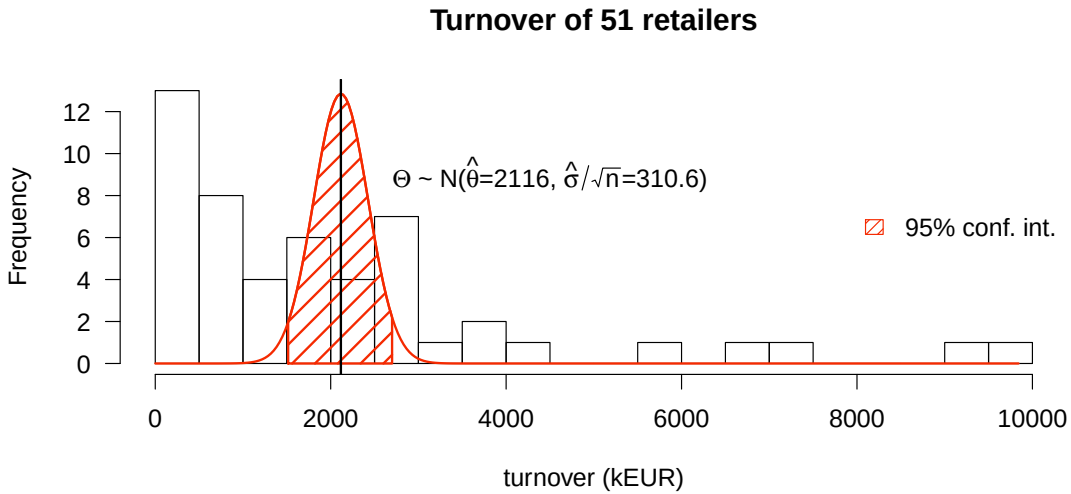
   1. Derive a distribution of the estimate, in the limit of sampling over and over again.
   2. Based on a your actual sample, estimate the parameters of the distribution of your estimate over repeated samples.

2. **Computational (bootstrap) route**

   1. From your randomized sample, draw a new sample (SRSWR), and estimate.
   2. Repeat step 2 many times to estimate the distribution of your estimate over repeated samples.

3. **Use estimated sampling distribution for p values, CI, ...**

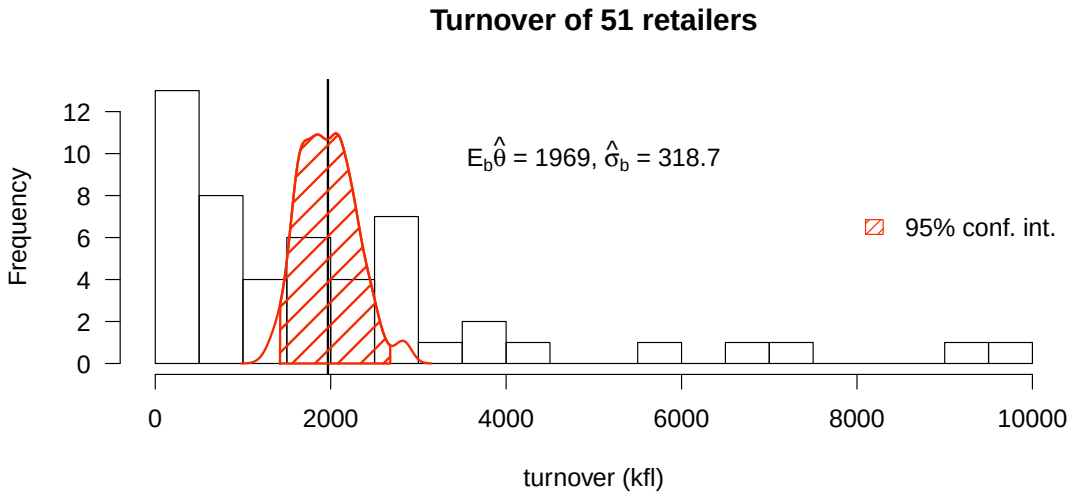
Example: estimate mean turnover from a sample of retailers



Example 2: Computational inference of windsorized mean

```
# Mean, windsorized at median + 3 times median absolute deviation
wins_mean <- function(x){
  limit <- median(x) +  3 * mad(x)
  x[x > limit] <- limit
  mean(x)
}
```

Estimate windsorized mean from sample of retailers



Reproducibility of data processing

Scholarship in computational science

*An article about computational science in a scientific journal is **not** the scholarship itself, it is merely **advertising** of the scholarschip. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.*[3]

[3] Buckheim and Donovo (1995)

Main idea of a reproducible analyses

The product of an analyses is not a written article, but rather a *compendium*[4] containing

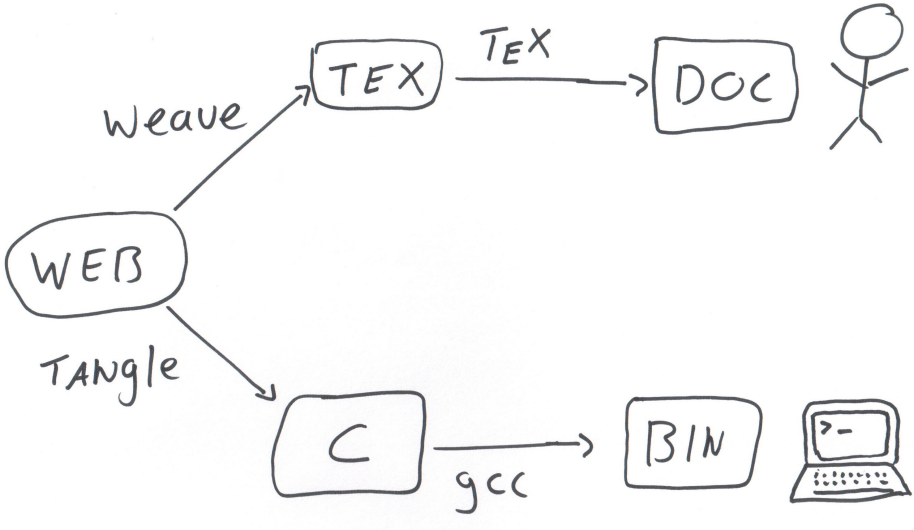
- The data being analyzed
- Runnable code used to compute the numbers
- The text explaining the problem, methods and interpretation.

Note

This requires significant programming/software engineering skills!

[4] Gentleman and Temple Lang (2004)

Dynamic documents: literate programming



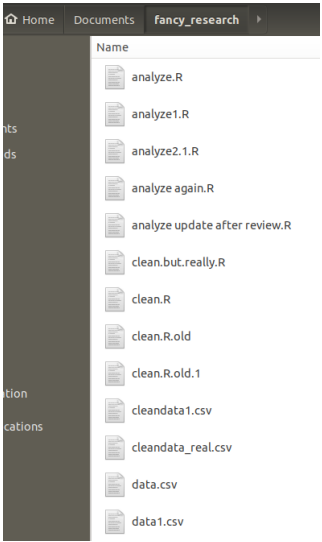
Knuth (1992)

Reproducible research using `rmarkdown`

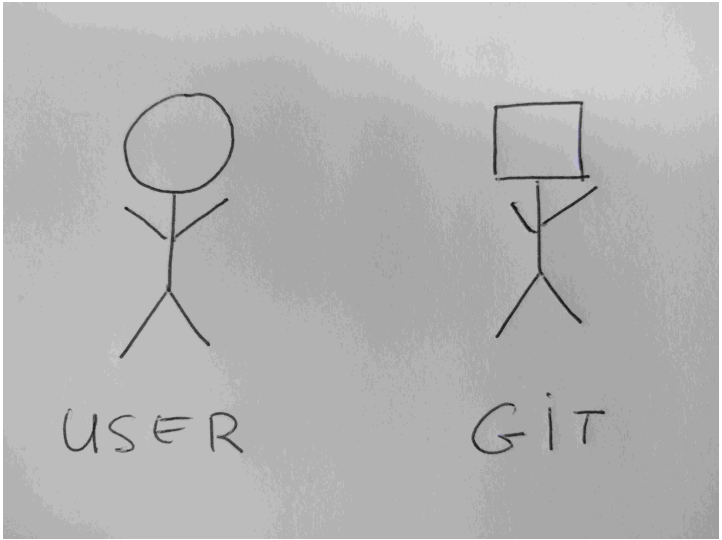
Demo

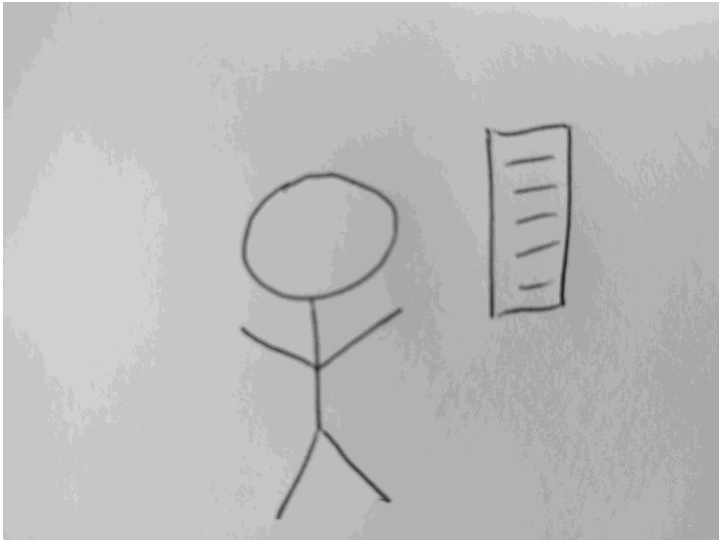
Keeping track of changes: version control

Why version control?

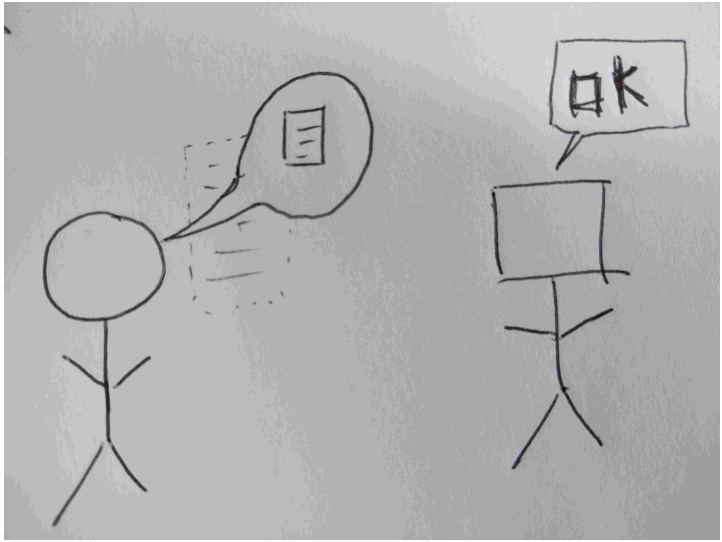


- Keeping track of updates manually is messy, difficult, and depends on discipline.
- Fortunately this can be automated.

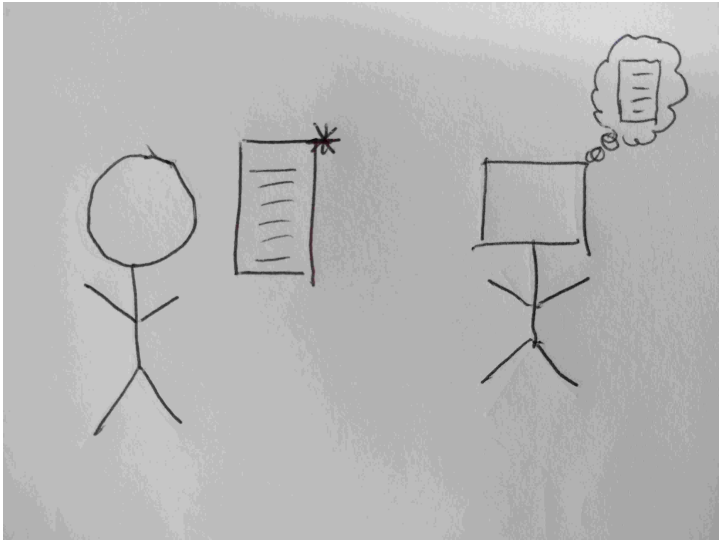

USER
GIT



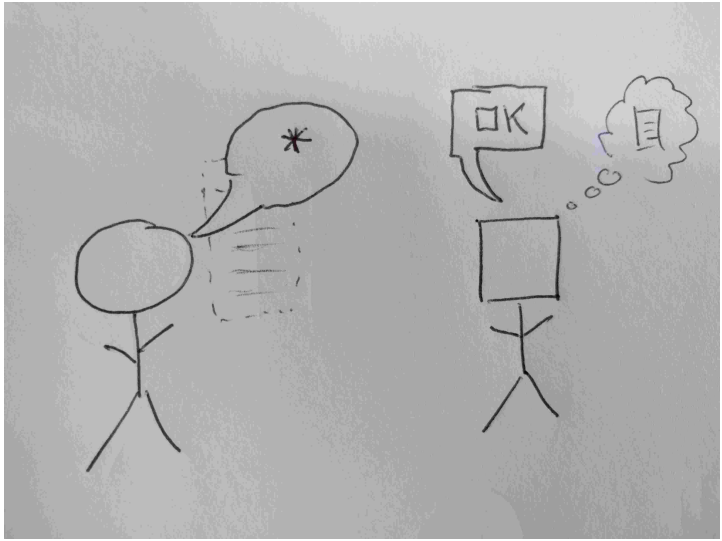
```
$> vi myfile.R
```



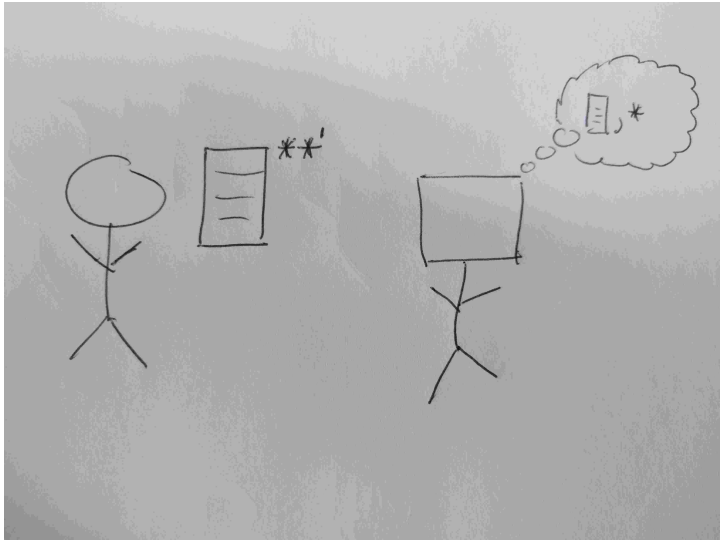

```
$> git add myfile.R
```



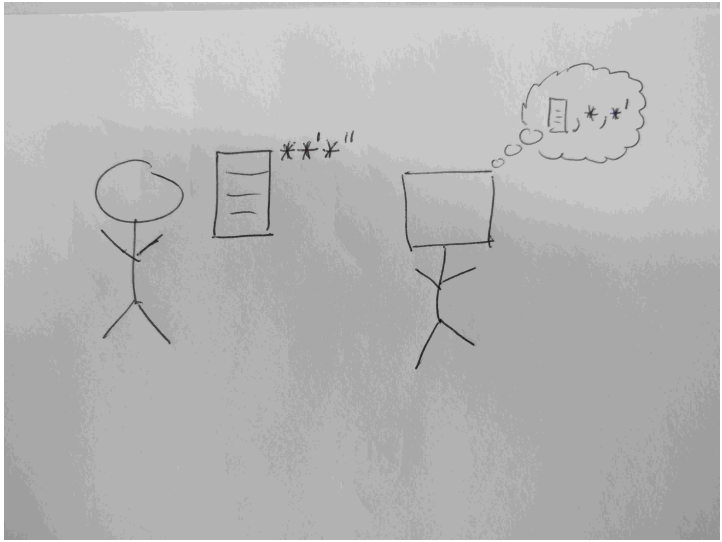
```
$> vi myfile.R
```



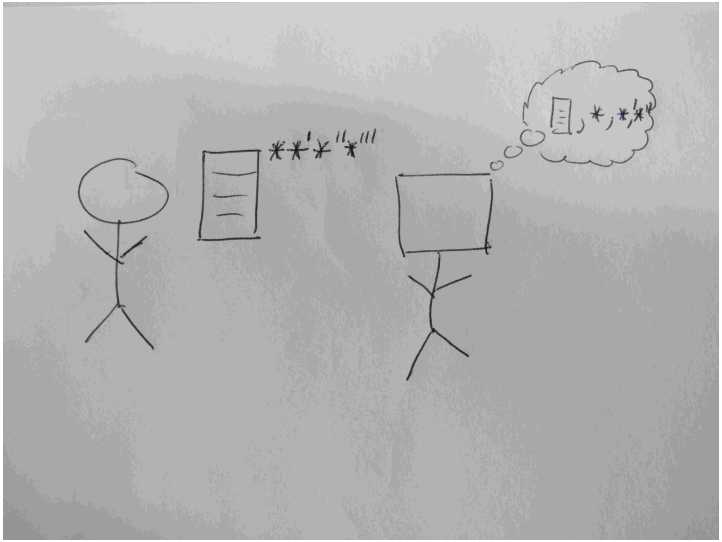

```
$> git add myfile.R
$> git commit -m "fixed a bug"
```



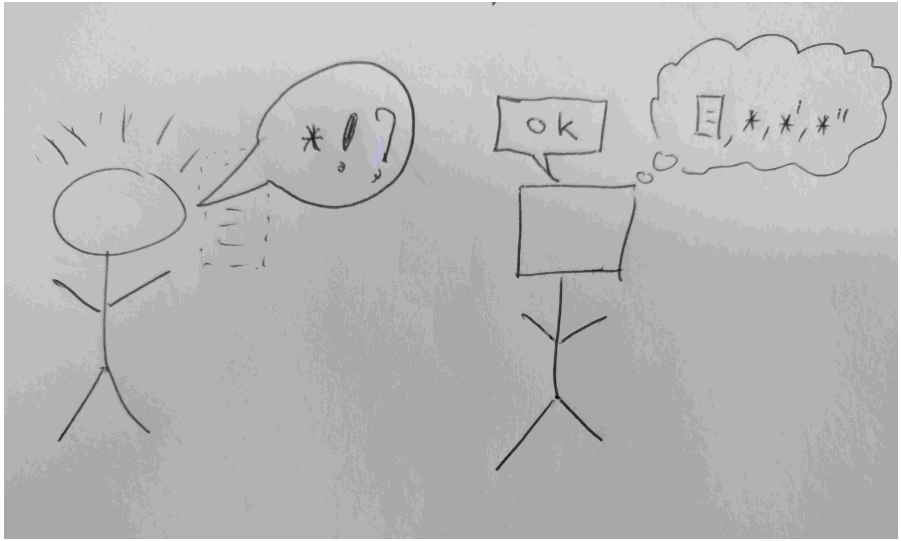
```
$> vi myfile.R
```



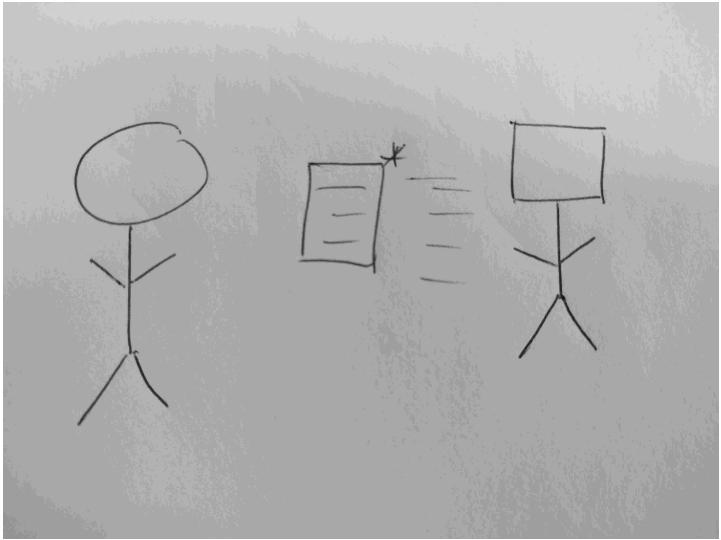
```
$> vi myfile.R
```

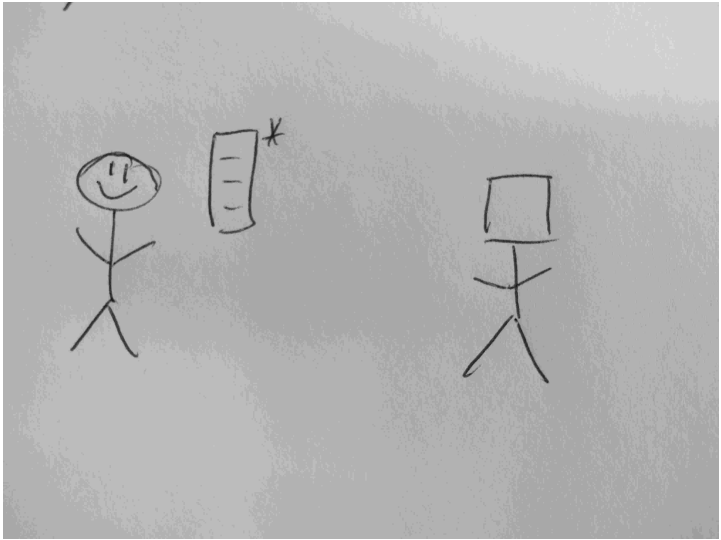


```
$> vi myfile.R
```




```
$> git log --oneline
$> git checkout be9055b .
```





Demo

https://github.com/markvanderloo/UFPEL2019