



Why it's hard to reproduce results in software

Bjørn Fjukstad
Department of Computer Science



What is reproducibility?





YouTube NO

how to remove drywall anchors

SIGN IN

Home

Trending

History

BEST OF YOUTUBE

- Music
- Sports
- Gaming
- Movies
- News
- Live
- 360° Video

Browse channels

Sign in now to see your channels and recommendations!

SIGN IN

FILTER

How to remove plastic dry wall anchors!!
Joe Average • 29K views • 1 year ago
Quick video on how to remove drywall anchors easy process decided to share with you guys enjoy thank you for watching.

Removing drywall anchors
Radar Really • 49K views • 2 years ago
I'm getting ready to paint my bathroom, and I've got several drywall anchors. I went to my toolbox and grabbed three tools: Needle ...

How to Remove Plastic Wall Anchors
HomeAdditionPlus • 18K views • 10 months ago
In this video Mark Donovan of <http://www.homeadditionplus.com> shows how to remove plastic wall anchors from drywall walls.

How to: Remove Wall Anchors In Less than One Minute!
ThunderDivine • 219K views • 3 years ago
Here I'm Removing Big Metal Wall Anchors In a Plasterboard Wall In Less Than A Minute. #wallanchors With this method You ...

YouTube NO

reproducibility software

SIGN IN

Home

Trending

History

Music

Sports

Gaming

Movies

News

Live

360° Video

BEST OF YOUTUBE

FILTER

Grand Rounds: Research Reproducibility 11-07-2017 with Robert Ricci

Eccles Health Sciences Library Digital Publishing • 149 views • 1 year ago

Spencer S. Eccles Health Sciences Library Grand Rounds: Research Reproducibility (e-channel) - Software for Reproducibility ...

Singularity: Containers for Science, Reproducibility, and HPC

RichReport • 3.4K views • 1 year ago

In this video from the 2017 HPC Advisory Council Stanford Conference, Greg Kurtzer from LBNL presents: Singularity: Containers ...

Measurement Systems Analysis - Repeatability & Reproducibility Study

Symphony Technologies • 21K views • 4 years ago

The R&R Study in MSA is used to evaluate the Repeatability & Reproducibility of a Variable Measurement System. Explore how a ...

a reproducible workflow

bartomeuslab • 7.7K views • 2 years ago

Reproducible science not only reduce errors, but speeds up the process of re-running your analysis and auto-generate updated ...

A NON REPRODUCIBLE WORKFLOW

1:45

10:39

11:58

48:24

6

Definitions

Cornerstone of Science

- The principal goal of scientific publications is to teach new concepts, show the resulting implications of those concepts in an illustration, and provide enough detail to make the work reproducible.[3]

Different definitions

- **Replicate:** carry out the same task as the original researcher, and get the same results
 - Same code, same compiler, same hardware and same operating system (OS), gives the same result.
- **Reproduce:** carry out tasks that are similar to the original and get the same result
 - Same data, similar code, similar compiler, similar hardware, similar OS, gives the same result.

Different definitions

- **Preproducibility:** An experiment is preproducible if it has been described in adequate detail for others to undertake it.[4]

Different definitions

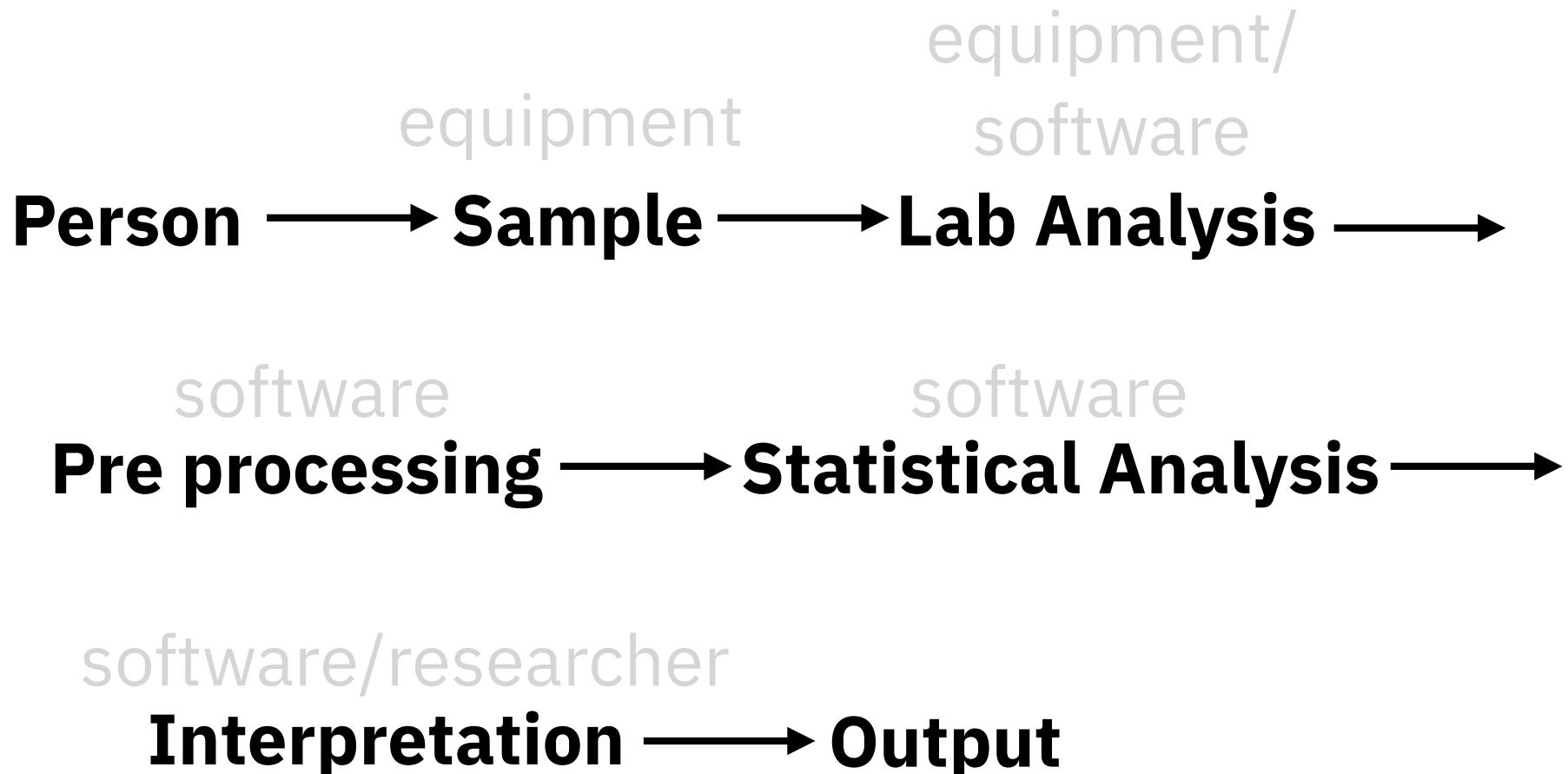
- **Verify:** The task of replicating an experiment to see if it yields the same results
- **Validate:** *the task of evaluating a result to see if the researcher's conclusions are warranted.*

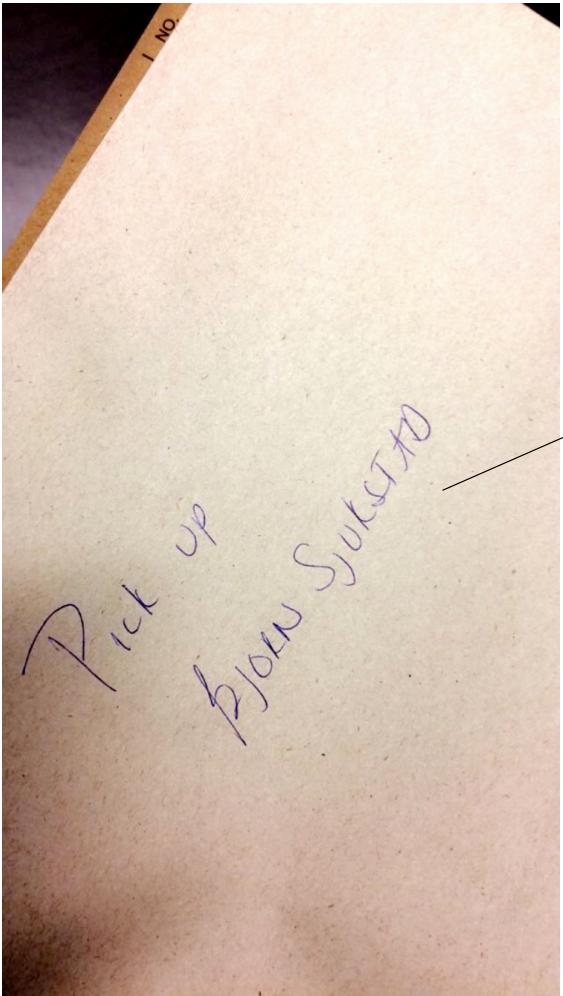
Consequences

- Science moves forward when researchers verify other's results. Science advances faster when people waste less time pursuing false leads [3]
- The society and companies can potentially waste hundreds of millions on failed drug development programs [4]
- If science is not useful, why should we fund or support it? Why should people donate blood to a study that can't be reproduced?

Where things (can) go wrong

The Trajectory of a Biomedical Research Project

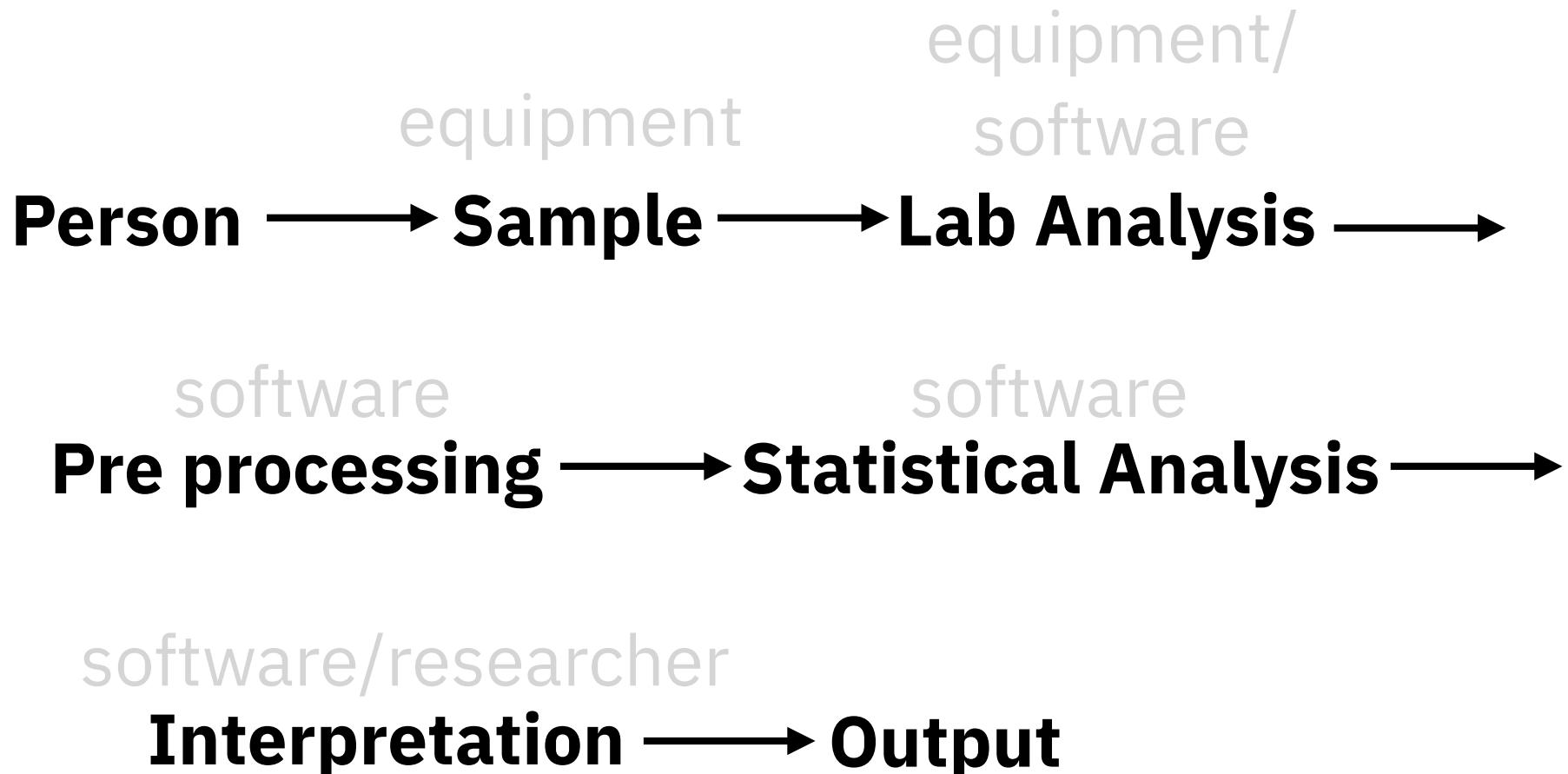




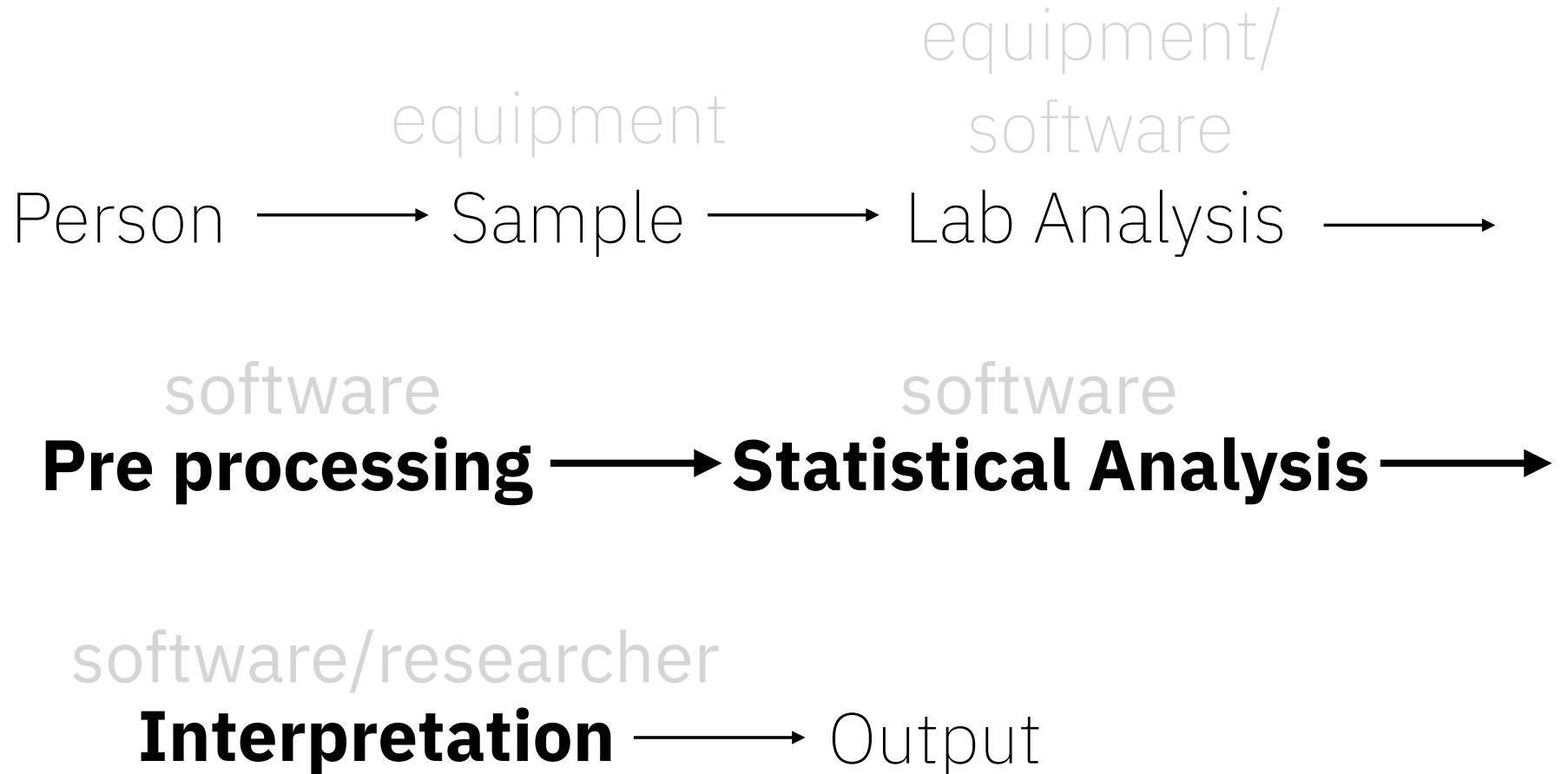
Sjukstad ≈ Sicktown

14 October 2016, pick up of biological sample at McGill University Health Centre

The Trajectory of a Biomedical Research Project



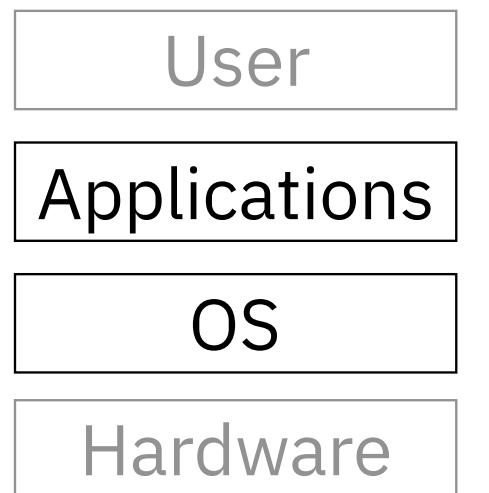
The Trajectory of a biomedical research project



Software

Software

- A collection of instructions that tell the computer (hardware) to work
- From high-level programming languages to low-level machine instructions
- Share: code or binary



Equal Experiment Results

- Reproducing an experiment results in **the same bits**
 - Example: Files are identical and yield the same hash.
- Reproducing an experiment yields in **the same result**
 - Example: The updated software now stores the results as a csv file rather than tsv file
- Reproducing an experiment results in **equivalent results**
 - Example: Updating the software may cause results to be represented with fewer decimal points

Why it's hard to reproduce results in software

- Focus on the difficulties of generating the same results using similar software in biomedical research
- Assume that the data exists, and that we do not have to re-collect it

This lecture

- *Definitions*
- Detail the problems with reproducibility in different fields
- Give an example of what difficulties may arise when reproducing an experiment
- Highlight some techniques for simplifying reproducibility
- Current trends in science

The Problem

2016 Nature Survey

- Nature asked **1,576** researchers to respond to a brief online questionnaire on reproducibility in research[2]
- **70%** responded that they have tried and failed to reproduce another scientist's experiment.
- More than **50%** have failed to reproduce their own experiments
- **52%** said it's a significant reproducibility crisis, **38%** said it's a slight crisis

Cancer Research

- Biotechnology company Amgen tried to reproduce 53 “landmark” papers in cancer research : could only reproduce 11% of the findings [6]
- Bayer HealthCare surveyed 67 in-house projects, but could only reproduce 21% of the projects [7]

“Reproducing” Computer Science Papers

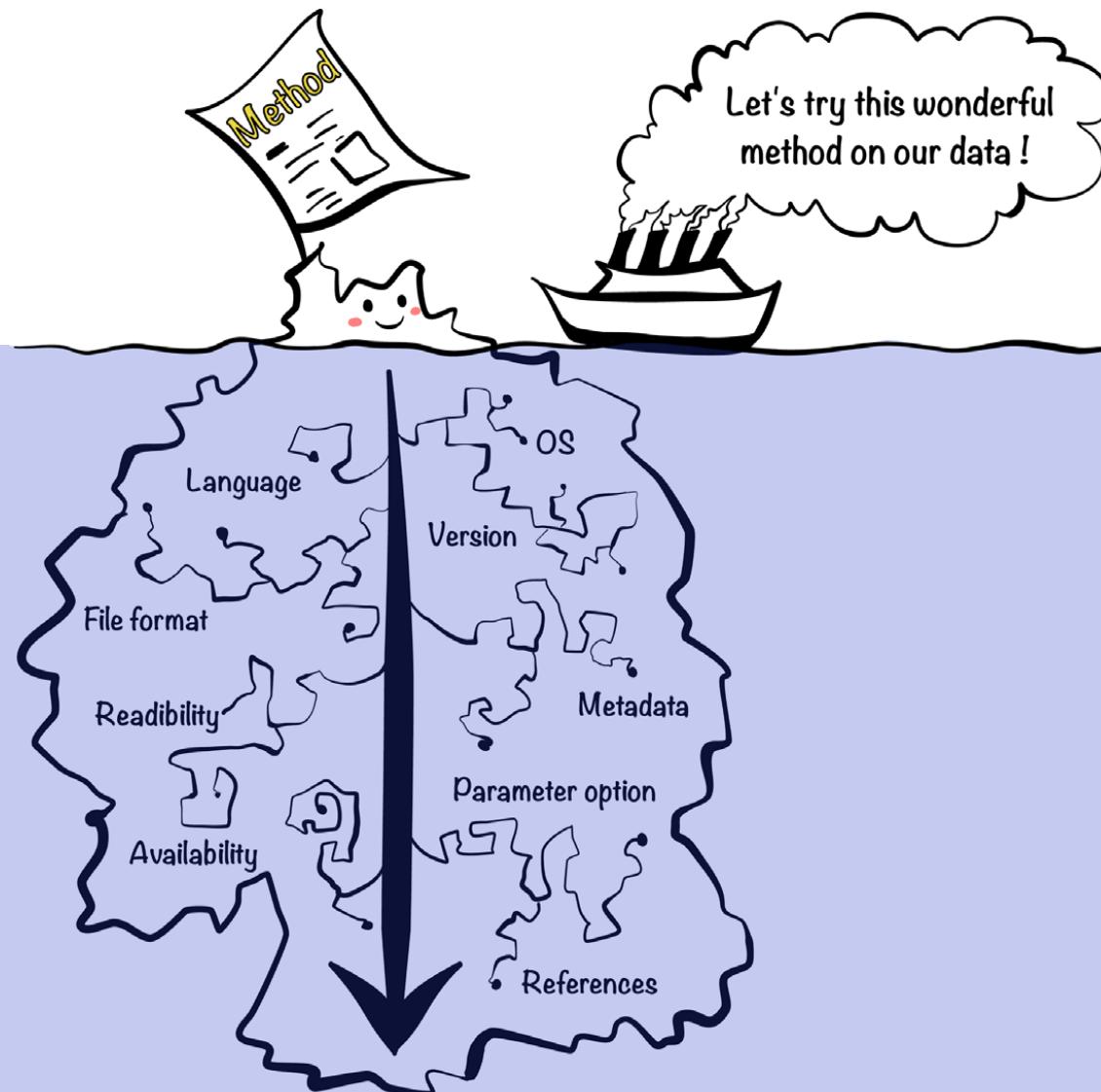
- Study from University of Arizona downloaded **601** papers from eight ACM conferences (including SOSP, SIGMOD, OSDI) and five journals. [8]
- Investigated if the code was available, and if they could build it
 - Excluded **199** papers
 - Did not find, or receive, code for **176** papers.
 - **226** papers backed by code: **130** built within 30 minutes; **64** after more than 30 minutes; **23** not built, but authors claim it will; **9** cannot be built by team or original authors

The need for reproducibility

“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”

–Jon Claerbout



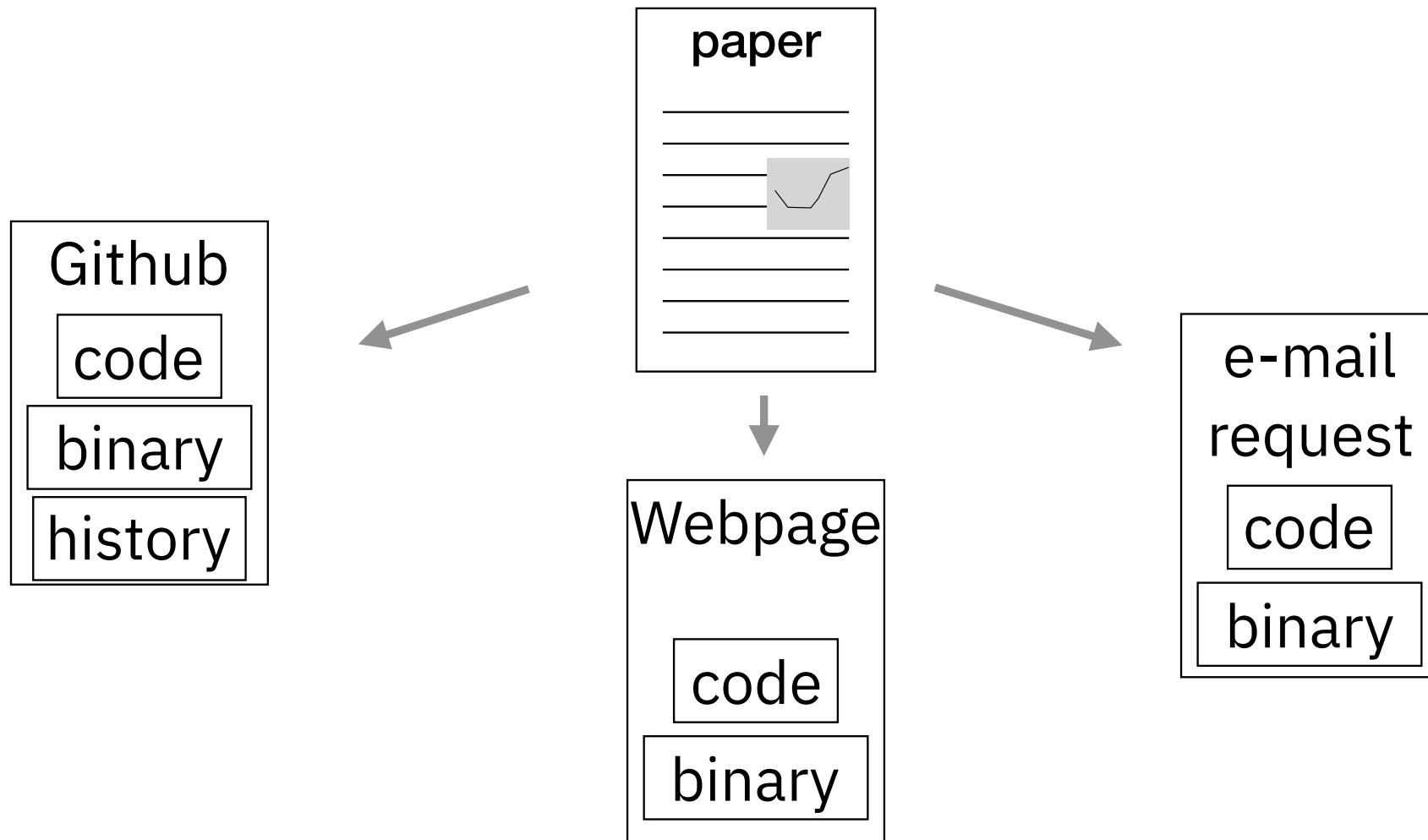


Why we should strive for easily reproducible research

- Verify and reuse other's results and methods
- Verify and reuse our own results and methods
- Improve one's own productivity
- To encourage collaborative maintenance

The technical and practical difficulties of reproducing results

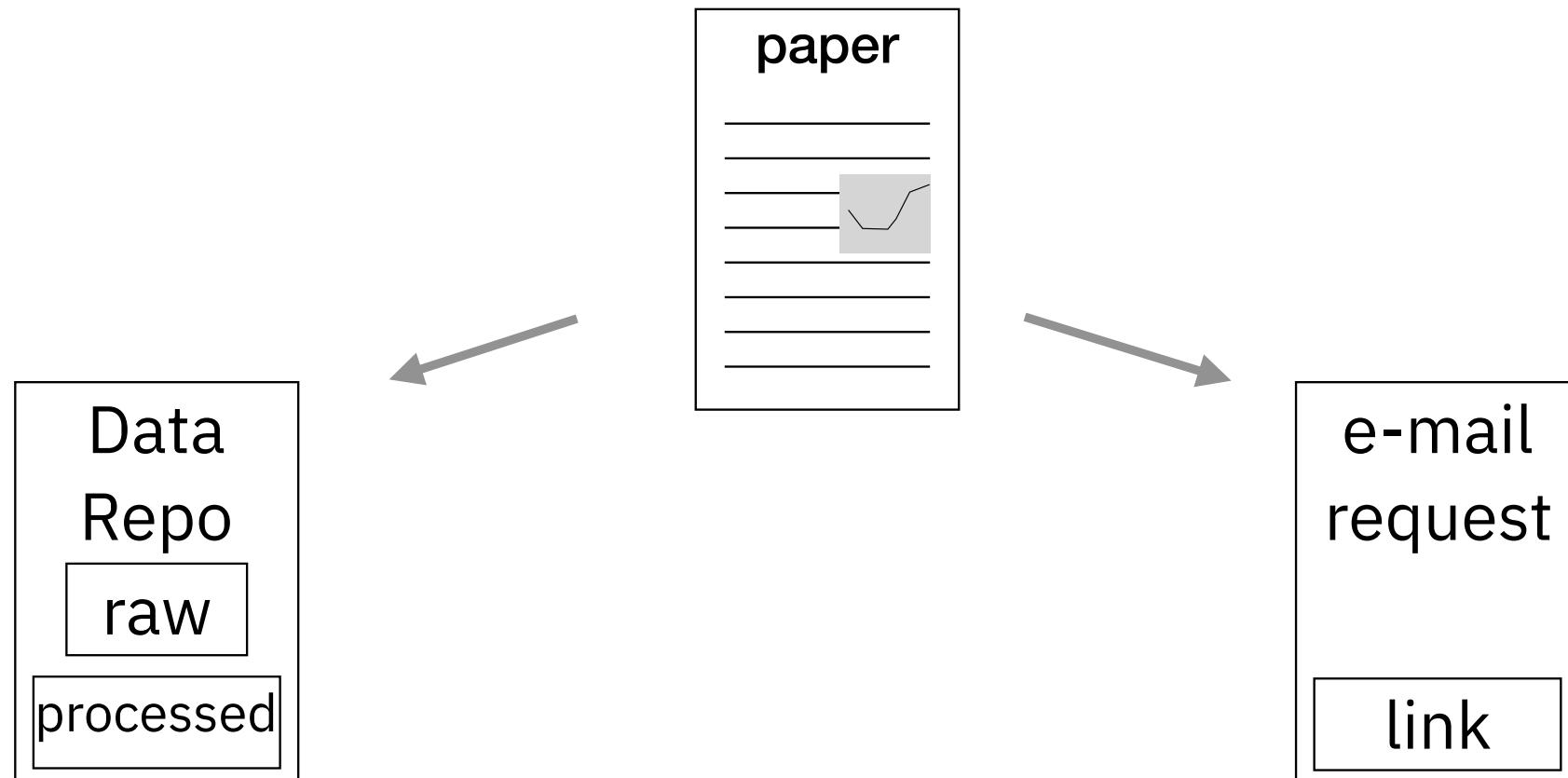
Simply getting the required program



Why researchers can't/won't share code

- *Versioning: Don't know if this version is the final one*
- *Available soon*
- *We don't intend to make the code available, ever*
- *The programmer left*
- *The system physically ceased to exist, got stolen or crashed*
- *The code is commercial or proprietary, and won't be open-sourced or given a permissive license*
- *Too busy to help out*

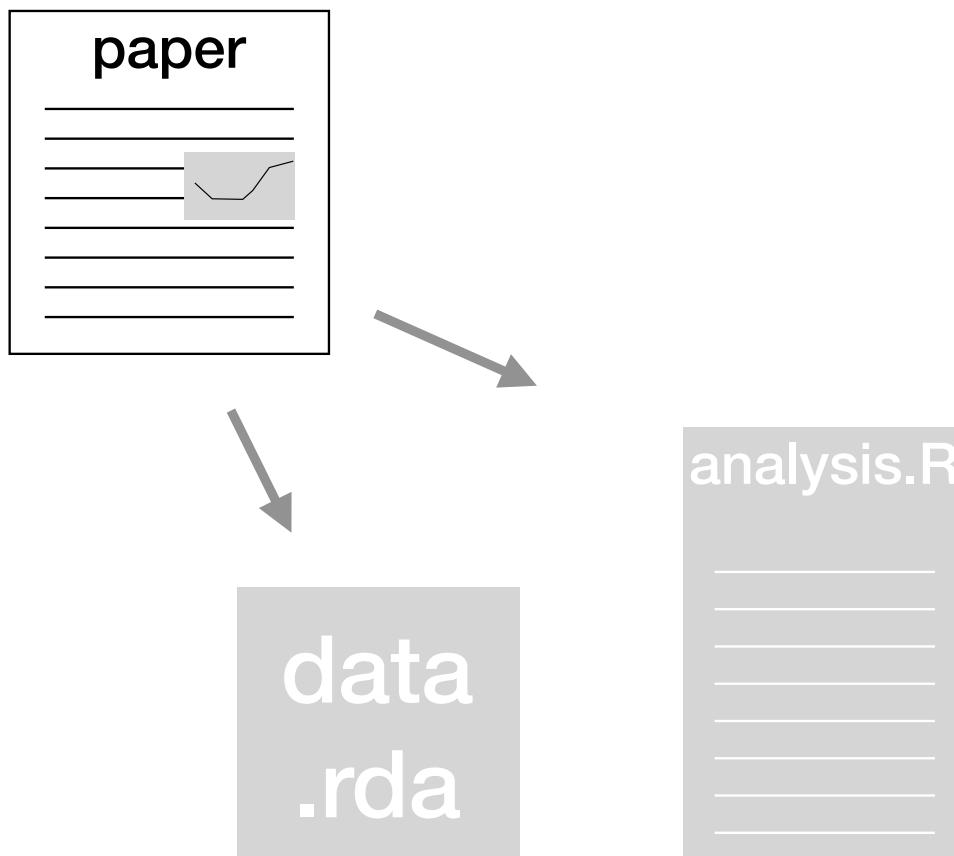
Get a hold of the data



Why researchers can't/won't share data

- *The data contains sensitive information*
- *The data is proprietary*
- *GDPR*

Example



paper

Example



```
> Rscript analysis.R --method="a" --input=data.rda --output=output.csv
```

**data
.rda**



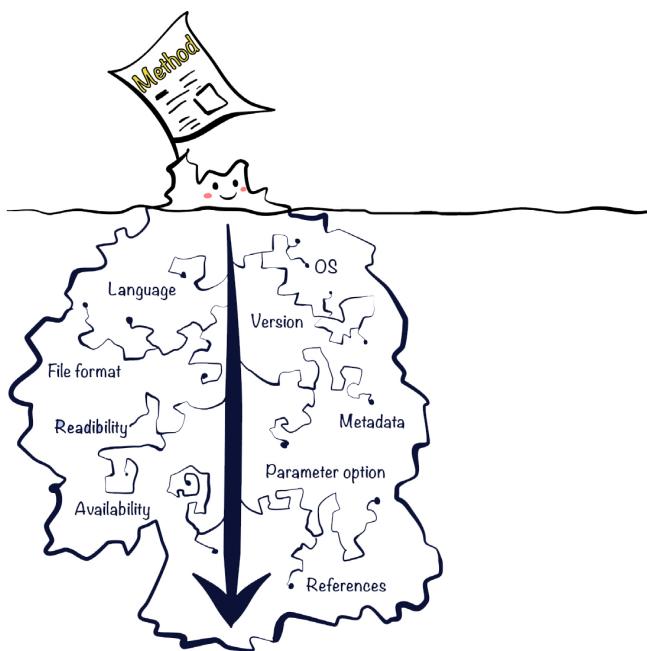
analysis.R



output.csv

Difficulties all the way down

```
> Rscript analysis.R --method="a" --input=data.rda --output=output.csv
```

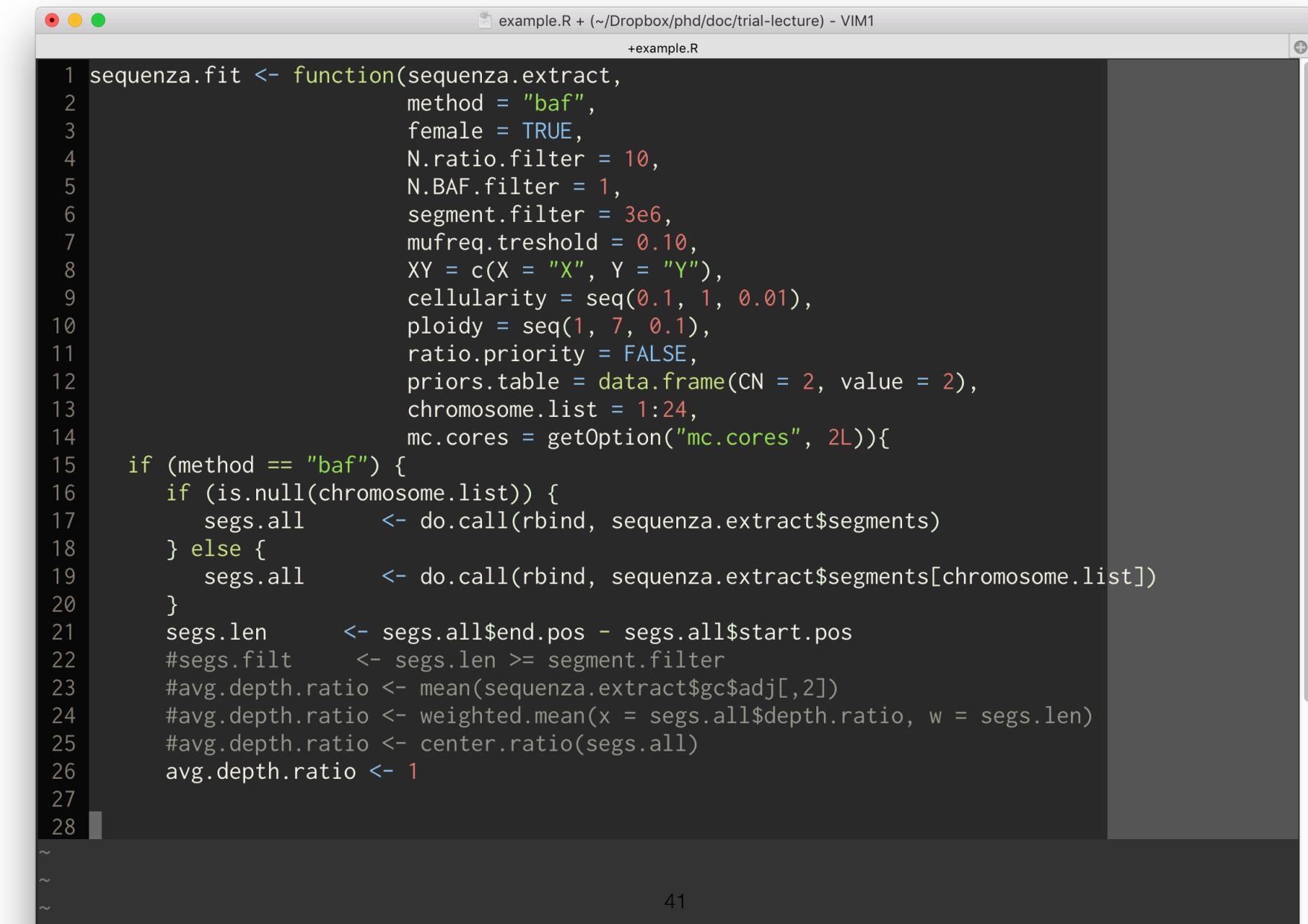


Command, data, and parameters	a, data.rda, output.csv
Software	R/Rscript 3.40, plotly 3.4.1, ...
Operating System	OS X High Sierra 10.13.6
Kernel	Darwin Kernel Version 17.7.0 xnu-4570.71.2~1/RELEASE_X86_64 x86_64
Hardware	Macbook Pro (Retina, 13-inch, Mid 2014), 3 GHz Intel Core i7, 16 GB RAM, 250 GB SSD

Command, script, and parameters

```
> Rscript analysis.R --method="a" --input=data.rda --output=output.csv
```

- All information needed to reproduce the results is hidden in the `analysis.R` script
- Parameters may improve reusability of the analysis, but could overcomplicate an experiment



The screenshot shows a VIM1 window with the title "example.R + (~/Dropbox/phd/doc/trial-lecture) - VIM1". The buffer contains an R function definition:

```
sequenza.fit <- function(sequenza.extract,
                           method = "baf",
                           female = TRUE,
                           N.ratio.filter = 10,
                           N.BAF.filter = 1,
                           segment.filter = 3e6,
                           mufreq.threshold = 0.10,
                           XY = c(X = "X", Y = "Y"),
                           cellularity = seq(0.1, 1, 0.01),
                           ploidy = seq(1, 7, 0.1),
                           ratio.priority = FALSE,
                           priors.table = data.frame(CN = 2, value = 2),
                           chromosome.list = 1:24,
                           mc.cores = getOption("mc.cores", 2L)) {
  if (method == "baf") {
    if (is.null(chromosome.list)) {
      segs.all <- do.call(rbind, sequenza.extract$segments)
    } else {
      segs.all <- do.call(rbind, sequenza.extract$segments[chromosome.list])
    }
    segs.len <- segs.all$end.pos - segs.all$start.pos
    #segs.filt <- segs.len >= segment.filter
    #avg.depth.ratio <- mean(sequenza.extract$gc$adj[,2])
    #avg.depth.ratio <- weighted.mean(x = segs.all$depth.ratio, w = segs.len)
    #avg.depth.ratio <- center.ratio(segs.all)
    avg.depth.ratio <- 1
  }
}
```

sequenza / sequenza / issues

Atlassian, Inc. [US] | https://bitbucket.org/sequenzatools/sequenza/issues/3/why-avgdepthratio-1

Create issue

Sequenza / R_packages / sequenza / Issues

Why avg.depth.ratio = 1?

Issue #3 **RESOLVED**

Bjørn Fjukstad created an issue 2016-12-14

Hi,

Just browsing the source code and I'm wondering why in `workflows.R`: `sequenza.fit` `avg.depth.ratio` is set to 1? E.g. [here](#). In the package documentation for both `baf.model.fit` and `mufreq.model.fit`, it says that the `avg.depth.ratio` is calculated from the `gc.stats`, so why is it set to 1 inside `sequenza.fit`? Isn't `sequenza.fit` just a wrapper to these other functions?

Thanks, Bjørn

Comments (5)

Francesco Favero
Hi Bjørn,

Thanks for bringing this up. This was a quick fix to the fact that oddly distributed (biased) data can lead to a consistently wrong estimation. The normalization doesn't work as expected in some cases. I suspect the `avg.depth.ratio` is influenced by some heavily covered region - and the quick solution in order to still get a decent result out of those cases was to set the average depth ratio to a fixed value. In general, this shouldn't be a big deal, since the workflow takes care to normalize the samples. If the samples weren't normalized then the `avg.depth.ratio` would be dependent on the library size of the two samples.

As you can see the code expects to have the `avg.depth.ratio` calculated from the data, instead of being hardcoded to 1, which is, as I mentioned, the expected value after the normalization. I'm slowly moving toward another normalization method, which handles the two samples separately (and hopefully handles additional bias other than GC-content). This might also be part of the not-efficient-normalization (the current implementation normalizes the `depth.ratio`, instead of the tumor and normal coverage separately).

I hope to revise this in future releases.

Atlassian, Inc. [US] | https://bitbucket.org/sequenzatools/sequenza/commits/c51079285c5cc841baa14ce95...

Create issue

Sequenza / R_packages / sequenza / Commits

Commit

Francesco Favero committed c510792
2017-07-06

Compute and use real avg depth ratio rather than hardcode it to 1

53a7f12
cleanup
No tags
Pull requests

[View raw commit](#)
[Watch this commit](#)
[Run pipeline](#)

Comments (0)

What would you like to say?

Files changed (3)

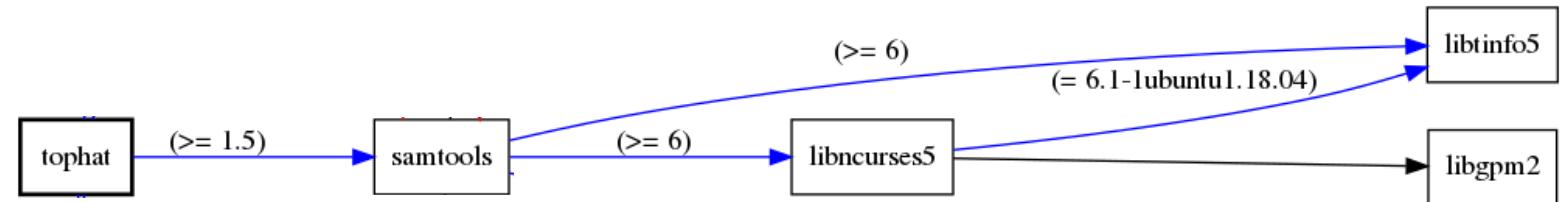
+22	-1	[M]	R/extract.R
+16	-21	[M]	R/fit.R
+1	-1	[M]	R/results.R

Software

R/Rscript 3.40, plotly 3.4.1, parallel 3.4.0

- Dependencies can be difficult to obtain and install
- Different versions of software *may* be compatible, but will include some changes
- *Software collapse*

TopHat Dependencies



TopHat Dependencies



Operating System and Kernel

OS X High Sierra 10.13.6 Darwin Kernel Version 17.7.0 xnu-4570.71.2~1/RELEASE_X86_64 x86_64

- The underlying operating system and kernel will impact how software is run
- The software might not be available on your operating system
- These (potentially) small changes may be hard to detect

Hardware

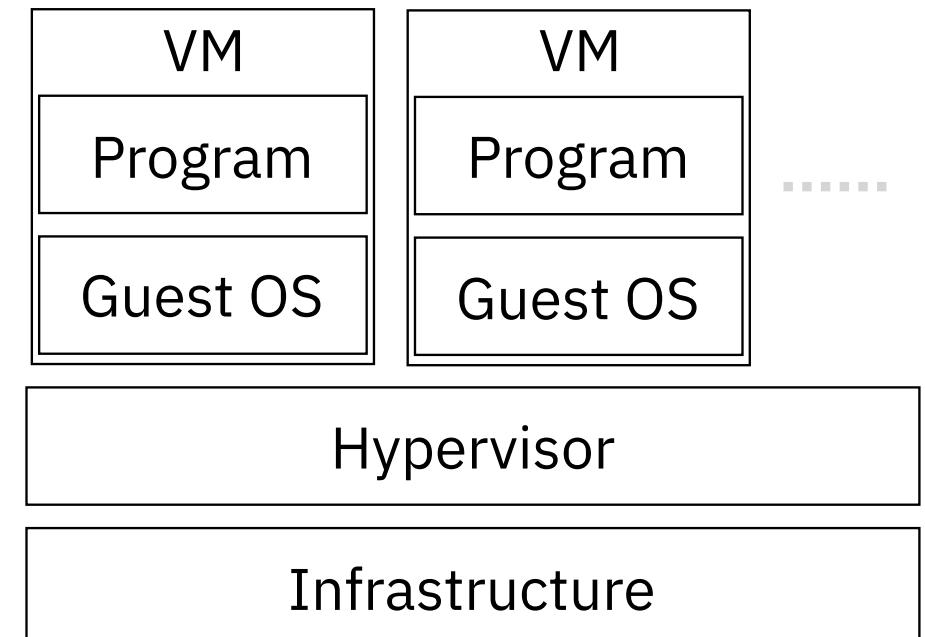
OS X High Sierra 10.13.6 Darwin Kernel Version 17.7.0 xnu-4570.71.2~1/RELEASE_X86_64 x86_64

- The software may require specialized hardware
- It may not be possible to run on a desktop computer

Reproducing the Environment

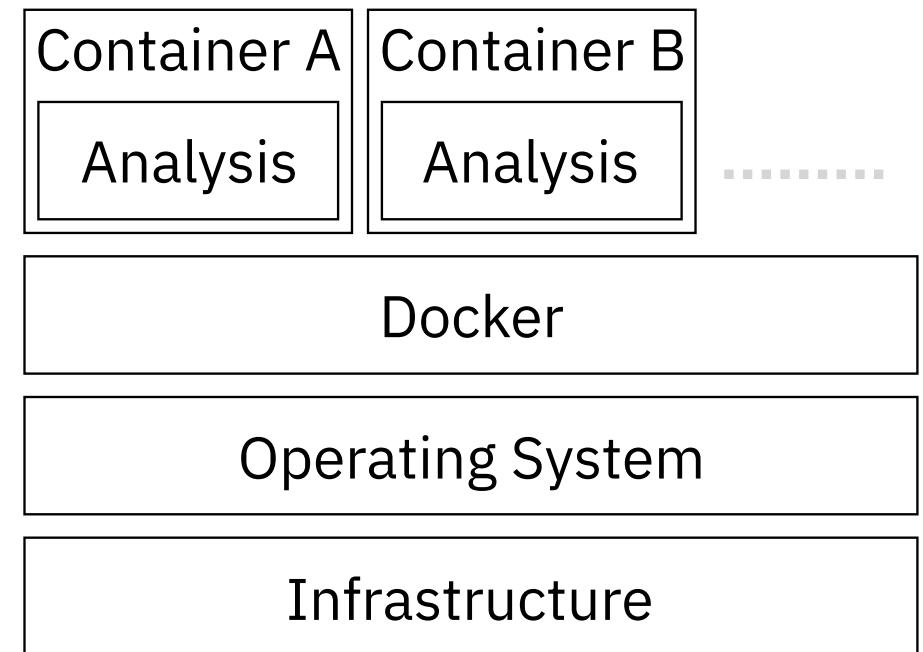
Virtual Machines (VMs)

- Capture the full environment including the OS
- Are big in size, and potentially hard to construct



Containers

- Allows packaging of an application along with its dependency
- Provides users with the same environment
- Lightweight, small in size, and declaratively specified



What about the rest of the
paper?

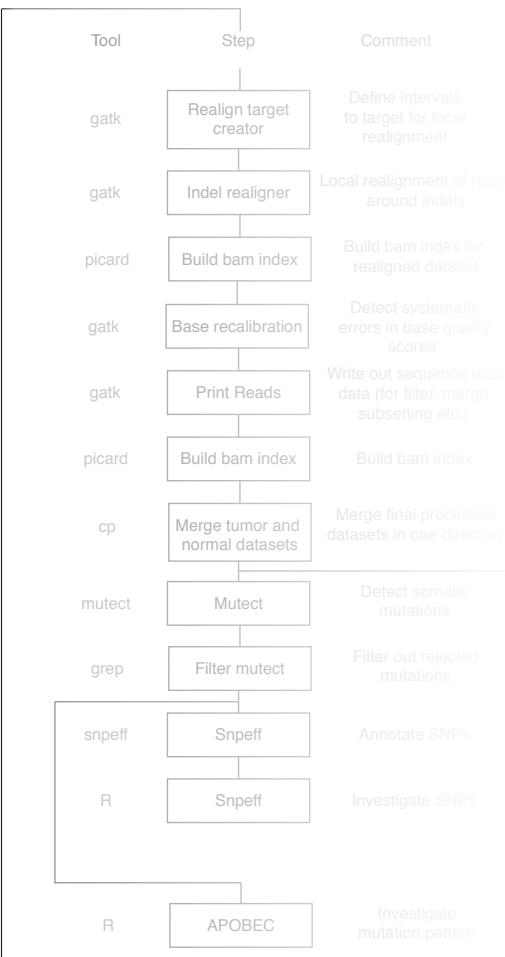
Workflows

- Specify the full workflow, e.g. the chain of different commands and parameters
- Many workflow management systems, and no common or accepted format for specifying or sharing these

Workflow granularity

- Workflow (e.g. BWA-GATK)
- Domain Tasks (e.g. simulate(x), analyze(y))
- Middleware Operations (e.g. split/merge, map/reduce)
- System Calls (sys_open, sys_gettimeofday)

Tool	Step	Comment
cp	1	Read datasets into system
picard	2	Convert files into a format fastqc can read
fastqc	3	Run quality control of the input data
cp	Get Illumina Adapters	Get illumina adapters
trimmomatic	Trim reads	Remove illumina adapters from reads
fastqc	FastQC trimmed	Quality control trimmed datasets
cp	Input reference genome	Read reference genome into system
bwa	Align reads	Align reads to reference genome
samtools	Sam to Bam	Convert files from sam to bam and sort it
picard	Build bam index	Build Bam index for downstream tools
cp	Merge with index	Place Bam and index files in same dir
picard	Add read groups	Assign read groups to dataset for downstream tools
picard	Mark duplicates	Identify duplicate reads
cp	Input intervals	Get target intervals
cp	Input indels	Get known indels
picard	Build bam index	Build bam index for dataset with marked duplicates



Integrating statistics and the publication

Executable Papers

- Jon Claerbout started in the early 90's to merge scientific software with word-processing software
- Started with magnetic tape, then CD-ROMs
- Popularized with IPython/Jupyter notebooks

Standard operating procedure Outlier Removal In Large-Sample Epidemiological Transcriptomics Datasets[11]

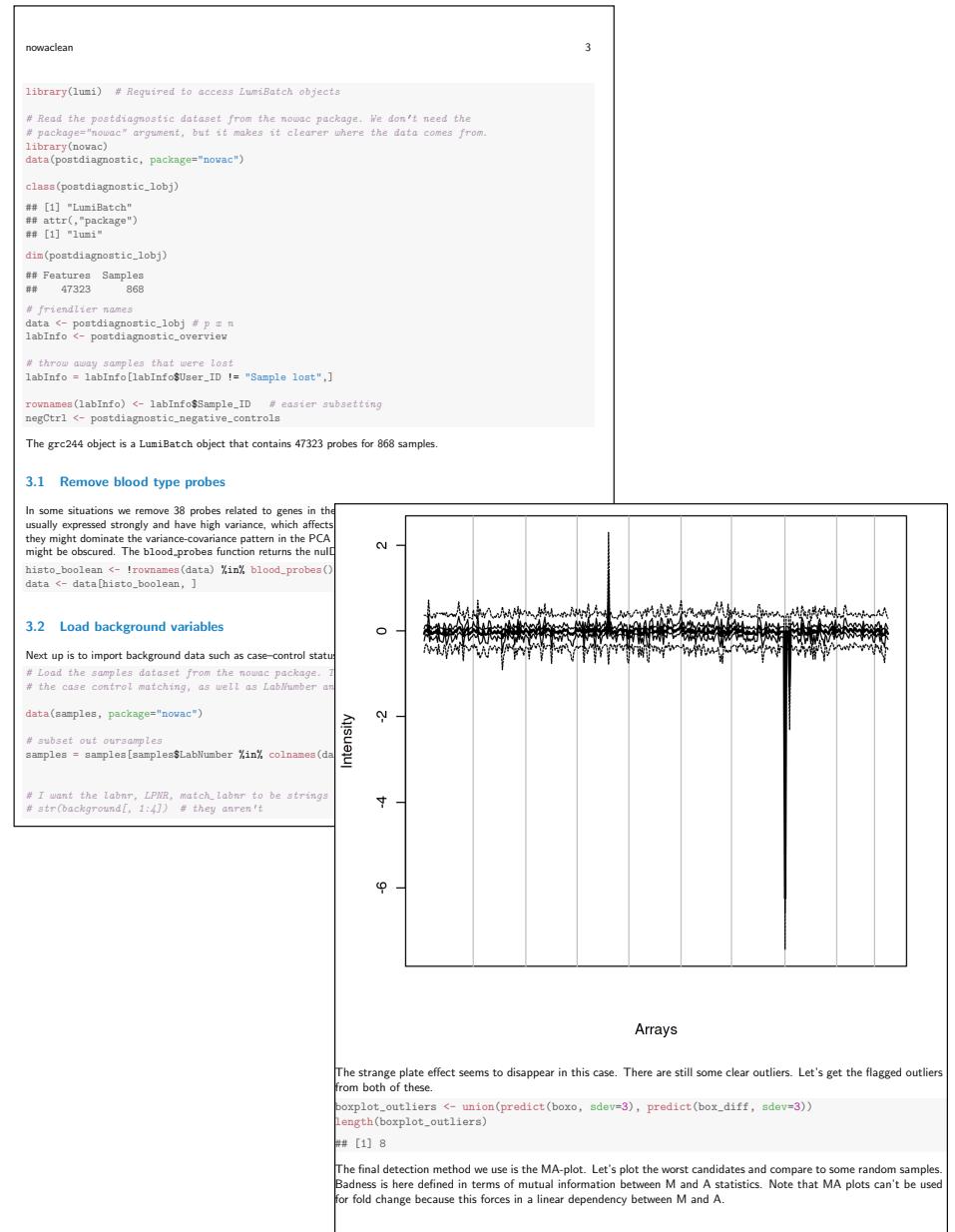
```

71 library(lumi) # Required to access LumiBatch objects
72
73 # Read the postdiagnostic dataset from the nowac package. We don't need the
74 # package="nowac" argument, but it makes it clearer where the data comes from.
75 library(nowac)
76 data(postdiagnostic, package="nowac")
77
78 class(postdiagnostic_lobj)
79 dim(postdiagnostic_lobj)
80
81 # friendlier names
82 data <- postdiagnostic_lobj # p x n
83 labInfo <- postdiagnostic_overview
84
85 # throw away samples that were lost
86 labInfo = labInfo[labInfo$User_ID != "Sample lost",]
87
88 rownames(labInfo) <- labInfo$Sample_ID # easier subsetting
89 negCtrl <- postdiagnostic_negative_controls
90 @
91 The \Rcode{grc244} object is a \Rcode{LumiBatch} object that contains 47323
92 probes for 868 samples.
93
94 \subsection{Remove blood type probes}
95 In some situations we remove 38 probes related to genes in the
96 human leukocyte antigen (HLA) system. These are usually expressed strongly and
97 have high variance, which affects multivariate analyses. Specifically we
98 have seen that they might dominate the variance-covariance pattern in the PCA
99 transformation of the data, and as such other patterns might be obscured.
100 The \Rfunction{blood_probes} function returns the
101 nuIDs of these probes.
102 <<names and hist>>
103 histo_boolean <- !rownames(data) %in% blood_probes()
104 data <- data[histo_boolean, ]
105 @
106
107 \subsection{Load background variables}
108 Next up is to import background data such as case--control status and similar:
109 <<load bg>>
110 # Load the samples dataset from the nowac package. This dataset contains
111 # the case control matching, as well as LabNumber and LPNR that we need.
112
113 data(samples, package="nowac")
114
115 # subset out oursamples
116 samples = samples[samples$LabNumber %in% colnames(data), ]
1:1 (Top Level) ▾

```

R Sweave ▾

57



Current Trends

Reproducibility Project: Cancer Biology

- Provide evidence about the reproducibility of preclinical cancer research, and what influences reproducibility [10]
- Aim to reproduce the findings of high-profile papers in cancer research
- While the first five replication studies are complete, it is too early to draw any conclusions

SC Reproducibility Initiative

- Artifact Description (AD) mandatory for all papers submitted to the SC19 Technical Program, Artifact Evaluation (AE) optional [9]
- AD should describe the computing environment. Details of software, hardware, and data.
- AE describes how the authors verified and validated their computational results

Challenges and Conclusion

Challenges and Conclusion

- Scientific Community
 - Incentivize reproducibility studies and efforts by researchers who strive for reproducible research
 - Science should be "Show me," not "trust me"
- Practical and technical
 - Getting hold of data and software behind a scientific publication
 - Capture, share, and reuse the full and correct execution environments

References

- [1] Ivie, P. and Thain, D., 2018. Reproducibility in Scientific Computing. *ACM Computing Surveys (CSUR)*, 51(3), p.63.
- [3] Claerbout, J.F. and Karrenbach, M., 1992. Electronic documents give reproducible research a new meaning. In *SEG Technical Program Expanded Abstracts 1992* (pp. 601-604). Society of Exploration Geophysicists.
- [2] Baker, M., 2016. 1,500 scientists lift the lid on reproducibility. *Nature News*, 533(7604), p.452.
- [3] Challenges in irreproducible research. 2018. *Nature*. nature.com/collections/prbfkwmwvz
- [4] Stark, P.B., 2018. Before reproducibility must come preproducibility. *Nature*, 557(7707), p.613.
- [5] Kim, Y.M., Poline, J.B. and Dumas, G., 2018. Experimenting with reproducibility: a case study of robustness in bioinformatics. *GigaScience*, 7(7), p.giy077.
- [6] Begley, C.G. and Ellis, L.M., 2012. Drug development: Raise standards for preclinical cancer research. *Nature*, 483(7391), p.531.
- [7] Prinz, F., Schlange, T. and Asadullah, K., 2011. Believe it or not: how much can we rely on published data on potential drug targets?. *Nature reviews Drug discovery*, 10(9), p.712.
- [8] Collberg, C., Proebsting, T. and Warren, A.M., 2015. Repeatability and benefaction in computer systems research. *University of Arizona TR 14*, 4.
- [9] SC Reproducibility Initiative sc19.supercomputing.org/submit/reproducibility-initiative/ and github.com/SC-Tech-Program/Author-Kit
- [10] Nosek, B.A. and Errington, T.M., 2017. Reproducibility in cancer biology: making sense of replications. *Elife*, 6, p.e23383.
- [11] Bøvelstad, H.M., Holsbø, E., Bongo, L.A. and Lund, E., 2017. A Standard Operating Procedure For Outlier Removal In Large-Sample Epidemiological Transcriptomics Datasets. *bioRxiv*, p.144519.



Why it's hard to reproduce results in software

Bjørn Fjukstad
Department of Computer Science



Testbeds

- Easily reproducible environment mainly for computer science systems research
- Example: **CloudLab** developed by the Flux Research Group at University of Utah
- Possible future trends could be to deploy Jupyter or Shiny on top of these testbeds