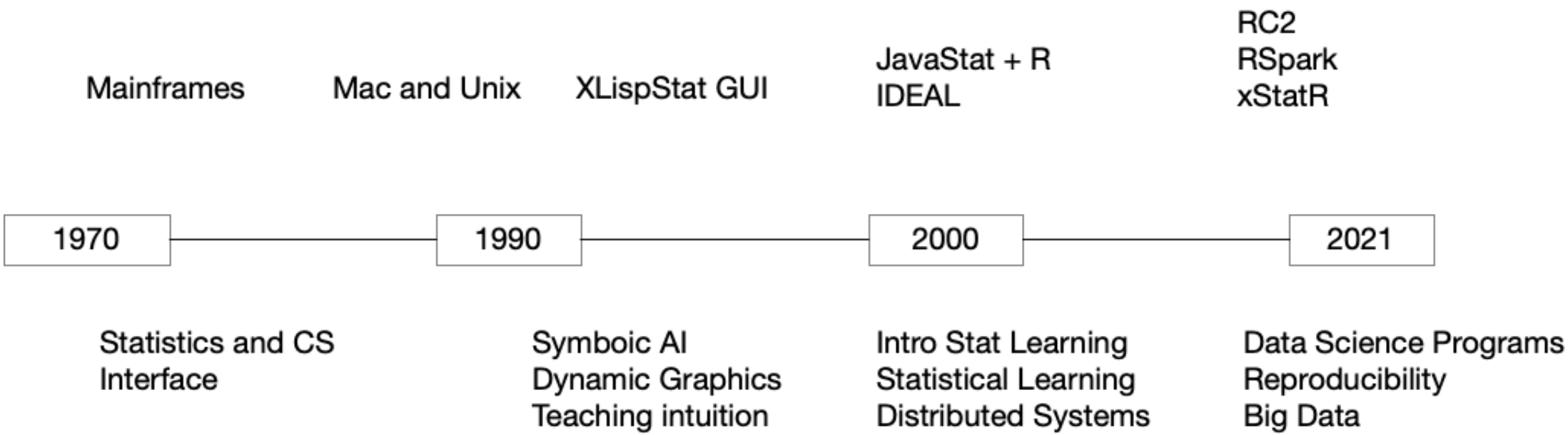
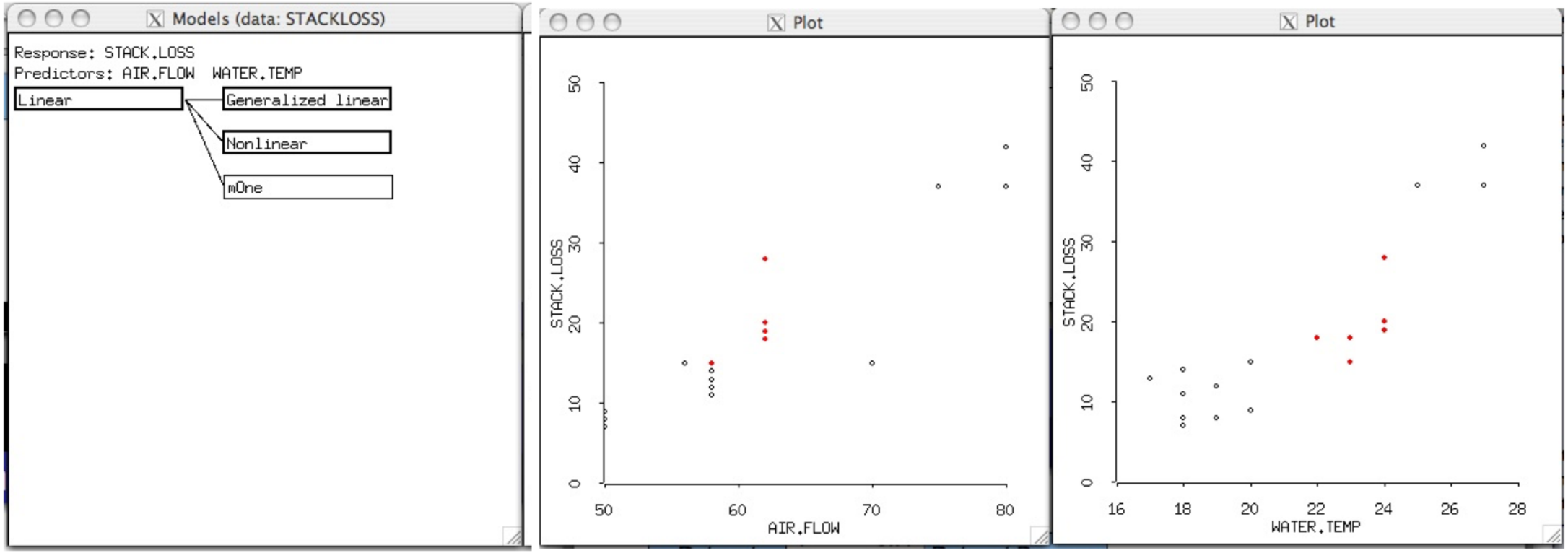
How Usability and Reproducibility in Software Improves Teaching and Research

Soren Harner, RC2AI

Jim's 50 years in statistical computing



1990s Interactive Graphics in XLispStat



2000s IDEAL and JavaStat: GUI with R Backend

JavaStat: a Java/R-based statistical computing environment

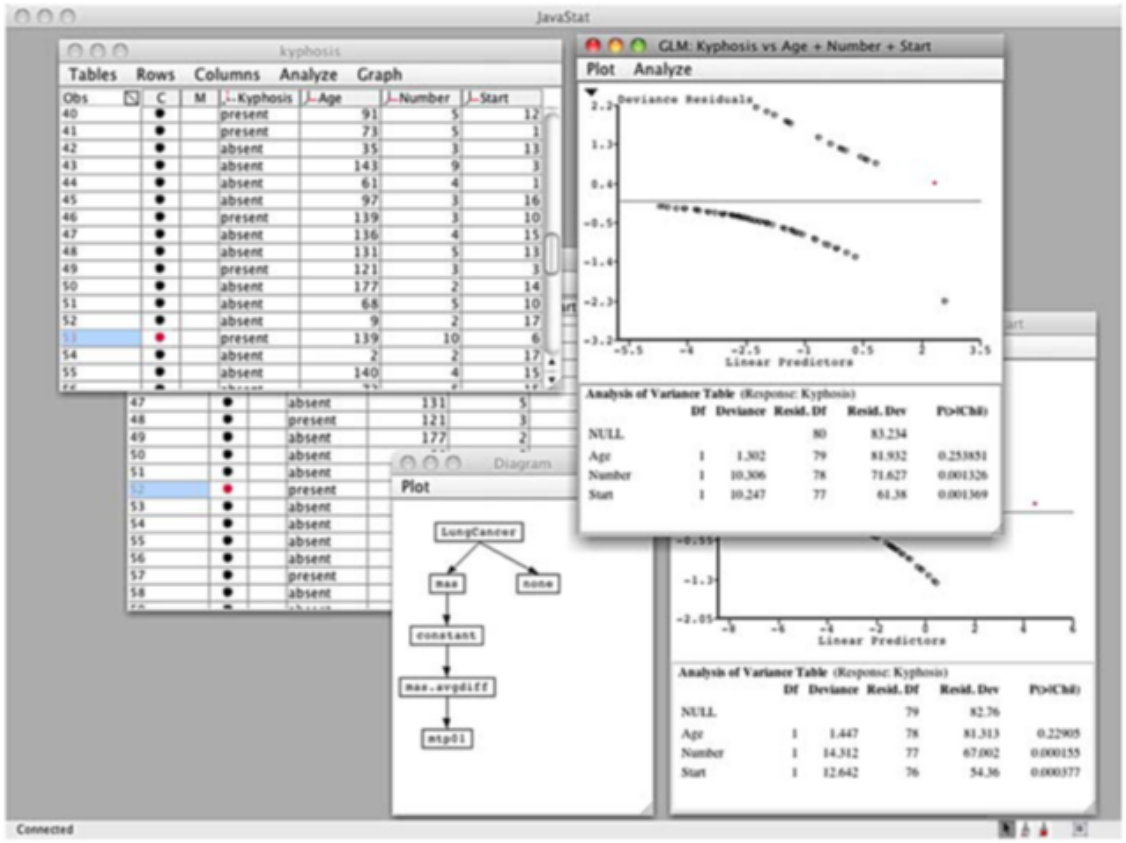
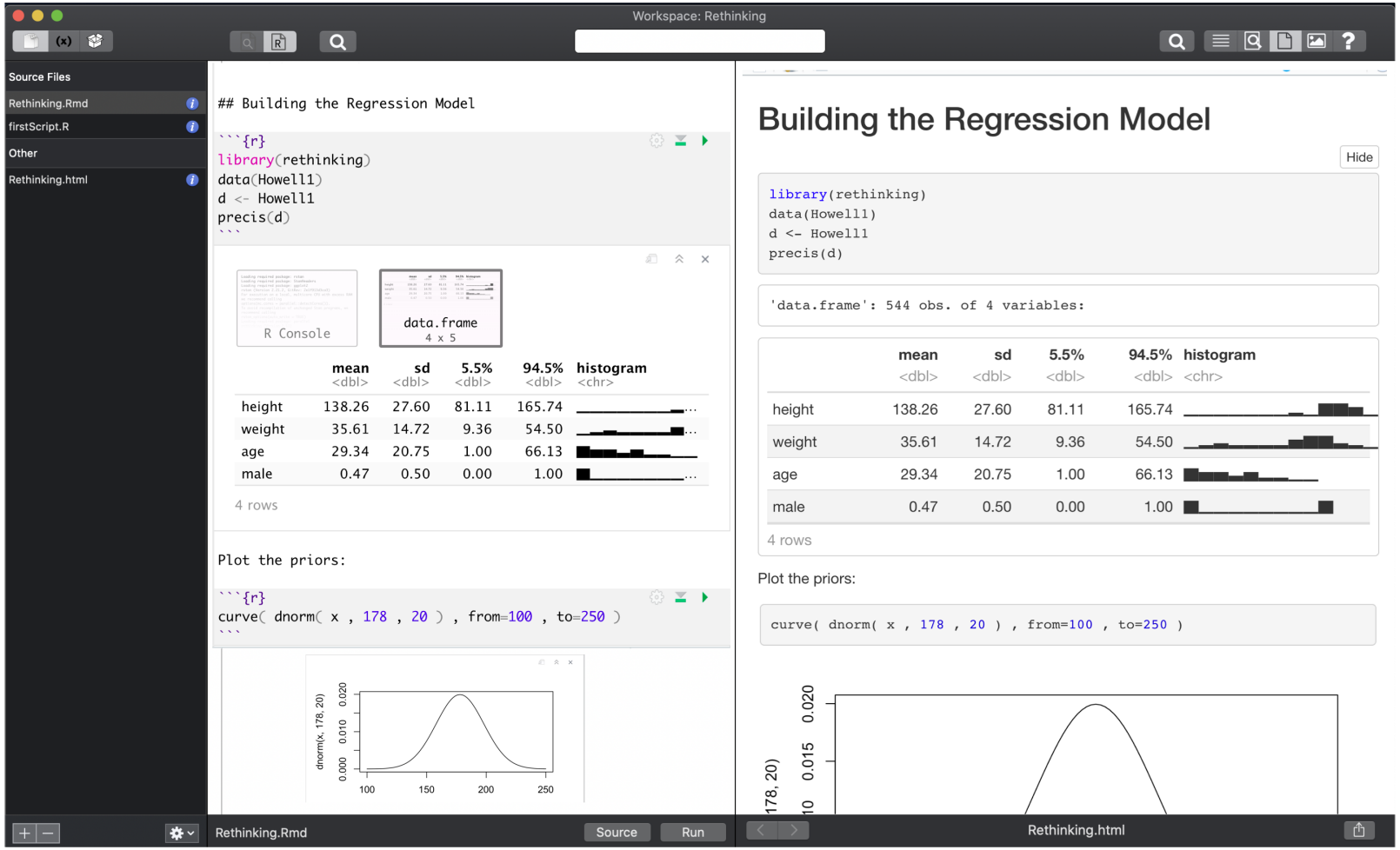


Fig. 3 JavaStat Java application

After 2015: RSpark and Reproducibility

- Teach data science courses and seminars
- Introduce SQL, DataFrames, MapReduce, Streaming
- Emphasis on reproducibility with Docker and Git
- Built on R, Spark, Apache Arrow, and Postgres
- Building on rocker, R on Docker
- Brought back XLispStat with xStartR

Rc^2 Reproducibility for Everyone



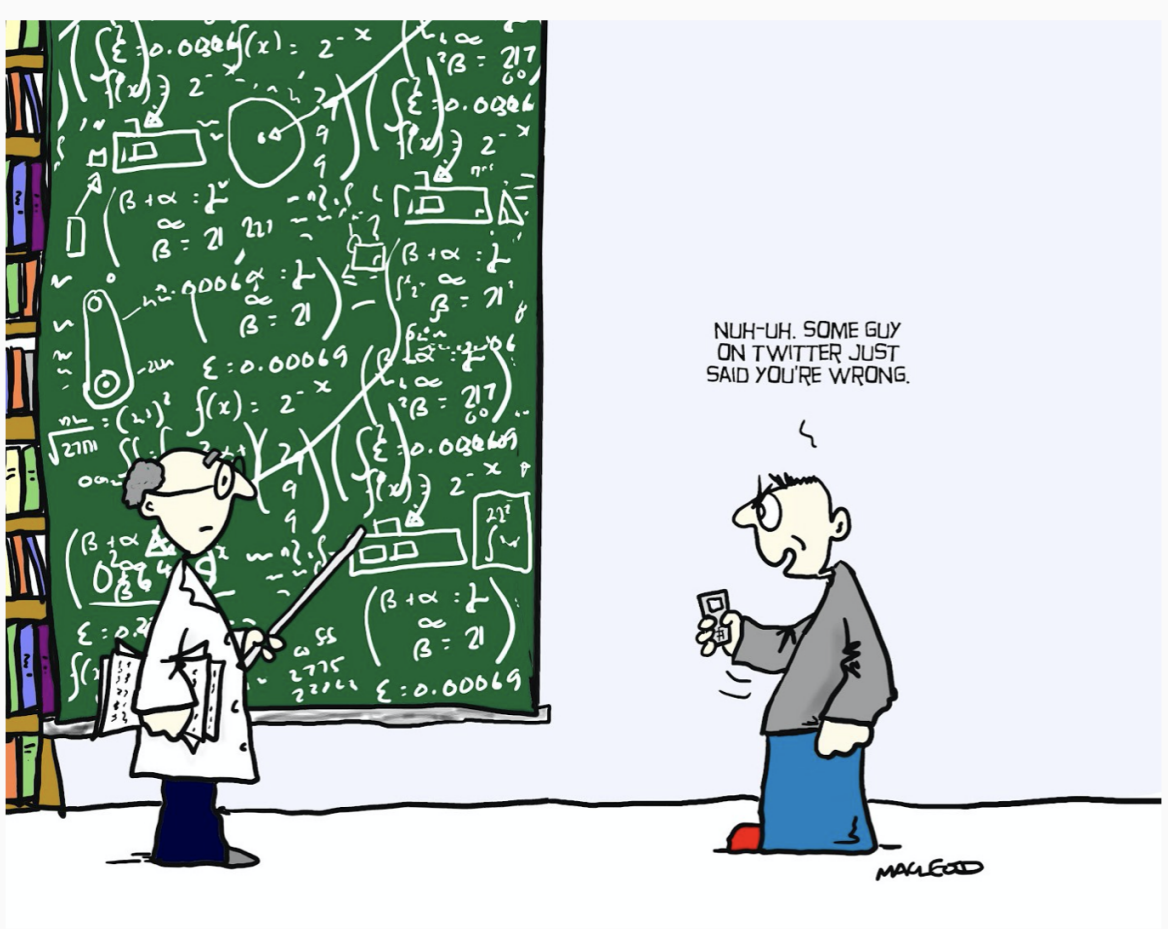

LIFE
THE UNIVERSE
42
EVERYTHING

Reproducibility

Given the same raw data, can you follow the steps and understand the assumptions of how the authors arrived at their conclusion?

Why does it matter?

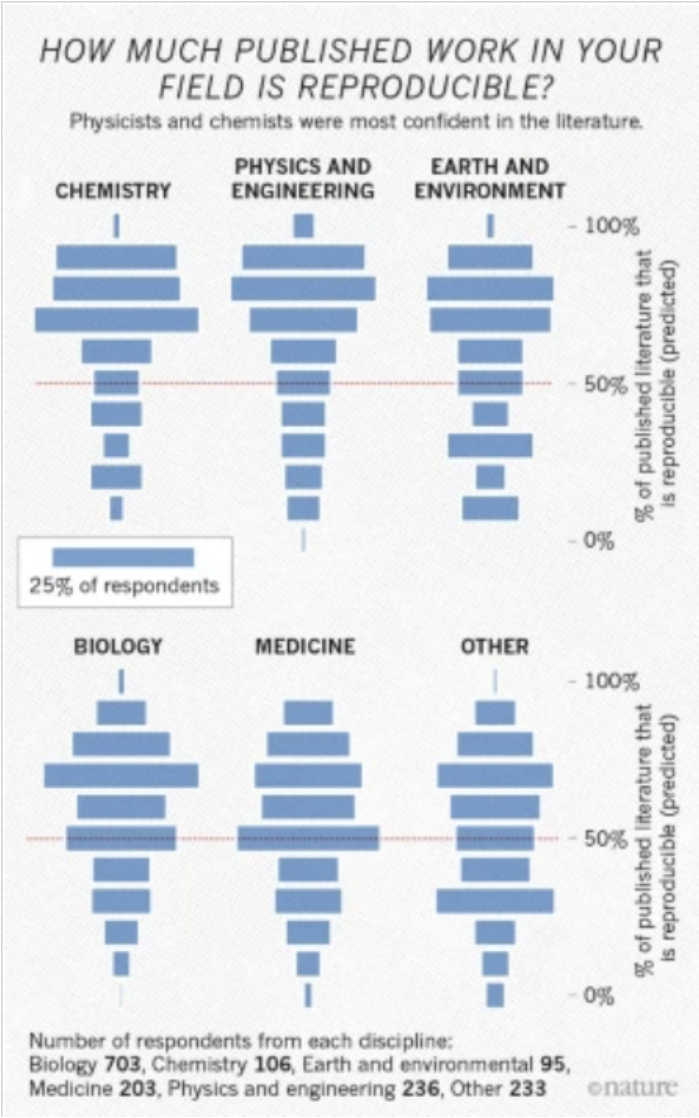
- Building on others' work
- Transparency over authority



Acceptable use from Clipart Library. 2017.

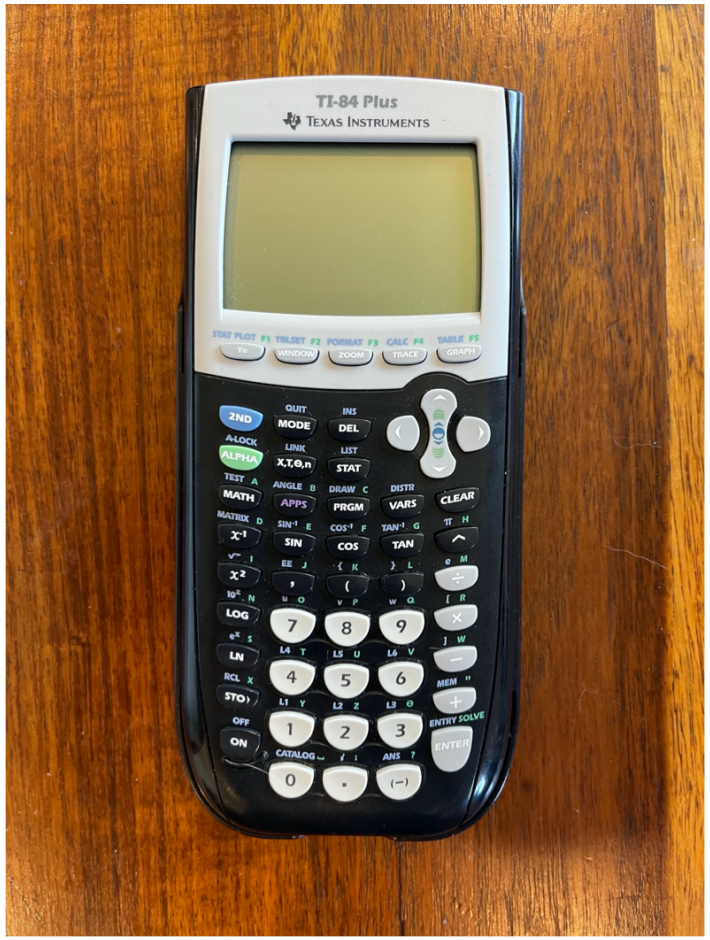
Reproducibility in research

- Nature 2016 Survey 52% say there is a crisis
- Nature 2021 Survey confidence in science leads to vaccination



Reproducibility in teaching

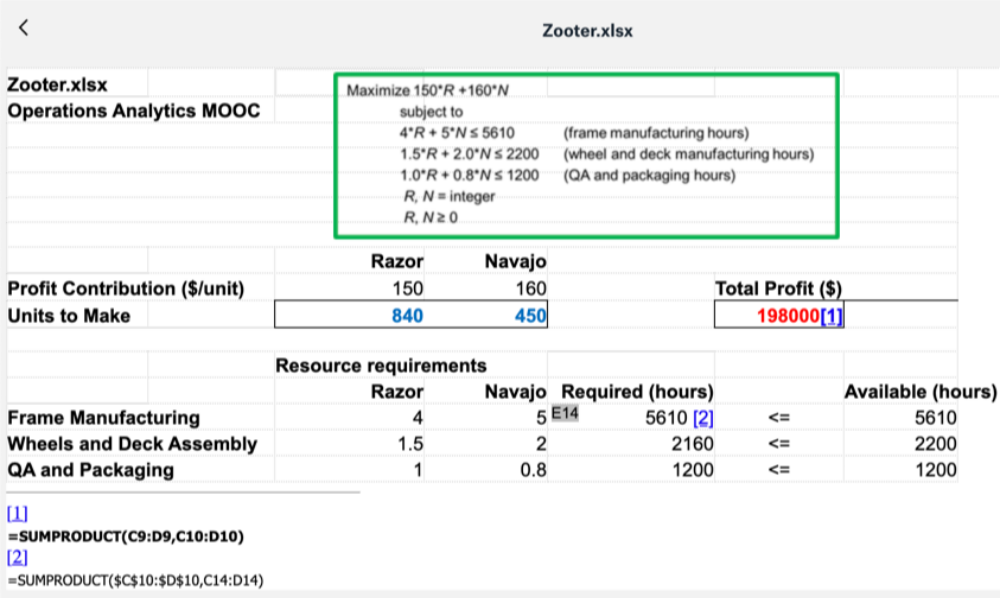
- Data plays great role in all fields
- Rise of computational sciences
- Larger models and big data



In 2021, really?

Why is it hard?

- Methods and habits
- Accessible tools and data
- Everything is versioned
- Sharing and verifying identity



Zooter.xlsx

Zooter.xlsx
Operations Analytics MOOC

Maximize 150*R +160*N
subject to
4*R + 5*N ≤ 5610 (frame manufacturing hours)
1.5*R + 2.0*N ≤ 2200 (wheel and deck manufacturing hours)
1.0*R + 0.8*N ≤ 1200 (QA and packaging hours)
R, N = integer
R, N ≥ 0

	Razor	Navajo		
Profit Contribution ($/unit)	150	160	Total Profit ($)	
Units to Make	840	450	198000[1]	

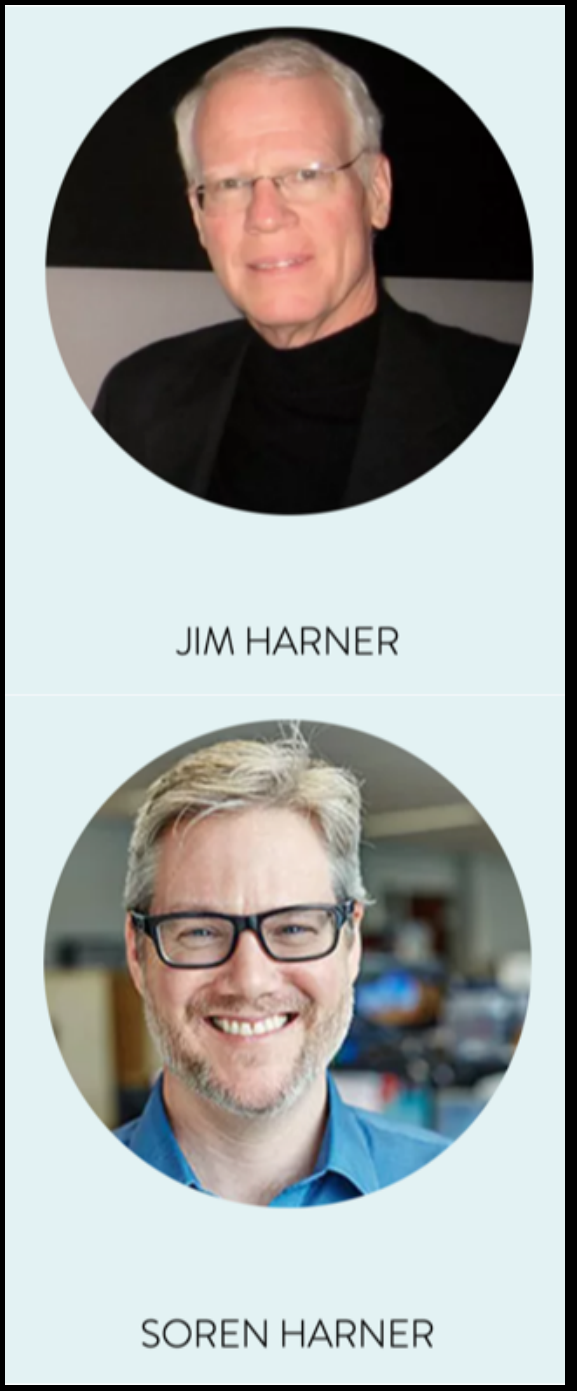
Resource requirements					
	Razor	Navajo	Required (hours)		Available (hours)
Frame Manufacturing	4	5 E14	5610 [2]	<=	5610
Wheels and Deck Assembly	1.5	2	2160	<=	2200
QA and Packaging	1	0.8	1200	<=	1200

[1]
=SUMPRODUCT(C9:D9,C10:D10)
[2]
=SUMPRODUCT($C$10:$D$10,C14:D14)

Excel fails on reproducibility

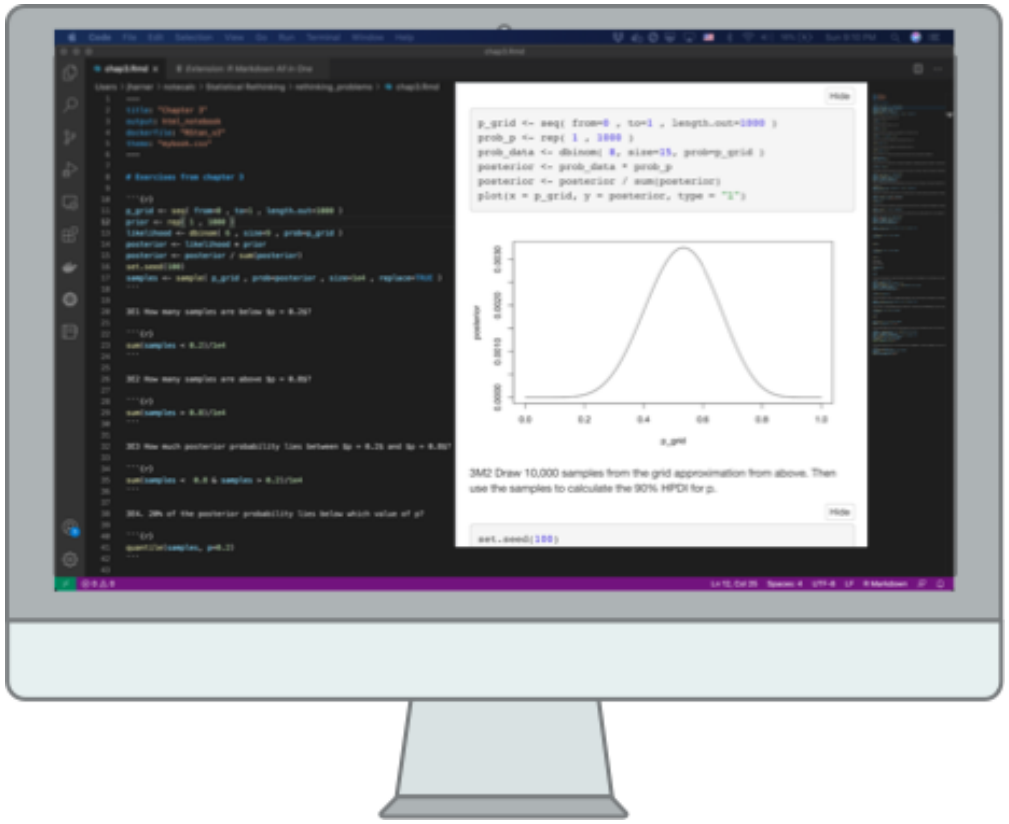
Continuing Jim's Work

Reproducibility helps teaching through sharing while it engengers skills and habits



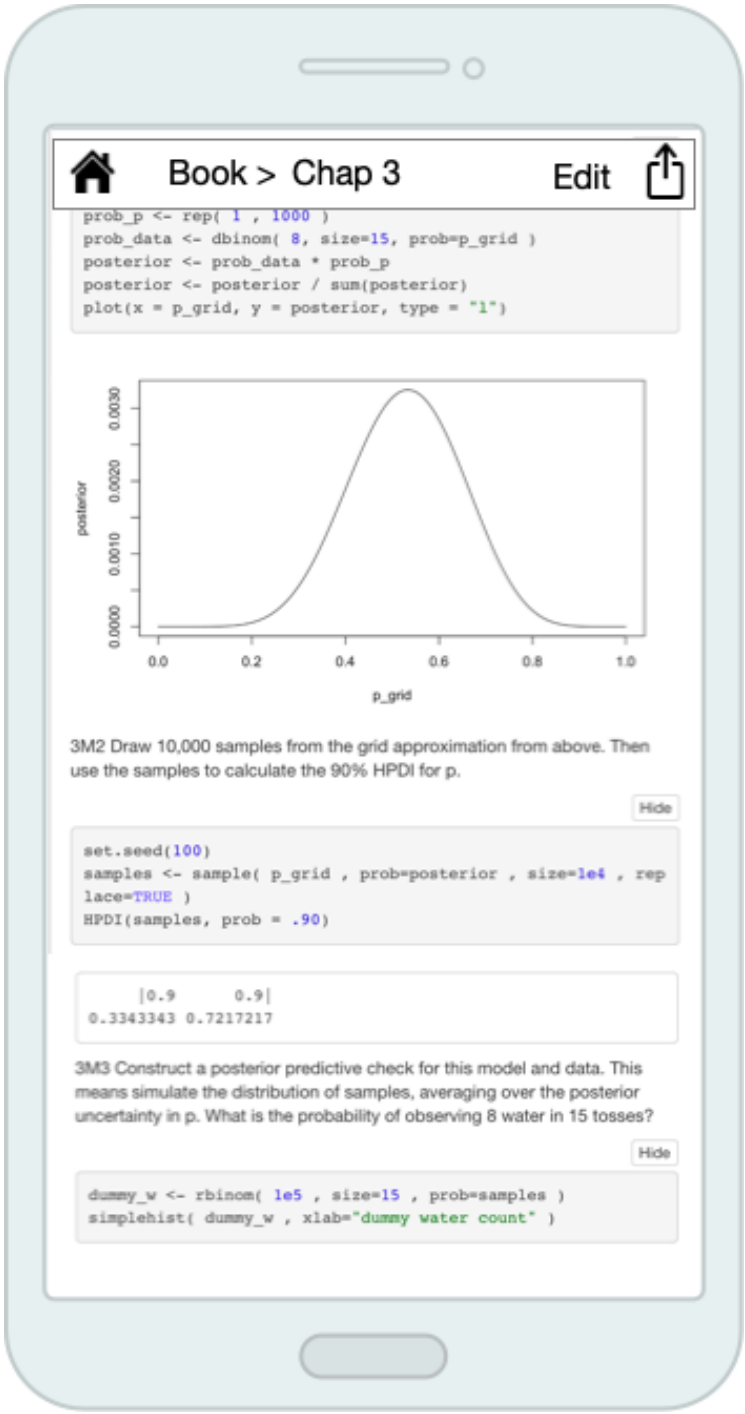
Author reproducibly

- VS Code Extension
- Versioned markdown
- Remotely executed code chunks
- Github, Bitbucket integration
- R, Python, Julia, etc.
- Javascript, CSS, vegalite
- Pull request to publish
- CLI and API

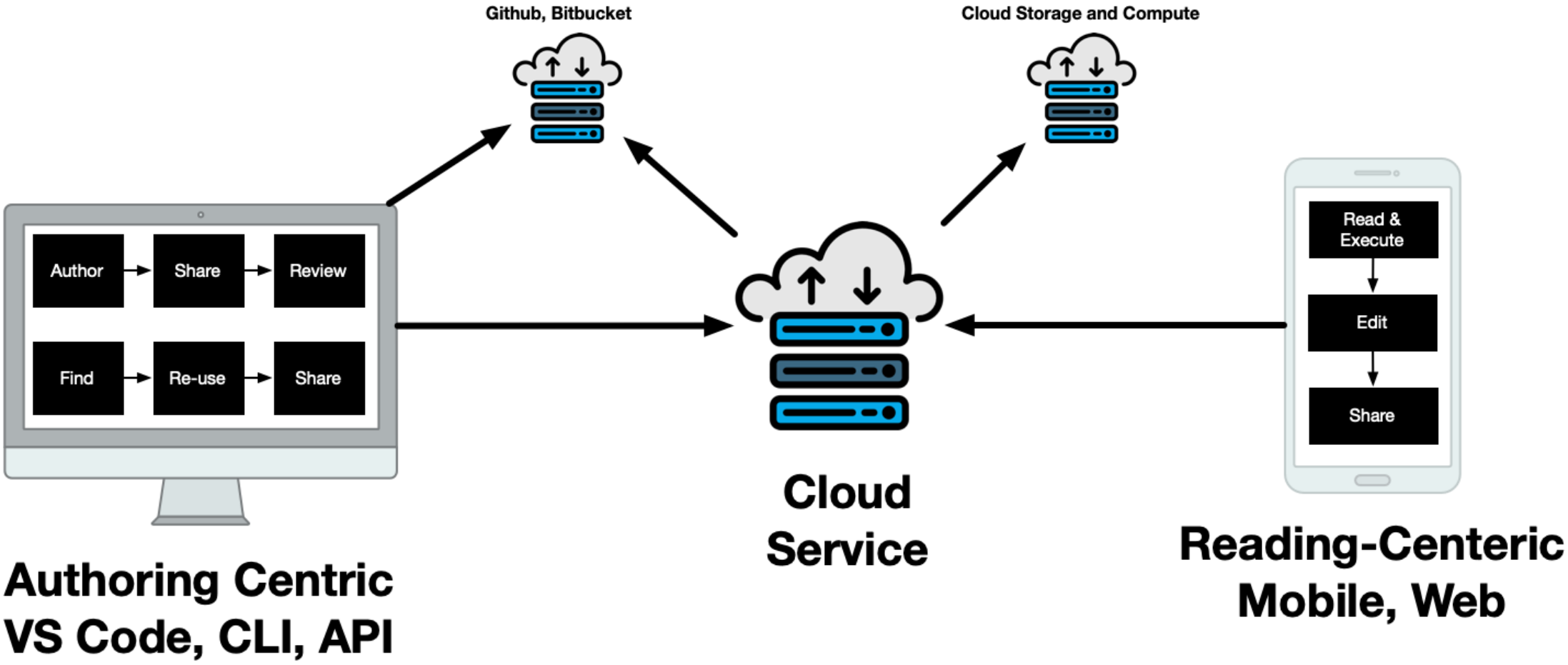


Share, read, tinker

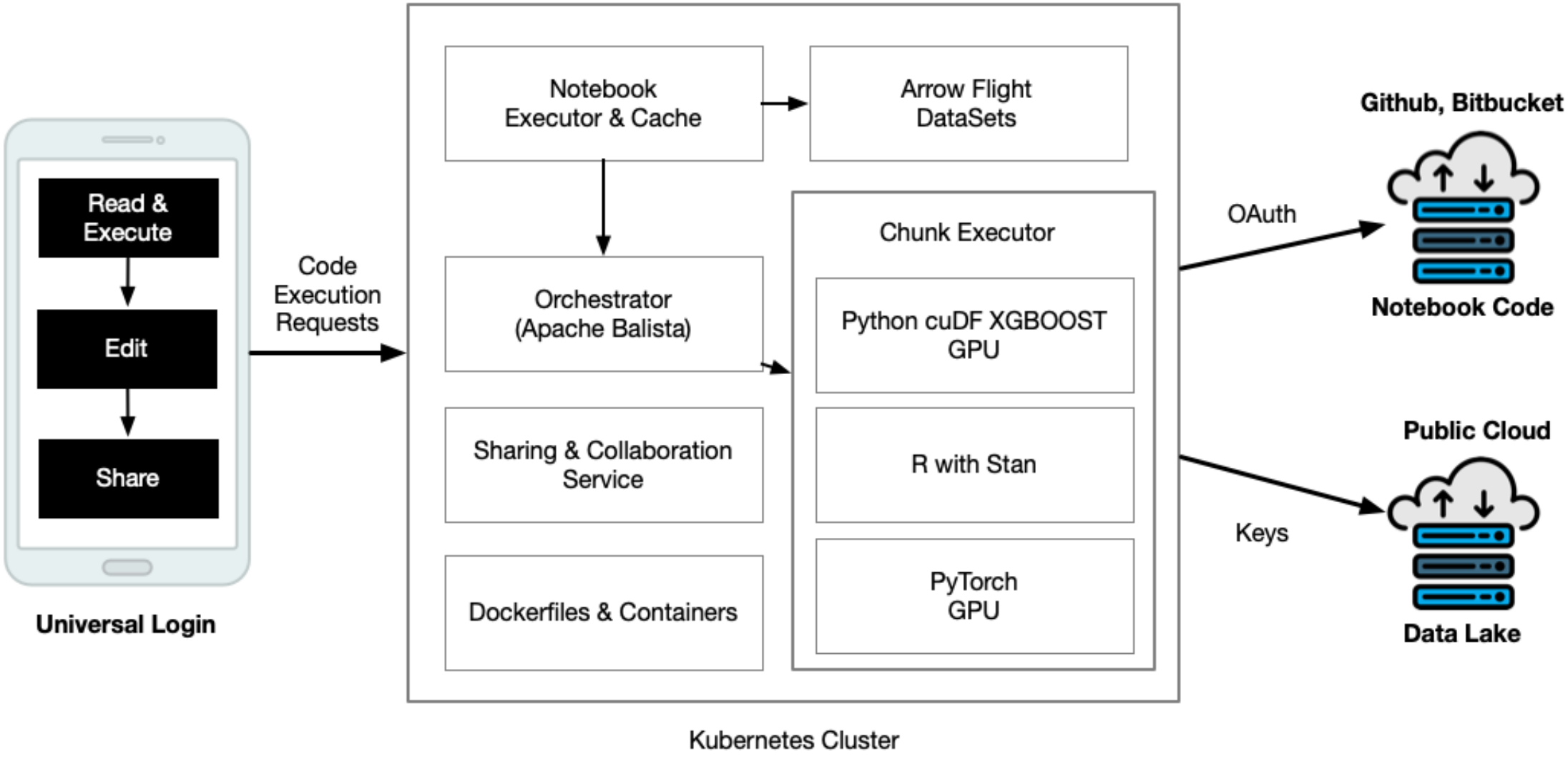
- Read, share, re-use interactive notebooks
- Mobile and Web
- Obviate the calculator
- Better than static books and reports
- Verifiable, tamperproof



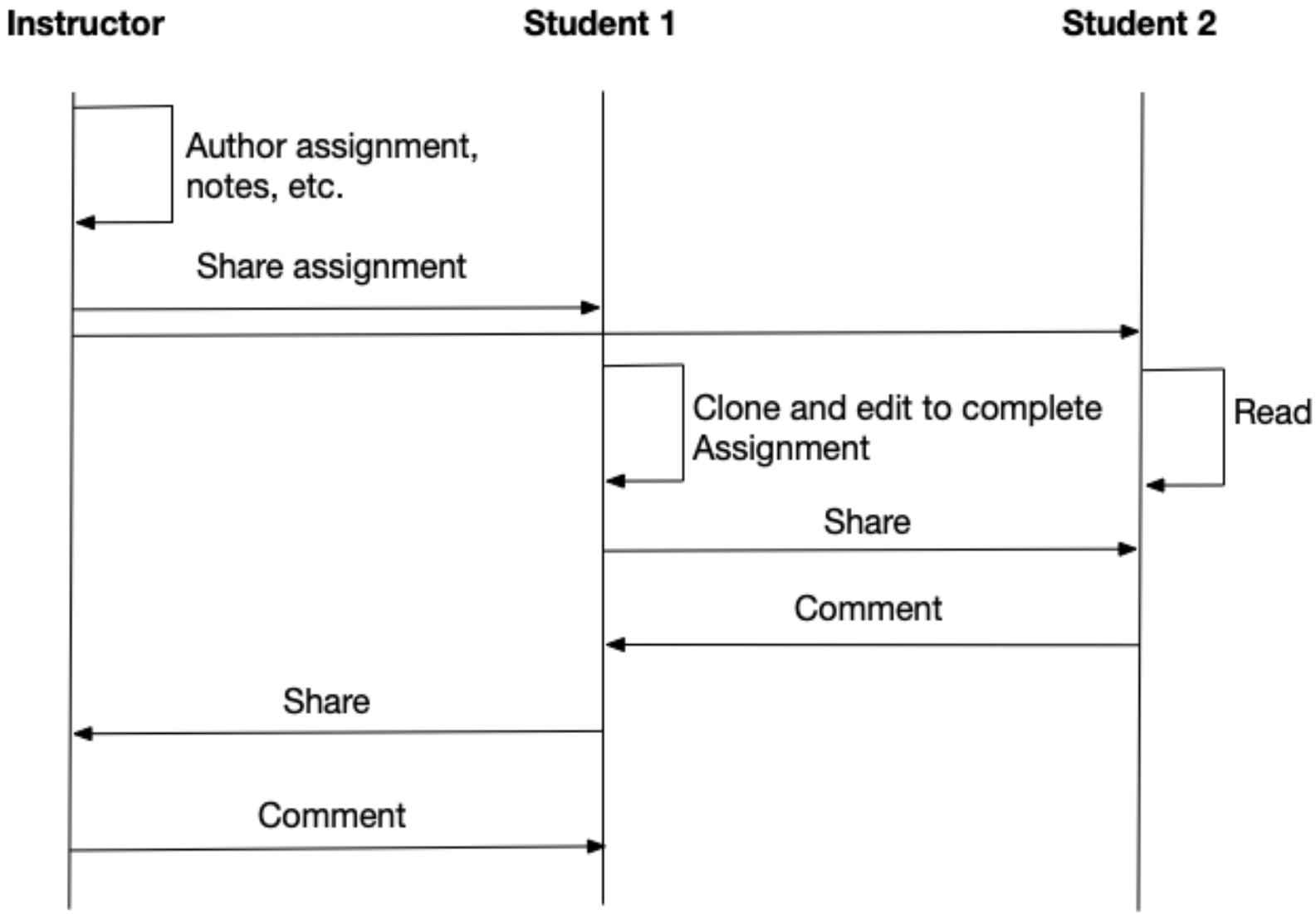
Hosted as a cloud service



Accessible, scalable cloud containers



Use-case: instructor student sharing



Next steps

- Now: Collect feedback from the community
- Sept 2021: early user testing
- Early 2022: conduct pilots in representative courses
- Develop as Open Source
- Finalize commercial component

Please get involved

- Join the pilot or give feedback (feedback and sign up)
- Looking for open source collaborators
- Some funding available for developers

Thank you

Link to presentation, references, etc.

Feedback and sign-up

Feedback Form

Presentation