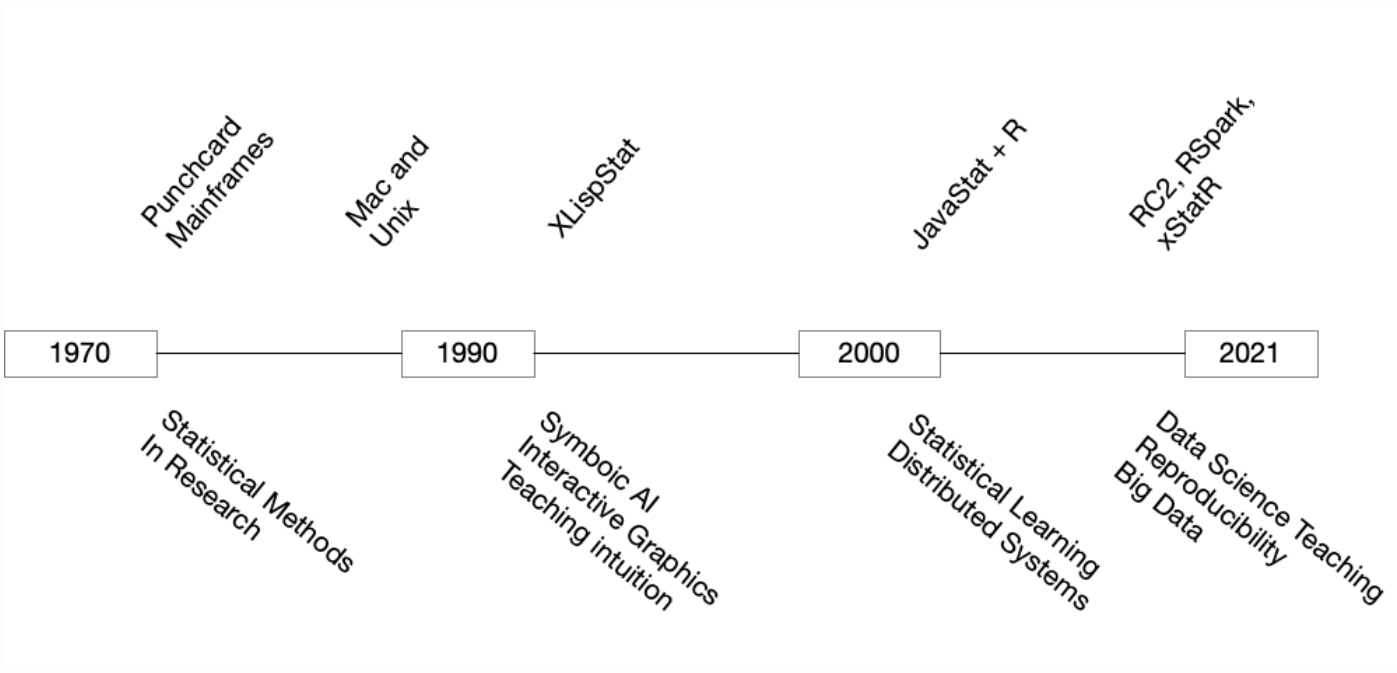
How Usability and Reproducibility in Software Improves Teaching and Research

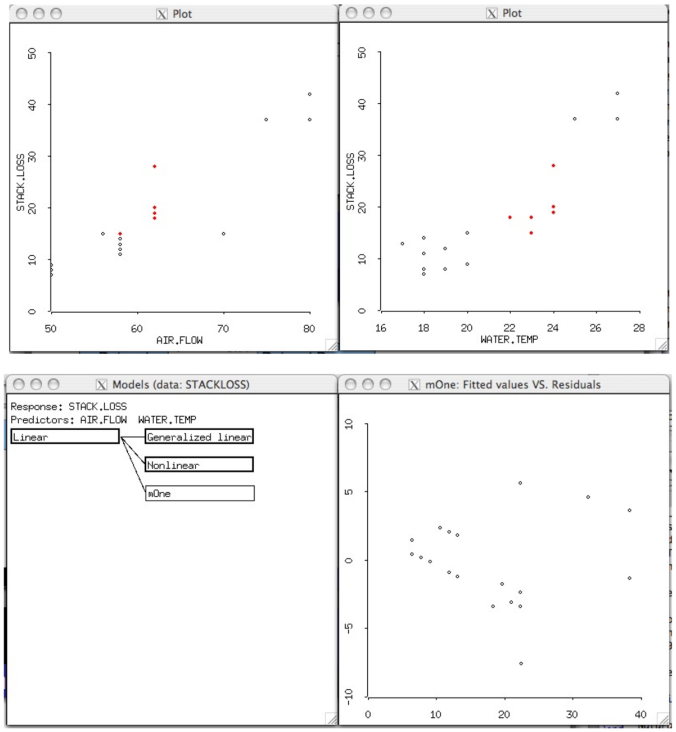
Soren Harner, RC2AI

SDSS 2021, © Soren Harner, RC2AI

Jim's 50 years in statistical computing



1990s Interactive Graphics in XLispStat



2000s JavaStat: GUI with R Backend

JavaStat: a Java/R-based statistical computing environment

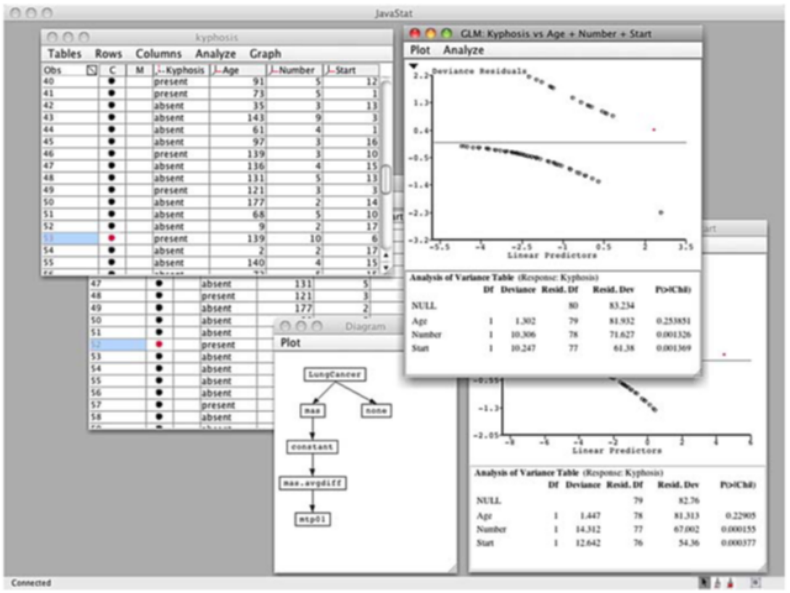
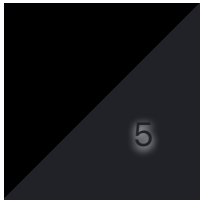


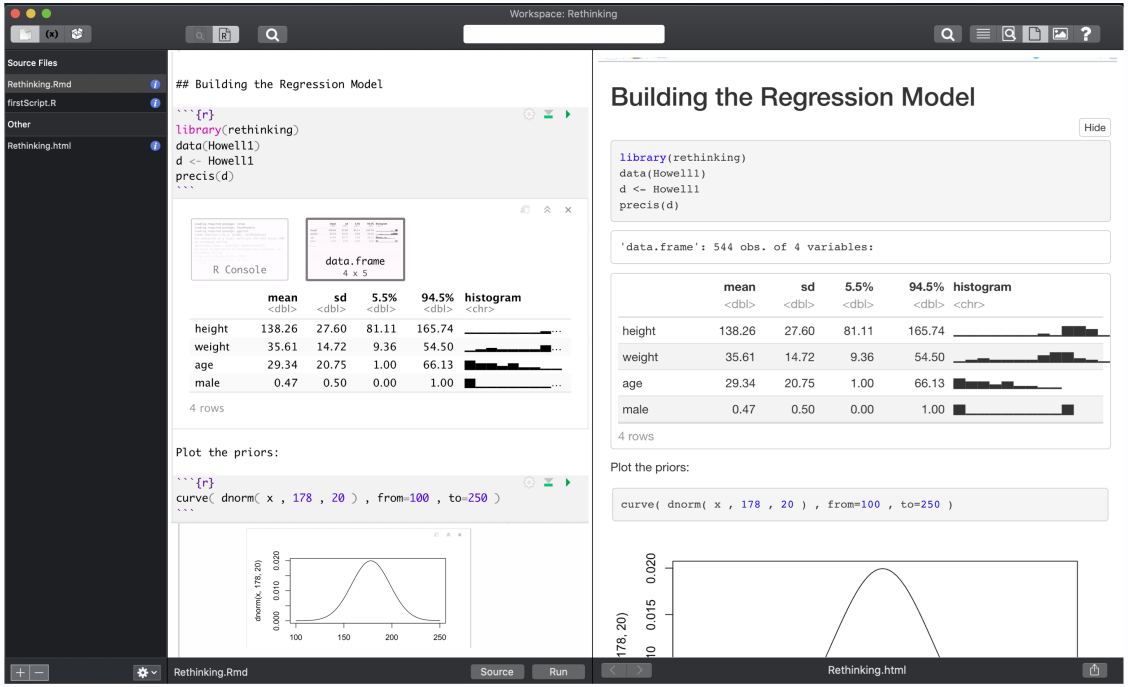
Fig. 3 JavaStat Java application

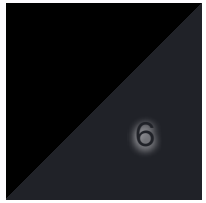
After 2015: RSpark and Reproducibility

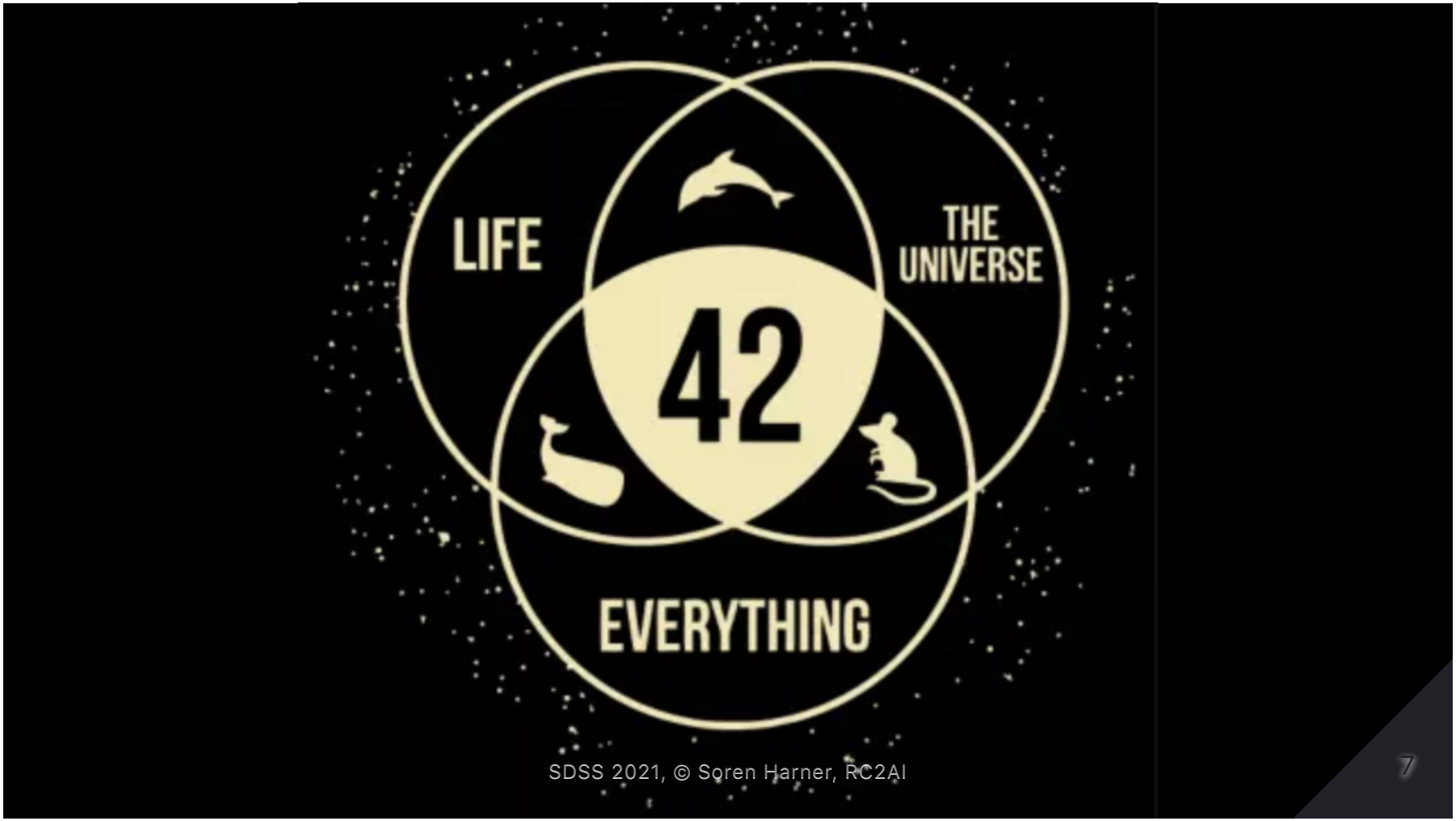
- Teach data science courses and seminars
- Introduce SQL, DataFrames, MapReduce, Streaming
- Emphasis on reproducibility with Docker and Git
- Built on R, Spark, Apache Arrow, and Postgres
- Building on rocker, R on Docker
- Brought back XLispStat with xStartR



Rc^2 Reproducibility for Everyone



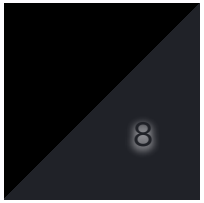



LIFE
THE UNIVERSE
42
EVERYTHING
SDSS 2021, © Soren Harner, RC2AI
7

Reproducibility

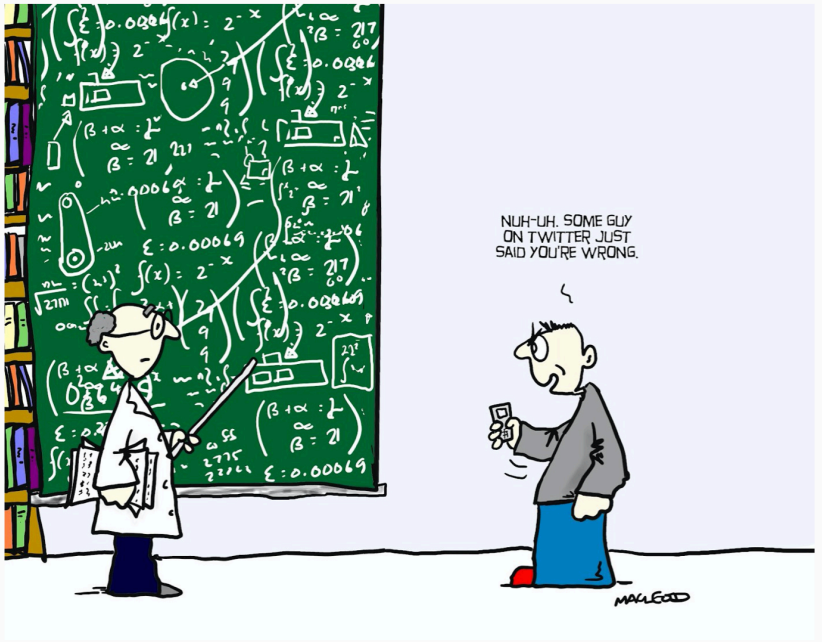
Given the same raw data, can you follow the steps and understand the assumptions of how the authors arrived at their conclusion?

SDSS 2021, © Soren Harner, RC2AI



Why does it matter?

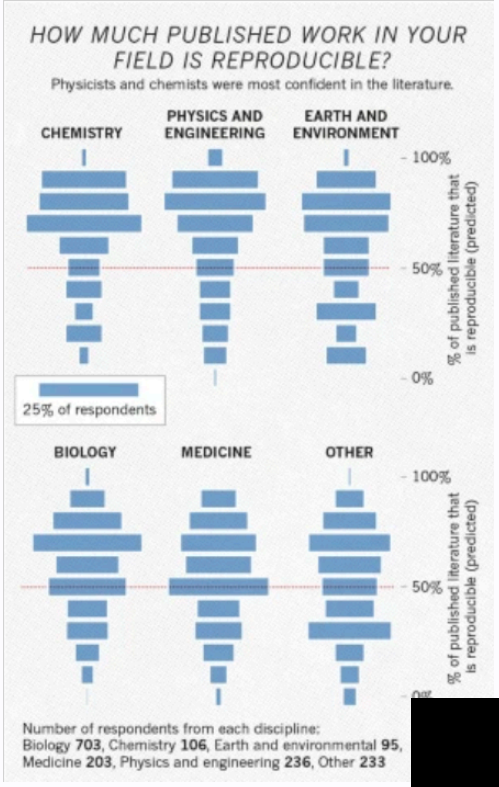
- Building on others' work
- Transparency over authority



Acceptable use from Clipart Library. 2017.

Reproducibility in Research

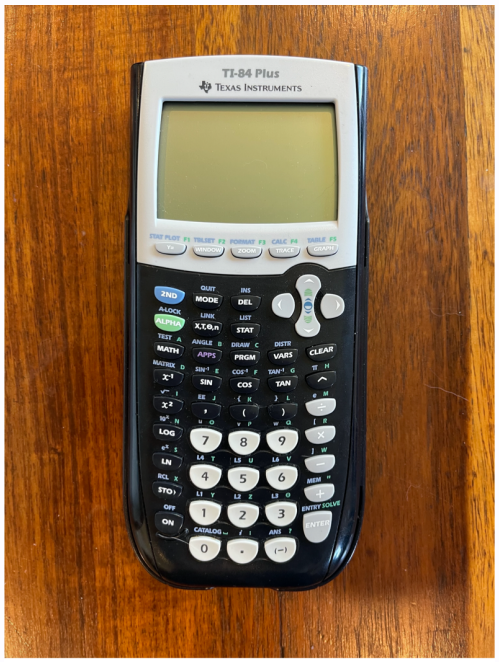
- Nature 2016 Survey 52% say there is a crisis
- Nature 2021 Survey confidence in science leads to vaccination



SDSS 2021, © Soren Harner, RC2AI

Teach reproducibility

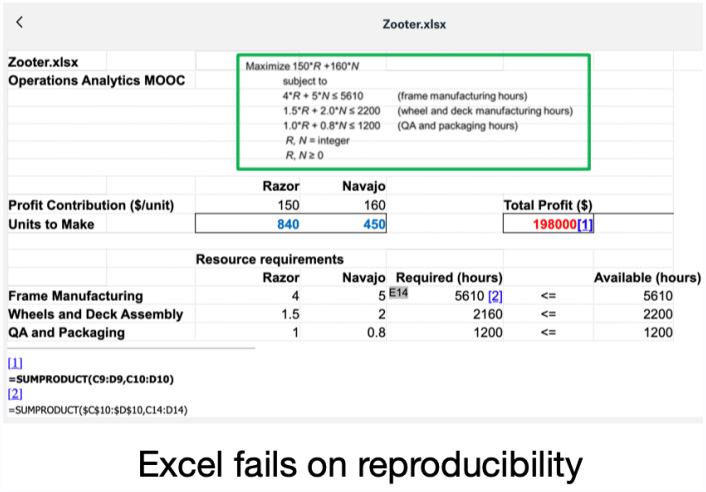
- Data used in all fields
- Computational sciences
- Larger models and big data



In 2021, really?

Why is it hard?

- Methods and habits
- Accessible tools and data
- Version everything
- Sharing and verify identity



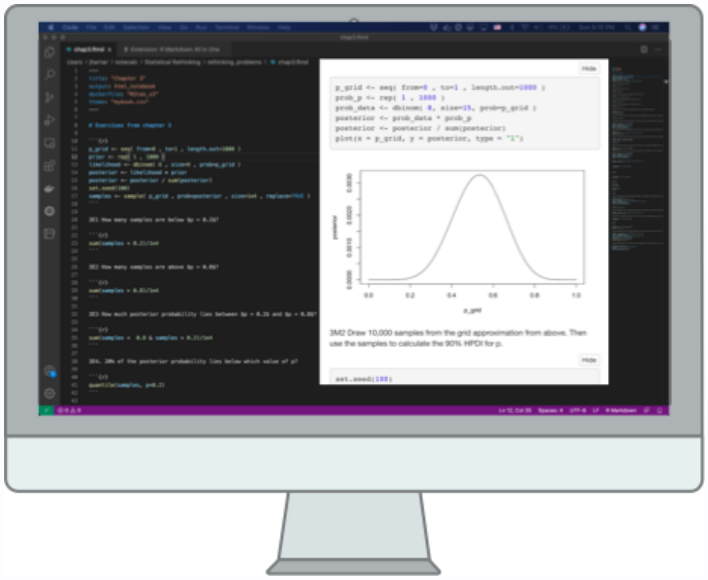
Excel fails on reproducibility

Making reproducible methods more teachable

Continuing Jim's Work

Author reproducibly

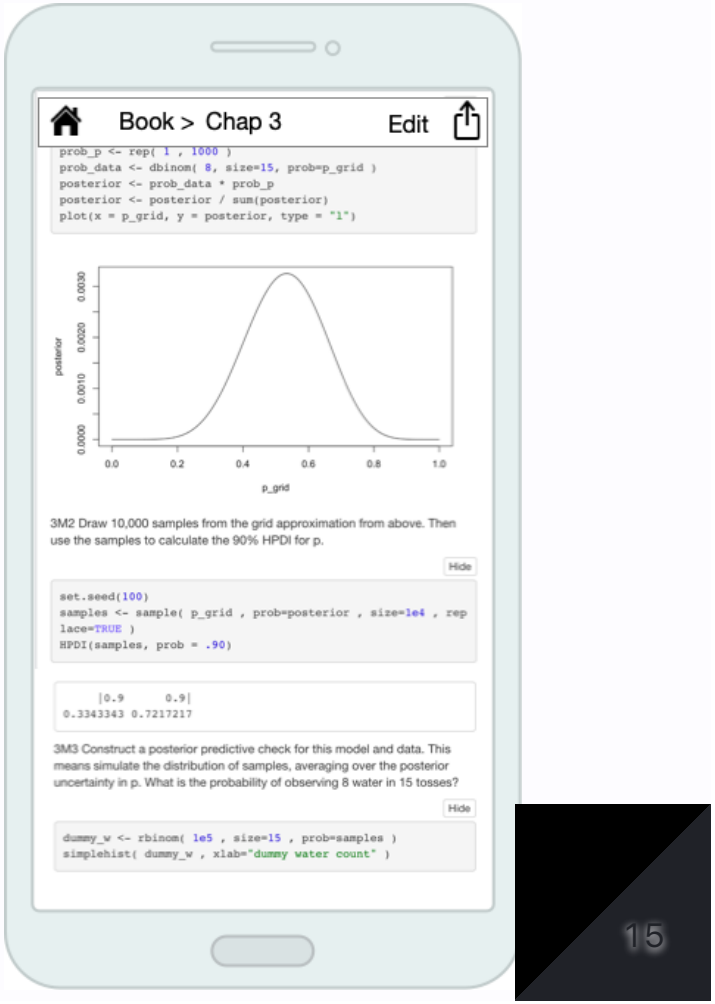
- VS Code Extension
- Versioned markdown
- Remotely executed code
- Github, Bitbucket integration
- R, Python, Julia, etc.
- Javascript, CSS, vegalite
- Pull request to publish



SDSS 2021, © Soren Harner, RC2AI

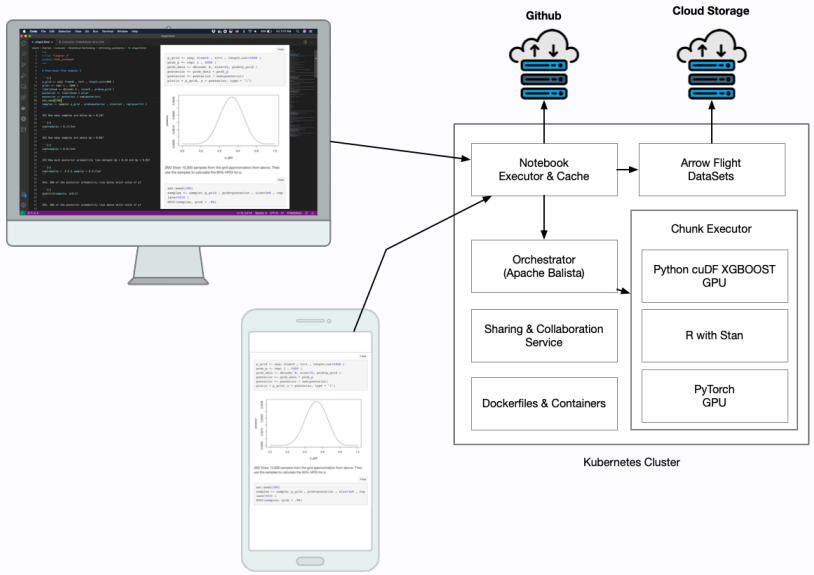
Read, tinker

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



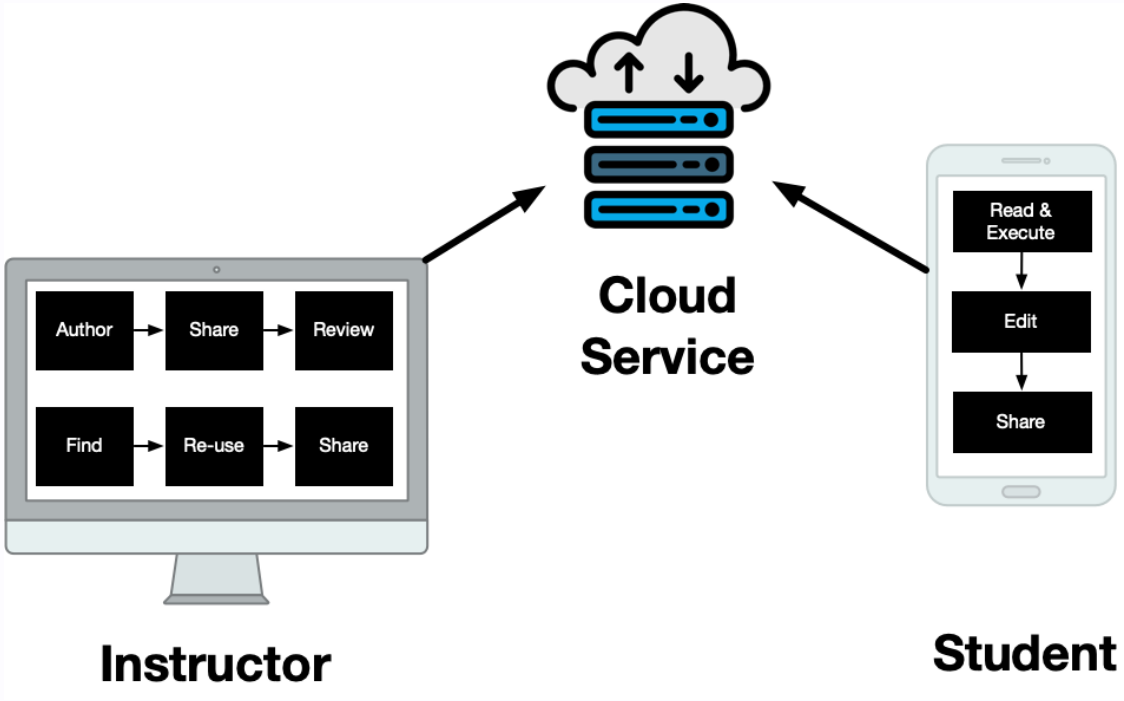
SDSS 2021, © Soren Harner, RC2AI

Accessible, scalable cloud containers

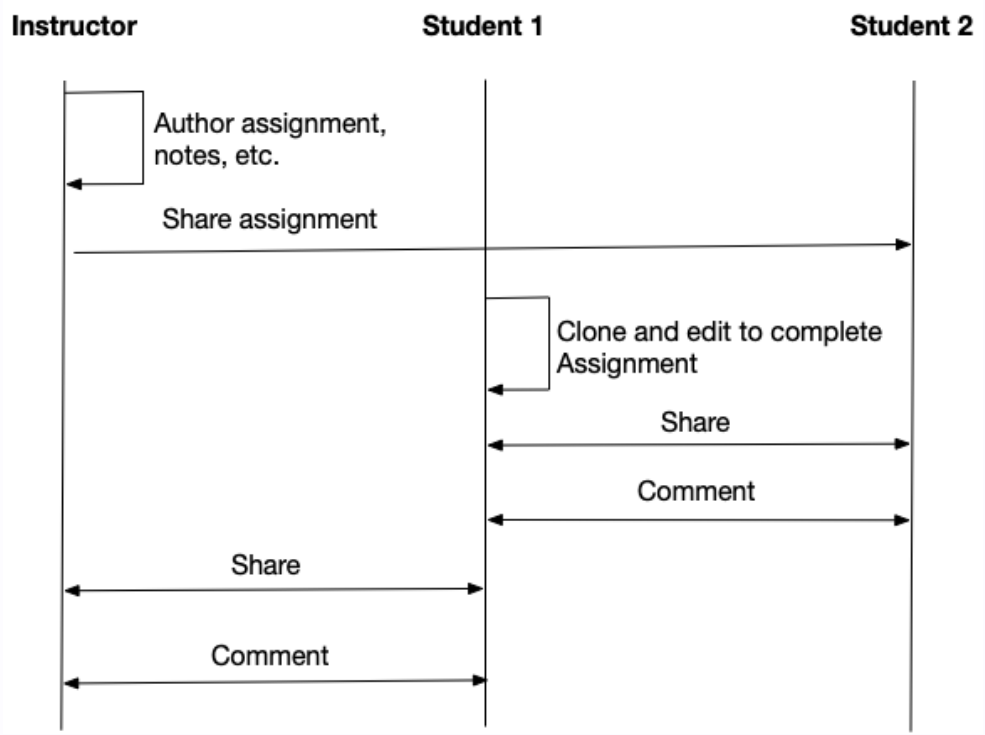


SDSS 2021, © Soren Harner, RC2AI

Integrated collaboration



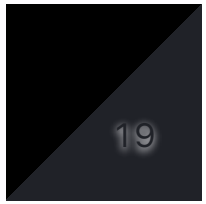
Use-case: instructor student sharing



SDSS 2021, © Soren Harner, RC2AI

Next steps

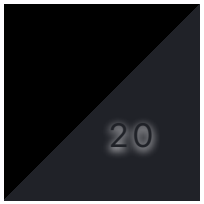
- Collect feedback
- Conduct pilots (sign up)
- Continue active development
- Finalize commercial component



You can get involved

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

SDSS 2021, © Soren Harner, RC2AI



Thank you

Link to presentation, references, etc.

SDSS 2021, © Soren Harner, RC2AI