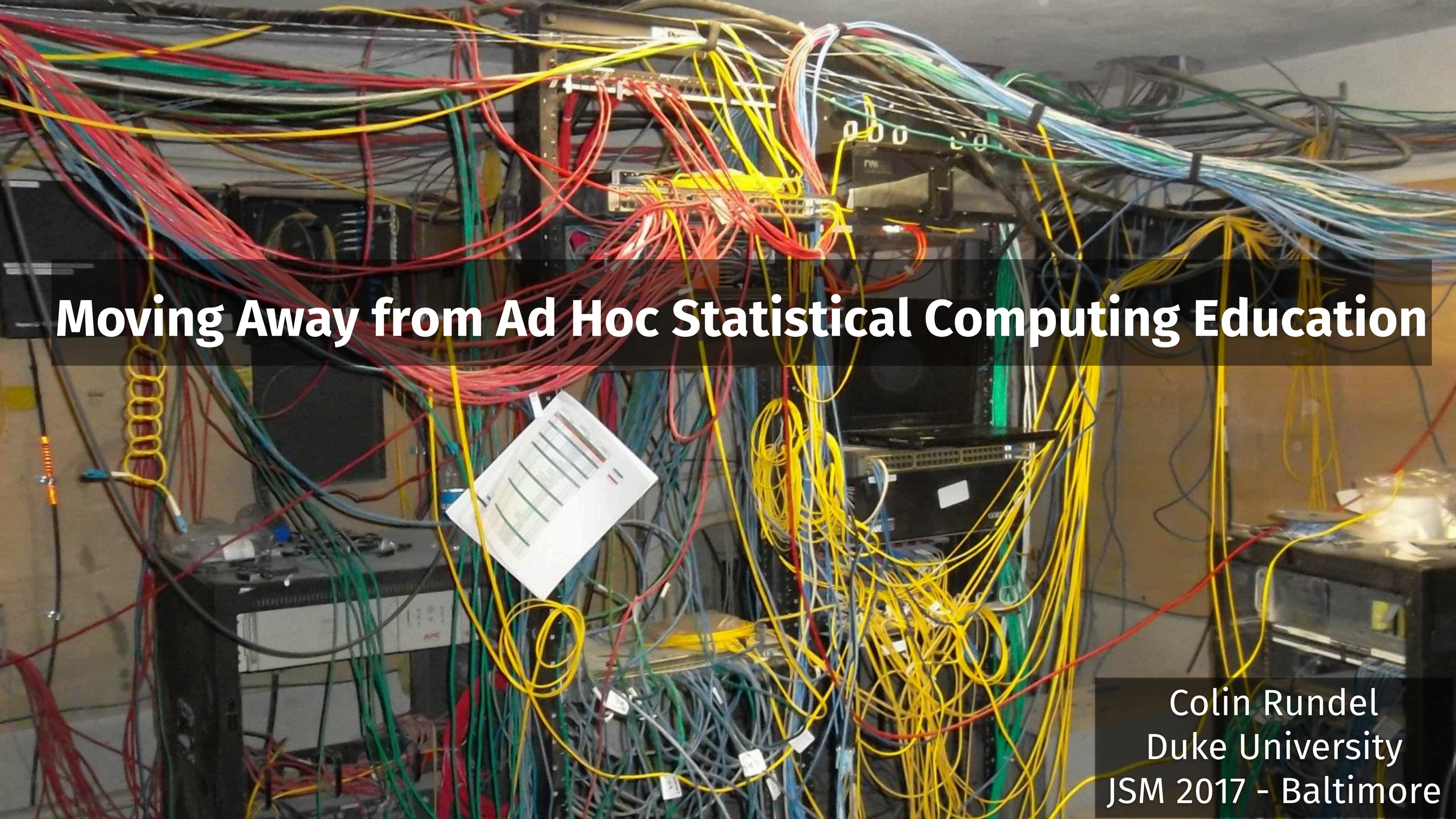


Moving Away from Ad Hoc Statistical Computing Education

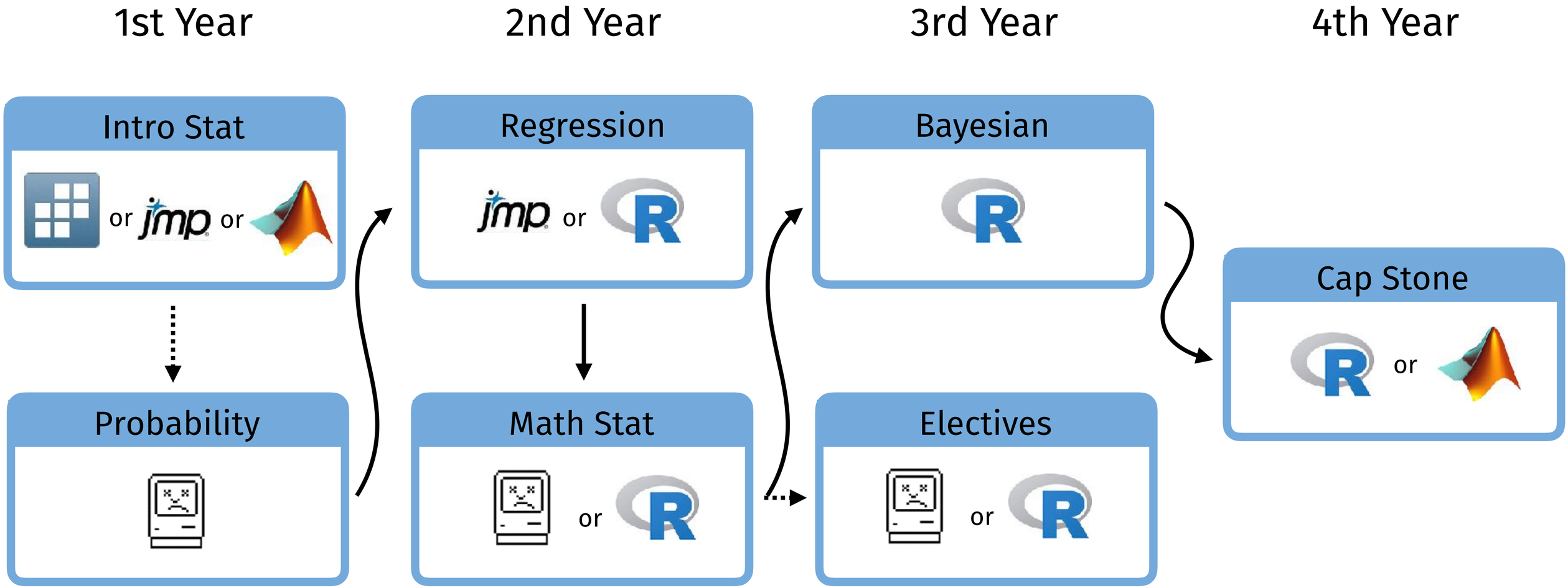
Colin Rundel
Duke University
JSM 2017 - Baltimore

Claim:

Computation is fundamental to (modern) statistics, but we often don’t teach it that way.

Where we started:

Duke StatSci Curriculum circa 2011/2012



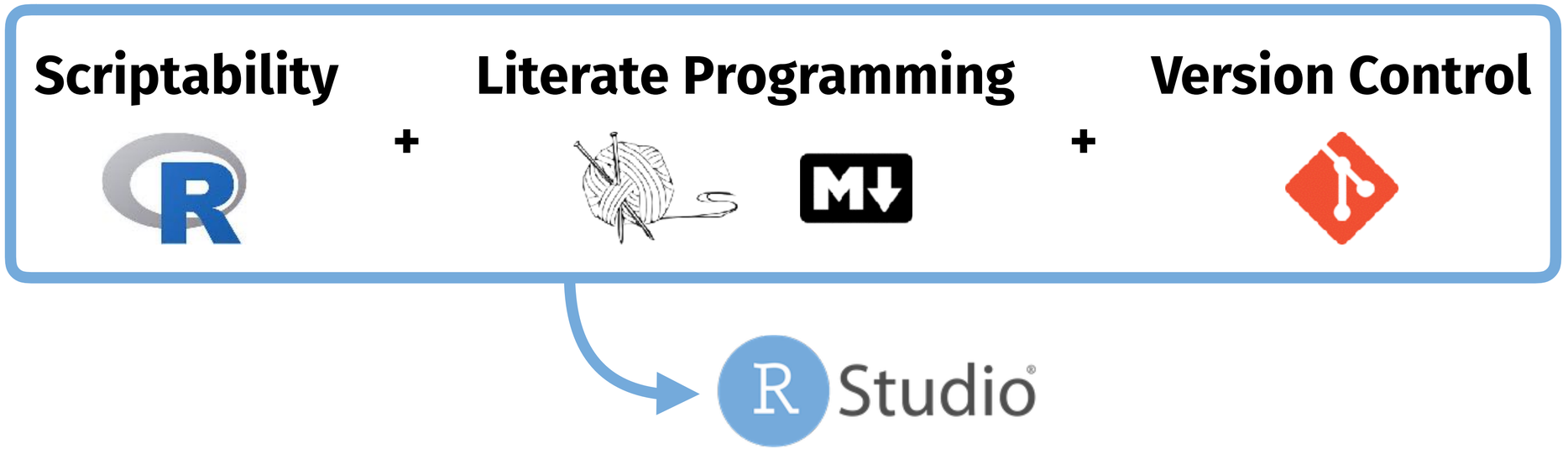
An initial foray into teaching computation ...

- OpenIntro R labs created in Fall 2011
  - Used by two Intro Stat courses (Sta 10 & Sta 101) that year,
  - and Sta 102 the next Fall.

- Right place at the right time for R,
  - RStudio server beta (Fall 2011)
  - knitr & R Markdown (Summer 2012)

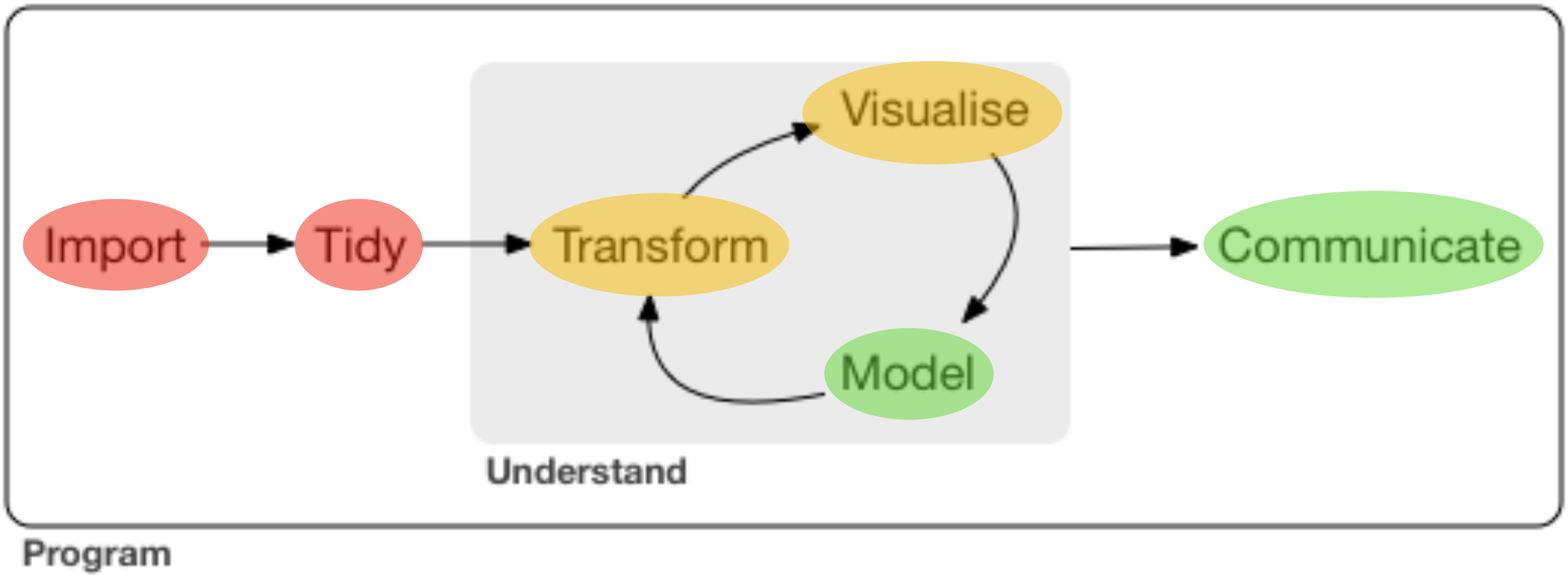
Reproducible computation

- Audience are non-stat majors coming from social and life sciences
  - Teach tools and methods that can grow with the students
  - Teach workflows that are inherently reproducible



Where we were:

How are we doing?



From *R for Data Science* by Garrett Grolemund and Hadley Wickham

Statistical Computing (Sta 323)

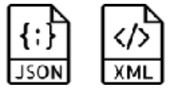
- *Optional* 2nd / 3rd year elective with *minimal* prereqs
- Offered once a year in the Spring, ~30-40 students
- Focus on team based open-ended problem solving

A programming course with statistics

vs.

A statistics course with programming

Computational Skills in Sta 323

- Programming ()
- Reproducibility ()
- Version control ()
- Unix shell ()
- Data munging ()
- Web scraping ()
- Web APIs ()
- Visualization ()
- Databases and SQL ()
- Distributed computing ()
- Interactivity ()
- ...

Rethinking gateways

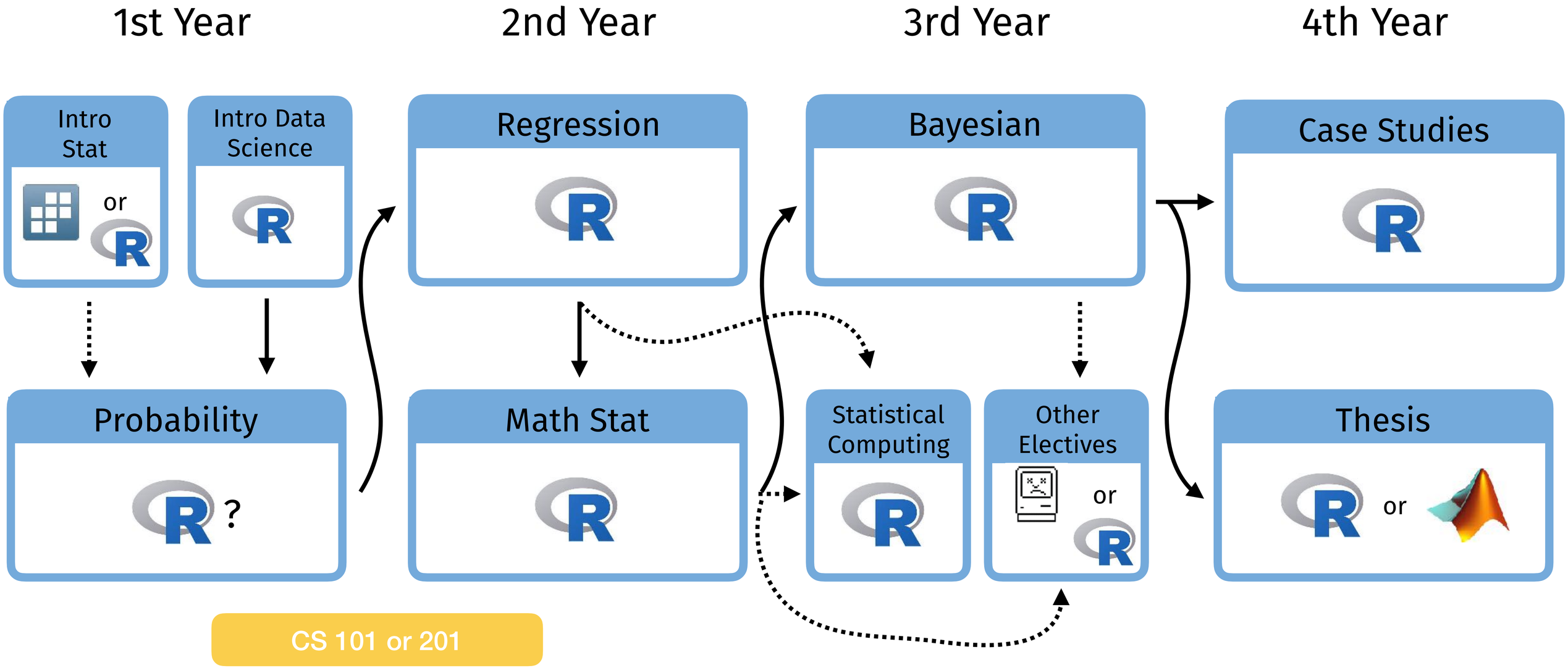
- Students are desperate to learn “*Data Science*”

- Identify the skills students find useful / exciting while still preparing students for later requirements (theoretical and computational)
  - Teach the *good* stuff in R (visualization, munging, scraping, etc.)
  - Computation complements theory

- See Mine’s slides from yesterday (bit.ly/first-ds)

Where we are:

Duke StatSci Curriculum circa 2017/2018



Lessons Learned / Future Thoughts

- There was never a master plan, mostly an exercise in identifying what we could do better / filling in the gaps
- Immediate value + downstream effects
- Healthy cross-pollination between our undergraduate and master’s curricula - demand for the same shared core skills
- Staying current is hard, but immensely valuable
- Coordination between courses is still a work in progress

Questions / Comments?

rundel@gmail.com

@rundel

github.com/rundel

github.com/rundel/Presentations/

- Sta 101 - Data Analysis and Statistical Inference
- Sta 102 - Intro Biostatistics
- Sta 112 - Better Living through data science
- Sta 323 - Statistical computing