



# Moving Away from Ad Hoc Statistical Computing Education

Colin Rundel  
Duke University  
JSM 2017 - Baltimore

# **Statistics as a Computational Discipline**

# Duke StatSci Curriculum - Circa 2011/2012

1st Year

2nd Year

3rd Year

4th Year

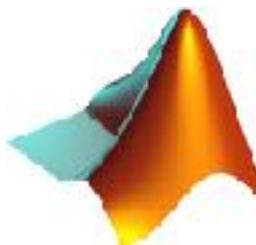
Intro Stat



or

*jmp*

or



Regression

*jmp*

or



Bayesian



Probability



Math Stat



or



Electives



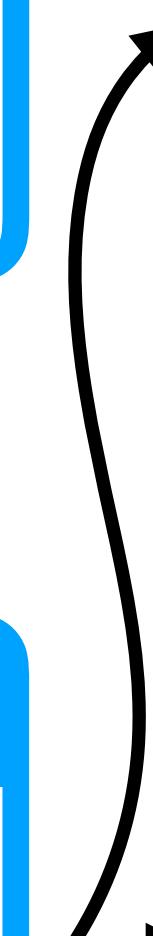
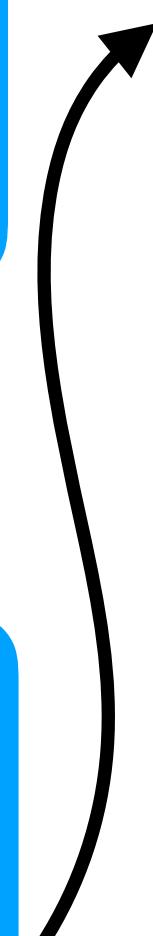
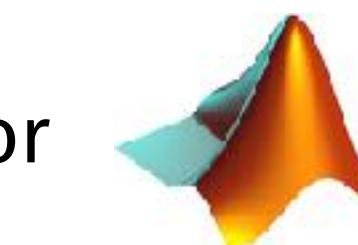
or



Cap Stone



or



# Duke StatSci Major - Circa 2011/2012

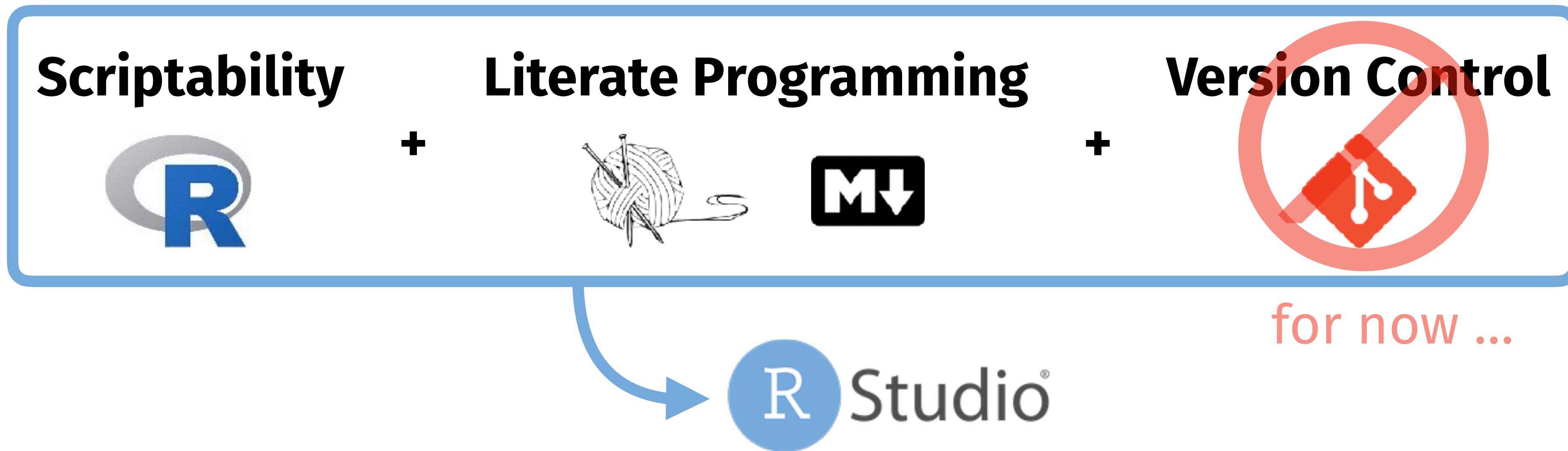
- Rapid growth of the major
  - 2009: 5 2010: 6 2011: 10 2012: 18 ... 2016: 36
- Possible for students to reach the Bayesian course without seeing R or learning any programming
- Exit interviews:
  - *The most useful course I took was a matlab programming course in the engineering department*

# **Proof of concept - 2011 to 2013**

- Creation of OpenIntro R labs 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,
  - 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
  - Critical to reduce on-boarding friction and get student buy-in



# **Non-scientific results**

- Course enrollments went up, reviews improved
- Increased recruitment into the major
- ...

# Duke StatSci Curriculum - Circa 2013/2014

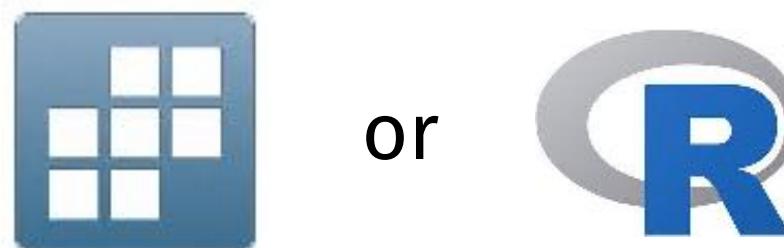
1st Year

2nd Year

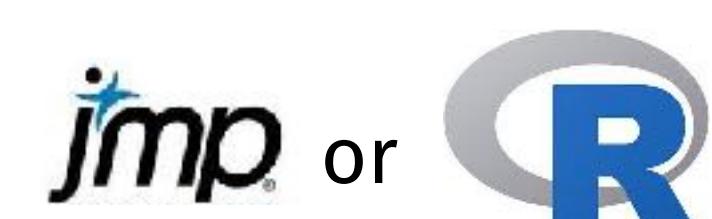
3rd Year

4th Year

Intro Stat



Regression



Bayesian



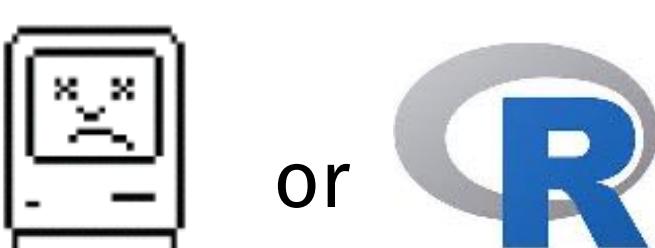
Case Studies



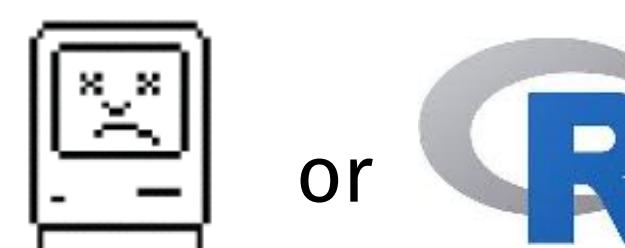
Probability



Math Stat



Electives

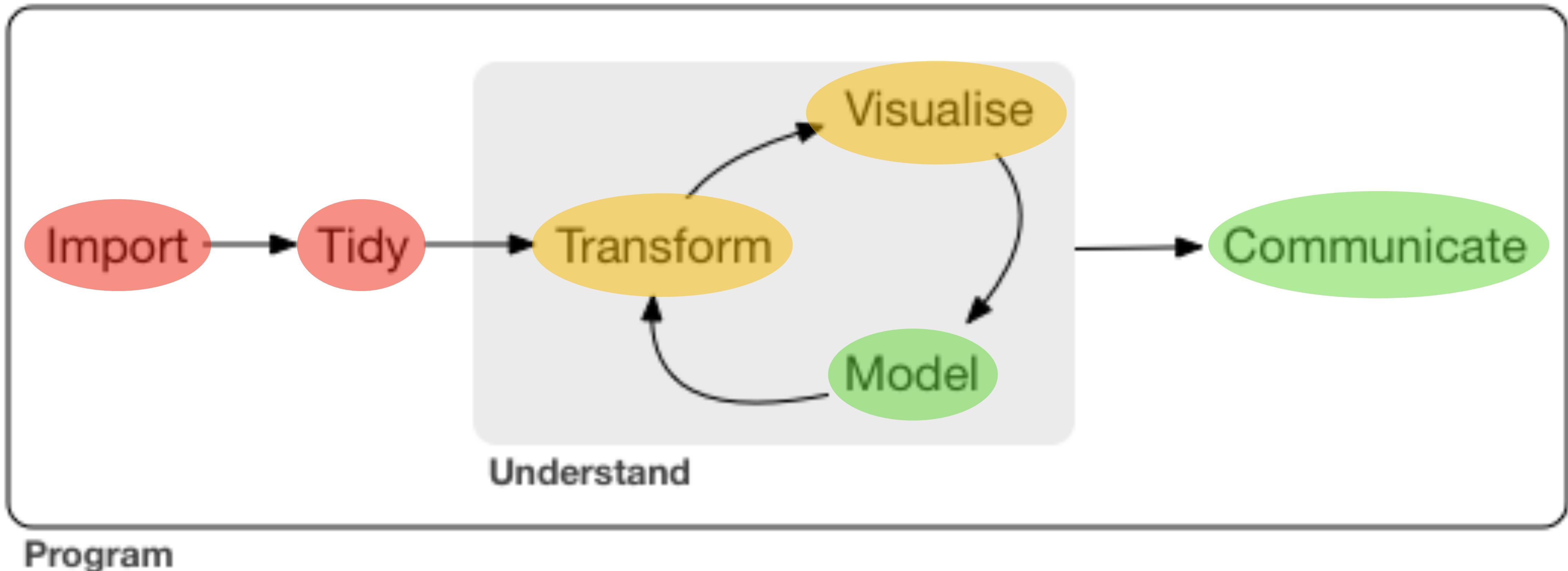


Thesis



CS 101 or 201

# What are we teaching?



# Statistical computing as an elective

- If there was room in our existing courses we would be doing this already
- Optional 2nd / 3rd year course with *minimal* prereqs - no assumed programming background
- Added in 2016, offered once a year in the Spring, ~30-40 students
- Build on the successes of R in the intro classroom ( + )
- Focus on computational skills (shell, git, scraping, munging, ...)

# **More non-scientific results**

# Duke StatSci Curriculum - Circa 2015/2016

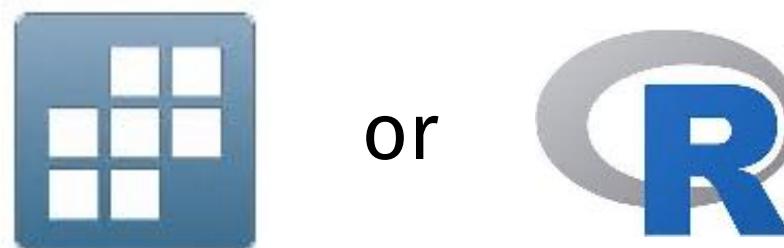
1st Year

2nd Year

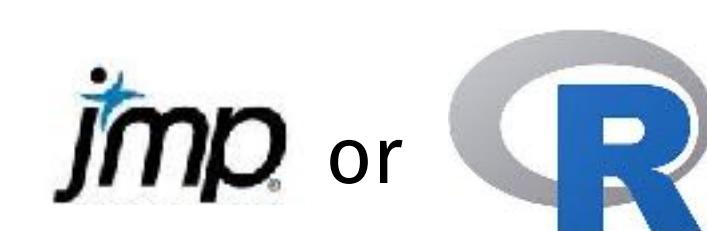
3rd Year

4th Year

Intro Stat



Regression



Bayesian



Case Studies



Probability



Math Stat



Statistical Computing



Other Electives



Thesis



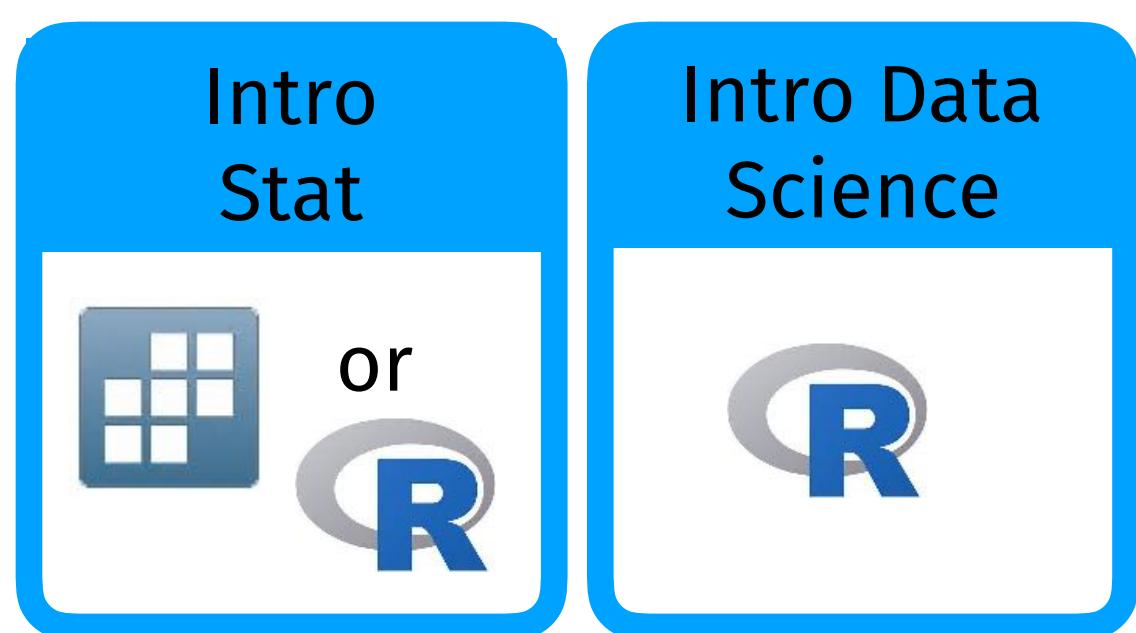
CS 101 or 201

# Rethinking gateways

- Students are desperate to learn “*Data Science*”
  - Our discipline loses when we tell them: “go take probability first”
- Identify the skills students find useful / exciting while still preparing students for later requirements (theoretical and computational)
  - Teach the *good* stuff in R (visualization, scraping, etc.)
  - Computation complements theory
- A version of the course has been taught, since 2014, as a small freshman seminar (~16 students)
  - will be scaled up to a large intro lecture (50-100 students).

# Duke StatSci Curriculum - Circa 2017/2018

1st Year



2nd Year



3rd Year



4th Year



Probability



Math Stat



Statistical Computing



Other Electives  
or



Thesis



CS 101 or 201

# **Lessons Learned / Future Thoughts**

- There was never a master plan, mostly an exercise in identifying what we could do better
- Healthy cross-pollination between our undergraduate and master's curricula - demand for the same shared core skills
- Staying current is hard, but immensely valuable

# Questions / Comments?



[rundel@gmail.com](mailto:rundel@gmail.com)



[@rundel](https://twitter.com/rundel)



[github.com/rundel](https://github.com/rundel)



[github.com/rundel/Presentations/](https://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](#)

