



# We'll start with Data Science

Mine Çetinkaya-Rundel  
Duke University + RStudio

[bit.ly/swds18](https://bit.ly/swds18)

@minebocek   
mine-cetinkaya-rundel   
mine@stat.duke.edu 

# three ingredients



theory



application



computation

The average number of AP/IB course credits earned in high school by Asian students (4.5 credits) was higher than the average earned by students of any other racial/ethnic group. Additionally, White students earned a higher number of total AP/IB credits in high school (3.1 credits) than Black students (2.7 credits).

Musu-Gillette, Lauren, et al. "Status and Trends in the Education of Racial and Ethnic Groups 2016. NCES 2016-007." National Center for Education Statistics (2016). <https://nces.ed.gov/pubs2016/2016007.pdf>

[...] while controlling for academic preparedness, career intentions, and instruction, the odds of a woman being dissuaded from continuing in calculus is 1.5 times greater than that for a man. [...] a lack of mathematical confidence, rather than a lack of mathematically ability, may be responsible for the high departure rate of women. [...] if women persisted in STEM at the same rate as men starting in Calculus I, the number of women entering the STEM workforce would increase by 75%.

Ellis, Jessica, Bailey K. Fosdick, and Chris Rasmussen. "Women 1.5 times more likely to leave STEM pipeline after calculus compared to men: lack of mathematical confidence a potential culprit." *PloS one* 11.7 (2016): e0157447.

A photograph of a paved road stretching into the distance through a hilly, grassy landscape under a cloudy sky.

BETTER

[bit.ly/swds18](http://bit.ly/swds18)

# start with data science

a course that provides  
a common (gateway) experience  
to students wanting to get started with stats,  
and that is

- modern
- places data front and center
- quantitative (but not mathematical)
- different than HS stats
- challenging (but not intimidating)

# pedagogical choices

- computing education begins early in the curriculum and weaved throughout the curriculum
- backwards design starts with an “impressive” final data analysis project and breaks it down into bite sized chunks
- on boarding friction minimized with computing infrastructure that “just works” -> server/cloud setup with all required packages and system libraries pre-installed
- uniformity of computational tools help ease the pain (and anxiety) -> tidyverse

# course overview

- **curriculum:**

- **part 1 - diving into data:** data gathering + wrangling + exploratory data analysis + visualization (5 weeks)
- **part 2 - making rigorous decisions:** multivariate modeling + basic inference (6 weeks)
- **part 3 - looking forward:** interactive visualizations + web scraping + bayesian inference + [insert something timely, useful, and fun here] (3 weeks)

- **learning goals:** modern and multivariate EDA + data visualization, reproducible computation, version control and collaboration, effective communication

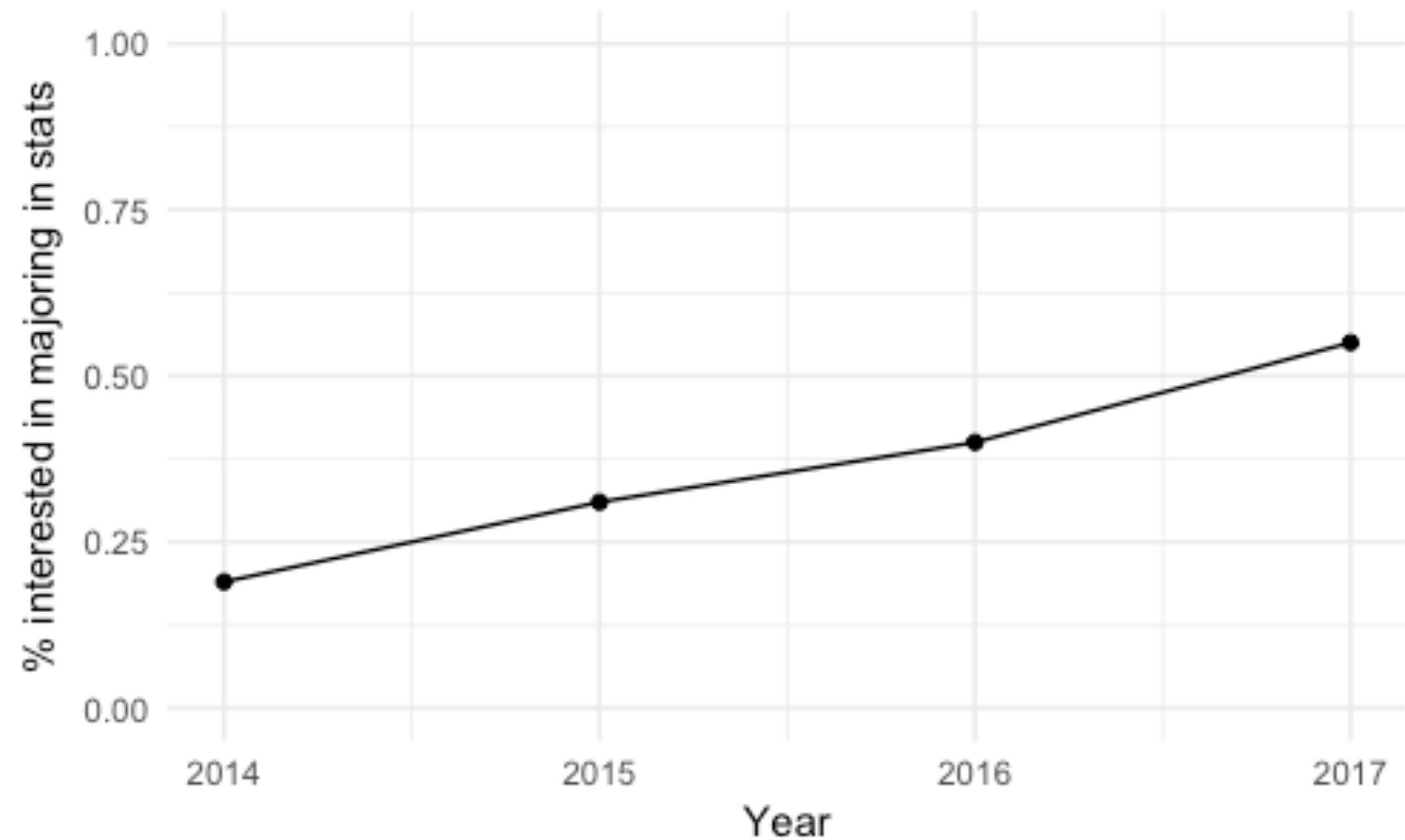
- **assessment:** not just final work but also the process, peer evaluations, and contribution diagnostics

[bit.ly/sta199-s18](http://bit.ly/sta199-s18)

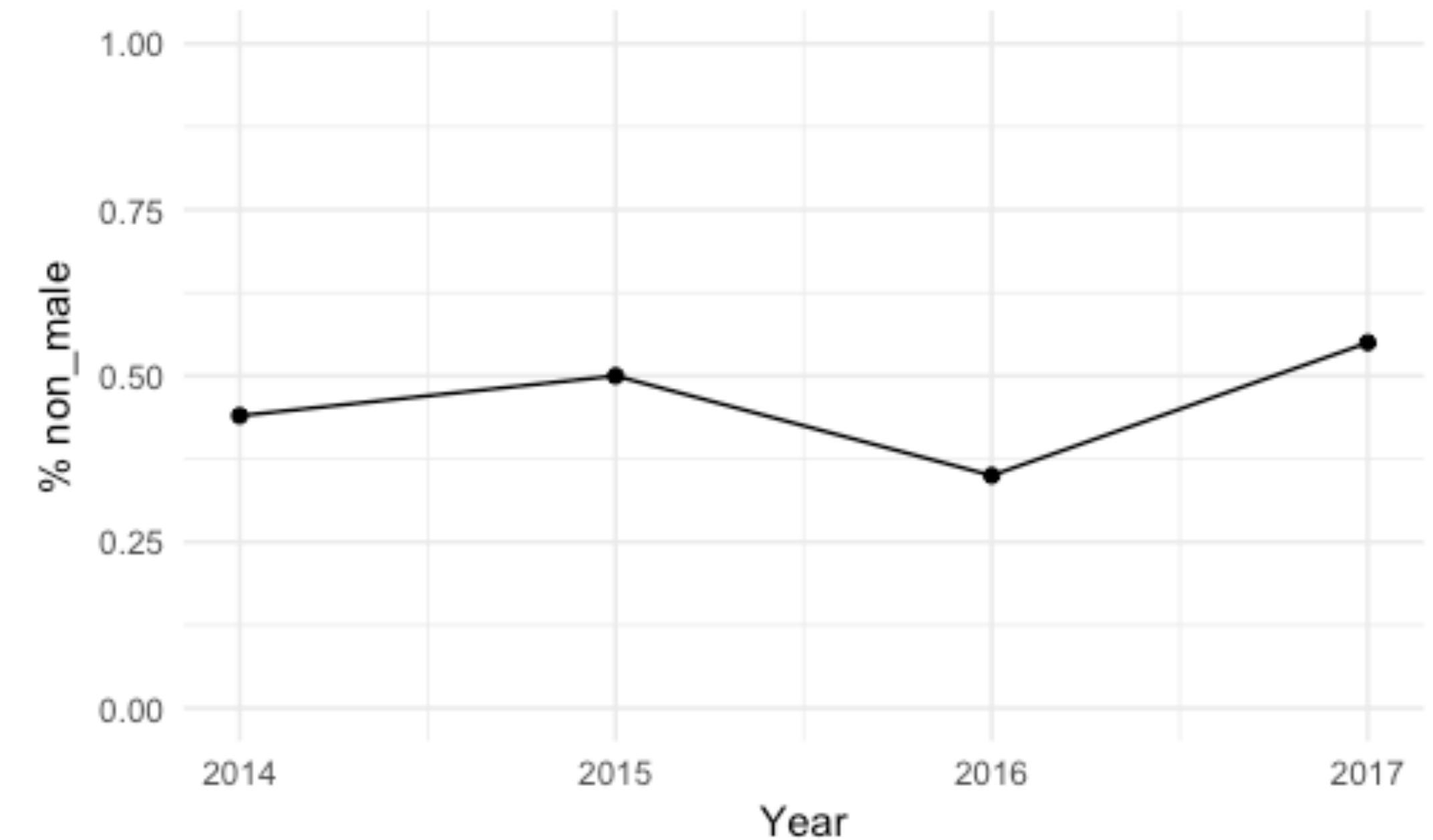
[bit.ly/swds18](http://bit.ly/swds18)

# impact

Pipeline for stats major



Increasing non-male audience



[bit.ly/swds18](http://bit.ly/swds18)

# We'll start with Data Science

Mine Çetinkaya-Rundel  
Duke University + RStudio

[bit.ly/swds18](https://bit.ly/swds18)

@minebocek   
mine-cetinkaya-rundel   
mine@stat.duke.edu 

