

# Data science methods for online teaching

Matthew Brett

## What is data science – not?

*The now-contemplated field of Data Science amounts to a superset of the fields of statistics and machine learning which adds some technology for ‘scaling up’ to ‘big data’. This chosen superset is motivated by commercial rather than intellectual developments. Choosing in this way is likely to miss out on the really important intellectual event of the next fifty years.*

(Donoho 2015)

## What is data science – really?

*An approach to the practice and understanding of data analysis that is founded in code.*

## What is data science?

## Data science education

- ▶ "... embracing a reinvention of statistical education in the era of pervasive computation." ([2015 Berkeley faculty report](#)).
- ▶ Teaching statistics "assuming computers exist, rather than assuming they don't exist." ([John Denero](#)).
- ▶ "Express in code what we would otherwise express in equations." ([John Denero](#)).
- ▶ "**Visualize First** - represent the data graphically in order to motivate questions about inference and concepts of statistics" ([Eric Van Dusen](#)).

See also (Cobb 2015, 2007; Horton 2015; Hardin et al. 2015).

# Data science at Berkeley



~1500 students / semester. No requirements for computing or maths.

# To the online system

<https://github.com/matthew-brett/talmo>

With caveats.