

What is Data Science? Does It Matter?

Matthew Brett

Handout

Go to <https://github.com/matthew-brett/bham-turing>

Look for `bham_turing_th1_handout.pdf` in the `townhall-1` directory.

About me

- ▶ Analysis of functional brain images, since 1996;
- ▶ Heavy user of, contributor to scientific Python stack since 2006;
- ▶ At UC Berkeley for 10 years or so, returning in 2017.

A definition of data science . . .

*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'.
– Donoho (2015)*

... that would be a disastrous error

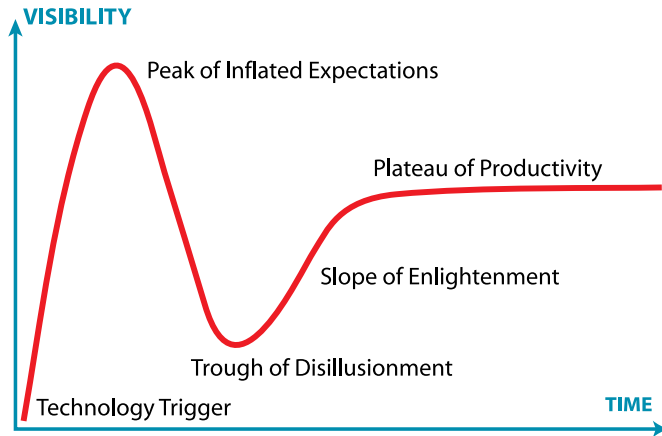
The now-contemplated field of Data Science amounts to a superset of the fields of statistics and machine learning which adds some technology for 'scaling up' to 'big data'.

...

Choosing in this way is likely to miss out on the really important intellectual event of the next fifty years. – Donoho (2015)

Donoho calls this “Lesser data science”.

Missing out



Graphic by Jeremykemp at English Wikipedia CC BY-SA 3.0.

See: <http://asterisk.dynevor.org/after-data-science.html>.

No definitions? Hype bubble.

A marketer said: "We use AI to predict fraud so our customers can spend with confidence."

The marketer meant: "Our fraud score is based on a logistic regression model that has been working well for years."

<https://blog.dominodatalab.com/avoiding-data-science-hype-bubble>

DATA

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

(Davenport and Patil 2012)

The data scientist in industry

... what data scientists do is make discoveries while swimming in data ... At ease in the digital realm, they are able to bring structure to large quantities of formless data and make analysis possible. ... Data scientists' most basic, universal skill is the ability to write code.

(Davenport and Patil 2012) - "Who Are These People?"

Who are these people?

Some of the best and brightest data scientists are PhDs in esoteric fields like ecology and systems biology. George Roumeliotis, the head of a data science team at Intuit in Silicon Valley, holds a doctorate in astrophysics.

(Davenport and Patil 2012) - “Who Are These People?”

What does this describe?

Data science is an approach to data analysis with a foundation in code and algorithms.

What is data science?

What does this look like for teaching?

... academic institutions should encourage the development of a basic understanding of data science in all undergraduates.

National Academies of Sciences and Medicine (2018)

Teaching at Berkeley

- ▶ Massive (~1500 student) course Foundations of data science. No requirements in mathematics or programming. Running since 2015.
- ▶ Large (~900 student) intermediate course Principles and techniques of data science with further requirements in Python programming and linear algebra.
- ▶ 27 Connector courses using teaching methods from the foundation course.
- ▶ The greatest change in undergraduate teaching in a generation.

What does this look like for research?

- ▶ Careers
- ▶ Education and Training
- ▶ Tools and Software
- ▶ Reproducibility and Open Science
- ▶ Physical and Intellectual Space
- ▶ Data Science Studies

Moore-Sloan Data Science Environments: Themes.

<http://msdse.org/themes>. (\$37.8M).

Recruiting for data science research

- ▶ Berkeley Division of data science.
- ▶ New recruitment half-time division of data science, half in their home department (of astronomy, psychology etc).

What should we do?

We should:

- ▶ reflect on the nature of the change that “data science” represents;
- ▶ study success and failure of other universities in their “data science” initiatives;
- ▶ consider our own barriers to fruitful “data science environments”;
- ▶ plan for huge changes in teaching and research.

Conclusion

- ▶ “Lesser data science” will go the way of “e-science” and other fashions.
- ▶ “Greater data science” is going to transform teaching and research.
- ▶ Business as usual means “lesser data science”.
- ▶ “Choosing in this way is likely to miss out on the really important intellectual event of the next fifty years”

The end

See: [https://github.com/matthew-brett/bham-turing townhall-1](https://github.com/matthew-brett/bham-turing-townhall-1)
for slides, handouts, source.