A Widening Horizons Module for Data Science

Matthew Brett

April 9, 2018

Source

https://github.com/matthew-brett/apsig

The object in question

"Data Science for Everyone"

Widening Horizons Module (WHM) for 2018 / 19, based in Life and Environmental Sciences.

Berkeley Fundamentals of Data Science course

- http://data8.org
- "[The course] combines three perspectives: inferential thinking, computational thinking, and real-world relevance." (link)
- "The course is designed for entry-level students from any major. It is designed specifically for students who have not previously taken statistics or computer science courses." (link).
- Subject-specific connnector courses.
- Since 2015. Projected enrollment of 2000 students in 2017 (link).
- Written as interactive Jupyter Notebooks, hosted on Berkeley servers.
- Textbook at https://www.inferentialthinking.com.

Running the book

```
git clone https://github.com/data-8/textbook
cd textbook
pip install -r requirements.txt
cd notebooks
jupyter notebook
```

- ▶ Plotting the classics.ipynb.
- ▶ Literary characters.ipynb.
- ▶ Nearest_Neighbors.ipynb.

Curriculum

- Exploring, cleaning, visualizing data.
- Modeling data with simple algorithms
- Simulation and resampling for statistical inference
- Classification and prediction with machine learning

How can you help?

- ▶ What do you think of the idea? What could go wrong?
- ▶ Do y'all have any datasets you think would be interesting?
- ▶ Would you consider giving a guest lecture?
- ▶ Would you like to collaborate on writing some teaching material?
- ▶ How can I best go about getting hosted Jupyter Notebooks on campus?

Is this the end?

Yes, it's the end of the talk.