Data 8 Adoption Panel

Taylor R. Brown PhD

Department of Statistics
University of Virginia

June 24, 2020

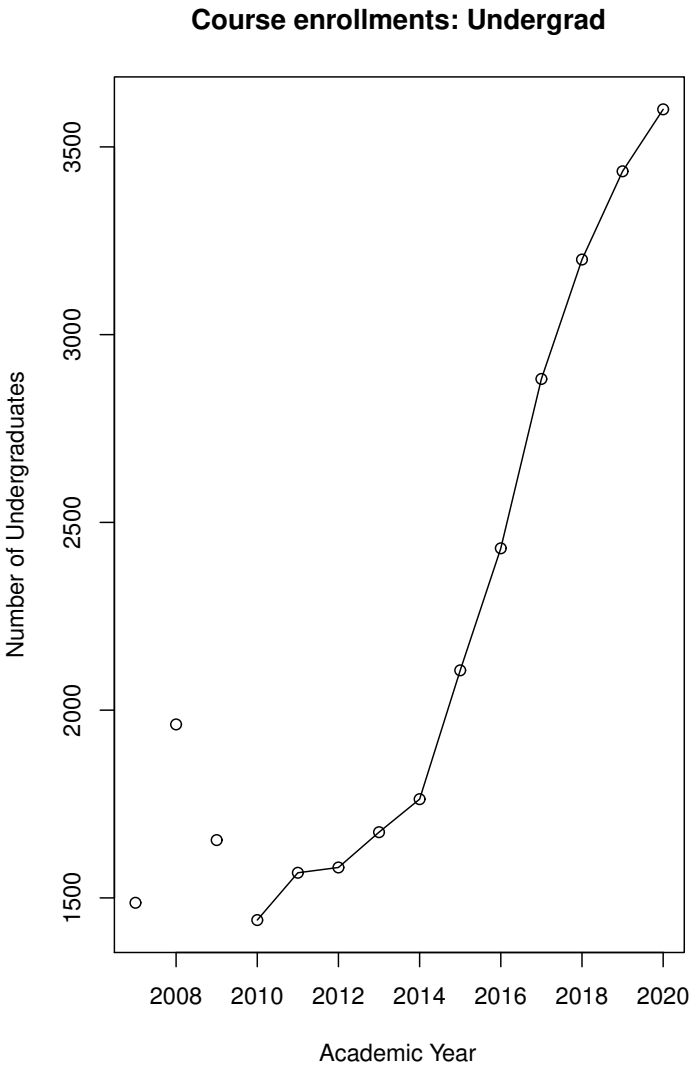
Overview

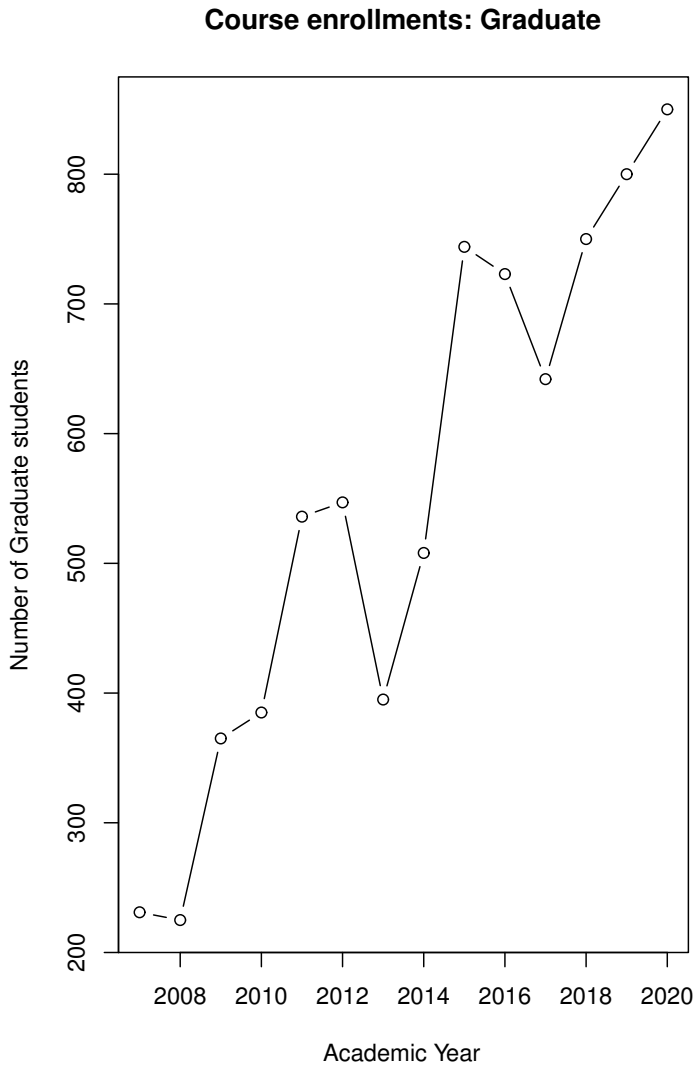
UVA and Me

Me (Taylor)

http://people.virginia.edu/~trb5me

Stats Dept. @ UVA





Post-Semester Reflections

Reflections on the...

1. textbook,
2. exercises,
3. slide decks,
4. datascience library, and
5. OK autograder and submission system

provided by Data 8

Python

From students

Pros :

- It's in fashion, useful in many areas, relatively easy to learn.

Cons :

- I've never programmed before!
- I don't "get" ——-
- Please don't ask me to install libraries from the command line.

Jupyter

From students

Pros :

- Everything is all in one place (embedded markdown instructions, videos, pictures, diagrams).
- The terminal is avoidable.

Cons :

- How do I switch this cell back to Code type from Markdown type?
- I can’t open this .ipynb file by double clicking on it.

Textbook and Lecture Materials

From students

Pros :

- I like how it's more "data science-y" than "statistics-y."
- Textbook examples are reinforced in class with demo code.
- It's free and online!
- There isn't an overwhelming amount of mathematical notation.

Cons :

- The search bar in the textbook doesn't always work :/

Textbook and Lecture Materials

From the instructor

Pros :

- Everything's bundled (students don't get lost).
- The organization is very good (programming before inference, delaying iteration, classification right after regression, simulation over closed-form expressions etc.).

Cons :

- There is light mathematical notation (mostly good, but e.g. CLT got dicey).
- There is a bias towards Berkeley-based examples.

datascience Python library

From students

Pros :

- This is way more “legit” than Excel.

Cons :

- I don’t get .join or .group

datascience Python library

From the instructor

Pros :

- There is no row indexing or slicing.
- Table methods have fewer parameters than Pandas analogs (.where, .join, .apply, .group, .plot, ...).
- Students never really complained about not getting to learn Pandas.

Cons :

- TAs would probably prefer Pandas.
- A couple functions have sparse documentation.
- Function signatures change (I can't think of many examples).

OK Autograder

From students

Pros :

- Code checks are great reassurance.

Cons :

- I thought I submitted, but I didn’t...

OK Autograder

From instructor

Pros :

- Easier than manual grading or writing your own autograder software.

Cons :

- It didn't work as advertised ("Grading with one click").

Key Takeaways

Key Takeaways

Interests continues to grow quickly–this is important to get right, and get right fast.

1. Sharing/co-developing/open-sourcing intro course materials is a very good idea.
2. Goals for an intro class: hold interest and set students up for success later.
3. Perspectives on statistics vs. data science (or ML or AI or ...):
   - students: DS = more jobs and less math!
   - job market: ?
   - emphasizing DS from within a statistics department...

What I would change

...if I had unlimited resources and I teach this again:

1. use different autograding software,
2. take Jupyter to the cloud,
3. change some examples (e.g. regression, two-sample tests),
4. slightly more generous with mathematical notation,
5. mention population versus sample statistics,
6. add more UVA-based examples and default plotting color-schemes ;)

The End

Thanks!