

Responses to questions on STATS 401 and the future of undergraduate data science

- Since we are all figuring out this future together, all the responses are copied below.
- Some comments on common themes follow.

Question 1. Should STATS 401 in future follow the data science perspective outlined above?

- Yes, it already does; combines stats with computing well. More real examples (i.e., forecasting?)
- We liked how “applied” the homeworks are but there’s sometimes a disconnect between lectures and the HW.
- Yes.
- Yes, especially if it is more similar to how statistics is utilized in the real world.
- Yes that would make the class more applicable.
- Yes, However we think adding a project component would be worthwhile. It would give students a more intimate relationship with the data (if group project, get exposed to Git). Aside: make HW difficulty more consistent throughout. Remove the abstraction and show the application. Maybe don’t be so pedantic.

Question 1. Continued

- Yes, it's important to keep a modern approach, and learn about the applicability of the topics covered, rather than just memorize different concepts and methods.
- Yes. For HW, suggest to need more example of code. We feel like the HW has been graded a bit differently from the syllabus's requirement of based on "effort". Piazza has not been good at answering questions clearly.
- Yes. "Applied statistics" is turning into data science, makes it more applicable for the future.
- Yes - we think applying stats with technology is useful for a variety of future career fields.
- Yes, but we feel like in contrast to this semester, more emphasis should be placed on teaching students with various levels of preparation.

Question 2. Has this version of STATS 401 developed math/stats/computing topics at a suitable level (challenging but not unreasonable)?

- Yes. Stats 306 built early skills, this class expanded on the modeling part of data science in a challenging way.
- Homeworks are harder than we think they should be. We also think the syllabus shouldn't say "HW is for completion if there's sufficient effort" because we haven't found that to be the case.
- Reasonable. More computing perhaps? Cohesion between the three. So far, we have covered a good amount of basic linear algebra. Learning about RREF's might help a lot [reduced rank echelon form of a matrix]. It is simple but extremely powerful.
- Not a suitable level, more preparation for graded assignments, more time spent on examples.

Question 2. Continued

- Even though most of the topics are challenging and at a suitable level, some of it has been unreasonable because the hw is very different from lecture at times.
- Clearly define notation before explanation to avoid misinterpretation. Add more coding exercises.
- Yes, there are many topics to cover so it is hard to go in depth with each of them, but the class has done a good job of combining those topics.
- Yes, but the lecture and HW have been very abstract; hard to apply theory to actual coding question. Need more clear expectations for grading, more examples to model homework/tests off of. Disconnect between prerequisites and level expected to be at to do HW/code. Usually don't finish material for HW until Wednesday but due on Friday.

Question 2. Continued

- Yes, but we don't really know how to apply the topics. We wish the class was project-based.
- The workflow could be more defined. Use more examples so we know what steps to follow once we get the homework.
- No. The topics that were taught in class and on the homework/practice exam were not continuous with what was on the midterm (i.e., the midterm was much more application & higher level / unreasonable difficulty). The class was not adequate prep for the exam.

Question 3. Has today's discussion helped to clarify the goals of this version of STATS 401?

- Yes.
- A little bit.
- Yes.
- A little bit.
- No, goals are still unclear, practice problems in lecture would make more goals easier to see. Homework and lecture are completely unrelated.
- Yes!
- Yes, it has helped to find our common questions and suggestions to improve the content of the course.
- Yes.
- Yes.
- Yes, but the class structure does not accomplish these goals.

Some conclusions from the feedback

- Learning data analysis is like learning a musical instrument. Spend too long on scales and technical exercises and you lose motivation. Focus on playing songs, to the exclusion of exercises, and you greatly limit your longer-term potential. Not everyone thinks this course has the right balance.
- HW grading needs discussion and changes.
- More support is needed for those with the minimum prerequisites (STATS 250 and Calc I, and little or no previous computing experience.)
- Some disconnect between class and homework is natural for learning applied statistics: class covers the principles and homework lets you investigate how they play out in practice. **The bigger this disconnect, the faster and more challenging the class can be.** If the scope of the class is to remain similar, extra support is needed. Idea: additional structured teaching of R could help bridge the gap between the introductory material (swirl and some foundations of R) and the level of R skills needed later in the course.