Herding cats:
Pros and Cons of a large-team approach to
data science at a major research university

David R. Hunter
Penn State University

JSM 2018, Vancouver

History of data sciences at Penn State

- 2014: The Department of Information Sciences and Technology (IST) proposes an undergraduate major in data science
  - IST wanted statistics courses among the requirements
  - Computer Science balks due to "new" courses proposed that overlap with existing CS courses.
  - IST's proposal goes nowhere in original form.
- 2015: IST, CS, and Statistics propose an undergraduate major in data sciences, to be jointly administered among 3 different colleges.
- Important: This DS major is not the only DS game in town. More on this later.

Don't let the perfect be the enemy of the good

- To get the DS major started, we required existing courses in STAT, CS, and IST.
- I strongly favor specially designed DS courses—such as Berkeley's Data 8—but such development takes a time.
- I think it's better to get a program up and running than to delay while courses are developed: Above all, waiting doesn't serve students.

Data Science is an evolving field

Success, Opportunities, and Challenges for Statistics and Biostatistics in the Data Science Era

A Report of the July 2016 NSF-Sponsored Workshop for Chairs of Departments of Biostatistics and Statistics

ASA

- ▶ Whatever DS is, anyone knowledgeable these days will say that it overlaps statistics.
- ▶ As support for this position, see the 2016 white paper along with many, many discussions like this panel.

http://www.amstat.org/asa/files/pdfs/Chairsworkshop/
WhitePaper.pdf

Just as there was a time when statistics was nervous about being shut out by CS and others, and we rightly thought this was improper, I think it would be a mistake to try to claim DS as part of statistics.

That many want to get into DS is a good thing

- At Penn State, IST (Information Sciences and Technology) has been pushing hard to start not only an undergraduate major but also a minor, master's degree, and graduate certificate.
- On one hand, I'd like to see more statistics and CS faculty excited to help with this.
- On the other, this work takes time and I've come to appreciate the enthusiasm of IST for taking it on. Cultivating a working relationship is very worthwhile.
- Many, many others have expressed interest. . .

That many want to get into DS is _{usually!} ^ a good thing

- ▶ Social Data Analytics: Undergraduate major that predated DS at Penn State.
- ▶ Administered by our political science department
- ▶ Knowledgeable colleagues who involved stats from the start

That many want to get into DS is ^(usually!) a good thing

- ▶ Social Data Analytics: Undergraduate major that predated DS at Penn State.
- ▶ Administered by our political science department
- ▶ Knowledgeable colleagues who involved stats from the start

- ▶ Even in a couple cases where other groups held negative or inaccurate views of statistics, working together has turned out well eventually.

Summary: Expanding the Tent

- ▶ Data Science at a large university like Penn State is a complicated, multi-party affair.
- ▶ If that's what "Expanding the Tent" means, then it's good to embrace it. . .
- ▶ . . . and know that you can turn to the statistics community for advice and moral support!