

Some thoughts on hosting a DataFest

Nathan Taback, University of Toronto

May 29, 2019

What is ASA DataFest?

TEAMS OF STUDENTS GET A PROBLEM AND DATASET ON THE FIRST DAY, and work on the problem until the last day where they present their results. After 48 hours of intense data wrangling, analysis, and presentation design, each team is allowed a few minutes to present a few slides to dazzle a panel of judges. Prizes are given for various categories.

Why hold a DataFest?

DATAFEST IS ONE WAY TO PROVIDE A CO-CURRICULAR EXPERIENTIAL LEARNING OPPORTUNITY TO MANY STUDENTS.¹ Teaching statistics should involve integrating real data with a context and purpose and active learning. “Our job is to help [students] use data to answer a question that matters ... This may be the biggest exogenous challenge to our profession, the least explored in our undergraduate curriculum, and the most promising for rethinking what we teach.”²

Experiential education is not simply “learning by doing”, and “simple participation in a prescribed set of learning experiences does not make something experiential.”³ For example, a student running a script and examining statistical output is not experiential learning.

Where to start?

<https://ww2.amstat.org/education/datafest> has all the information on hosting an official ASA DataFest.

ALTERNATIVELY, FIND A DATA DONOR AND PROBLEM that can be used to hold a data science contest.

If your institution is a member of CANSSI then apply for up to \$1000 in funding.

Judges, Mentors, and Prizes

The ASA DataFest@UofT has teams submit three slides, not including a title slide, to be presented in 5 minutes to a panel of judges.

Data Science Community

DataFest can bring together a data science community. Undergraduate students do the work, but they are assisted by mentors who are graduate students, faculty, and industry professionals. Professionals find ASA DataFest to be a great recruiting opportunity – they get to watch talented undergraduate students work under pressure in a team and examine their thinking processes.

¹ Taback [2018]

² Cobb [2015]

³ Chapman et al. [1992]

Key planning milestones:

- Inform ASA of your intention to host a local DataFest by the first week in January.
- Host the DataFest during the specified time period (usually mid-March to early May).
- Keep the data source secret before the event and until the last DataFest.
- Fund raise to cover costs such as food and prizes.
- Recruit judges and mentors.

ASA DataFest@UofT prize categories: Best In Show; Best Use of External Data; Best Visualization; and Honourable Mention.

Recruiting Judges and Mentors

The majority of judges at ASA DataFest@UofT are external. It's a good idea to start thinking about recruiting as early as possible.

Judging Process

2019 ASA DataFest@UofT had 43 teams with 150 students. Teams were split into 4 separate rooms with separate judging panels for the first round of judging. Each panel selected 2-3 teams that would move to the final round, where all the judges evaluated the presentations. This process allows the presentation part of the contest to be 2-3 hours.

Judging Rubric

PROVIDE JUDGES WITH A SIMPLE RUBRIC that can be used to evaluate presentations.

Judges are not typically given time or resources to investigate if a team's work is reproducible (cf. a prediction contest).

Student Profile and Feedback

112 (70%) completed participants a registration survey for 2019 ASA DataFest@UofT.

- UofT programs of study: 67% statistics; 27% computer science; 29% math; and 94% at least one of statistics, computer science, or math.
- Gender: 48% Female; 52% Male.

References

Steve Chapman, Pam McPhee, and Bill Proudman. What is experiential education? *Journal of Experiential Education*, 15(2):16–23, 1992.

George Cobb. Mere renovation is too little too late: We need to rethink our undergraduate curriculum from the ground up. *The American Statistician*, 69(4):266–282, 2015.

Nathan Taback. Do you have experience? incorporating experiential learning opportunities into statistics education is messy but important. ICOTS10, 2018.

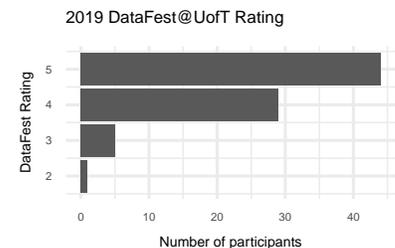
Alumni might be a great source of external judges. The majority of mentors are graduate students and post-docs.

Many judges are not statisticians or data scientists.

The rubric evaluates teams for: Creativity (1-5); Technical (1-5); and Presentation (1-5).

Try to emphasise that the experience of struggling to develop a question, analysis, and presentation is **as important** as winning.

- 94% registered to “gain experience applying knowledge that they learned in courses”.
- 44% described thier ability to present the results of a data analysis as less than adequate or have never presented a data analysis.



Based on 79 participants rating 2019 ASA DataFest@UofT (1–5)