

Sta 771S: DataFest

- Discuss the questions assigned to your team: 20 mins
- Prepare a brief presentation that summarizes the key points of your discussion: 2-3 mins
- Lead a discussion on your topic: 5 mins

Tentative teams are noted below. Team assignments might be updated if there are absences.

Team 1: Michael, Mao, Christine

1. What would the possible preparations/guidelines be for a team of students who intend to better participate the DataFest event?
2. How can we lead students into thinking more about statistics after DataFest? Ideally, we don't want DataFest being a one off event but rather an entrance into the broader statistics field. But, because DataFest is held over a weekend, many will not have enough time to explore and learn what they want to do. Should there be a written component, to be submitted later, where they can explore these ideas, for example?
3. Should we consider increasing the frequency of DataFest to twice or thrice per year to enable students experience "thinking with data" more than currently?
4. Instead of making students work on a certain dataset for 2-3 days and forget about it afterwards, how to use the DataFest to positively influence their learning of statistics in the future?

Team 2: Kyle, Lu, Princeton, Shaobo

1. Should presentation groups only be limited to two slides? Does this make presentations more concise or prevent valuable content from being delivered?
2. What is the ideal size for this event? Feedback is important in these exercises in order to get a sense of accomplishment and recognize your shortcomings, but with too many attendees, this might be a problem. On the other hand, having a broader appeal and bringing in people from outside is also important. How can we balance this, and if we need to cut down on attendees, how can we do it?
3. Since students' are interested in different problems, how to evaluate students' work in a Datafest?
4. What are some ways that DataFest can be improved from its current state?

Team 3: Matt, Willem, Ken

1. Since it turns out that data description is the most importance skill in activities like DataFest, should we reduce traditional inference in curriculum of statistics and emphasize more on description and interpretation of data?
2. To make DataFest more effective, how can we begin to prepare students to think about how to analyze data?
3. The last observation points out that students were quick to find confidence intervals for parameters in the L.A. crime data, but that they were sort of useless, and that it would be better to “explain” and predict crime. What topics could be taught to undergraduates that would have been more useful?
4. Taking DataFest as a mirror, what are the essential parts missing from current undergraduate statistics curriculum, and how shall we make improvements?
5. Is causality relevant in DataFest or should participants be steered away from it since they are (supposedly) dealing with observational data? If a VIP consultant wants to help a team obtain 'causal' results, what could s/he explain, assuming the participants have little to no background on causal inference?

Additional questions:

1. Should we instead consider complementing the current DataFest set up with a class or seminar similar to STA723 (case studies) where instead undergrads get to work on one dataset every two to three weeks?
2. How do you encourage students to be creative in using “external” data? In one year, most groups used the population data from the Census Bureau.
3. Part of the appeal of DataFest is that the data and question are a surprise, and that the type of work done differs between groups and doesn’t nicely fit into the classroom. How do you build a class around that?
4. Should DataFest be the first introduction to self analysis of data for students?
5. What knowledge do students need in exploratory data analysis (EDA)?
6. Is it a good idea to provide an award exclusively for first time undergraduate participants in DataFest? Why or why not?