

Course Development and Enhancement Grant Application (2013-2014)

1. Proposal Title

"Analyzing Large Data Sets"

2. Name of the Proposer

Bill Schellhorn
Assistant Professor of Mathematics

3. A description of the proposed course for which funding is being sought if it is a new course, or the catalog description if it is an existing course.

I plan to teach a Simpson Colloquium course in Fall 2014 that ties into my sabbatical research. The broad topic of the course will be "large data sets", which we will define as data sets that are difficult to analyze because of their size. Large data sets are being created at an increasing rate due to technological advances in gathering and storing information ["Data, data everywhere"]. Large data sets are common in genomics, medical records, sensor networks, internet search indexing, social networking, e-commerce, astronomy, atmospheric science, and military surveillance ["Big Data"]. The course will introduce students to techniques in statistics and data mining, that are key to dealing with large data sets, for example determining characteristics of the entire data set and detecting trends within the data set.

4. Information about when and how often you intend to teach the course.

I plan to teach this particular SC course in Fall 2014. I may not offer the exact same course afterwards, but I anticipate using the materials in parts of other courses as well as in undergraduate research projects.

5. A description of how you intend to use the funding. This should include a specific description of the purchases and activities, including location, schedules or itineraries. Provide a justification for your use of the funds.

I intend to use the funding from this grant to attend the PProfessional Enhancement Program (PREP) described below, which is sponsored by the Mathematical Association of America (MAA). If I can receive funding early (before May 6, 2013), I will use \$275 for my registration fee and the remaining \$225 to pay for part of my airfare expenses. Otherwise I will use \$350 for registration and \$150 to help offset the cost of airfare.

The PREP topics change every summer, so this will be the only opportunity for me to attend this particular program. I need to travel because bringing mathematics educators together is an integral part of the experience, which includes hands-on activities and group projects.

PREP: "Big Data" and Data Mining for Mathematicians

Organizers: Dick De Veaux and Bernhard Klingenberg (Williams College)

Dates: June 17-19, 2013

Location: Williamstown, Massachusetts

Registration Fee: \$275 per person (\$350 for registrations received after May 6, 2013)

Description: This workshop will give mathematics educators experience solving real 'big data' problems in science, industry and government. Participants will learn the methodology and algorithms used by statisticians to analyze large data sets. They will learn the models most commonly used, how to implement them and how to deal with the practical data issues such as data typing, missing data and data quality. Most of the techniques, which are often grouped under the rubric of 'data mining,' are based on methods (e.g., regression) introduced in elementary statistics courses and will seem familiar and intuitive, even to those with limited statistics experience other than teaching the first course. Participants will learn the techniques interactively, performing and discussing the analysis with each other after learning each method. For the final component, they will work in groups on projects and report on their findings. [De Veaux]

6. A description of the student learning outcomes that will be met through the use of these funds and a description of how these learning outcomes will be assessed.

The format of the PREP workshop is similar to the way I want to design my SC course. Therefore I based the following student learning objectives on the ones in the workshop description.

- Students will ask appropriate research questions about large data sets.
- Students will solve problems related to large data sets.
- Students will describe the methodology and algorithms used by statisticians to analyze large data sets.
- Students will implement models commonly used to analyze large data sets.
- Students will deal with practical data issues such as data typing, missing data and data quality in acceptable ways (given the nature of the data).
- Students will discuss their analyses of large data sets with others in various settings.
- Students will report on their findings after working on a data analysis project.

These student learning outcomes will be assessed in my courses and undergraduate research projects in a variety of ways. The learning outcomes will be assessed through homework assignments and technology labs using statistical software packages. They will also be assessed through projects in which students analyze a large data set, write a paper about their findings for a general audience, and make a presentation summarizing their results.

7. Please clarify how your proposal supports the College's mission and strategic plan.

In my courses and undergraduate research projects, students will be required to work "hands-on" with large data sets, by which I mean they will be presented with a large data set to analyze using statistical

and data mining techniques. This will require the students to develop intellectual skills (ask appropriate research questions, solve problems related to large data sets, describe commonly used methodology and algorithms) as well as practical skills (implement models commonly used, deal with practical data issues in acceptable ways, discuss their analyses with others in various settings, report on their findings).

Large data sets arise in a wide variety of areas, including all branches of science, technology, economics, and operations research. This diversity of information will result in integrative learning opportunities for students. For example, suppose students are analyzing a large set of medical data. In reporting their findings, the students will need to provide context for the data: the nature of the data, how the data set was collected, reasons why the data was gathered, and responsible uses for the data.

8. A complete budget for the project. The budget should be submitted using the standardized budget form available from the Faculty Development Office. This form includes information about budget restrictions.

Refer to the Budget Proposal Form attached.

9. Please indicate whether, if your proposal is funded, we may share your proposal with others. Your answer to this question will not influence whether your proposal is funded.

Yes; you may share my proposal with others.

Works Cited

"Big Data." *Wikipedia: The Free Encyclopedia*. Wikimedia Foundation, Inc. 2 Feb. 2013. Web. 3 Feb. 2013. <http://en.wikipedia.org/wiki/Big_data>

"Data, data everywhere" *Economist.com*. The Economist Newspaper Limited. 25 Feb. 2010. Web. 1 Feb. 2013. <<http://www.economist.com/node/15557443>>

De Veaux, Dick, and Bernhard Klingenberg. "Big Data and Data Mining for Mathematicians." *PREP 2013*. The Mathematical Association of America. 2013. Web. 1 Feb. 2013. <<http://www.maa.org/prep/2013/data.html>>

SIMPSON COLLEGE
 BUDGET PROPOSAL FORM
 Course Development and Enhancement Grant
 2013 - 2014

ITEM		AMOUNT
Equipment		\$
1:	Cost:	
2:	Cost:	
3:	Cost:	
Materials		\$
1:	Cost:	
2:	Cost:	
3:	Cost:	
Travel Costs (partial airfare)		\$ 225 / \$150
Airfare: \$ 225 if before May 6, 2013 / \$150 afterwards		
Mileage: Number of miles _____ @ \$.565/mile (or current mileage rate)		
Lodging		\$
Number of days _____ @ \$ _____ /day		
Other Expenses		\$ 275 / \$350
1: Registration	Cost: \$ 275 if before May 6, 2013 / \$350 afterwards	
2:	Cost:	
3:	Cost:	
TOTAL EXPENSES		\$ 500 / \$500
AMOUNT REQUEST (not to exceed \$500)		\$ 500 / \$500