SQLShare Information

Biology 340: Biometrics

Instructor: Robin Kodner, robin.kodner@wwu.edu

In the new world of "Big Data", technology is enabling all fields to collect large-scale data sets that require a specific set of skills to be able to analyze the data. New advances in DNA sequencing technology has made it easy and affordable to generate millions and billions of short sequence reads on bench-top instruments, widespread ecological data bases makes access to data from around the world much easier, environmental survey instruments and satellites are constantly logging data and making it available via the web, and so on. Our students, no matter the discipline, will be at an advantage in most fields if they can learn to 1) access the data 2) manipulate the data files, and 3) share data and results efficiently. However, these large data files often cannot be manipulated via traditional software (typically Microsoft Excel). If Excel or other spreadsheet-based software can't open the file because it's too big, most of our students would be stuck at that point, and would need to start taking computer science classes to find a way forward.

In response to this, my colleague at the University of Washington eScience Institute and Department of Computer science developed a cloud-hosted relational database called SQLShare aimed at making large scale data management and manipulation easy for scientists and students. See:
http://escience.washington.edu/sqlshare

> "SQLShare is a database service aimed at removing the obstacles to using relational databases: installation, configuration, schema design, tuning, data ingest, and even application design. You simply upload your data and immediately start querying it."

In addition to bringing service-learning to BIOL 340 this year, I also began a pilot of using SQLShare and teaching the basics of the SQL language in my class. As I introduced relational databases to students, some thought it would be a useful skill for a biologist to learn.

I think learning to use a relational database will help me learn to do research in a biological field.

Answered: 36   Skipped: 0
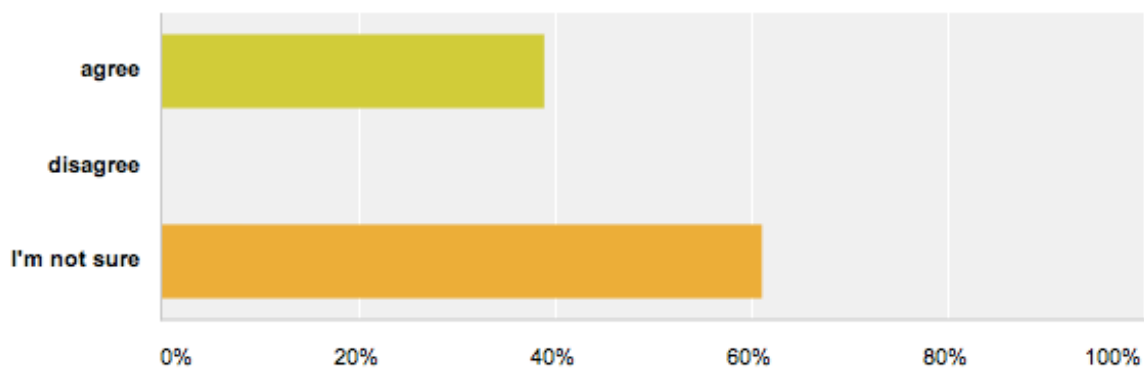
Many groups (21%) of students used SQLShare for their service-learning projects, 56% of students agreed that SQLShare would be helpful for them in analyzing data in the future, and 87% said that using SQLShare helped them understand how they could analyze large data sets. I have also had many other research students using SQLShare, including all my students in my Environmental Genomics class.

Students said:

> *"SQL was easier to learn than R and just felt less complicated once you learned what you were entering. I also think I would use SQL if I were going to do any research and had to organize and analyze data. It's definitely a useful program to know if you ever do go off and work in a lab. "*

> *"SQL was very interesting and fun to use"*

> *"I enjoyed using SQLShare the most because the queries made sense, almost like sentence structure. It was easy to figure out without much instruction."*