EINDHOVEN UNIVERSITY OF TECHNOLOGY

What are common mistakes that novices make when learning query languages?
Daphne Miedema (Database Research Group)

Problem statement

Databases are powerful tools, both for storing data and for performing data analytics. Interaction with databases requires dedicated query languages, so any user who wants to pose a query to the database needs to learn this language. Unfortunately, such languages are hard to learn to use effectively. Mitrovic [2] poses that this may be due to the high cognitive load involved in writing queries. The goal of this poster is to examine the errors that are made by novice learners when learning SQL, the most popular and common database query language.

Case study

I ran an experiment to examine the querying process by novice SQL users based on the homework exercises of second-year bachelor students taking an introductory course on Databases[1]. In these exercises, students attempted to solve six query formulation questions. All commands that the students executed were logged in a file.
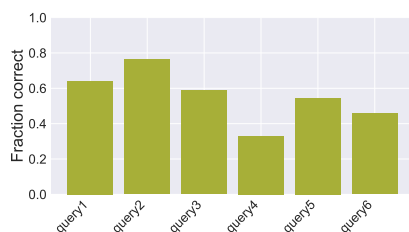

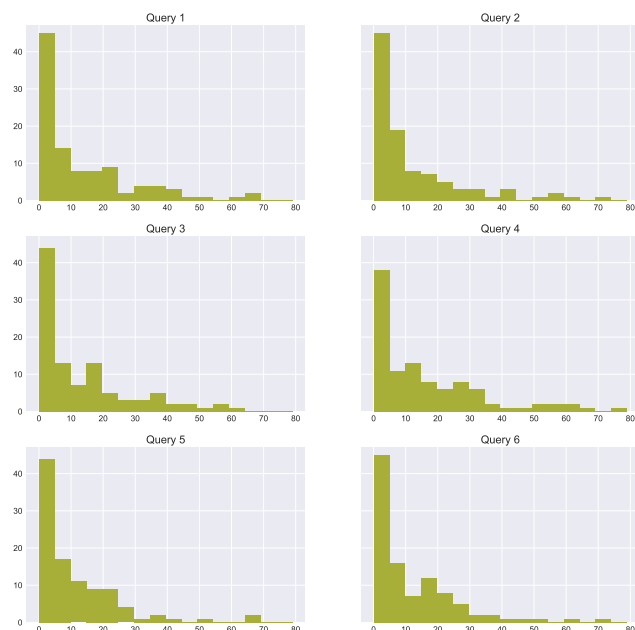**Figure 1:** Fraction of students that answered the question correctly.

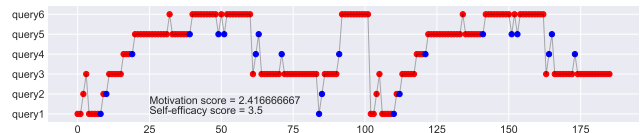**Figure 2:** Histogram of the number of attempts per person per query.

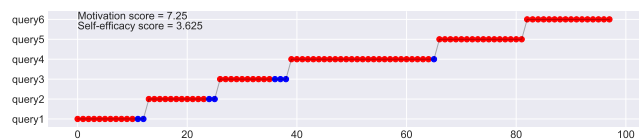

Figure 3: Trace recording for participant 127.

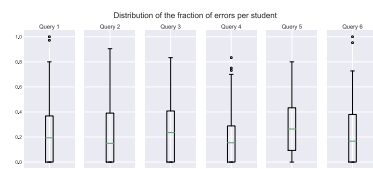**Figure 4:** Trace recording for participant 277.

Figure 5: Fraction of attempts that were erroneous.

Results

From the figures we see that students have *different ways of working*. Some work in a neat and structured manner, others do not. We also conclude that many of the attempts that were made contain errors, which may reflect a *lack of understanding*.

Besides the analysis displayed on the poster, I also analyzed the raw query texts as submitted [1]. In this analysis, I found more proof of lack of understanding as students were *writing overly complex queries*. In addition, it seems that many students repeatedly pose *exploratory queries*, perhaps to recall the structure of the database schema.

Discussion

In this research we found that many mistakes were made by novice learners. Many examples of *unproductive behavior* were found, as well as different types of transgressions. We also found support for the hypothesis of *high cognitive load*.

These findings can be used to support learners in their querying process, by teaching, documenting and implementing solutions.

References

[1] Miedema, D. (2019). *Towards successful interaction between Humans and Databases*. MSc Thesis, Eindhoven University of Technology.

[2] Mitrovic, A. (1998). Learning SQL with a computerized tutor. In *ACM SIGCSE bulletin*, 307–311.