



jumping rivers

# An Introduction to SQL (with R)

---

Using databases is a fundamental part of a data scientist's role. As they quickly discover, a poorly written SQL query can cost both time and money. The main focus of this training course is to introduce SQL databases and how R can be used to retrieve and manipulate data stored in a relational database. We cover the fundamental database concepts and illustrate how we can leverage R to work alongside them.

We use the *PostgreSQL* database as an example for public courses. For in-house training, we are happy to adapt the course to match your database requirements.

## Course Outline

---

- Introduction to databases: An introduction to relational databases that implement the SQL standard. This will cover standard SQL table queries, as well as a brief discussion into indexing.
- Data entry and retrieval: Using R as a means to query and modify data in a SQL database.
- Data aggregation: A dplyr front end to data manipulation on a database.
- Multiple tables: Dealing with and joining data that resides in multiple tables within a database.

## Learning Outcomes

---

By the end of the day, participants will understand...

- how to access relational databases via R
- that R can directly query and modify data in a SQL database
- how to use dplyr to access a database
- that dplyr can run commands directly in the data.

## Prior Knowledge

---

No knowledge of database software is assumed however familiarity with R programming and use of packages and functions is required. Successful completion of the Introduction to R course offered by Jumping Rivers is sufficient background.