An Introduction to SQL with R

Level: Intermediate
Duration: 6 hours

Using databases is a fundamental part of a data scientist's role. The main focus of this training course is to introduce SQL databases, write your first SQL queries, and show how R can be used to retrieve and manipulate data stored in a relational database. The course uses both the {DBI} and {dbplyr} packages.

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.
- **Standard SQL commands**: Typical SQL commands when working with a database
- **Data aggregation**: A {dplyr} and {dbplyr} 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

Session 1:

By the end of session 1 participants will...

- understand the concepts of relational database management.
- be introduced to the PostgreSQL dialect.
- understand how to form a connection using {DBI}.
- have learned how to run basic SQL commands from R:
  - extracting data with SELECT statements
  - filtering and organising data
  - creating tables and inserting data

Session 2:

By the end of session 2 participants will...

- understand how to access a database using {dplyr} and {dbplyr}.
- have learned how to run {dplyr} commands within the database.
- understand how to join tables by matching variables.

This course does not include:

- A thorough introduction to {dplyr}, see our Data Wrangling in the Tidyverse course for this.
- Database services by cloud providers like AWS and Azure, see our Intro to SQL course for an introduction.
- Connecting to databases through other programming languages like Python, see our Intro to SQL with Python course for this.

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. In particular, basic knowledge of {dplyr} is assumed.

Attendee Feedback

- "Great insight into SQL databases in R. Content delivery was excellent and the knowledge acquired will be very useful in my job."

Contact

hello@jumpingrivers.com