

Lab 5

Statistical Computing & Programming

06-16-20

Getting started

- Navigate to your team repo, `lab5-[github_teamname]`
- Open an RStudio (Pawn or Rook) session; then go to
 - `File > New Project`
 - `select Version Control`
 - `select Git`
 - `paste the repository URL - available at your GitHub repo lab5-[github_teamname] when you click Clone or download and then Clone with HTTPS`
 - `Click Create Project`
- This is a team lab.

You may do this on your local machine if you have git configured with R/RStudio.

Introduction

A database is a structured set of data. The terminology is slightly different when working with a database management system compared to working with data in R.

- field: variable or quantity
- record: collection of fields
- table: collection of records with all the same fields
- database: collection of tables

The relationship between R terminology and database terminology is explained below.

| R terminology | Database terminology |
|---------------------------|-----------------------------|
| column | field |
| row | record |
| data frame | table |
| types of columns | table schema |
| collection of data frames | database |

SQL (structured query language) allows you to directly interact with a database and perform tasks such as pulling data and making updates.

Verb connections

| SQL | dplyr |
|----------|--|
| SELECT | <code>select()</code> |
| table | data frame |
| WHERE | <code>filter()</code> pre-aggregation/calculation |
| GROUP_BY | <code>group_by()</code> |
| HAVING | <code>filter()</code> post-aggregation/calculation |
| ORDER BY | <code>arrange()</code> with possibly a <code>desc()</code> |
| LIMIT | <code>slice()</code> |

SQL arithmetic and comparison operators

SQL supports the standard +, -, *, /, and % (modulo) arithmetic operators and the following comparison operators.

| Operator | Description |
|----------|--------------------------|
| = | Equal to |
| > | Greater than |
| < | Less than |
| >= | Greater than or equal to |
| <= | Less than or equal to |
| <> | Not equal to |

SQL logical operators

| Operator | Description |
|----------|--|
| ALL | TRUE if all of the subquery values meet the condition |
| AND | TRUE if all the conditions separated by AND is TRUE |
| ANY | TRUE if any of the subquery values meet the condition |
| BETWEEN | TRUE if the operand is within the range of comparisons |
| EXISTS | TRUE if the subquery returns one or more records |
| IN | TRUE if the operand is equal to one of a list of expressions |
| LIKE | TRUE if the operand matches a pattern |
| NOT | Displays a record if the condition(s) is NOT TRUE |
| OR | TRUE if any of the conditions separated by OR is TRUE |
| SOME | TRUE if any of the subquery values meet the condition |

Today's objectives

- Follow along as the TA gets you started.
- Complete Lab 5 (24 hours to submit this team lab)
 - Work with those in your group in a breakout room
 - Grade is for effort and completion
 - This lab will be helpful for the final homework assignment - Homework 4
- If you finish early, ask questions about your project