Databases and the role of SQL

Databases in detail:
What is a database? DBMS? SQL?

https://www.youtube.com/watch?v=FR4QIeZaPeM

If you want to learn more about SQL take a free online class, e.g. at academy.vertabelo.com

Discover our Interactive Courses

SQL   MS SQL Server   Python   R language   Data Visualization   Statistics   All courses

SQL Basics

BASIC

Start here if you're new to SQL. Learn the basics of SQL, retrieve data from an SQL database and build simple reports.

FREE TEST DRIVE
LIFETIME ACCESS
56128 USERS ENROLLED   4.7/5 ★

SQL Practice Set

BASIC

Refine your SQL skills with 88 interactive exercises, ranging from simple tasks with SELECT FROM statements to more advanced problems involving multiple subqueries.

FREE TEST DRIVE
LIFETIME ACCESS
2260 USERS ENROLLED   4.7/5 ★

SQL JOINs

BASIC

Review and deepen your knowledge of SQL JOINs with 93 exercises. Practice common and less common ways of getting data from multiple tables.

FREE TEST DRIVE
LIFETIME ACCESS
1507 USERS ENROLLED   5/5 ★

A general introduction to SQL:
SQL syntax is analogous to data.table syntax

The `DT[i,j,by]` command has <u>three parts</u>: `i`, `j` and `by`. We talk about the command by saying "Take `DT`, subset the rows using `i`, then calculate `j` grouped by `by`". **❶**

R data.table	**i**	**j**	**by**
SQL	**WHERE**	**SELECT**	**GROUP BY**

If you think in SQL terminology, the `i` corresponds to `WHERE`, `j` to `SELECT` and `by` to `GROUP BY`. **❶**

For more information, see the data.table FAQ 2.16:

http://datatable.r-forge.r-project.org/datatable-faq.pdf

A general introduction to SQL:
SQL syntax is analogous to data.table syntax

The `DT[i,j,by]` command has <u>three parts</u>: `i`, `j` and `by`. We talk about the command by saying "Take `DT`, subset the rows using `i`, then calculate `j` grouped by `by`". **❶**

R data.table	i	j	by
SQL	WHERE	SELECT	GROUP BY

If you think in SQL terminology, the `i` corresponds to `WHERE`, `j` to `SELECT` and `by` to `GROUP BY`. **❶**

For more information, see the data.table FAQ 2.16:

http://datatable.r-forge.r-project.org/datatable-faq.pdf

A general introduction to SQL:
SQL syntax is analogous to data.table syntax

The `DT[i,j,by]` command has three parts: `i`, `j` and `by`. We talk about the command by saying "Take `DT`, subset the rows using `i`, then calculate `j` grouped by `by`". ❶

R data.table	`i`	`j`	`by`
SQL	`WHERE`	`SELECT`	`GROUP BY`

If you think in SQL terminology, the `i` corresponds to `WHERE`, `j` to `SELECT` and `by` to `GROUP BY`. ❶

For more information, see the data.table FAQ 2.16:

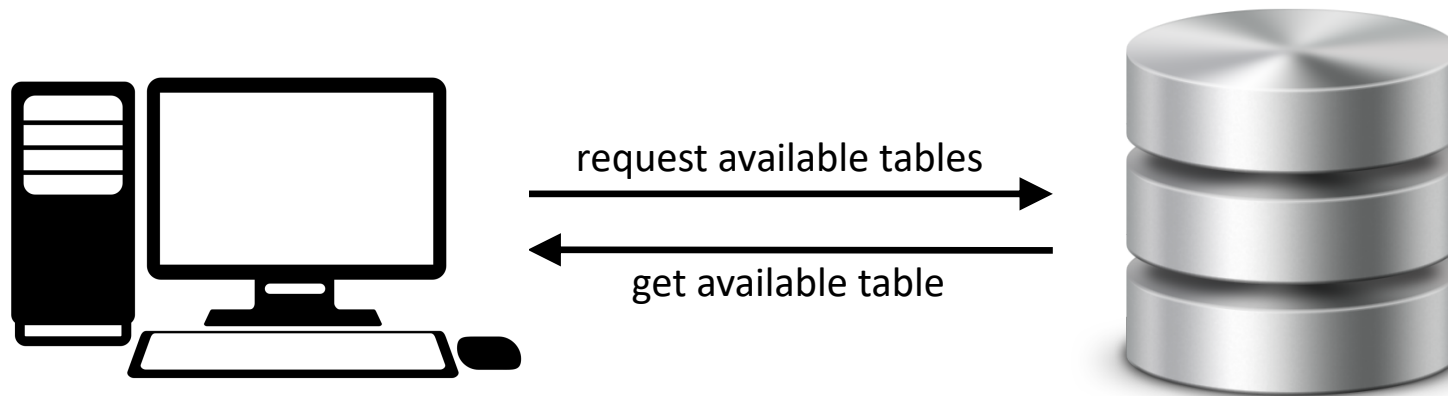http://datatable.r-forge.r-project.org/datatable-faq.pdf

Connecting to a database in 3 simple steps

1. Open connection

2. Get data / run commands

3. Close connection

A general introduction to SQL:
How to connect to a database

request available tables

get available table

1 Create a connection to the data

```
con <- dbConnect(drv=RSQLite::SQLite(),
dbname="database.sqlite")
```

2 Name of database

Recap Lecture 06:
Load Data

A general introduction to SQL:
How to execute SQL queries in R (1/2)

Customer	TransDate	Quantity	PurchAmount	Cost	…
149332	15-11-05 00:00	1	199.95	107.00	…
172951	29-08-08 00:00	1	199.95	108.00	…
120621	19-10-07 00:00	1	99.95	49.00	…
149236	14-11-05 00:00	1	39.95	18.95	…
149236	12-06-07 00:00	1	79.95	35.00	…
…	…	…	…	…	…

Specify the required database connection ❷

```
dbGetQuery(con, "SELECT * FROM transactions;")
```

❶ `dbGetQuery()` simultaneously submits and executes SQL queries to the database engine

❸ SQL query is separated by quotation marks

A general introduction to SQL:
How to execute SQL queries in R (1/2)

Customer	TransDate	Quantity	PurchAmount	Cost	…
149332	15-11-05 00:00	1	199.95	107.00	…
172951	29-08-08 00:00	1	199.95	108.00	…
120621	19-10-07 00:00	1	99.95	49.00	…
149236	14-11-05 00:00	1	39.95	18.95	…
149236	12-06-07 00:00	1	79.95	35.00	…
…	…	…	…	…	…

Specify the required database connection ❷

```
dbGetQuery(con, "SELECT * FROM transactions;")
```

dbGetQuery() simultaneously submits and executes SQL queries to the database engine ❶

SQL query is separated by quotation marks ❸

A general introduction to SQL:
How to execute SQL queries in R (1/2)

Customer	TransDate	Quantity	PurchAmount	Cost	...
149332	15-11-05 00:00	1	199.95	107.00	...
172951	29-08-08 00:00	1	199.95	108.00	...
120621	19-10-07 00:00	1	99.95	49.00	...
149236	14-11-05 00:00	1	39.95	18.95	...
149236	12-06-07 00:00	1	79.95	35.00	...
...	...	...	...	...	...

Specify the required database connection ❷

```
dbGetQuery(con, "SELECT * FROM transactions;")
```

`dbGetQuery()` simultaneously submits and executes SQL queries to the database engine ❶

SQL query is separated by quotation marks ❸

A general introduction to SQL:
How to execute SQL queries in R (2/2)

| Customer	TransDate	Quantity	PurchAmount	Cost	...
149332	15-11-05 00:00	1	199.95	107.00	...
172951	29-08-08 00:00	1	199.95	108.00	...
120621	19-10-07 00:00	1	99.95	49.00	...
149236	14-11-05 00:00	1	39.95	18.95	...
149236	12-06-07 00:00	1	79.95	35.00	...
...	...	...	...	...	...

`FROM` determines the table to which the query is applied. ❷

```
dbGetQuery(con, "SELECT * FROM transactions;")
```

`SELECT` specifies the columns that are queried. The asterisk says that the result should include all available columns. ❶

SQL queries always end with a semicolon ❸

DB step 3:
Close connection to free resources
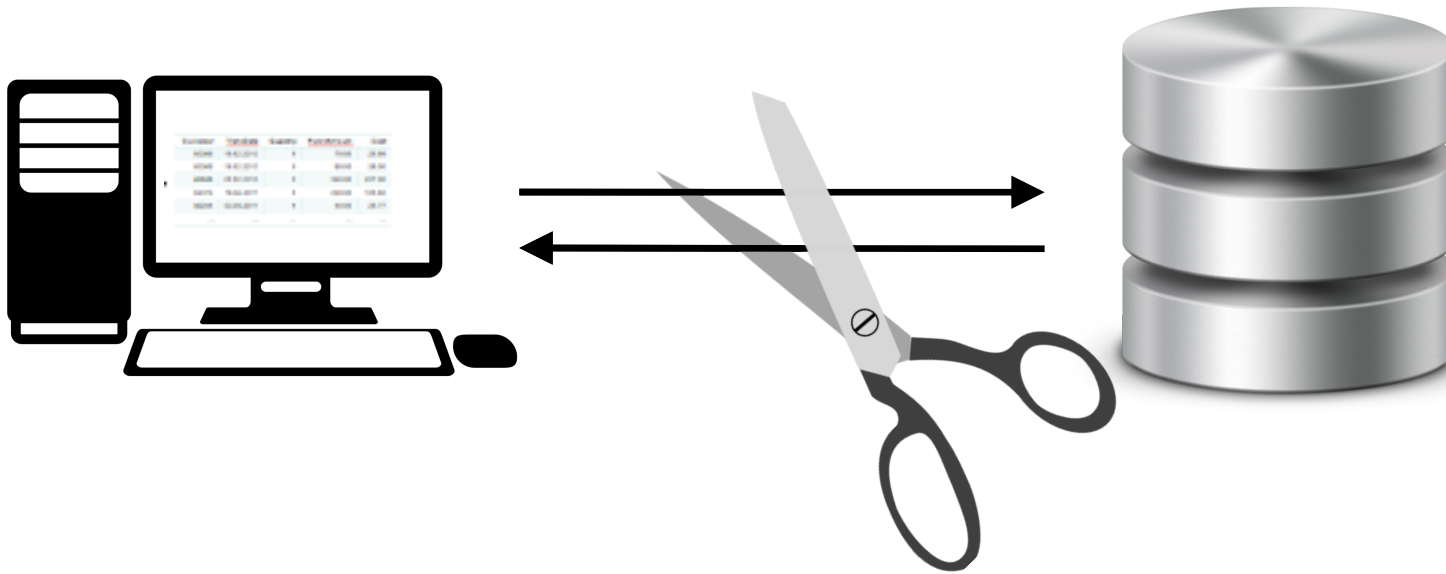
dbDisconnect(conn=con**)**

Recap Lecture 02:
Load Data