



PostgreSQL

 **Goal: install PostgreSQL and create your first Postgres database**

What is PostgreSQL?

- (PostgreSQL = postgres)
- A piece of software that lets you create databases and work with them
- This kind of software is called Relational Database Management System (RDBMS); OracleSQL, mySQL, (Microsoft) SQL Server, SQLite
- NoSQL ("not only SQL") is different from all these — more on that, and MongoDB, next week
- Two main reasons we talk about postgres / reasons it's popular: it is open source, and Amazon Redshift uses it

What is a relational database?

- Database: source of different kinds of data that you can *query* to get the information you need
- Relational database: data is stored in *tables*; connected with an identifier / *key*
- It can be queried and maintained by SQL (structured query language)

Why work with databases if we have csv and python/pandas?

- Automated updating of tables easier
- Centralized data storage, one ground truth dataset, simplifies collaboration
- Amount of data — can store much more data that fits into your computer's memory
- SQL enables people who don't necessarily know other programming languages to work with data

- (Table) constraints can serve as data pre-cleaning / data validation
- Security (authentication, permissions)
- APIs to connect to other languages / interfaces
- Backups, versioning
- Performance/speed

How do we work with a database?

- We need a client to connect to the database
 - `psql`: Postgres shell that allows you to interact with the database server via the command line
 - pgadmin/Postico/dbeaver: GUIs to interact with the database server
 - python + SQLAlchemy
- To connect to a database we need 5 pieces of information:
 - **host**: `-h` flag; today and tomorrow this will be *localhost*, don't need to specify
 - **port**: `-p` flag; default port is 5432, generally don't need to specify
 - **database**: `-d` flag; likely can do without / go with default
 - **username**: `-U` flag; either won't need it, or need to specify postgres user
 - **password**: most likely don't need (but if you do, we'll help you figure it out)
- We need to specify host (machine on which the program is running), port (postgres default: 5432), database, username, password

Getting started

- To connect to postgres via psql:
 - `psql`
 - `psql -h localhost -p 5432 -U postgres -d postgres`
- Your first psql commands!
 - To see all available databases: `\l`
 - To switch the database within your PostgreSQL server: `\c <name_of_database>`
 - To list all tables in the current database, write `\dt`

- To create a database:
 - `CREATE DATABASE <name_of_database>;` , from within psql
 - `createdb <name_of_database>` , from bash/terminal
- The use of uppercase in SQL commands is not mandatory but a convention; the use of the `;` is mandatory.
- To look around and do all the admin-stuff, we use these commands starting with a backslash; when we actually want to do something with databases and tables, we use SQL.