

# DATABASES USING SQL

**Ecology**

NEXT STEP IN THE  
DATA WORKFLOW ...

# DATABASES USING SQL

## Objectives

- Understand the benefits of using a relational database
  - Set up a small database from .csv files using SQLite
  - Understand SQLite data types
-

# DATABASES VERSUS EXCEL

## Excel

- Directly edit cells
- Use formulas based on other cells

## Database

- Send commands (called queries) to a database manager
- The manager does calculations and queries on our behalf
- Manager returns results in a tabular form

# WHY USE A DATABASE?

- Keeps data separate from analysis
  - Fast
  - Improves quality control of data
-

What files and information would I need for the following questions:

- How has the hindfoot length and weight of the animals belonging to the *Dipodomys* genus changed over time?
- What is the average weight of each species per year?
- What operations would I need to perform if I were doing these analyses by hand?

QUESTION

# GOOD DATABASE DESIGN

- Atomic values
  - One field per type of information
  - Split different classes of data into different tables
  - Relational databases connect the different classes of data together
-

# CONNECTING TABLES

The Relational Part of  
Relational Databases

- Use a shared column to connect tables
  - *Primary keys* uniquely identify a record
  - *Foreign keys* refer to a primary key in another table
-



# WRITING SQL QUERIES

Objectives

- Write and build queries
- Filter data given various criteria
- Sort the results of a query

Find the distinct years  
in which measurements  
were taken for the  
surveys.

QUESTION

Write a query that  
returns the day, month,  
and species\_id for  
individuals caught on  
Plot 1 that weigh more  
than 75 grams.

QUESTION

Write a query that  
returns year,  
species\_id, and weight  
in kilograms from the  
surveys table, sorted  
with the largest weights  
at the top.

QUESTION

# SQL AGGREGATION

## Objectives

- Apply aggregation to group records in SQL
  - Filter and order results of a query based on aggregate functions
  - Save a query to make new tables
  - Apply filters to find missing values in SQL
-

Write a query that  
returns: total weight,  
average weight, and the  
min and max weights for  
all animals caught over  
the duration of the  
survey.

QUESTION

Can you modify it so  
that it outputs these  
values only for weights  
between 5 and 10?

QUESTION

Write a query that  
returns the number of  
genus in each taxa from  
the species table.

Return only the taxa  
with more than 10 genus.

QUESTION



# JOINS

## Objectives

- Employ joins to combine data from two tables
  - Apply functions to manipulate individual values
  - Employ aliases to assign new names to items in a query
-

Write a query that  
returns 30 instead of  
NULL for values in the  
hindfoot\_length column.

QUESTION

# FINAL THOUGHTS

- SQL allows us to ask questions of our data
- Can you translate the following questions into SQL queries?

---

- How many of each plot type is there?
- What is the average weight of each taxa?
- How many specimens of each sex are there for each year?

QUESTION