Introduction to Programming

For Archaeologists

Part 5: SQL & Databases

2021-2022

Universiteit Leiden
The Netherlands

Topics of this lecture series

1. Introduction: Python, variables, comments
2. Lists & Loops
3. Loading and manipulating data
4. Graphs & Plots
5. **SQL & Databases**
6. Advanced methods: Machine Learning, QGIS integration

Assignment

Assignment deadlines

- Assignment 1: 22 April
- Assignment 2: 6 May
- Assignment 3: **20 May**

Assignment 3 can now be found on github!

Topics of this lecture

- Databases
- Entity Relationship Diagram (ERD)
- SQL queries
    - SELECT
    - Wildcards
    - INSERT
    - UPDATE
    - DELETE
    - INNER JOIN

After this session:

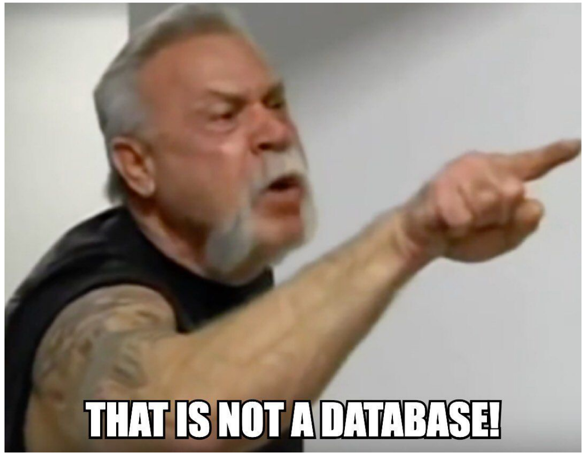
- You know what a database is
- You can interpret an ERD
- You know the basic syntax of SQL
- You know how wildcards function in SQL
- You can write SQL queries (SELECT, INSERT, UPDATE, DELETE, INNER JOIN)

What is a database?

- Stores information in columns / rows, but calls them fields / records

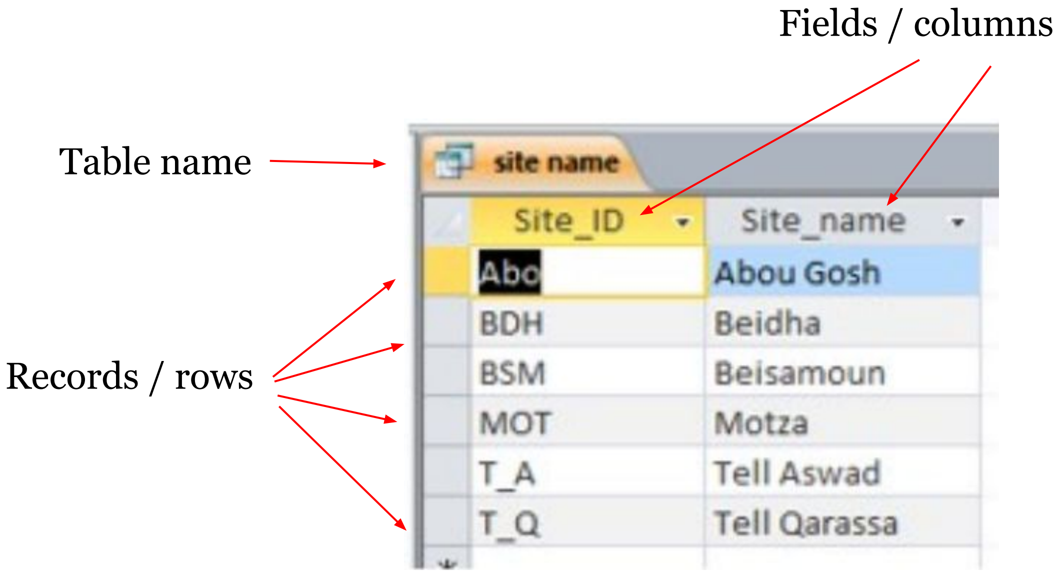
- Field names and types are defined in advance
- Can hold more data than spreadsheet
- Can have multiple linked tables
- Data integrity (data is checked on input)
- Forms for data entry
- Advanced querying

WHEN SOMEONE SAYS THE HAVE AN "EXCEL DATABASE"



What is a table?

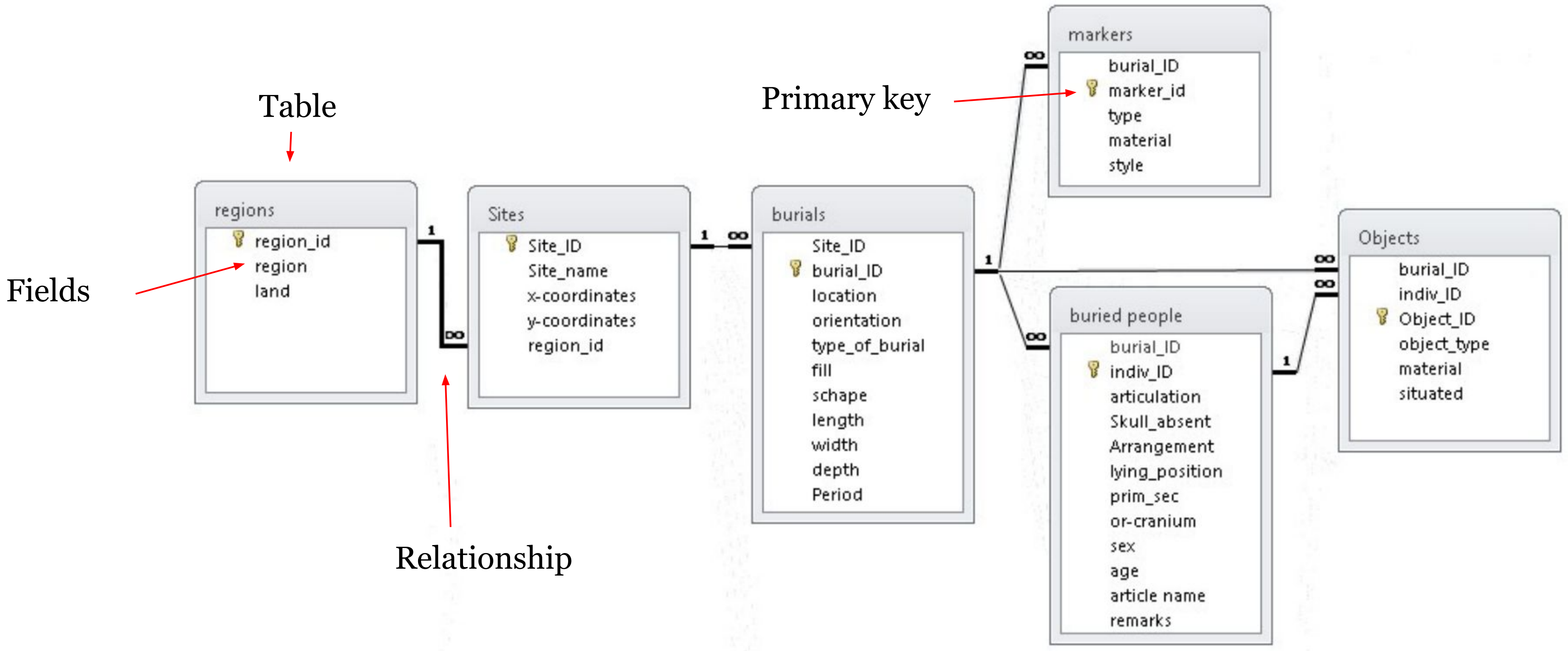
- A table is similar to a tab on Excel
- Contains data in columns and rows



site name

Site_ID	Site_name
Abo	Abou Gosh
BDH	Beidha
BSM	Beisamoun
MOT	Motza
T_A	Tell Aswad
T_Q	Tell Qarassa

Entity Relationship Diagram



How to use DB in Python

- Easiest way
  - Export table(s) as CSV, open in Pandas
  - Can’t edit the database
  - Good for analysis of single table
  - Difficult to use multiple tables

- Advanced method
  - Connect Python to database (Access / Mysql / SQLite / etc)
  - Allows editing
  - Allows use of SQL
  - Easier to use multiple tables

Structured Query Language (SQL)

- Lets you access and manipulate databases
- Uses ‘queries’ to get, insert, update, delete information
- MS Access uses GUI query builder
- Python has no GUI, need to learn syntax

SQL Syntax

- Need to define:
  - What type of query we want
  - Which table(s) to query
  - Which fields we want in the result
  - A criteria (true/false statement)
- Tradition: type in ALL CAPS

SQL Syntax

```
SELECT Site_ID, Site_name FROM Sites WHERE region_id == ‘1’
```

Query type

Fields to show in result

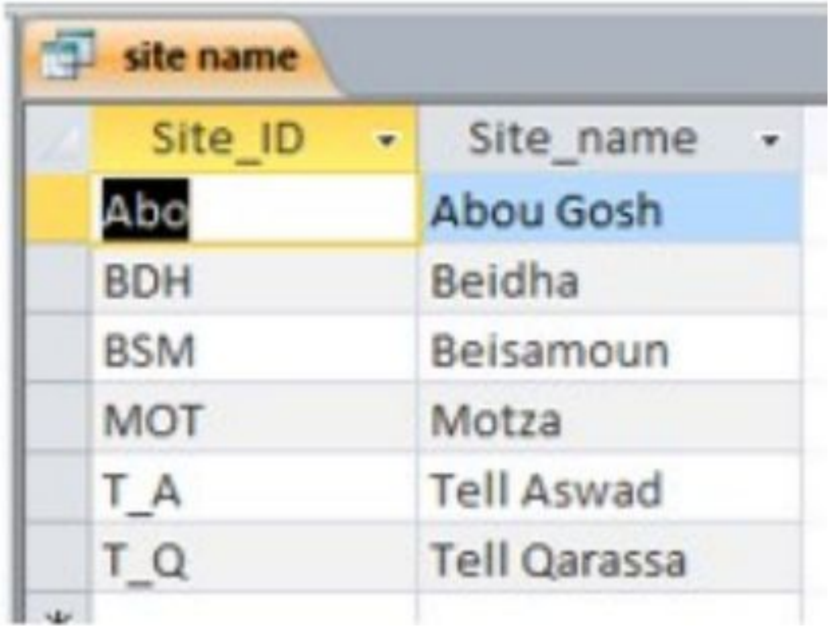
Table name

Criteria

Keywords in ALL CAPS
Table / field names as they are defined in db

Result

```
SELECT Site_ID, Site_name FROM Sites WHERE region_id == 1
```



site name

Site_ID	Site_name
Abo	Abou Gosh
BDH	Beidha
BSM	Beisamoun
MOT	Motza
T_A	Tell Aswad
T_Q	Tell Qarassa

Query types

- SELECT: select data from table(s) and return the relevant records
- INSERT: insert a new record into an existing table
- UPDATE: update 1 or more fields of an existing record
- DELETE: delete a record

WHERE

A condition to select only certain rows

```
SELECT * FROM Sites WHERE region_id == 1
```

Similar to Python `if` statements!

Wildcards / LIKE

A placeholder that can mean ‘anything’

- `SELECT * FROM graves`
  - Select every column from this table

- `SELECT * FROM graves WHERE type LIKE “a%”`
  - Select only records where grave type starts with ‘a’

AND

Just like Python if statements, you can combine conditions

```
SELECT *
FROM Sites
WHERE
    region_id == 1 AND
    dating == ‘bronze age’
```

(get all sites in region 1 dated to bronze age)

OR

```
SELECT *
FROM Sites
WHERE
    region_id == 1 OR
    region_id == 2
```

(get all sites located in region 1 or 2)

GROUP BY

Similar to DataFrame `groupby()`

```
SELECT region_id, COUNT(region_id)
FROM Sites
GROUP BY region_id
```

(Count number of sites grouped by region)

After this session:

- You know what a database is
- You can interpret an ERD
- You know the basic syntax of SQL
- You know how wildcards function in SQL
- You can write SQL queries (SELECT, INSERT, UPDATE, DELETE, INNER JOIN)

Questions?

- **Any questions about any of the subjects?**

- Contact me at
    - a.brandsen@arch.leidenuniv.nl

Slides are available on Brightspace

Exercises

github.com/alexbrandsen/Introduction-to-Programming-for-Archaeologists

- Go to github
- Click on ‘modules’
- Click click on the 5th module
- Right click ‘raw’, then select ‘save link as’ or ‘download as’
- Save the file in the ‘modules’ folder within your own Scripts folder
- **Also download the ‘grafheuvels.sqlite’ file from the ‘data’ folder**
- Start Anaconda
- Start Jupyter Notebook
- Navigate to the notebook file and run it