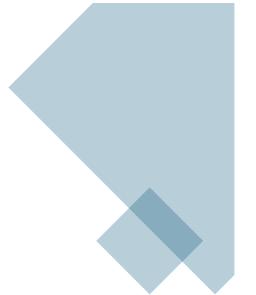


INTRODUCTION TO SQL

QUANTITATIVE COMPUTING LAB

DANIEL PEREZ
GRADUATE FELLOW



After this session you'll be able to:

- Use DBeaver and SQL statements
 - Create a database
 - Define table elements
 - Import tables and data into your project





PRE-REQUISITES

<https://dbeaver.io/download/>

TODAY'S DATA

Documentation:

<https://www.sqlite.org/>

<https://github.com/CMC-QCL/Introduction-to-SQL-DBeaver-Pt1>

State Crime CSV

Information on the crime rates and totals for states across the United States for a range of years: 2010, 2014 and 2019

https://github.com/CMC-QCL/Introduction-to-SQL/blob/main/state_crime.csv <<https://corgis-edu.github.io>>

State Demographics CSV and SQL Files

[state_computer_data.sql](#), [state_workforce.csv](#), [state_people.sql](#)

Summarized information obtained about state demographics in the US between 2015 - 2019 via the United States Census Bureau

https://corgis-edu.github.io/corgis/csv/state_demographics/

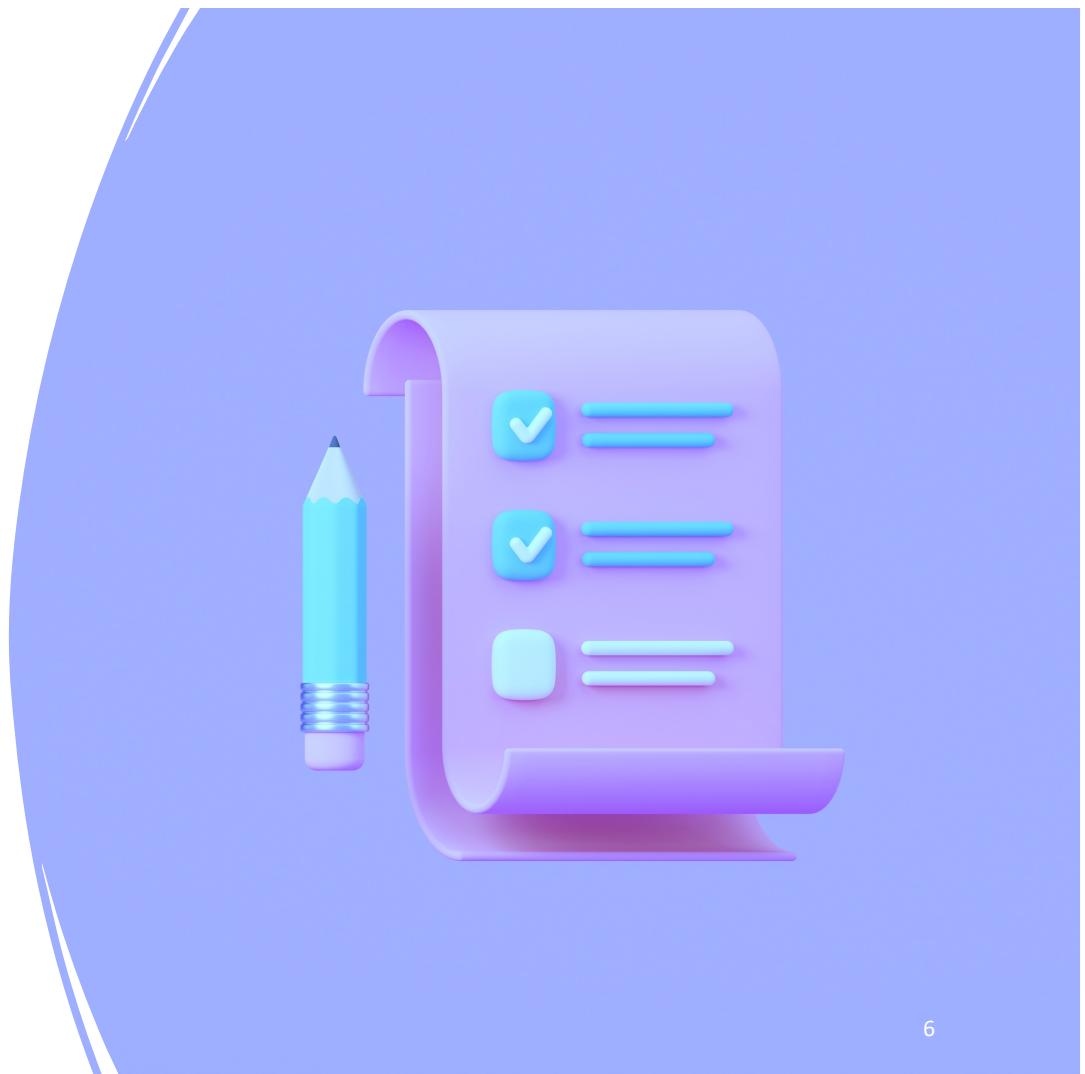
Databases

A database is a collection of related data with an implicit meaning.



IMPLICIT PROPERTIES

- Represents some aspect of the real world
- Has a logically coherent collection of data with some inherent meaning
- It has specific purpose. Preconceived application with a user target.



DATABASE ADVANTAGES



BETTER DATA INTEGRATION
Improves data handling and reduces redundancy

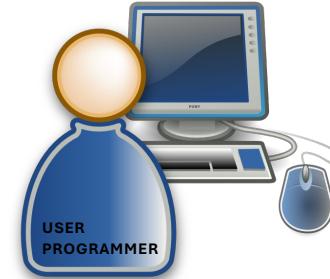


STORAGE IS MORE SECURE
Provides better privacy and security policies



FASTER DATA ACCESS Produce quick answers to data queries

<The three-level schema architecture>



<External level>

<Conceptual level>

Database System

- Application Programs/Queries

<Internal schema>

DBMS software

Software to process Queries/Programs

Software to Access Stored data

Stored database definition (Meta-data)

Stored database

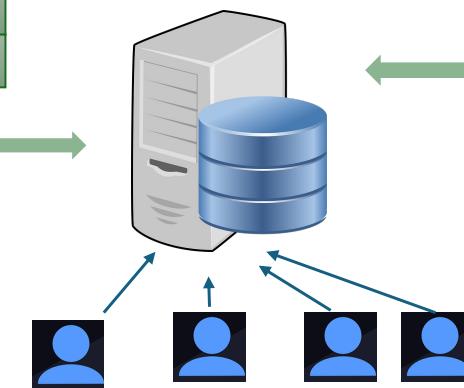
DATABASE MANAGEMENT SYSTEM (DBMS)

Computerized system that enables users to create, protect, maintain and share a database

Defining

- Data Types
- Data structures
- Data constraints

Constructing



Manipulating



RELATIONAL DATABASE MANAGEMENT SYSTEMS

(RDBMS)



Postgres Database





The screenshot shows the DBeaver 24.0.5 interface with several panes highlighted by orange boxes:

- CONNECTION**: Located at the top left, showing connection details for "QCLWorkshop".
- EDITOR**: The central pane where SQL queries are written and executed.
- RESULTS**: The pane below the editor displaying the results of a query. The results table has columns: #, State, Persons_per_household, Households_with_computer, and Households_with_internet. The data includes rows for Alabama through Missouri.
- PROJECT**: The bottom-left pane showing project structure and files.
- EDITOR CONTROLS**: The bottom-right pane containing various controls and settings for the editor.

Editor Content (SQL Query):

```
SELECT * FROM state_computer_data
-- Create a table
CREATE TABLE state_computer_data (
    State TEXT,
    Persons_per_household REAL,
    Households_with_computer REAL,
    Households_with_internet REAL
)
-- Insert Values
INSERT INTO state_computer_data
VALUES("Alabama", 2.55, 85.5, 76.4)
```

DBeaver Community is a free cross-platform **database tool** for developers, database administrators, analysts, and everyone working with data. It supports all popular SQL databases like MySQL, MariaDB, PostgreSQL, SQLite, Apache Family, and more.

RELATIONAL DATA MODEL AND TABLE REPRESENTATION

Concepts such as

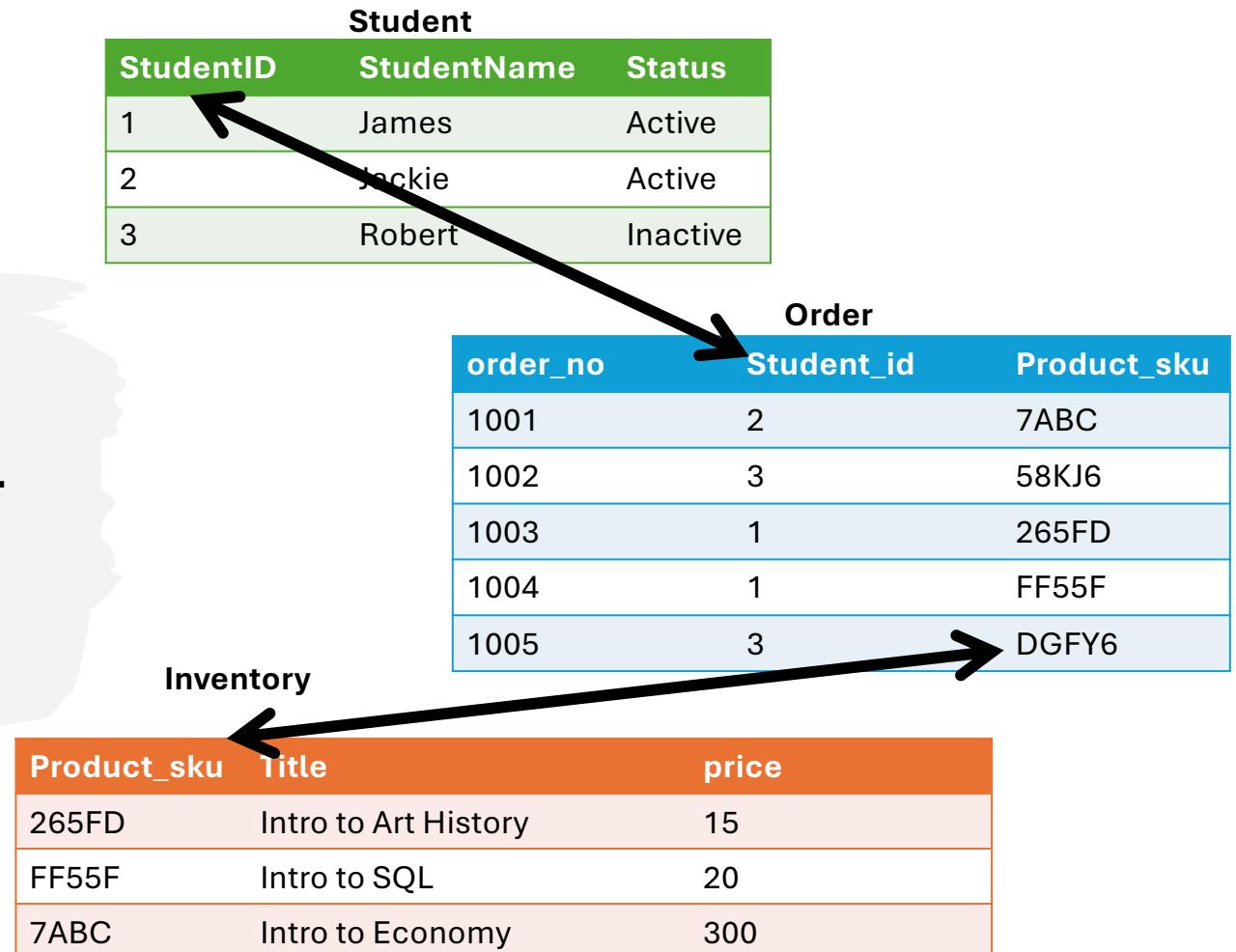
- Entities
- Attributes
- Relationships



	StudentID	StudentName	Status
Row => Tuple/Records	1	James	Active
	2	Jackie	Active
	3	Robert	Inactive

Column =>
Attribute/field

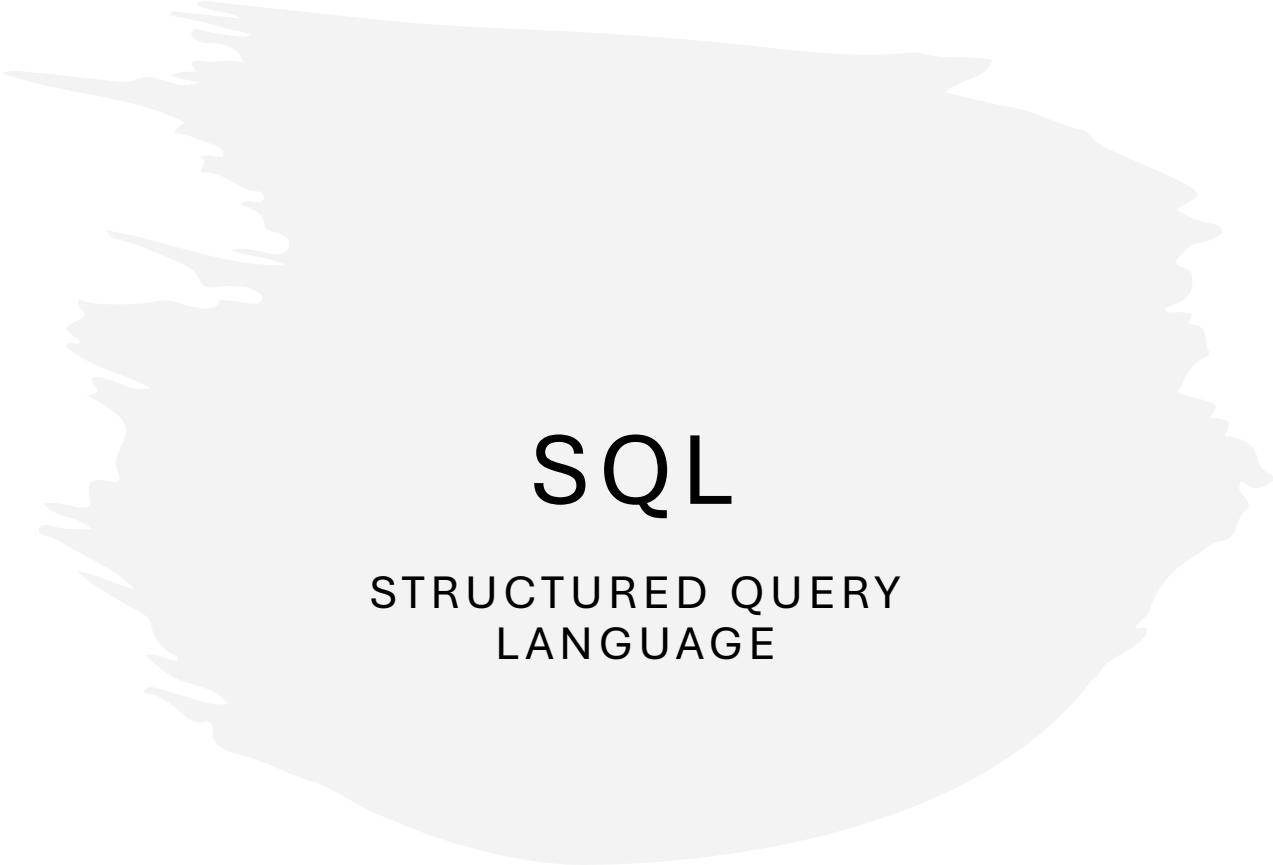
RELATIONAL DATABASES



STORAGE CLASSES AND DATA TYPES

- NULL
- INTEGER
- REAL
- TEXT
- BLOB

Integer	Character	Boolean / Flag	Date and time	Float / Decimal
StudentID	StudentName	StatusActive	DOB	GPA
1	James	1 - True	2002-03-12	2.2
2	Jackie	1 - True	2005-01-19	3.6
3	Robert	0 - False	2000-06-06	3.8



SQL

STRUCTURED QUERY
LANGUAGE

QUERY

An application program
accesses the database by
sending queries for the data to
the DBMS

Transactions
Data
**Read Written**

STRUCTURED QUERY LANGUAGE

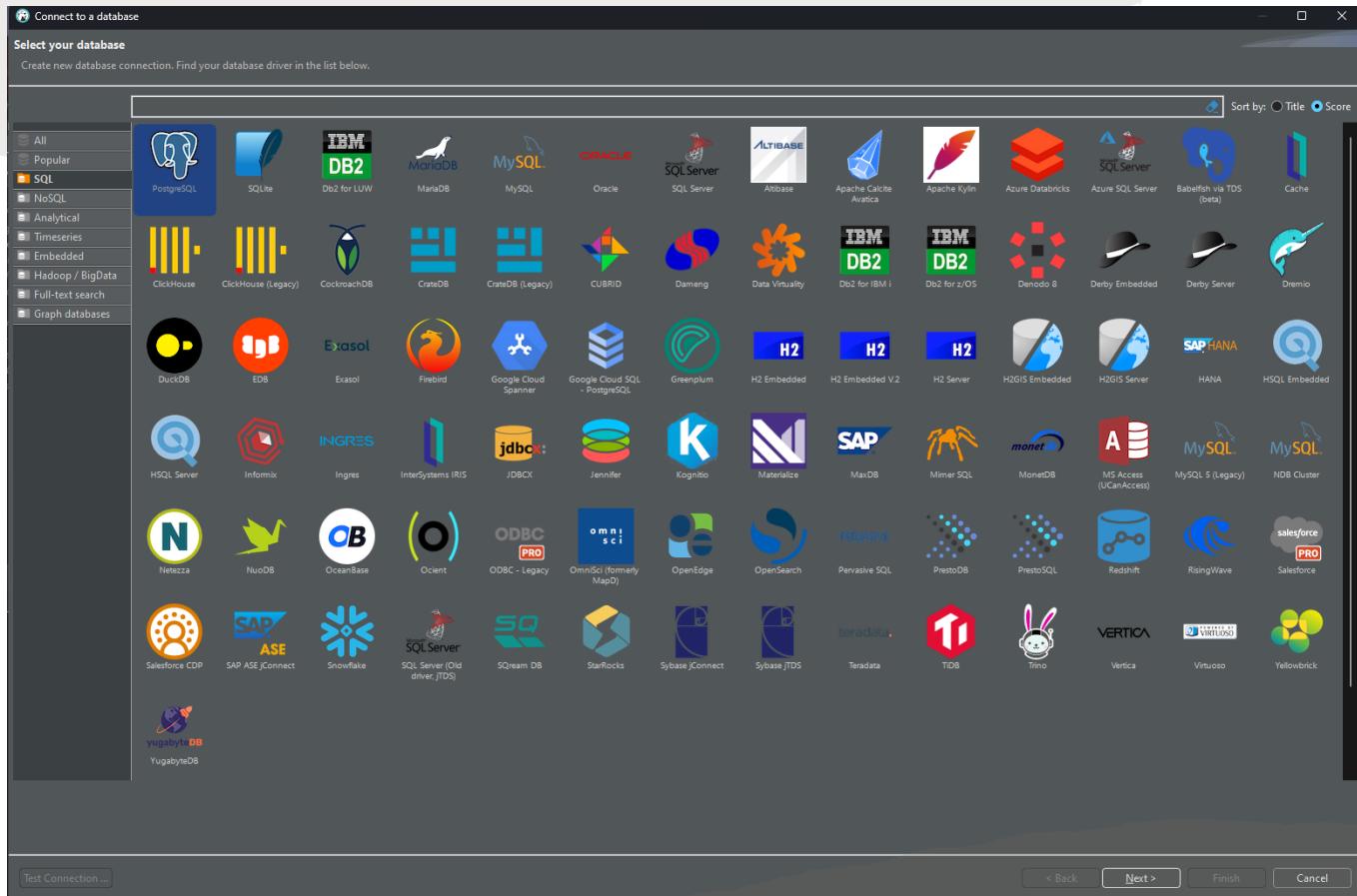
(SQL) is a programming language for storing and processing information in a relational database.

Why do we use it?

Integration

Embed SQL queries with Java
Fairly easy to learn as it uses common English keywords in its statements

SQL SYNTAX



DATA TRANSACTION LANGUAGE

Definition (DDL)

CREATE, DROP, ALTER, TRUNCATE

Manipulation (DML)

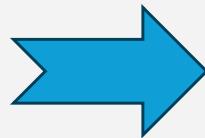
INSERT, UPDATE, DELETE

Query (DQL)

SELECT, JOIN

Data Control or transaction control (DCL/TCL)

GRANT, REVOKE



Create
Read
Update
Delete

DATA DEFINITION & CONSTRAINTS

DEFINE SQL DATA TYPES
COLUMN<ATTRIBUTE> | DATA_TYPE

DEFINE KEYS
COLUMN| DATA_TYPE| PRIMARY KEY

DATA DEFINITION & CONSTRAINTS

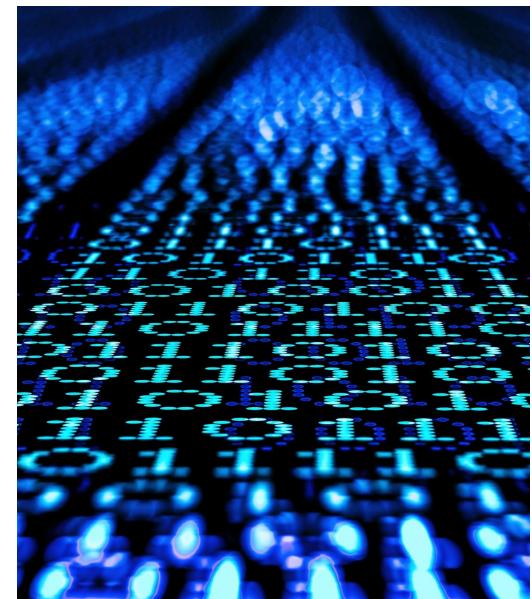
DEFINE CONSTRAINTS FOR DATA TYPES

```
STATE VARCHAR(50) PRIMARY KEY,  
PERSONS_PER_HOUSEHOLD INT CHECK(PERSONS_PER_HOUSEHOLD > 0),  
HOUSEHOLDS_WITH_COMPUTER DECIMAL(10,2),  
HOUSEHOLDS_WITH_INTERNET DECIMAL(10,2))
```

CREATE & DROP TABLES

```
CREATE TABLE TABLE_NAME(  
DATA DEFINITION  
)
```

```
DROP TABLE TABLE_NAME
```



Create the table **state_computer_data**

Name	Value
Updated Rows	50
Query	<pre>CREATE TABLE state_computer_data (State TEXT PRIMARY KEY, Persons_per_household DECIMAL(10,2), Households_with_computer DECIMAL(10,2), Households_with_internet DECIMAL(10,2))</pre>
Start time	Mon Jun 10 18:26:34 PDT 2024
Finish time	Mon Jun 10 18:26:34 PDT 2024

QCLWorkshop > QCLWorkshop state_computer_data

Properties Data ER Diagram

state_computer_data Enter a SQL expression to filter results (use Ctrl+Space)

Grid Text Record

	State	Persons_per_household	Households_with_computer	Households_with_internet
1	Alabama	2.55	85.5	76.4
2	Alaska	2.8	94.1	85.5
3	Arizona	2.68	91.7	84.1
4	Arkansas	2.52	86.2	73
5	California	2.95	93	86.7
6	Colorado	2.56	93.9	87.6
7	Connecticut	2.53	90.8	85.5
8	Delaware	2.57	91.6	85
9	District of Colur	2.3	91.8	82.6
10	Florida	2.65	91.5	83
11	Georgia	2.7	90.2	81.3
12	Hawaii	3	91.2	84.8
13	Idaho	2.68	91.8	82.7
14	Illinois	2.57	89.9	82.7
15	Indiana	2.52	88.7	80.1
16	Iowa	2.4	89	80.8
17	Kansas	2.51	90	81.8
18	Kentucky	2.49	86.4	78.4
19	Louisiana	2.61	85.6	75.5
20	Maine	2.32	89.7	82.1
21	Maryland	2.67	92.4	86.4
22	Massachusetts	2.52	91.4	86.4
23	Michigan	2.47	89.6	81.5
24	Minnesota	2.49	91.6	84.8
25	Mississippi	2.62	83.8	71.5
26	Missouri	2.46	89	80.2
27	Montana	2.39	88.9	80.7
28	Nebraska	2.45	90	83.4
29	Nevada	2.67	92.5	83.2
30	New Hampshire	2.46	93	87.7
31	New Jersey	2.69	91.4	85.8
32	New Mexico	2.63	85.9	74.6
33	New York	2.59	89.6	82.8
34	North Carolina	2.52	89.1	80.7
35	North Dakota	2.3	89.8	80.7
36	Ohio	2.43	89.1	82
37	Oklahoma	2.58	88.6	78.6
38	Oregon	2.51	93	85.9
39	Pennsylvania	2.45	88	81.5
40	Rhode Island	2.47	89.1	84
41	South Carolina	2.54	88.3	78.2

Refresh Save Cancel Export data 51 200 51 51 row(s) fetched - 0.001s, on 2024-06-09 at 23:17:08

HANDS-ON

CREATE TABLE AND INSERT DATA

state_people		
ABC	state	TEXT
123	employment_firms_total	INTEGER
123	age_percent_under_18_years	REAL
123	age_percent_65_and_older	REAL

Hint: Insert data into the new table with the file
`state_people_insert.sql`

HANDS-ON

IMPORT FILES

Import file named *state_workforce*

Import file named *state_crime*

- What are the table's attributes?
- Data Types
- Constraints
- Keys

state_computer_data	
ABC	State
123	Persons_per_household
123	Households_with_computer
123	Households_with_internet

state_people	
ABC	State
123	employment_firms_total
123	age_percent_under_18_years
123	age_percent_65_and_older

state_crime	
ABC	State
123	Crime_Year
123	Population
123	Rates_Property_Theft
123	Rates_Violent_Robbery
123	Totals_Property_Theft
123	Totals_Violent_Robbery

state_workforce	
ABC	State
123	Population_Percent_Change
123	Population_2014
123	Population_2010
123	Education_High_School_or_Higher
123	Education_Bachelors_Degree_or_Higher
123	Sales_Retail_Sales
123	Mean_Travel_Time_to_Work
123	Income_Median_Household_Income
123	Income_Per_Capita_Income
123	Income_Persons_Below_Poverty_Level