SQLite on the command line, via Hillary Clinton's emails

Hillary Clinton emails (2016), 50MB, .sqlite database

- https://www.kaggle.com/kaggle/hillary-clinton-emails
- http://graphics.wsj.com/hillary-clinton-email-documents/
- https://www.wikileaks.org/clinton-emails/

The commands below assume

- You have sqlite installed
- You are in the same directory as the clinton.sqlite database

WRITE	GET
sqlite3	Start running sqlite 3
.open clinton.sqlite	Open a connection to the database
sqlite3 clinton.sqlite	Start running sqlite3 and open connection to db
.help	List all the .commands
.database(s)	Show the database(s) we have open
.tables	Show the tables in the db
.schema Persons	Show the cols & data type in Persons table
.mode column	Display data prettily, in columns
.mode table	Display data prettily, in a table
.mode box	Display data prettily, in a box[1]
.headers on	Display column names
.exit	Close and exit
.quit	Close and exit
.show	Show all the display options
.shell clear	This runs 'clear' on sqlite so you can unclutter

---

[1] *markdown*, *line* and *quote* are also available for .mode

From here we are writing SQL syntax and need to finish our queries with a semicolon ; CAPITALS are not required here, but it's good practice to respect them	
SELECT name FROM sqlite_master WHERE type ='table' AND name NOT LIKE 'sqlite_%';	Show all tables in the db
PRAGMA table_info (some_table);	Show cols & data type for the table
SELECT COUNT(*) FROM table;	Count the rows in the table
SELECT * FROM table LIMIT 5;	Show table, only first 5 rows
SELECT column FROM table WHERE column = "some text";	Filter a column, returning only that column
SELECT column_one, column_two FROM table WHERE column = 'some text';	Filter a column returning it and another column
SELECT * FROM table WHERE column LIKE 'W%';	Filter a column for words beginning W
SELECT COUNT(*) FROM table WHERE column LIKE '%are%';	Count rows, having filtered column for words containing -are-
SELECT COUNT (DISTINCT column) FROM table;	Count unique rows in column

Questions

How many tables are there in this database?

What shape is each table? (rows, columns)

Who is the 13th person listed in the *Persons* table?

Which of the two Milibands appears in the *Persons* table?

How many email subjects mention "Benghazi" in the *Emails* table?

What's in the text of the email (*Emails*) that has "New York Times" as MetaDataFrom?

(Select just one column)

Who sent it ? (Select just two columns)

How many different email recipients are there in the ExtractedTo column of *Emails*?

Answers

.schema

SELECT COUNT(*) FROM Emails;

SELECT * FROM Persons LIMIT 15;

SELECT * FROM Persons WHERE Name LIKE 'D%';

SELECT COUNT(*) FROM Emails WHERE ExtractedSubject LIKE '%Benghazi%';

SELECT ExtractedFrom, ExtractedBodyText FROM Emails WHERE MetaDataFrom = 'New York Times'

SELECT ExtractedBodyText FROM Emails WHERE MetaDataFrom = 'New York Times'

SELECT COUNT(*) FROM Emails WHERE ExtractedSubject LIKE '%Benghazi%';

SELECT COUNT (DISTINCT ExtractedTo) FROM Emails;