

Overview

“A little learning is a dangerous thing.”

“IF the only tool you have is a hammer,
THEN every problem looks like a nail”

You (Smith students) know more about R than I do.

You did homework, workshops, projects, etc.

Whereas, I merely audited 😊

Be skeptical. “Push back” on my observations.

Part I: Introduction to SQL

Getting Started

Introduction: SQL

Introduction: Embedding SQL in R

Part II: R and SQL

More about SQL

More about Embedding SQL in R

Dealing with “Large” Data

Other Stuff (Maybe)

“Pop Quiz”

Procedural Languages

Real-World

FORTRAN

COBOL

C

C++

Java

Python

Educational

BASIC

PASCAL

Languages with Mathematical Heritage

SQL

APL

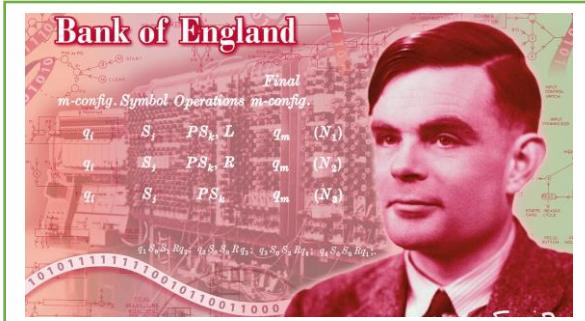
Prolog

Datalog

R

Another Pop Quiz: History

Who are these People?



Grace Hopper (“Amazing Grace”)

- “Bug” in your program

She gets credit for coining the name of a ubiquitous computer phenomenon: the bug. In August 1945, while she and some associates were working at Harvard on an experimental machine called the Mark I, a circuit malfunctioned. A researcher using tweezers located and removed the problem: a 2-in. long moth.

Hopper taped the offending insect into her logbook. Says she: "From then on, when anything went wrong with a computer, we said it had bugs in it." (The moth is still under tape along with records of the experiment at the U.S. Naval Surface Weapons Center in Dahlgren, Va.) Smithsonian – 12/2013

- U.S. Navy Destroyer: USS Hopper
 - Presidential Medal of Freedom (2016)
 - Honored by Nvidia's Hopper GPU & Grace CPU
 - Optimism: Future of Winner of Turing Award

Why Have So Few Women Won the Most Important Award in Computing?
Since 1966, 70 computer scientists have won the Turing Award. Only three
have been women. Susan D'Agostino, Slate 6/2020

Alan Turing

Alan Turing, Computing Genius & WWII Hero On U.K.'s New 50-Pound Note NPR, ALL THINGS CONSIDERED 7/15/19

Alan Turing, the **father of computer science and artificial intelligence** who **broke Adolf Hitler's Enigma code system** in World War II — but who died an outcast because of his homosexuality — will be featured on the Bank of England's new 50-pound note.

- Hero (unknown for many years)

Movie (2014): Imitation Game

- “Father of CS and AI”

Turing Machine

[Alan Turing \(Stanford Encyclopedia of Philosophy\)](#)

COMPUTING MACHINERY & INTELLIGENCE, Mind 49

[computing machinery and intelligence - a.m. turing, 1950 \(abelard.org\)](#)

- Interesting aside

[Meet the female codebreakers of Bletchley Park | Work & careers | The Guardian](#)

R

```
Result <- flights |>
  filter( ) |>
  select( )
```

R is functional

Above pipeline could be recoded as a function.

SQL

```
SELECT ___, ___
FROM FLIGHTS
WHERE ___
```

SQL is declarative

Above SELECT can be understood as the definition of a subset.

Not Procedural - No (explicit) iteration
(Advantages: Baumer's Chapter 7)

However, both languages are still “somewhat” procedural

1. Specify table (data frame)
2. Pull out some rows
3. Pull out some columns

SQL at Smith College

Smith Courses that may “touch on” SQL

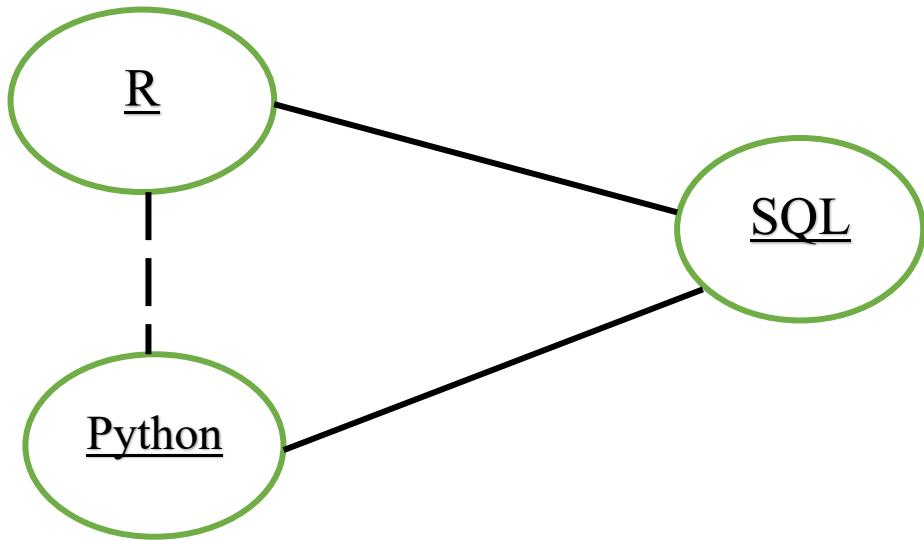
SDS 192 Introduction to Data Science

SDS 261 SQL for Data Science (1 credit)

SDS 270 Advanced Programming for Data Science

CSC 230 Introduction to Database Systems Databases

Data Science: Computer Languages



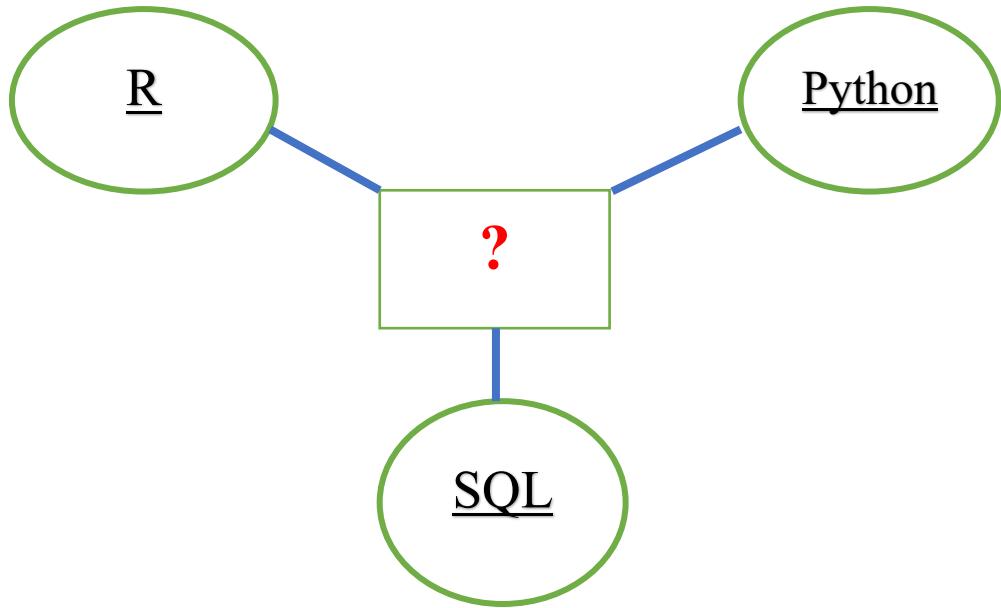
Observations & Question:

R and Python are relatively new programming languages

SQL is a very old programming language (late 1970s).

What accounts for SQL's longevity and popularity?

Maybe Someday?

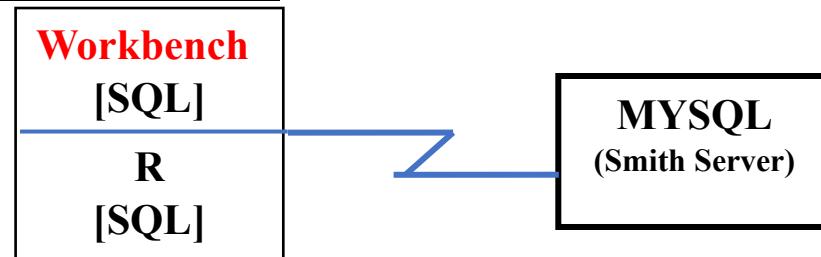


“Hands On” – Four Architectures

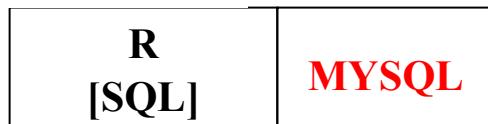
CS-1: No Installs for Smith Students



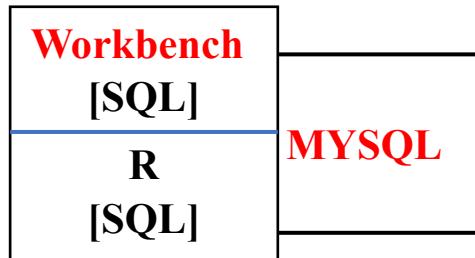
CS-2: Install Workbench



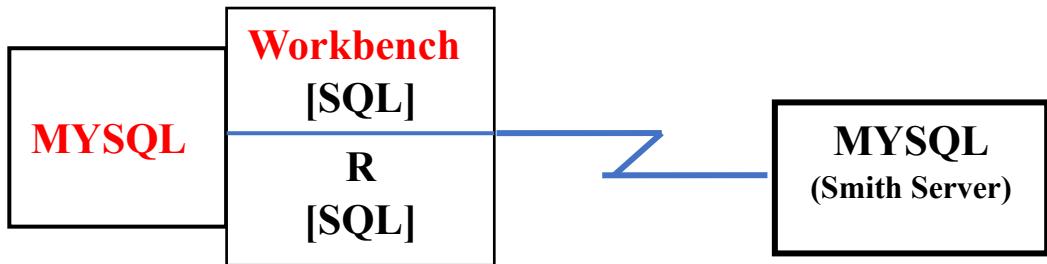
PC-1: Install MYSQL



PC-2: Install MYSQL & Install Workbench



“Everything” Architecture



PART I: Introduction to SQL

Objective: Learn *a little bit* about querying a database using SQL.

Reference: Baumer, Sec 15.4 – The SQL Data Manipulation Language

See related PDF document: Smith-NotesForVideos-0-4.pdf

Assumptions:

Tables have been designed. (Assume design is tidy.)

Data has been inserted into tables.

“Hello World”

Demo on Martyn’s PC: One-time getting-started stuff

Assume DBA privileges:

```
SHOW DATABASES;
```

```
CREATE DATABASE SMITHDEMO
```

```
SHOW DATABASES;
```

```
USE SMITHDEMO;
```

```
CREATE TABLE JUNKTAB
  (COL1 INTEGER,
  COL2 CHAR (2));
```

```
SHOW TABLES;
```

```
DESCRIBE JUNKTAB;
```

```
INSERT INTO JUNKTAB (100, 'HI');
INSERT INTO JUNKTAB (200, 'BY');
```

```
SELECT * FROM JUNKTAB;
```

```
DROP TABLE JUNKTAB;
```

```
SHOW TABLES;
```

```
USE freesql_db;
```

Introduction to Embedding SQL in R

```
library(tidyverse)
```

```
library(RMySQL)
```

```
library(dbplyr)
```

```
db <- dbConnect (MySQL(),  
host = "scidb.smith.edu",  
user = "sds192",  
password = "DSismfc@S")
```

```
class(db)
```

```
dbGetQuery(conn = db, "SHOW DATABASES;")
```

```
dbGetQuery(conn = db, "USE citibike;")
```

```
dbGetQuery(conn = db, "SHOW TABLES;")
```

```
dbGetQuery(conn = db, "DESCRIBE station_months;")
```

```
dbGetQuery conn = db, "SELECT *
  FROM station_months
  LIMIT 10;")
```

```
dbGetQuery (conn = db, "SELECT station_id, name, num_starts
  FROM station_months
  WHERE num_starts > 15000
  LIMIT 10;")
```

Optional Exercise: Fill in the blanks.

```
dbGetQuery(conn = db, " SELECT ___, ___
  FROM station_months
  WHERE _____
  LIMIT 10;")
```

Workshop: Embed following SELECTs in R

[Optionally make modifications to SQL]

Chapter 2 ORDER BY

```
SELECT station_id, name
FROM station_months
WHERE num_starts > 15000
ORDER BY station_id
LIMIT 10;
```

```
SELECT * station_id, name
FROM station_months
WHERE num_starts > 15000
ORDER BY station_id DESC
LIMIT 10;
```

Chapter 3 DISTINCT

```
SELECT DISTINCT station_id, name
FROM station_months
WHERE num_starts > 10000
LIMIT 10;
```

Chapter 4 AND-OR-NOT

```
SELECT station_id, name, num_starts
FROM station_months
WHERE num_starts >= 10000 AND num_starts <= 10050
ORDER BY station_id
LIMIT 10;
```