WORK WITH DATABASES DIRECTLY FROM R

Amy F. Szczepański • @amyszczepanski
Art of Problem Solving • amy@aops.com

TOOLS FOR TODAY

- Database software: SQLite

- R packages: DBI, RSQLite, stringr

- Database: Chinook (SQLite version)
`https://github.com/lerocha/chinook-database/blob/master/ChinookDatabase/DataSources/Chinook_Sqlite.sqlite`

AoPS Art of Problem Solving

AMYSZ ▾

PROGRAMS   BOOKSTORE   CLASSES   AOPS NEWS

Training Today's Brightest Minds to Solve Tomorrow's Problems

Art of Problem Solving creates educational materials for motivated students in grades 2–12. Our books, classes, and online resources help students become creative, successful problem solvers.

LEARN MORE

$$\lim_{n \to \infty} \left(\, \right)$$

$$a^2 - b^2 = (a+b)(a-b)$$

AoPS Online

Textbooks and Online Classes for Grades 5–12

▸ Complete math curriculum for middle and high school, as well as full courses for math contest preparation

▸ Real-time classes held in our unique digital classroom, led by accomplished math instructors

▸ Rigorous materials that teach in-depth problem-solving skills, preparing students for prestigious universities and STEM careers

▸ Engaging community full of motivated students—over 375,000 members strong and growing!

AoPS Beast ACADEMY

Books and Online Curriculum for Grades 2–5

▸ Engaging math curriculum that students love, with reading and storytelling elements that develop critical thinking skills

▸ Challenging problems presented in a way that encourages students to keep trying, teaching perseverance

▸ Online learning system for more practice, along with games, puzzles, and progress reporting for parents and teachers

▸ Suitable as a complete curriculum or as an in-depth supplement

AoPS Academy

After School Learning Centers for Grades 2–12

▸ Educational enrichment math and language arts courses taught by accomplished, passionate instructors

▸ Small class sizes, ensuring that students receive individual attention from teachers

▸ An outstanding peer group where students find friends, teammates, and positive role models

▸ Problem-solving education that helps students excel in school and on tests while also encouraging them to pursue a deep, focused understanding of the content

GENERAL REQUIREMENTS

- Database accepts connections from you: username/password set by admin

- Database accepts connections from your IP: privileges set by admin

- R packages: DBI and one of RMySQL, RPostgres, RSQLite, etc. to match your database

REMOTE CONNECTIONS

📇 **Structure**	🗒 **SQL**	🔍 **Search**	🗄 **Query**	📤 **Export**	📥 **Import**	🛠 **Operations**	👥 **Privileges**	❌ **Drop**

| | 👤✓ **Users having access** | | | | |

User		Host	Type	Privileges	Grant	Action
amy	🖉	%	global	SELECT	No	🖉
amy		127.0.0.1	global	SELECT	No	🖉
amy		localhost	global	SELECT	No	🖉

This means that I can connect to this MySQL database from anywhere.

Someone with admin access to the database needs to set this up.

CHINOOK DATABASE

SELECT *
FROM Album
WHERE ArtistId = 118

	AlbumId	Title	ArtistId
1	178	Live On Two Legs [Live]	118
2	179	Pearl Jam	118
3	180	Riot Act	118
4	181	Ten	118
5	182	Vs.	118

SELECT Title
FROM Album
WHERE ArtistId = 118

	Title
1	Live On Two Legs [Live]
2	Pearl Jam
3	Riot Act
4	Ten
5	Vs.

```sql
SELECT *
FROM Album
INNER JOIN Artist
USING (ArtistId)
WHERE ArtistId = 118
```

	AlbumId	Title	ArtistId	Name
1	178	Live On Two Legs [Live]	118	Pearl Jam
2	179	Pearl Jam	118	Pearl Jam
3	180	Riot Act	118	Pearl Jam
4	181	Ten	118	Pearl Jam
5	182	Vs.	118	Pearl Jam

```r
library(DBI)
library(RSQLite)

con <- dbConnect(RSQLite::SQLite(),
"~/Downloads/Chinook_Sqlite.sqlite")

on.exit(dbDisconnect(con))

sql <- "SELECT * FROM Album"

albums <- dbGetQuery(con, sql)

albums[albums$ArtistId == 118, ]
```

OTHER CONNECTIONS

```
library(DBI)
library(RMySQL)
library(RPostgres)

conMySQL <- dbConnect(RMySQL::MySQL(),
user="jane_doe",
password="CorrectHorseBatteryStaple",
dbname="killerapp", host="example.com")

conPostgres <- dbConnect(RPostgres::Postgres(),
user="jane_doe",
password="CorrectHorseBatteryStaple",
dbname="killerapp", host="example.com")
```

- More efficient than exporting CSVs from the database and then reading them into R. R script will always have access to fresh data.

- Access multiple databases from one R script, even if the databases are on different servers and use different technologies.

- Avoid running slow queries on a production server by doing simple queries on the server and then processing and combining the data in R.

- Complicated SQL queries can be replaced with a sequence of smaller, easier to understand (and debug!) steps in R.

AOPS IS HIRING
≈20 full time roles + many part time

https://artofproblemsolving.com/company/careers

STEM Education • Math, Physics, Computer Science
Software Development (various web technologies) • DevOps
English Language Arts/Writing • Math Blogger/Social Media
Customer Service • Education Management

San Diego • SF Bay Area • Washington, DC • Seattle, WA • Dallas, TX
Princeton, NJ • Morrisville, NC • Atlanta, GA

I will share my internal referral bonus with you.
Contact me for more info: amy@aops.com.