Strategies and examples for data acquisition from distributed and complex sources

A presentation at JMP Discovery Summit Europe 2021:
https://discoverysummit.jmp/en/2021/europe/home.html

Abstract

JMP offers several possibilities to query data from different sources and to connect them afterwards. In daily business user often need the same or similar data from distributed sources.

In real life and in case you haven't been able to define the IT-infrastructure by yourself, what usually is the standard, the data you need is localized at different platforms (databases, fileservers etc.). User then need a lot of time to query and combine the data in appropriate form for analysis.

In this presentation some examples for workflows are shown, how to effectively get data into your table(s), as specified and probably needed by the user every day. Methods used to accomplish these tasks are: Query Builder, SQL, JSL, Add-Ins, Virtual Join etc.

Introduction

About the author

- Georg Raming, process development silicon crystal growth at Siltronic AG
- Experience with statistical evaluation of process and product data, statistical education
- Responsible for JMP software at Siltronic AG and training activities (a few hundred users)
- Siltronic AG is one of the world's leading manufacturers of highly specialized hyper pure silicon wafers with diameters up to 300 mm (www.siltronic.com)

Technical hints

- I'm working with JMP sample data tables instead of database in this presentation, but the concepts shown are originally used for querying data from relational databases via ODBC connection
- JMP Query Builder is working the same way for both, visual query and custom SQL query

What is the target?

Target of the presented work is to establish some ways for getting data from database in an easy and efficient way.

Building Blocks

JMP data table from database

A very easy way to use database queries is to save a table that have been originally queried from database via ODBC connection by means of the query builder in JMP.

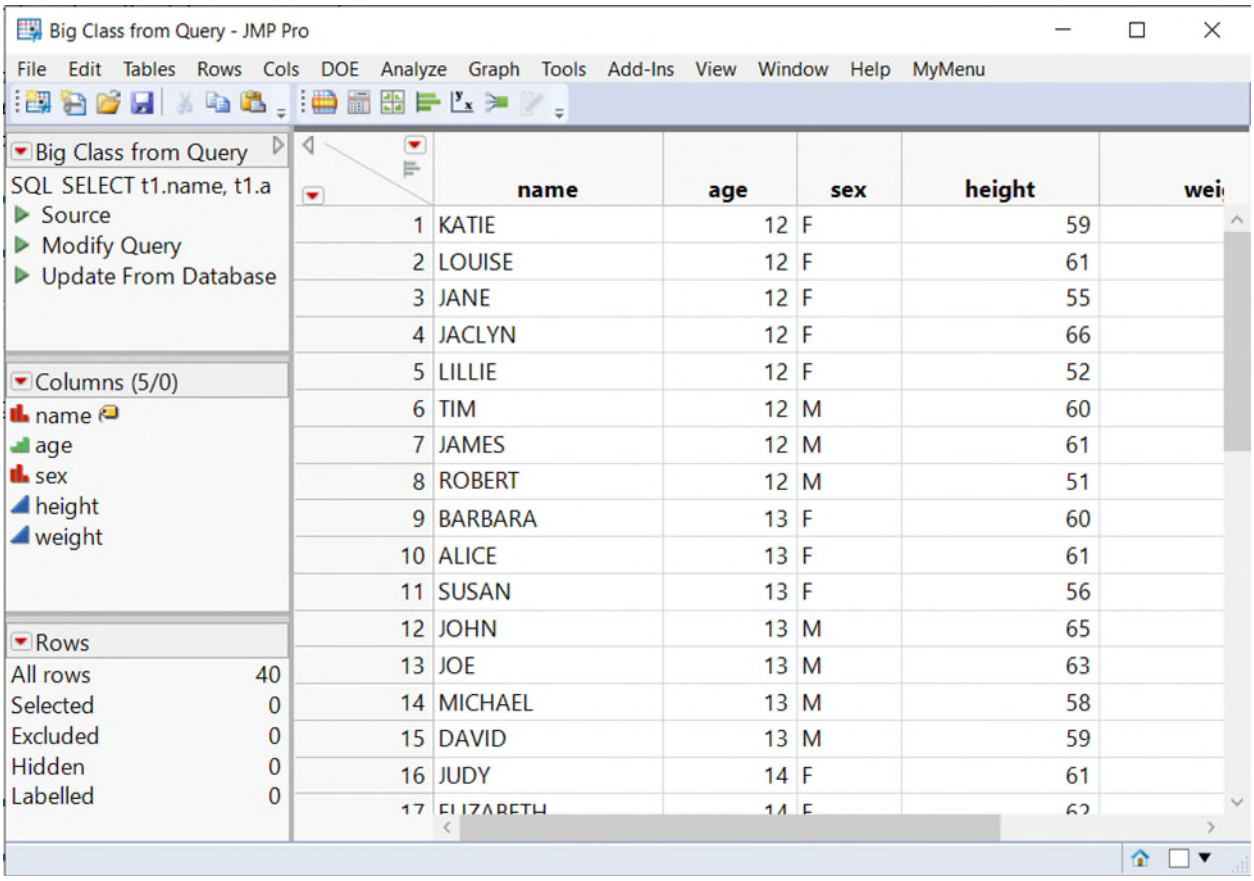


Figure 1: Sample data table Big Class generated by JMP Query Builder, see Scripts "Source", "Modify Query", "Update From Database"

In this table you can Update the data from e.g. database simply by pressing "Update From Database".

JMP table from database: JMP Query Builder

When one or more data tables are open in JMP, you can open JMP Query Builder to visually build a data query on that data tables.

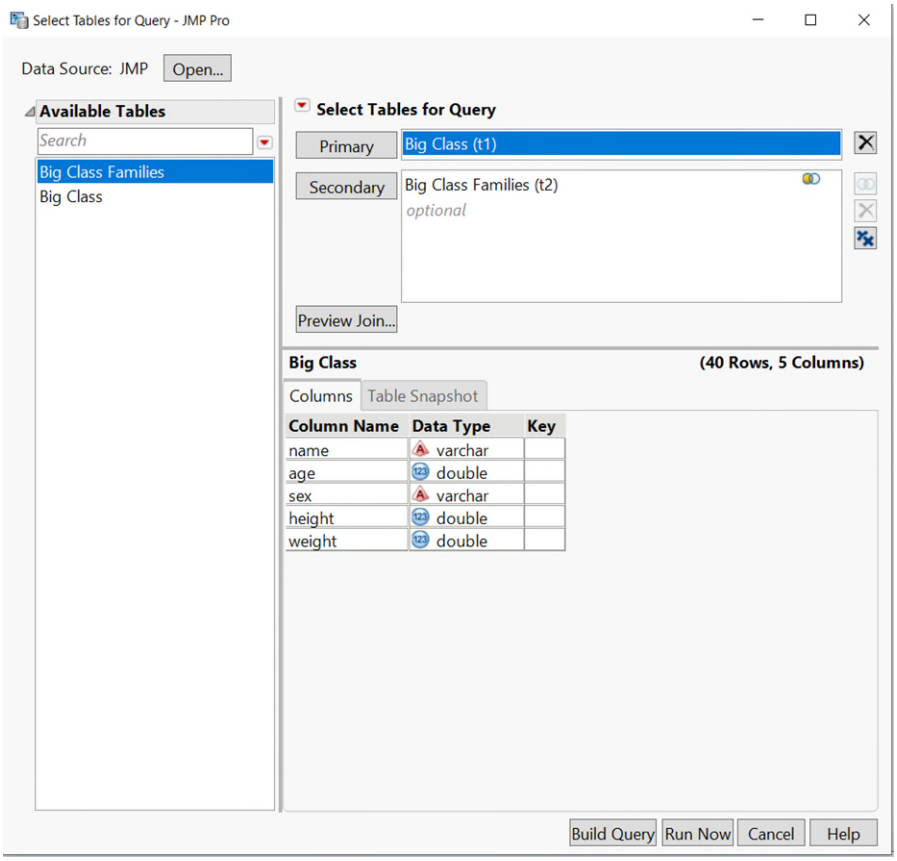


Figure 2: Query Builder with two opened JMP Data Tables (similar to data tables on database)

JMP tries to figure out the best way, how to join both tables. By “Edit Join” you can change the way both tables are connected.

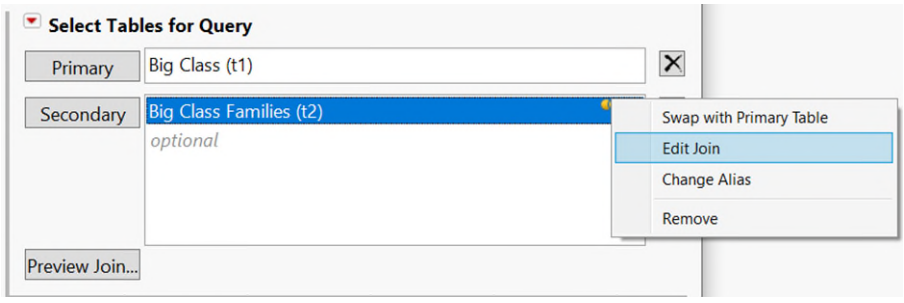


Figure 3: Edit Join lets you change the way both tables are connected

When pressing “Build Query” you get to the next step, selecting columns and rows.

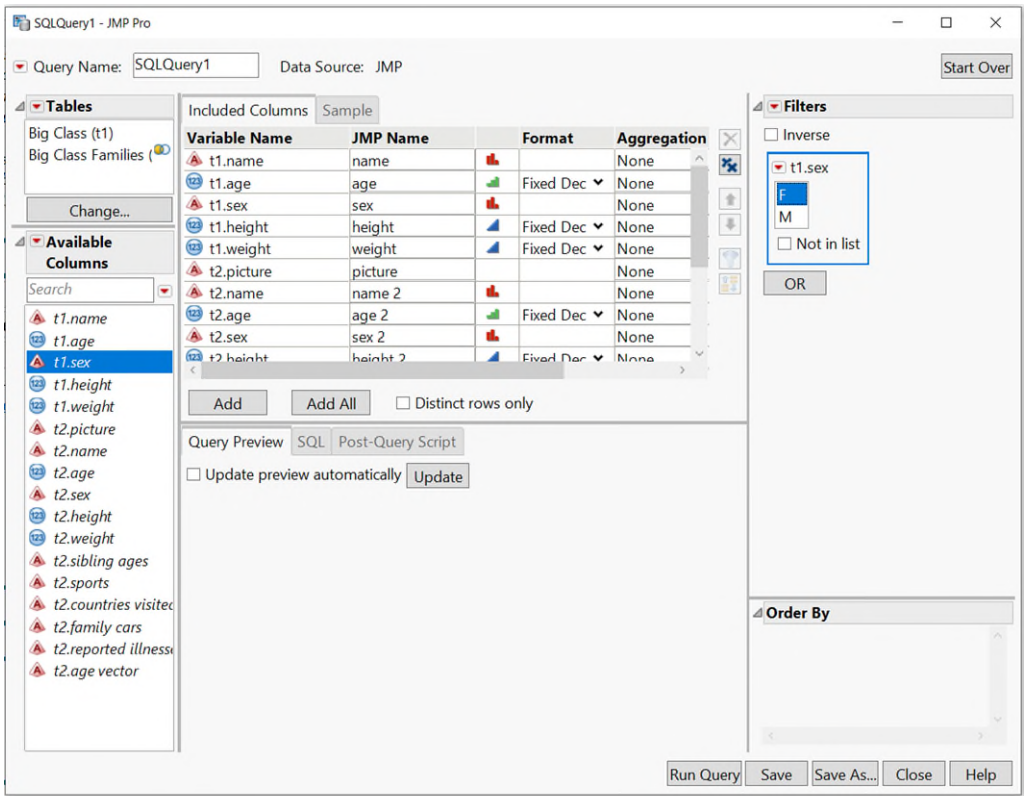


Figure 4: JMP Query Builder on data table "Big Class.jmp" joined to "Big Class Families.jmp"

Here you can also edit the type of join used:

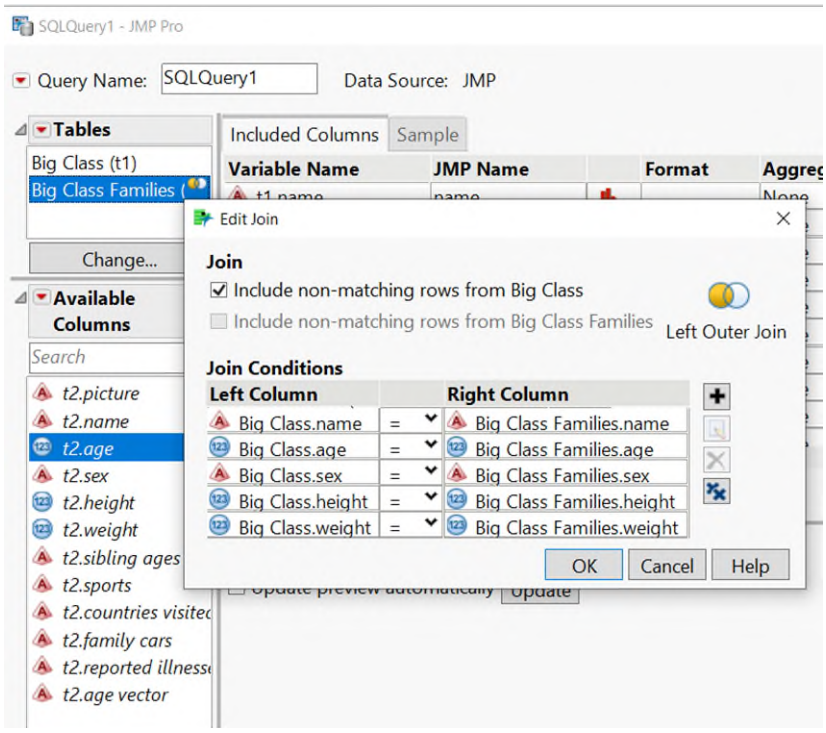


Figure 5: Type of join used

So far we edited the query in visual mode of JMP Query Builder. You can switch to custom SQL (i.e. typing the query in plain SQL) by the red triangle menu from Query Builder.

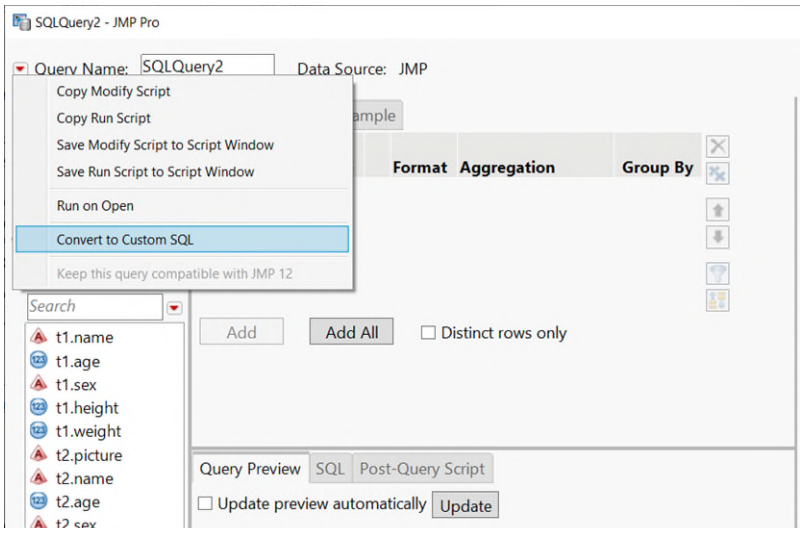


Figure 6: Converting query to custom SQL, please note that you can't reverse this step

Figure 7: SQL Custom Query

Data sources other than JMP tables (database etc.)

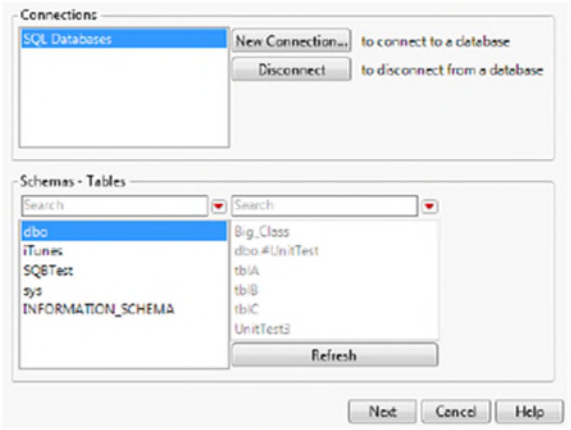
To perform above mentioned steps on database tables, usually an ODBC connection is needed. The details are well described in the manual “Using JMP” that comes with the program.

Connect to a SQL Database

Set up the ODBC connection through the Windows Control Panel or inside JMP.

1. Select **File > Database > Query Builder** to display the Select Database Connection window.
   The Connections box lists data sources that you connected to in the current JMP session.
2. If the desired data source is *not* listed in the Connections box, click **New Connection** to choose a data source. The method of choosing a data source depends on your operating system and the ODBC driver. See “Connect to a Database” on page 162.
3. Select a table or schema from the Schemas - Tables box and click **Next**.
   Query Builder examples are based on a table named SQBTest, which contains movie rental data.

Figure 3.34 Select the Database Schema



Tip: To find the table in a long list, enter the name in the search box above the schemas. You can also search for tables above the list of tables. The red triangle menu provides options for matching case and searching with regular expressions.

Figure 8: Screenshot from JMP Manual, Connect to a SQL Database

Usually your system administrator sets up your data sources and drivers, you may need to access the company database(s). You can check in Windows via app “ODBC Data Source Administrator (64-bit)”. Please note, that the bitness of drivers must fit the bitness of your JMP application, here 64-bit.

ODBC Connection

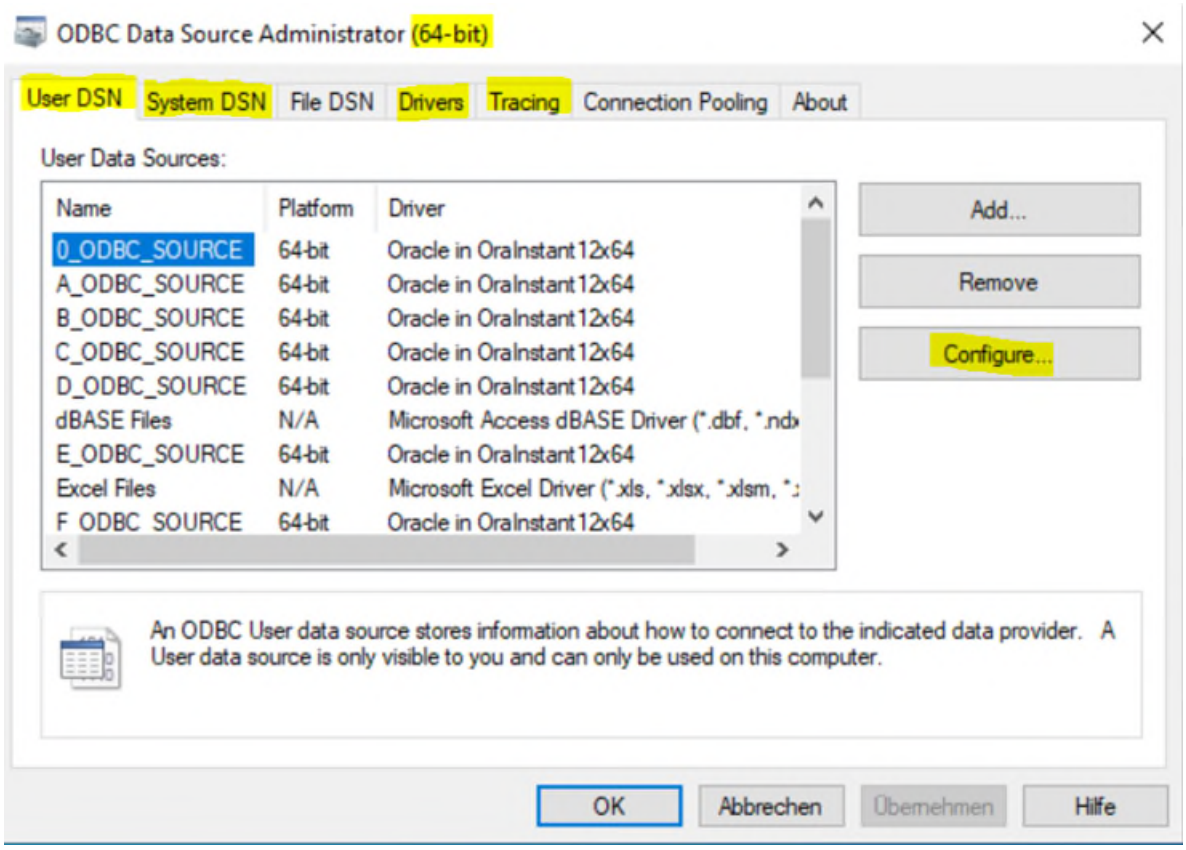


Figure 9: Windows App "ODBC Database Administrator (64-bit)". Here you can check installed data sources and drivers

For each database type an appropriate driver is needed (see tab "Drivers" in Figure 9). Please note, that there are two tabs "User DSN" and "System DSN". You can reach both from your JMP application, but usually "System DSN" are defined by the system administrator, and "User DSN" you can define on your own.

Three JSL methods to get data from database

In JSL there are basically three methods to script database queries:

- New SQL Query (method via Query Builder as shown below)
  - This is the most powerful command
  - It generates a new query object (.jmpquery)
  - Can generate a JMP Table directly
  - Documents very well the origin of data
  - Provides scripts for updating …
- Open Database (see Scripting Index or manual)
  - Perform query in just one command, most simple
- Database Connection (3 commands/steps needed)
  - Create database connection
  - Execute SQL, one or several tasks
  - Close database connection

Methods to install queries into JMP application

Once you have defined how to get your data, you can make the methods available via Add-In (Figure 10) or customized menu (Figure 11).

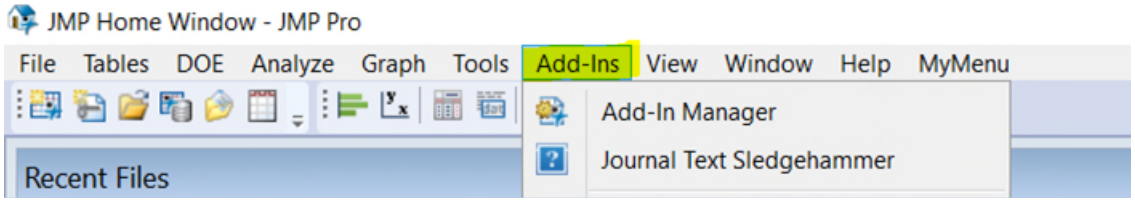


Figure 10: JMP Add-In

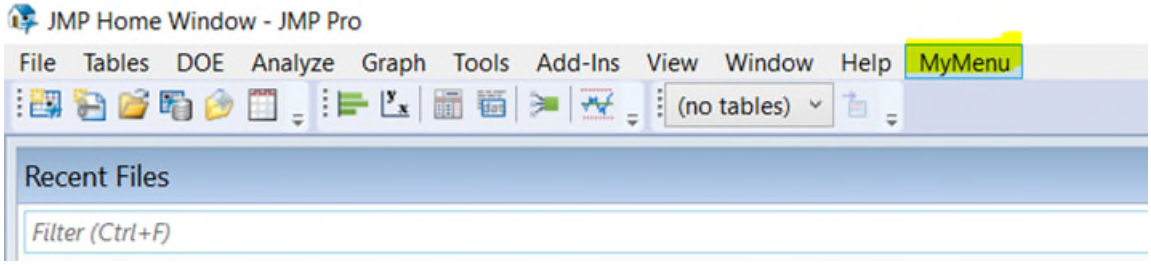


Figure 11: Custom menu

Examples

Using Table script to save table layout, and let data to be filled from database

If you have generated your table via Query Builder, you can place scripts for analysis into that table, and save table script without data via red triangle menu.

This script you can then use to place it into an Add-In, custom menu etc. to regenerate that table. The update script in a second step helps you to request the current data again from database etc..

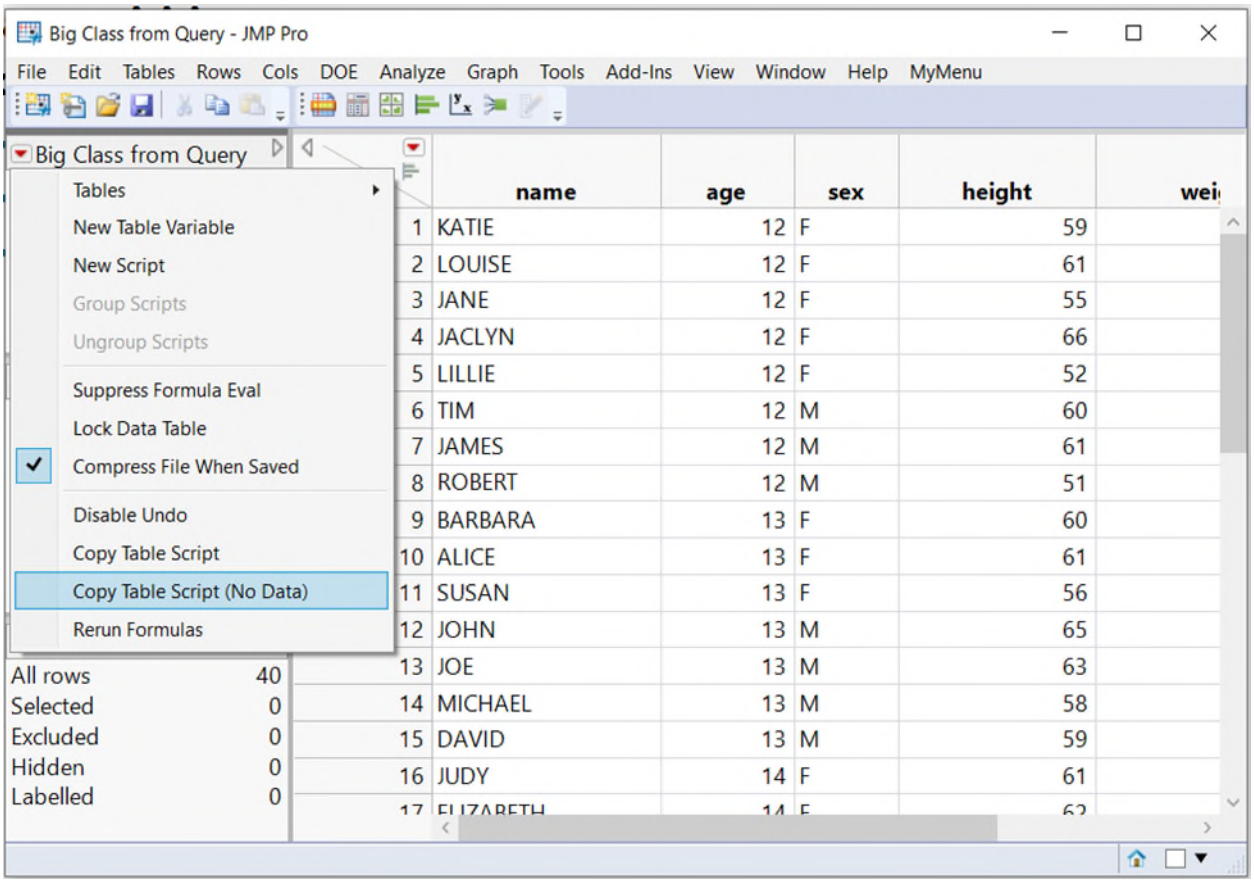


Figure 12: Get table script to use it in an Add-in, custom menu etc.

Query from inside a table

Let's assume that you have a few rows of data in your table (3 rows of Big Class data table), and you need some additional information from another data table corresponding to your first data table. The script in below screenshot (Figure 13) performs exactly that task. It takes the names from table "Big Class.jmp", puts it into a custom sql query to query the data from "Big Class Families.jmp".

If needed, both tables can be connected via virtual join, and e.g. the columns of "Big Class Families.jmp" can be used in evaluations of "Big Class.jmp".

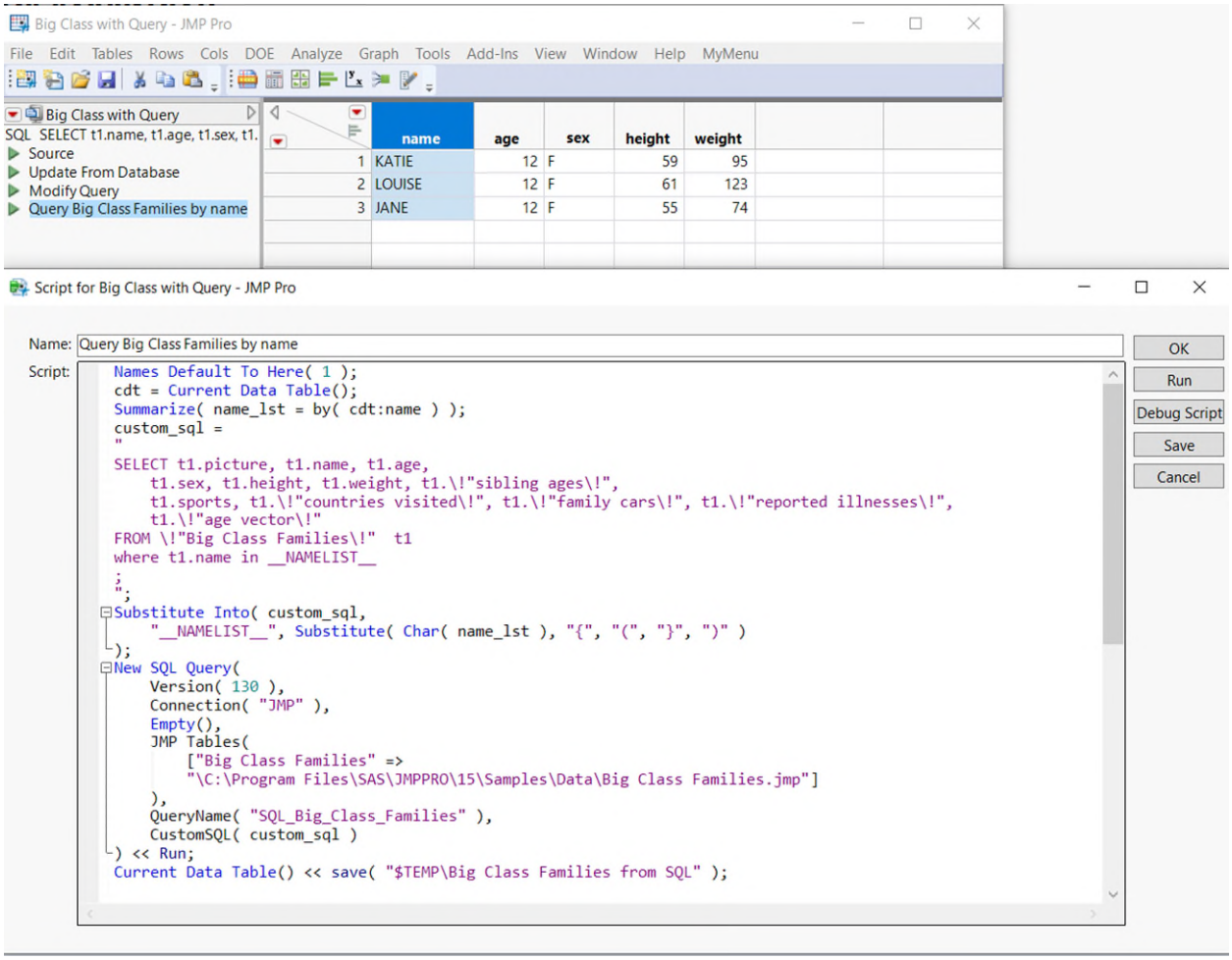


Figure 13: Script for query data depending on current table content

Custom query script

For small queries the Query Builder is quite handy to perform database selects. But if you have to query many tables with complex joins, you may want to hide that complexity from the user.

In this case you can

1. define the target table by means of a complex SQL query
2. before starting GUI dialog,
    a. query a filter table, that lets the user define the conditions (i.e. rows)
    b. query a column table, that lets the user define the columns
3. let the user define rows and columns to query (see Figure 14)
4. start the query and get the results

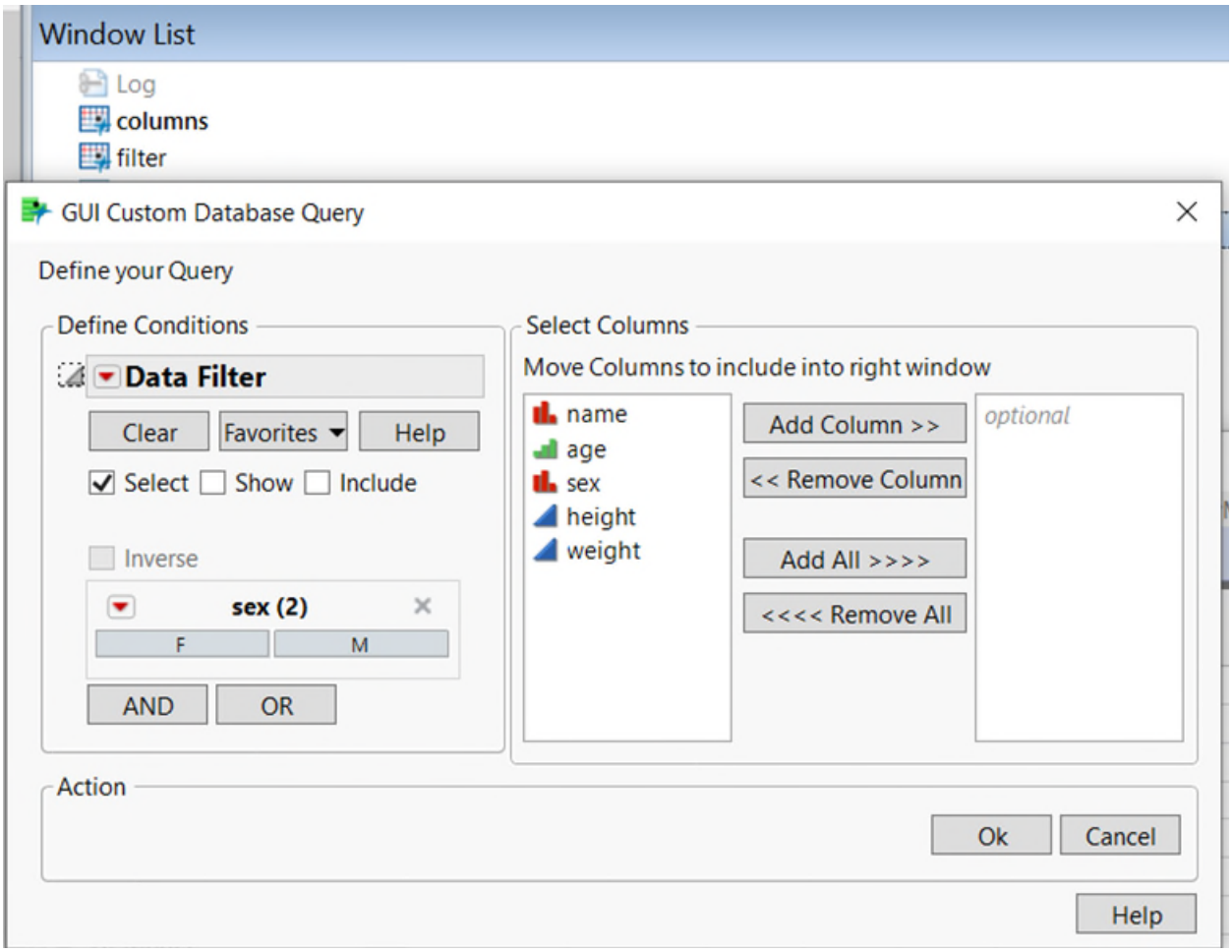


Figure 14: Custom query dialog with two background tables (columns + filter)

See also the script in the corresponding .jrn file.

Two step query

There are situations, when it does not make sense to put all data into one table, or database can not accomplish to query all in one run due to memory restriction, database design etc.. In this case it may be useful to do the job in two separate queries, and connect both tables at the end (virtual join).

In Figure 15 the result query gives 8 rows instead of 7, because the name “ROBERT” is listed twice in Big Class datatable. So to query the rows exactly, a more distinct condition (e.g. Name+Age) would be needed.

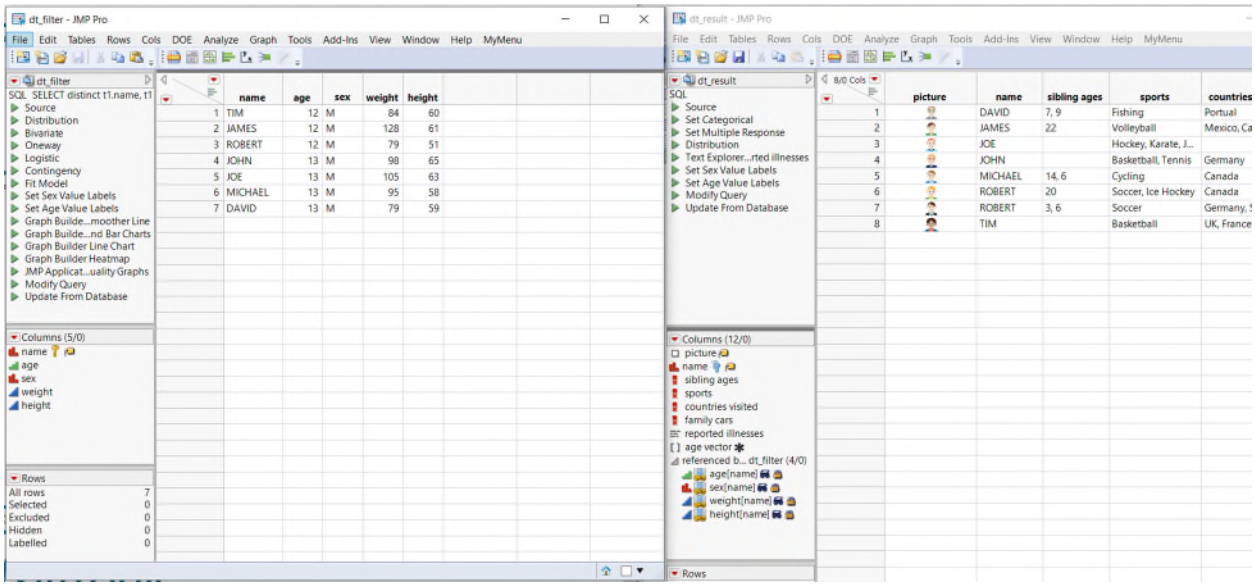


Figure 15: Two step query, in a first step the table dt_filter is drawn, names are put as condition to query dt_result; both tables are connected via virtual join, i.e. in dt_result the columns of dt_filter can be used as if they were in dt_result

Run several database queries in background

See discussion in JMP Community:

https://community.jmp.com/t5/Discussions/Run-several-database-queries-ODBC-in-background-and-combine-data/m-p/274960#M53349

Summary

Several building blocks and methods have been shown to query data from complex sources. Depending on your situation the one or the other method, or a combination of two may be appropriate for you.

Thanks to the community for discussion and to the developers of JMP for building and maintaining that great software.