QStudio/PRQL Quick Start #1 - Demonstration

Over the years, we have collected a ton of data - mostly from spreadsheets or PDF files, or retrieved from the Town, Vision or the Grafton County Register of Deeds. For example all the following files are in my Github repository.

- Lyme's Old-to-New values from the MS-1 reports
- Data "scraped" from Vision site
- Grafton County Register of Deeds

All these files have a similar format: a bunch of rows (usually representing parcels) with interesting data in the columns.

For example, I scanned all the Old-to-New documents for 2021 to 2024; they're in the "OldToNewValues" spreadsheet. If you look at that spreadsheet, you'll see these columns:

```bash
PID
Map/Block/Unit
Location (street address)
Owner
Code
OldValue
NewValue
Ratio
Difference
Year
```

Or look at the "Scraped Data" file. From time to time, I run a script that loads every page from Vision and plucks up the values for each property. It produces these columns:

```bash
1  PID
2  Owner
3  Street Address
4  Map/Block/Unit
5  Book/Page
6  Assessment
7  Appraisal
8  Lot size,
9  ...
```

The Grafton County Register of Deeds file includes columns for:

```bash
1   ID
2   Date&Time
3   Date
4   Time
5   Type
6   Book&Page
7   Book
8   Page
9   Party1 (seller)
10  Party2 (Buyer)
11  ...
```

The good side of spreadsheets

Spreadsheets are a very convenient way to view data. Everyone knows how to open and manipulate them. And lots of interesting data is reported in a spreadsheet.

You could consider putting each of the above into a separate tab of a spreadsheet. And if you got another source of data, you could add a new tab (with its columns and rows) to the spreadsheet.

The downsides of spreadsheets

But spreadsheets come with downsides, primarily because modifying a spreadsheet is (mostly) irreversible - it's very hard to get it back to its original state. For example:

- To see the difference or ratio between old and new values, you could create a new column. But each time there's a new column, the "underlying data" gets changed.

- Or sort by a column. But the act of sorting changes the underlying data of the spreadsheet. And putting the data back to its original state gets really hard.

- Or, you don't particularly care about a particular column, say last year's assessed land value: that column just gets in the way. You could delete it - but maybe you'll want it later on.

What if there were a way that you can select only the columns you want, and create new columns with their own calculations, and sort every which way?

Using a database

That's where SQL databases come in. Each of those spreadsheets becomes a *table* in the *database*. The table's columns and rows exactly match the spreadsheet's columns and rows. When you use a database, you create queries that produce new results without changing the underlying data.

The fundamental rule of databases (their "super power") is that every operation starts with a table and creates a new (modified) table. Subsequent operations make further modifications to create yet another table. That new table doesn't get "added to the database". Instead, the database program just displays the final result, without changing the underlying tables.

And finally, a *database query* is simply a series of instructions that tell how to manipulate the tables. That query can be saved and re-executed at any time, or modified to answer different questions.

QStudio demonstration

QStudio is an application that lets you view and manipulate the tables of a SQL database. It's the first tool that I think is good enough to criticize - that is, all the prior ones were just too hard to use, with too many steps to get useful results. I like QStudio because it:

- Runs on Windows, Mac, Linux.
- Provides a good enough user interface. (Although the QStudio interface has some sharp edges, they're easy to recognize and work around.)
- Uses PRQL (pronounced "prequel") to make it MUCH simpler to write queries (SQL is the pits, you won't need to know it)

So here's QStudio. *Demo follows, and ends Lesson #1*

Lesson #2 tells how to install QStudio on your computer, and execute the same queries as the demo.