2024
Indian Health Service
Partnership Conference

Power Up Excel

NICHOLAS LAUGHTON PHARMD MPH

AREA INFORMATICIST

14 AUG 2024

Outline

1. Introduction to Data Science
2. Tidy Data
3. Tables in Excel
4. PowerQuery
5. Next Steps

Invest in Your Self

"If this takes more than five minutes I don't have time for this"

-*Disgruntled Provider*

Be a Good Data Steward

Take all necessary precautions to protect IHS information assets

There is Rarely Only One Way

This presentation will share some thoughts and methods to accomplish some data wrangling tasks, but it is not the only way.

Data Science

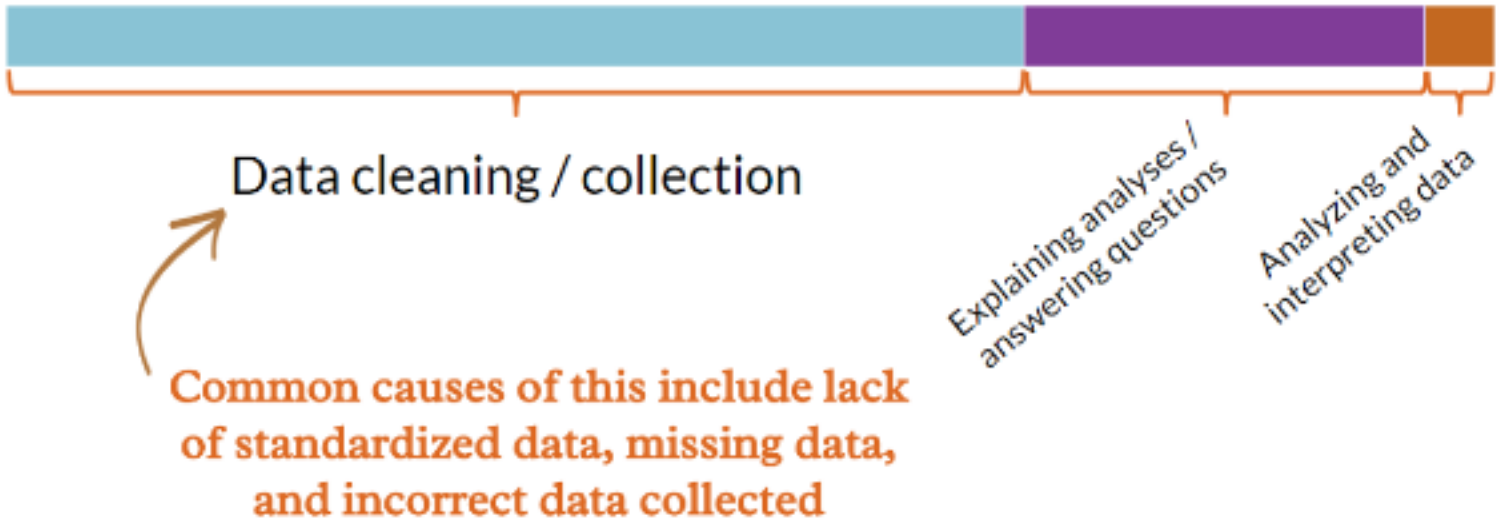
Data Science →

- math
- statistics
- specialized programming
- advanced analytics
- artificial intelligence (AI)
- machine learning
- specific subject matter expertise

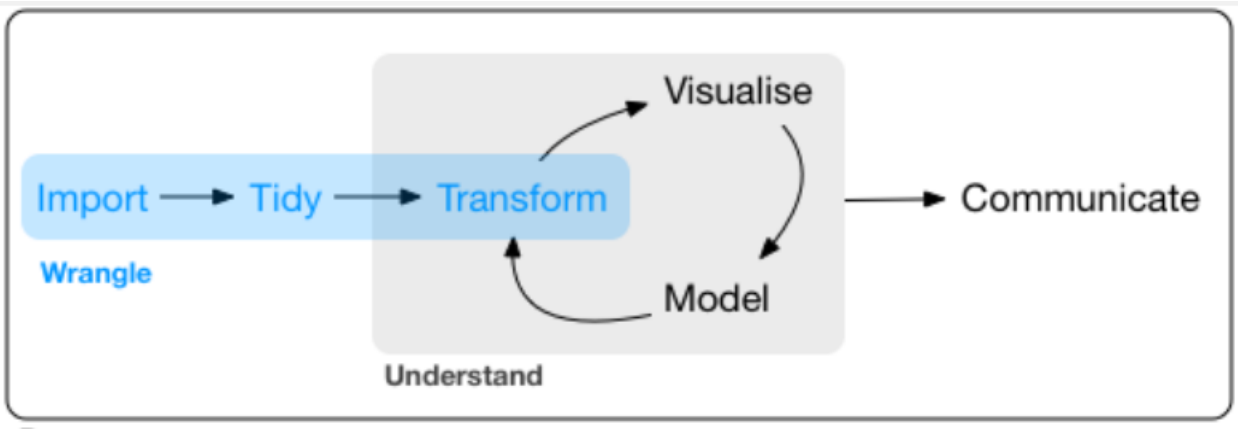
To uncover actionable insights hidden in an data.

How people think data scientists spend time

How data scientists actually spend time



Data Wrangle!



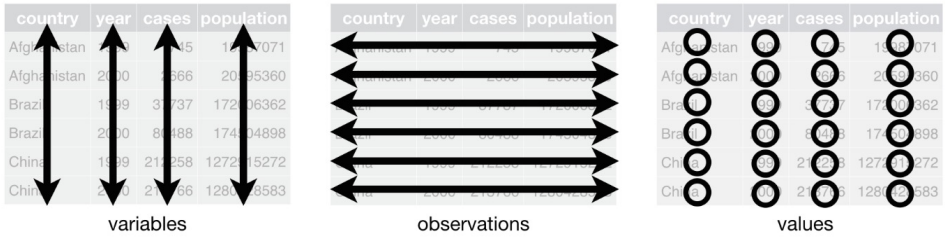
https://r4ds.had.co.nz/wrangle-intro.html

What is Tidy Data?

Each variable must have its own column.

Each observation must have its own row.

Each value must have its own cell.

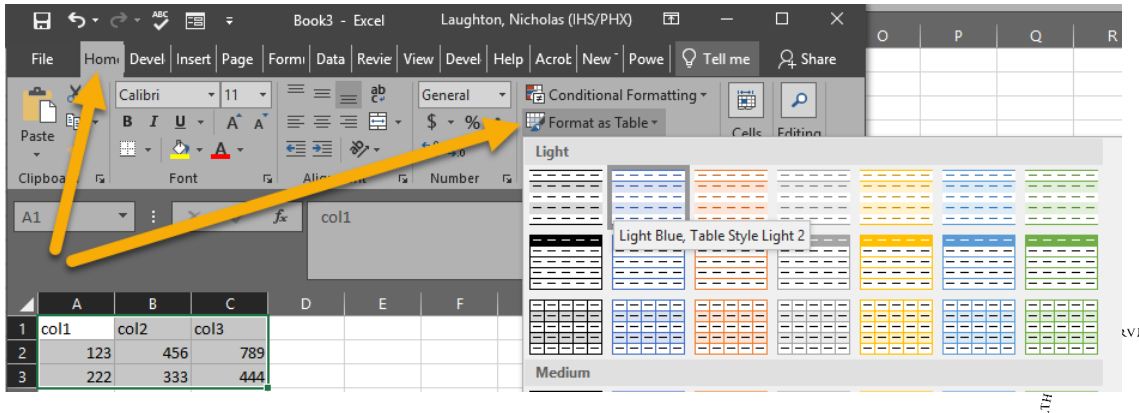


Excel Tables

Converting a range of cells to table makes working with data easier

Tables can contain

- **Header row**
- **Banded rows**
- **Calculated columns**
- **Total row**



Power Query Introduction

Power Query is part of Excel

Power Query is used by other applications (Power BI)

Power Query is a data transformation and data preparation engine.

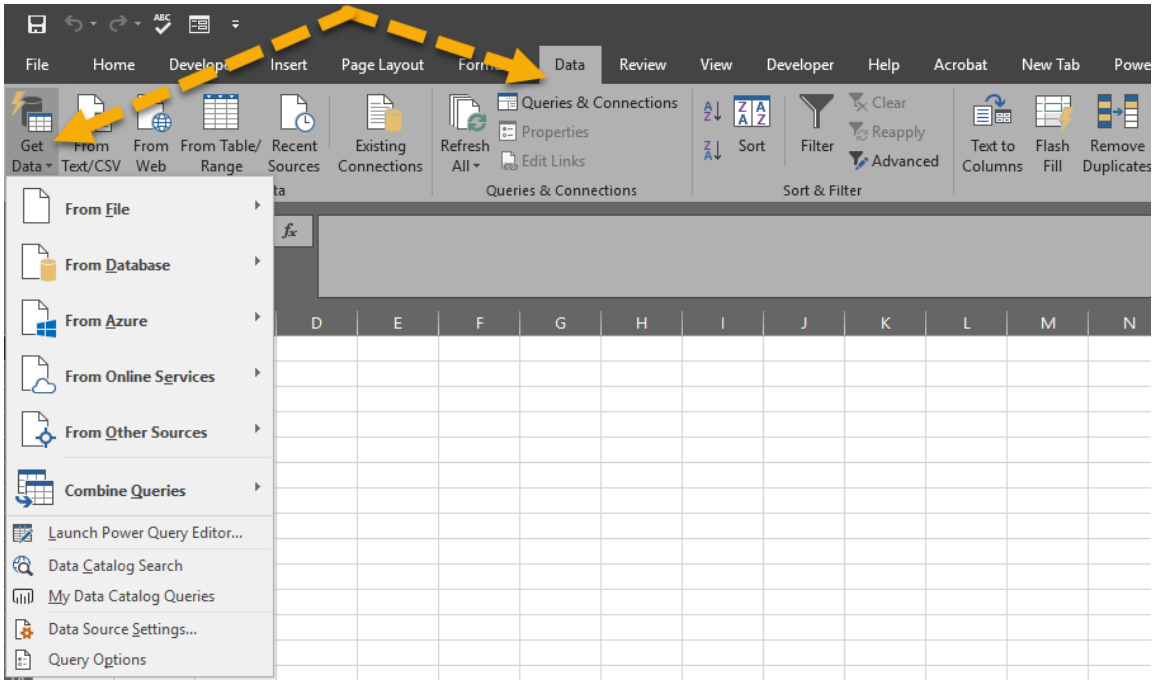
Power Query comes with a graphical interface for getting data from sources and a Power Query Editor for applying transformations.

What I Use Power Query for

Cleaning data

Combining data from multiple files into 1 excel table

How to Find Power Query



Power Query Example Wangling a Single File

SAMPLE_vgen1.TXT - Notepad

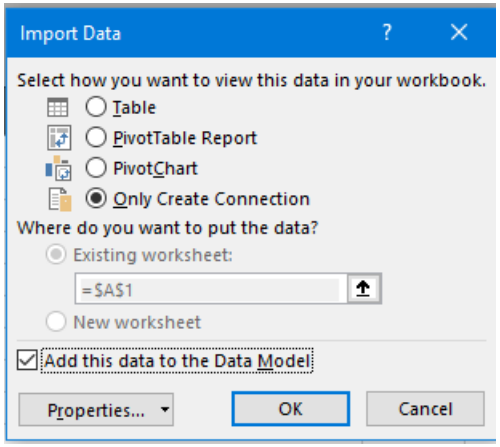
File Edit Format View Help

```
DEVICE: HOME// HOME;180;99999999999999  VT

                    PCC MANAGEMENT REPORTS VISIT LISTING
                                SUMMARY PAGE

REPORT REQUESTED BY: LAUGHTON,NICHOLAS

VISIT Selection Criteria
      Encounter Date range:  JAN 01, 2001 to DEC 31, 2001
      Any Immunization Administered?:   IMMUNIZATION ADMINISTERED

REPORT/OUTPUT TYPE
      Detailed Listing containing
      Visit Date  (10)
      Immunization Lot #  (12)
      Immunizations/Series  (40)
            TOTAL column width: 68

Visits will be SORTED by:  Visit Date



                        PCC VISIT LISTING                Page 1
DATE        LOT #         IMMUNIZATIONS
------------------------------------------------------------------------

01/01/2001  E71770LA      INFLUENZA, NOS
01/01/2001  --            INFLUENZA, NOS
01/01/2001  --            TD (ADULT)
01/01/2001  --            VARICELLA 1
                          IPV 3
01/01/2001  --            DTAP
01/02/2001  --            HEP B, NOS 2
01/02/2001  --            INFLUENZA, NOS
                          Pneumococcal, PCV-7
01/02/2001  71740LA       INFLUENZA, NOS
01/02/2001  E71770LA      INFLUENZA, NOS
01/02/2001  --            HEP B, NOS 3
01/02/2001  --            INFLUENZA, NOS
01/02/2001  --            DTAP 4
```


A	B	C
DATE	LOT #	IMMUNIZATIONS
01/01/2001	E71770LA	INFLUENZA, NOS
01/01/2001	Historical	INFLUENZA, NOS
01/01/2001	Historical	TD (ADULT)
01/01/2001	Historical	VARICELLA 1
01/01/2001	Historical	IPV 3
01/01/2001	Historical	DTAP
01/02/2001	Historical	HEP B, NOS 2
01/02/2001	Historical	INFLUENZA, NOS
01/02/2001	Historical	Pneumococcal, PCV-7
01/02/2001	71740LA	INFLUENZA, NOS
01/02/2001	E71770LA	INFLUENZA, NOS
01/02/2001	Historical	HEP B, NOS 3
01/02/2001	Historical	INFLUENZA, NOS
01/02/2001	Historical	DTAP 4
01/02/2001	Historical	DTAP 3
01/02/2001	Historical	HIB (HBOC) 3
01/02/2001	Historical	Pneumococcal, PCV-7 1

Navigating Power Query

1. Finding your Query again and making edits after loading
2. Switching Source
3. Close and Load
   - Changing the close and load options
   - When to use Only Create Connection
     - Large Data
     - You don't need to see it



Missing Data

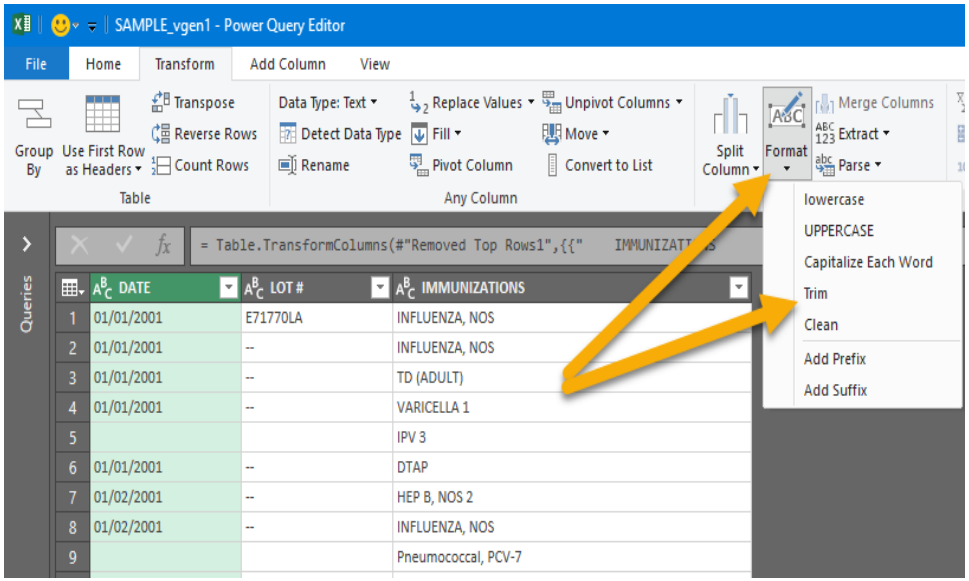
Before (ideally) cleaning your data you should decide what to do with missing data.

PCC VISIT LISTING Page 1

DATE	LOT #	IMMUNIZATIONS
01/01/2001	E71770LA	INFLUENZA, NOS
01/01/2001	--	INFLUENZA, NOS
01/01/2001	--	TD (ADULT)
01/01/2001	--	VARICELLA 1
		IPV 3
01/01/2001	--	DTAP

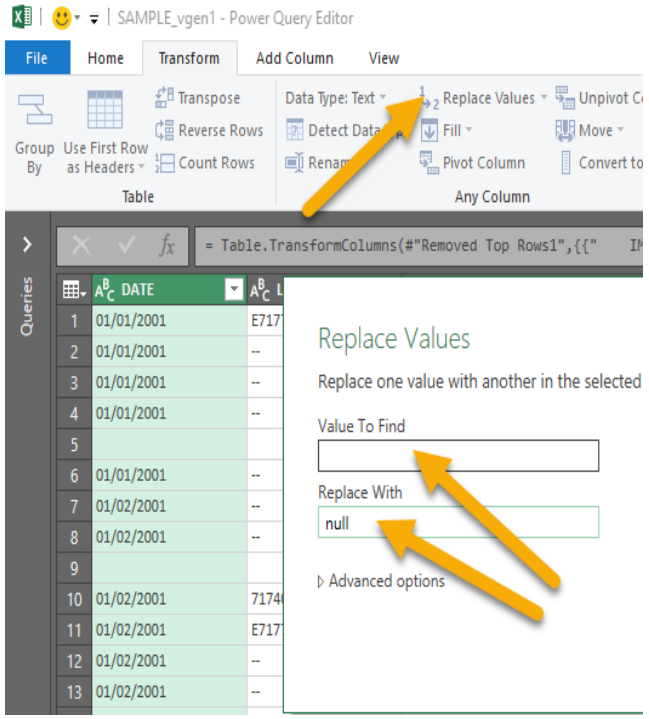
Filling Down Missing Data 1

1. Trim Text



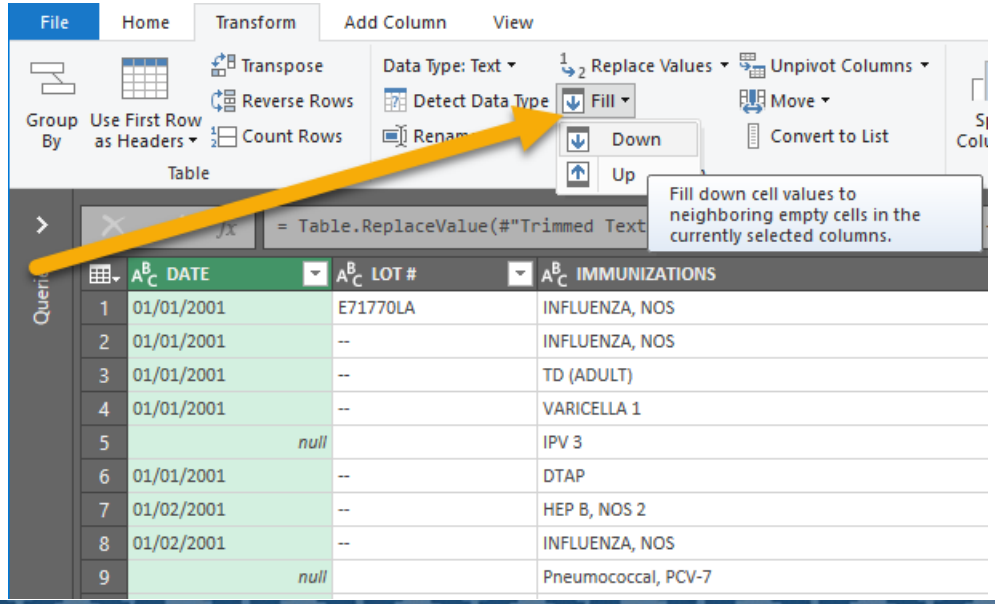
Filling Down Missing Data 2

2. Replace “” with “null” **(not NULL or Null)**



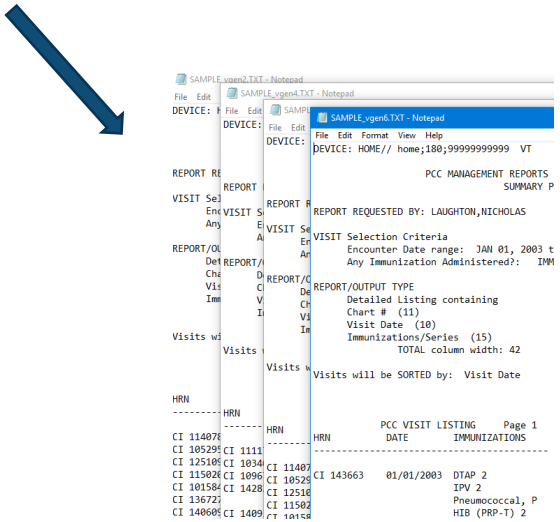
Filling Down Missing Data 3

3. Fill Down



Power Query Example Wrangling Multiple Files in a Folder

Name	Date modified	Type	Size
SAMPLE_vgen2.TXT	8/3/2023 10:00 PM	Text Document	226 KB
SAMPLE_vgen4.TXT	8/3/2023 10:11 PM	Text Document	3 KB
SAMPLE_vgen5.TXT	8/3/2023 10:12 PM	Text Document	3 KB
SAMPLE_vgen6.TXT	8/3/2023 10:12 PM	Text Document	3 KB



	A	B	C	D
	Source.Name	Chart	Imm_Date	IMM
	SAMPLE_vgen2.TXT	114078	1/1/2002	TD (ADULT)
	SAMPLE_vgen2.TXT	105295	1/1/2002	TD (ADULT)
	SAMPLE_vgen2.TXT	125109	1/1/2002	TD (ADULT)
	SAMPLE_vgen2.TXT	115020	1/1/2002	TETANUS TOXOID
	SAMPLE_vgen2.TXT	101584	1/2/2002	INFLUENZA, NOS
	SAMPLE_vgen2.TXT	136727	1/2/2002	INFLUENZA, NOS
	SAMPLE_vgen2.TXT	140609	1/2/2002	VARICELLA
	SAMPLE_vgen2.TXT	140609	1/2/2002	Pneumococcal, P

Getting Files from a Folder, Introduction

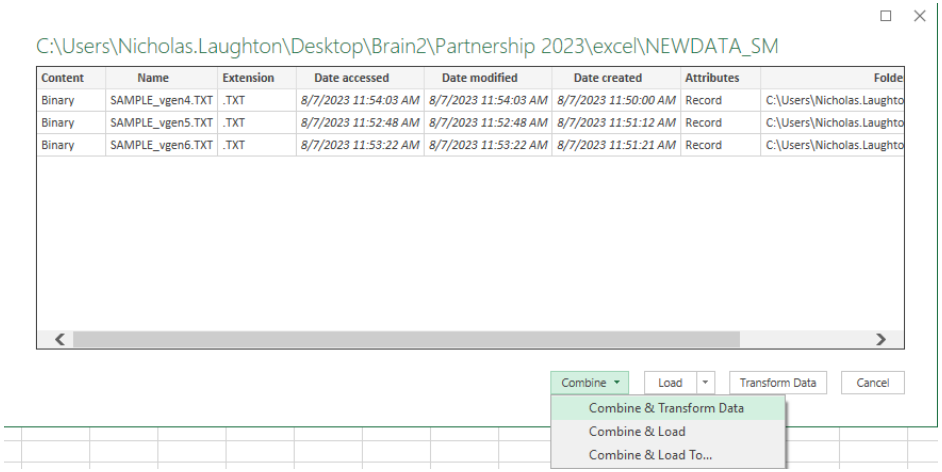
Keep only the files you want in the folder

Keep VERY consistent

- Same File Type
- Same Naming Convention
- Same Formatting

Getting Files from a Folder

1. Save files in folder
2. Get Data → From File→ From Folder
3. Navigate to the Folder
4. Combine & Transform Data

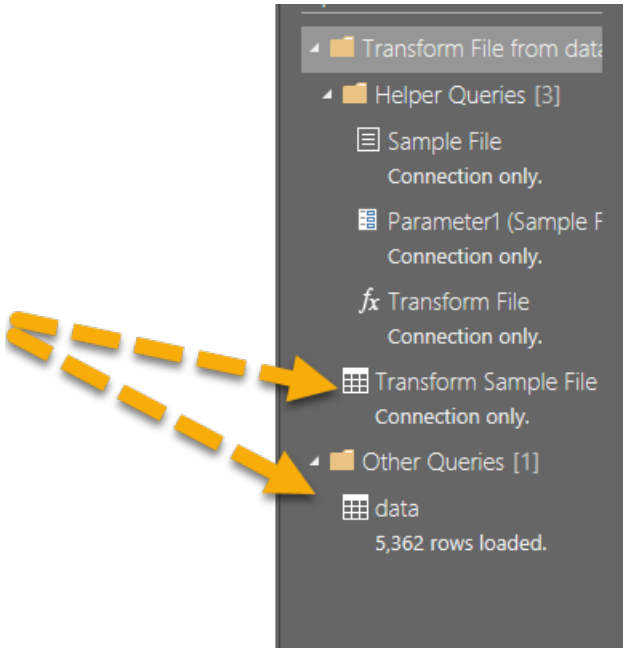


There are Two Important Queries

1. The query to process the files BEFORE combining
   - (‘Transform Sample File’ {almost always called this})

2 The query to process data After combining.
   - (‘data’ {will be named for the folder})

You can rename Queries



Practice working with Queries

Examine taking steps in

- Transform Sample File vs Other Queries

Next Steps, Power Pivot

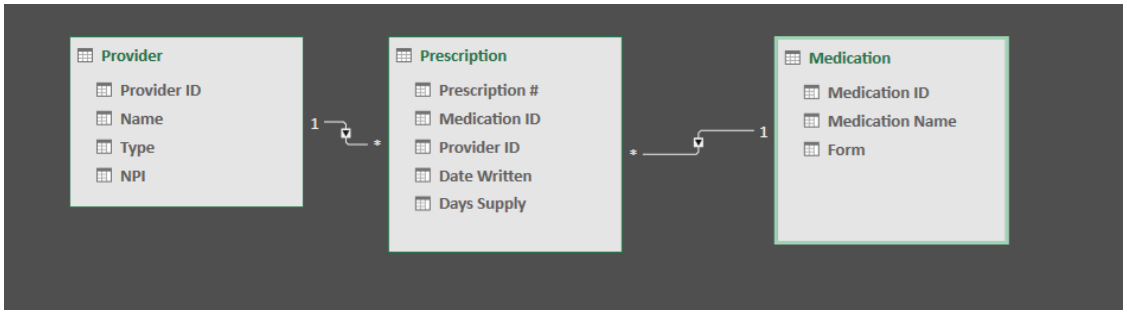
Power Pivot

- create data models,
- establish relationships,
- create calculations.

With Power Pivot you can work with large data sets, build extensive relationships, and create complex (or simple) calculations, all in a high-performance environment, and all within the familiar experience of Excel.

Power Pivot Example

Data science



Medication

Medication ID	Medication Name	Form
A12	Aspirin	Tablet
B23	Lisinopril	Tablet
C45	Tylenol	Capsule
D67	Ibuprofen	Oral Suspension

Prescription

Prescription #	Medication ID	Provider ID	Date Written	Days Supply
123	A12	1AB	7/1/2023	30
456	B23	2BC	7/2/2023	45
789	C45	3DE	7/3/2023	90
101112	D67	4FG	7/4/2023	10

Provider

Provider ID	Name	Type	NPI
1AB	Paulette	MD	123456
2BC	Shawnna	NP	789444
3DE	Sam	PA	687799
4FG	Carl	MD	454566


INDIAN HEALTH SERVICE
PHS · 1955 ·