



Please Read Below – Action Required!

- This is meant to be an interactive, hands on, follow-along session
- Have your laptops open and ready
- Go to <https://ihs.gov/partnershipconference/documents/> and download the ZIP file under “Leveraging Excel to Improve Operations”
- Extract the files and make sure they are easily accessible
- Open a fresh Excel Workbook
- Connect your mouse if you have one



Indian Health Service

Leveraging Excel to Improve Operations

NICHOLAS LAUGHTON PHARM.D, MPH

AREA INFORMATICIST

08/04/2023



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

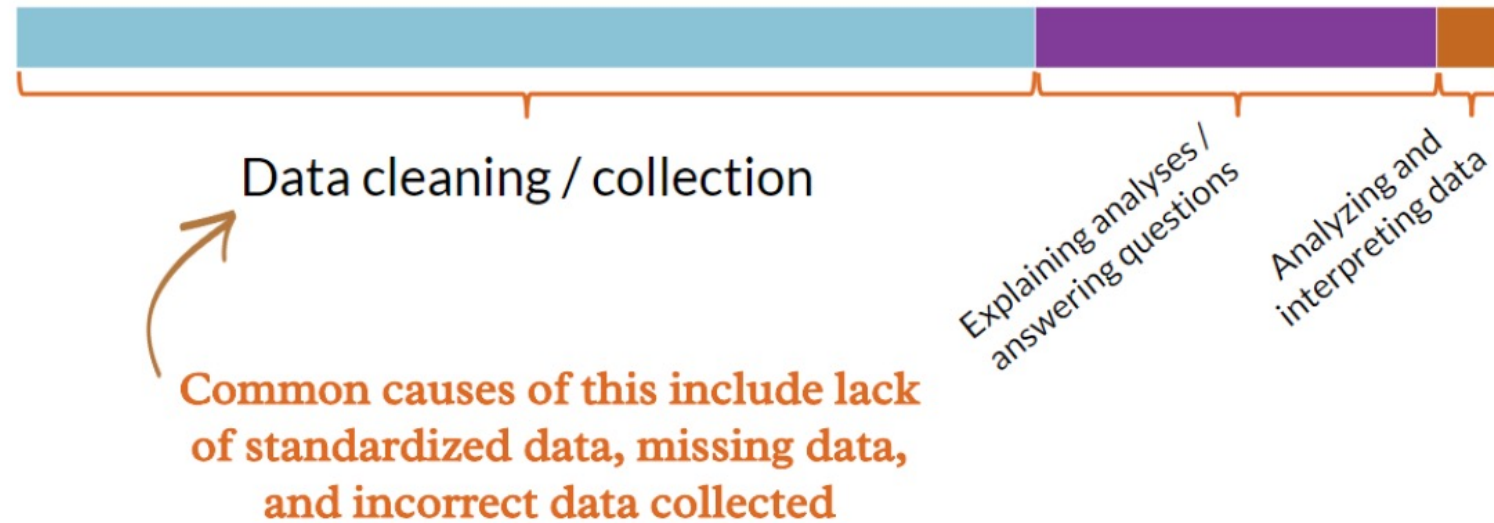
To uncover actionable insights hidden in an data.



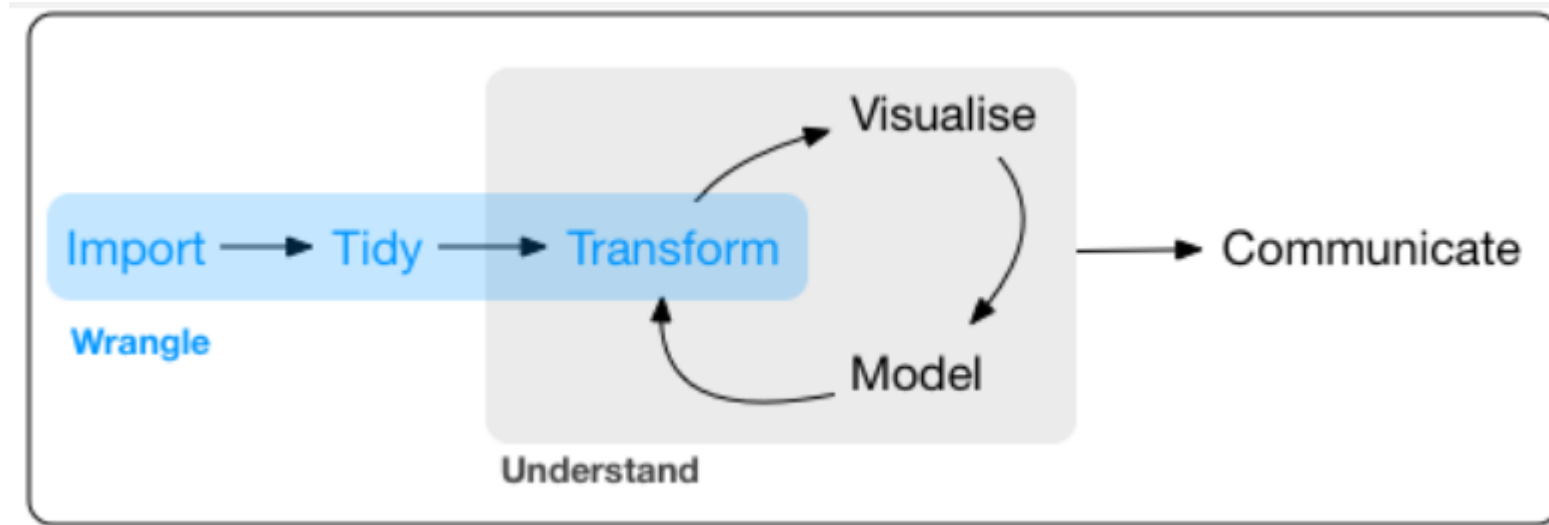
How people think data scientists spend time



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.

country	year	cases	population
Afghanistan	1999	15	1999071
Afghanistan	2000	1666	200095360
Brazil	1999	3737	17206362
Brazil	2000	8488	17404898
China	1999	21258	127915272
China	2000	21776	128042583

variables

country	year	cases	population
Afghanistan	1999	15	1999071
Afghanistan	2000	1666	200095360
Brazil	1999	3737	17206362
Brazil	2000	8488	17404898
China	1999	21258	127915272
China	2000	21776	128042583

observations

country	year	cases	population
Afghanistan	1999	15	1999071
Afghanistan	2000	1666	200095360
Brazil	1999	3737	17206362
Brazil	2000	8488	17404898
China	1999	21258	127915272
China	2000	21776	128042583

values

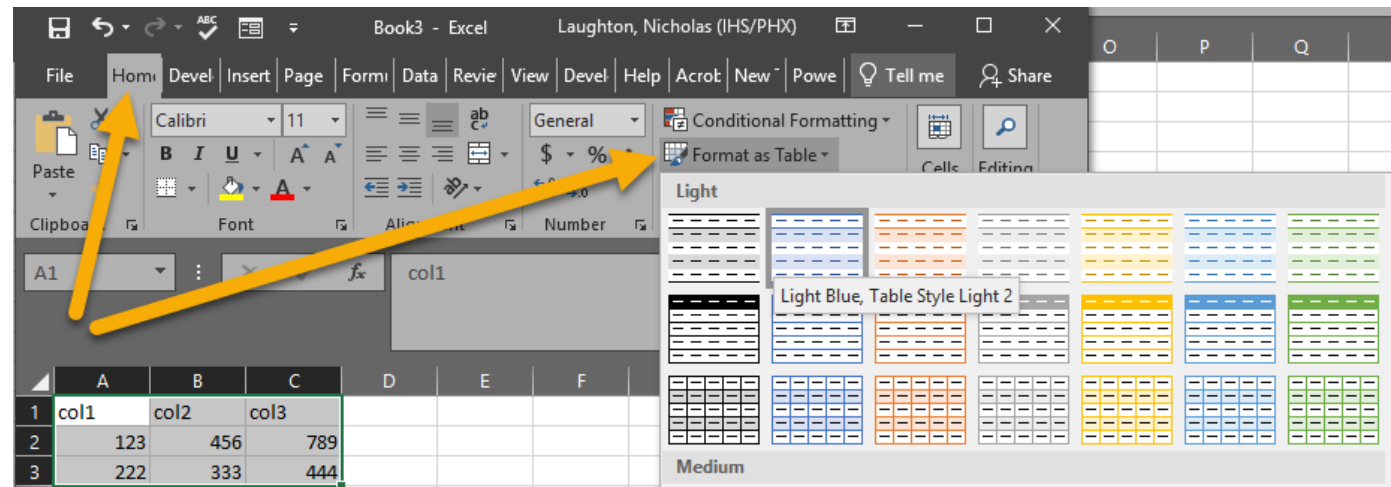


Excel Tables

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 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.



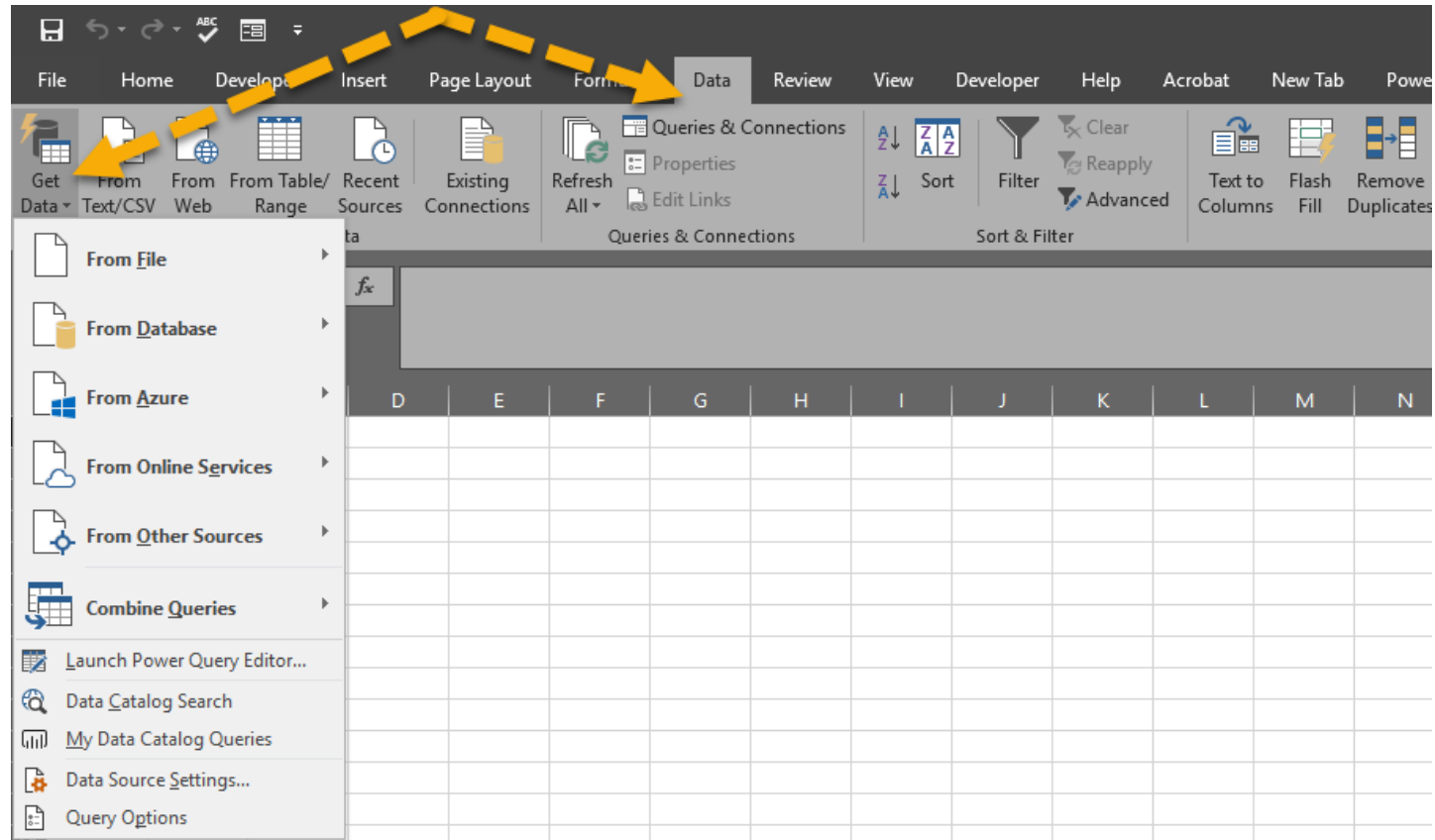
What I Use Power Query for

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;99999999999999999999 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

DATE          LOT #          PCC VISIT LISTING
-----
01/01/2001    E71770LA      INFLUENZA, NOS
01/01/2001    --            INFLUENZA, NOS
01/01/2001    --            TD (ADULT)
01/01/2001    --            VARICELLA 1
01/01/2001    --            IPV 3
01/01/2001    --            DTAP
01/02/2001    --            HEP B, NOS 2
01/02/2001    --            INFLUENZA, NOS
01/02/2001    --            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



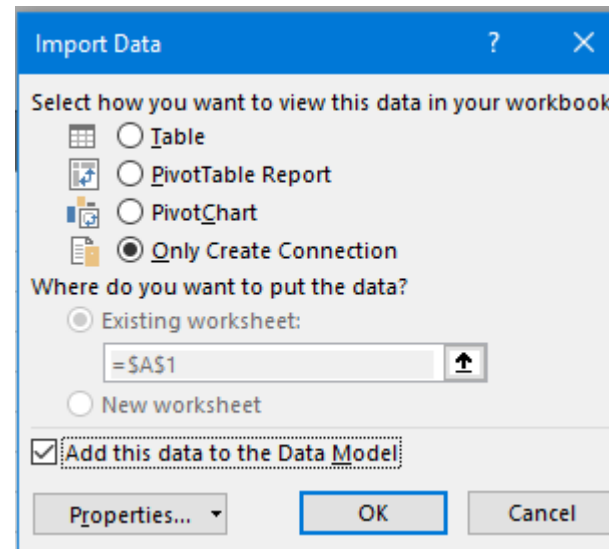
Navigating Power Query

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.

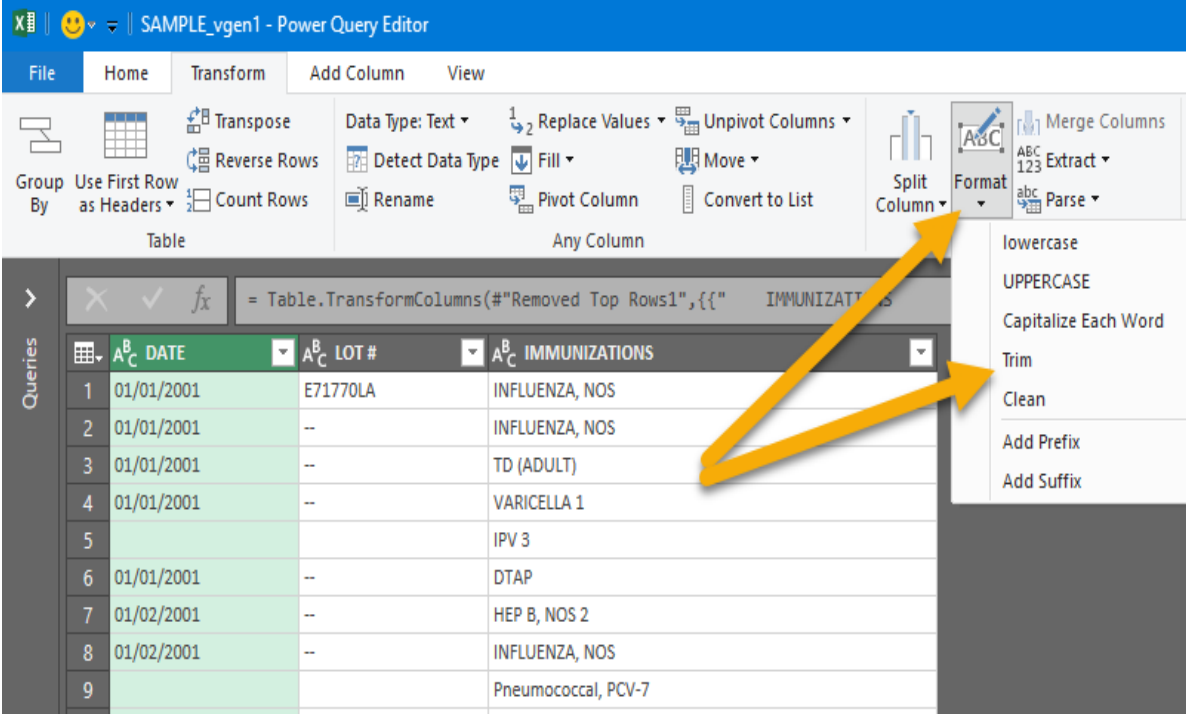
Page 1

DATE	LOT #	PCC VISIT LISTING 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



The screenshot shows the Power Query Editor interface. The ribbon includes 'File', 'Home', 'Transform', 'Add Column', and 'View'. The 'Format' dropdown menu is open, showing options like 'lowercase', 'UPPERCASE', 'Capitalize Each Word', 'Trim', 'Clean', 'Add Prefix', and 'Add Suffix'. Two yellow arrows point from the 'Format' dropdown to the 'Trim' option. The data table below shows columns for DATE, LOT #, and IMMUNIZATIONS.

	DATE	LOT #	IMMUNIZATIONS
1	01/01/2001	E71770LA	INFLUENZA, NOS
2	01/01/2001	--	INFLUENZA, NOS
3	01/01/2001	--	TD (ADULT)
4	01/01/2001	--	VARICELLA 1
5			IPV 3
6	01/01/2001	--	DTAP
7	01/02/2001	--	HEP B, NOS 2
8	01/02/2001	--	INFLUENZA, NOS
9			Pneumococcal, PCV-7



Filling Down Missing Data 2

2. Replace "" with "null" (not NULL or Null)

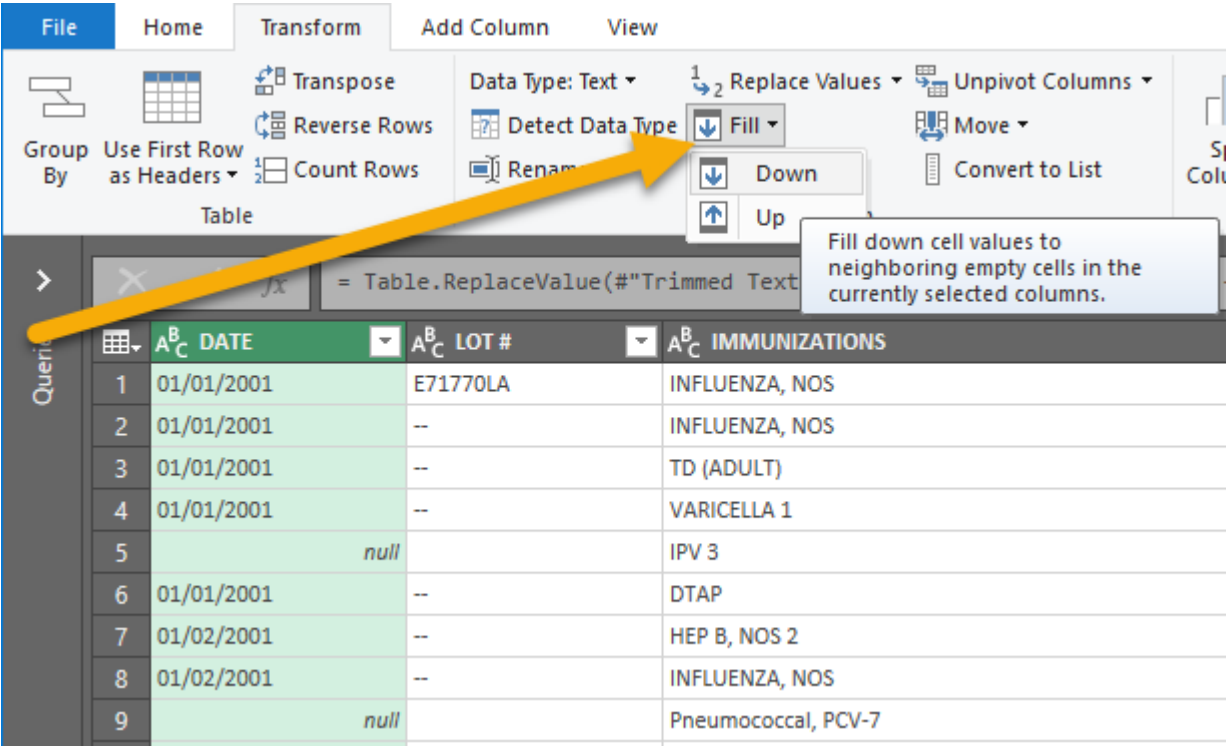
The screenshot shows the Power Query Editor interface. The ribbon is set to 'Transform', and the 'Replace Values' button is highlighted with a yellow arrow. Below the ribbon, a data table is visible with columns 'DATE' and 'L'. The 'Replace Values' dialog box is open, showing 'Value To Find' as an empty text box and 'Replace With' as 'null'. Two yellow arrows point to these input fields. The background shows a grid of data with some missing values (represented by '--').

	DATE	L
1	01/01/2001	E717
2	01/01/2001	--
3	01/01/2001	--
4	01/01/2001	--
5		
6	01/01/2001	--
7	01/02/2001	--
8	01/02/2001	--
9		
10	01/02/2001	7174
11	01/02/2001	E717
12	01/02/2001	--
13	01/02/2001	--



Filling Down Missing Data 3

3. Fill Down



The screenshot shows the Power Query 'Transform' ribbon with the 'Fill' dropdown menu open. A yellow arrow points to the 'Fill' button. A tooltip explains: 'Fill down cell values to neighboring empty cells in the currently selected columns.'

	A ^B C DATE	A ^B C LOT #	A ^B C IMMUNIZATIONS
1	01/01/2001	E71770LA	INFLUENZA, NOS
2	01/01/2001	--	INFLUENZA, NOS
3	01/01/2001	--	TD (ADULT)
4	01/01/2001	--	VARICELLA 1
5		null	IPV 3
6	01/01/2001	--	DTAP
7	01/02/2001	--	HEP B, NOS 2
8	01/02/2001	--	INFLUENZA, NOS
9		null	Pneumococcal, PCV-7



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



Multiple Notepad windows showing text files. The active window displays the following text:

```

PCC MANAGEMENT REPORTS
SUMMARY P
REPORT REQUESTED BY: LAUGHTON, NICHOLAS
VISIT Selection Criteria
Encounter Date range: JAN 01, 2003 t
Any Immunization Administered?: IMM
REPORT/OUTPUT TYPE
Detailed Listing containing
Chart # (11)
Visit Date (10)
Immunizations/Series (15)
TOTAL column width: 42
Visits will be SORTED by: Visit Date

HRN
-----HRN
CI 114078 CI 1111
CI 105295 CI 1034
CI 125109 CI 11407
CI 115026 CI 1096 CI 10529
CI 101584 CI 1428 CI 12510
CI 136727 CI 11502
CI 140609 CI 1409 CI 10158
  
```



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

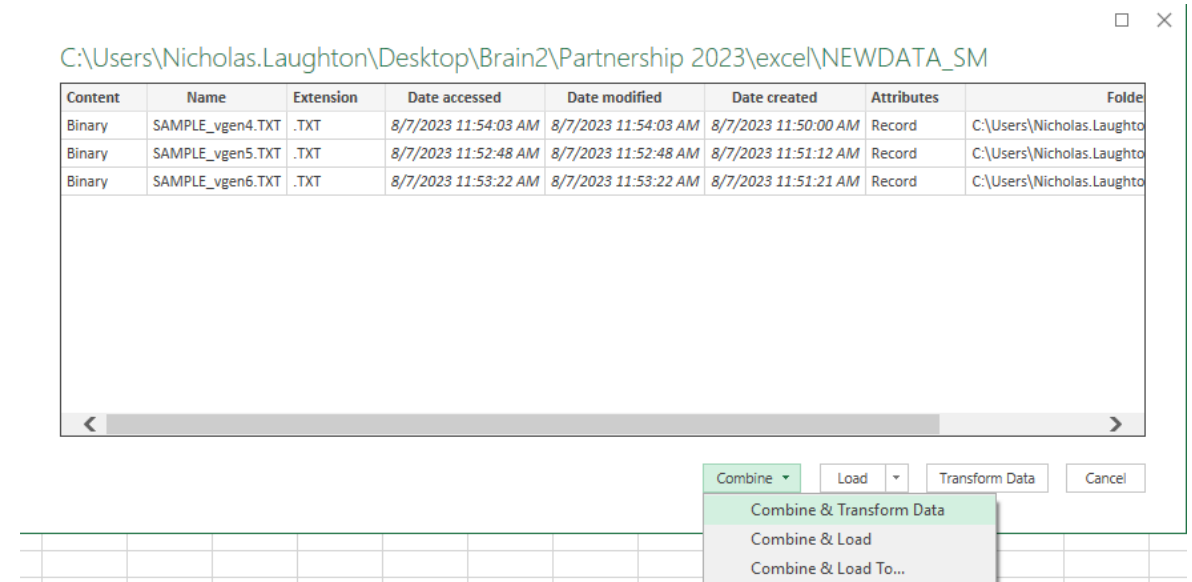
Keep only the files you want in the folder

Keep VERY consistent

- Same File Type
- Same Naming Convention
- Same Formatting

Getting Files from a Folder

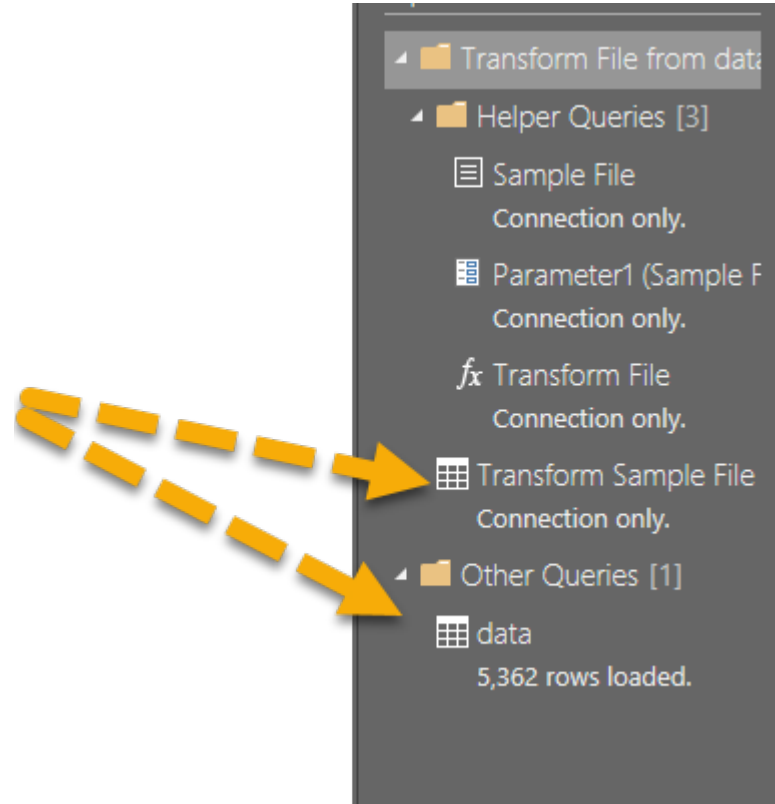
1. Save files in folder
2. Get Data → From File → From Folder
3. Navigate to the Folder
4. Combine & Transform Data



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

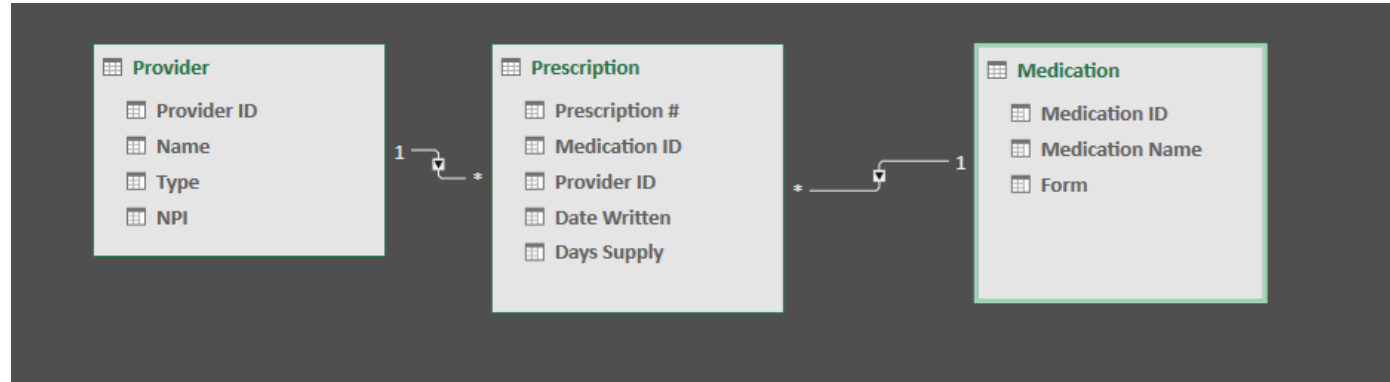
Power Pivot

- 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			Prescription				Provider				
Medication ID	Medication Name	Form	Prescription #	Medication ID	Provider ID	Date Written	Days Supply	Provider ID	Name	Type	NPI
A12	Aspirin	Tablet	123	A12	1AB	7/1/2023	30	1AB	Paulette	MD	123456
B23	Lisinopril	Tablet	456	B23	2BC	7/2/2023	45	2BC	Shawonna	NP	789444
C45	Tylenol	Capsule	789	C45	3DE	7/3/2023	90	3DE	Sam	PA	687799
D67	Ibuprofen	Oral Suspension	101112	D67	4FG	7/4/2023	10	4FG	Carl	MD	454566

