Clean Data from Excel, CSV, PDF, and Google Sheets with Data Interpreter

Applies to: Tableau Cloud, Tableau Desktop, Tableau Server

When you track data in Excel spreadsheets, you create them with the human interface in mind. To make your spreadsheets easy to read, you might include things like titles, stacked headers, notes, maybe empty rows and columns to add white space, and you probably have multiple tabs of data too.

When you want to analyze this data in Tableau, these aesthetically pleasing attributes make it very difficult for Tableau to interpret your data. That's where Data Interpreter can help.

Tip: Though Tableau's Excel add-in is no longer supported, Data Interpreter can help you reshape your data for analysis in Tableau.

What does Data Interpreter do?

Data Interpreter can give you a head start when cleaning your data. It can detect things like titles, notes, footers, empty cells, and so on and bypass them to identify the actual fields and values in your data set.

It can even detect additional tables and sub-tables so that you can work with a subset of your data independently of the other data.

After Data Interpreter has done its magic, you can check its work to make sure it captured the data that you wanted and identified it correctly. Then, you can make any necessary adjustments.

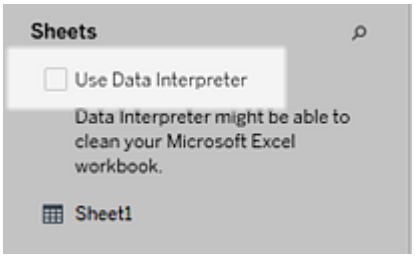
After you select the data that you want to work with, you might also need to do some additional cleaning steps like pivoting your data, splitting fields, or adding filters to get the data in the shape you want before starting your analysis.

Note: If your data needs more cleaning than what Data Interpreter can help you with, try Tableau Prep(Link opens in a new window).

Turn on Data Interpreter and review results

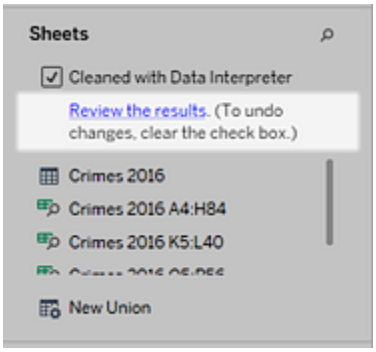
1. From the **Connect** pane, connect to an Excel spreadsheet or other connector that supports Data Interpreter such as Text (.csv) files, PDF files or Google sheets.

2. Drag a table to the canvas (if needed), then on the **Data Source** page, in the left pane, select the **Use Data Interpreter** check box to see if Data Interpreter can help clean up your data.

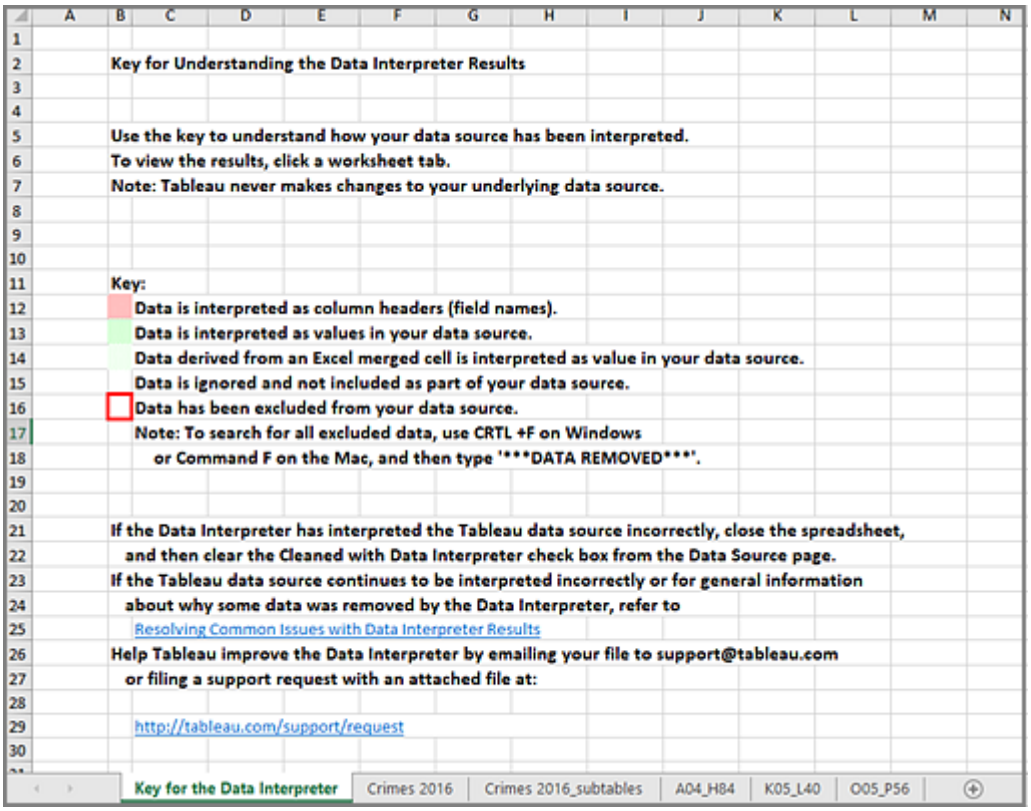


Note: When you clean your data with Data Interpreter, Data Interpreter cleans all the data associated with a connection in the data source. Data Interpreter does not change the underlying data.

3. In the Data pane, click the **Review the results** link to review the results of the Data Interpreter.



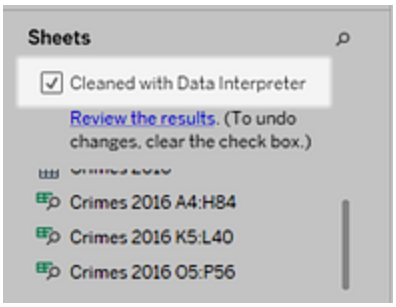
A copy of your data source opens in Excel on the **Key for the Data Interpreter** tab. Review the key to find out how to read the results.



4. Click each tab to review how Data Interpreter interpreted the data source.

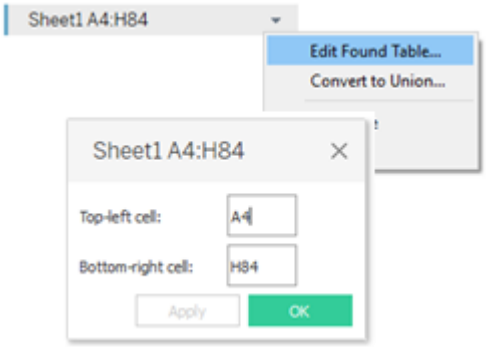
   If Data Interpreter found additional tables, also called found tables or sub-tables, they are identified in the <sheet name>_subtables tab by outlining their cell ranges. A separate tab is also included for each sub-table, color coded to identify the header and data rows.

If Data Interpreter does not provide the expected results, clear the **Cleaned with Data Interpreter** check box to use the original data source.



5. To replace the current table with any of the found tables, drag the current table off the canvas and then drag the found table that you want to use to the canvas.

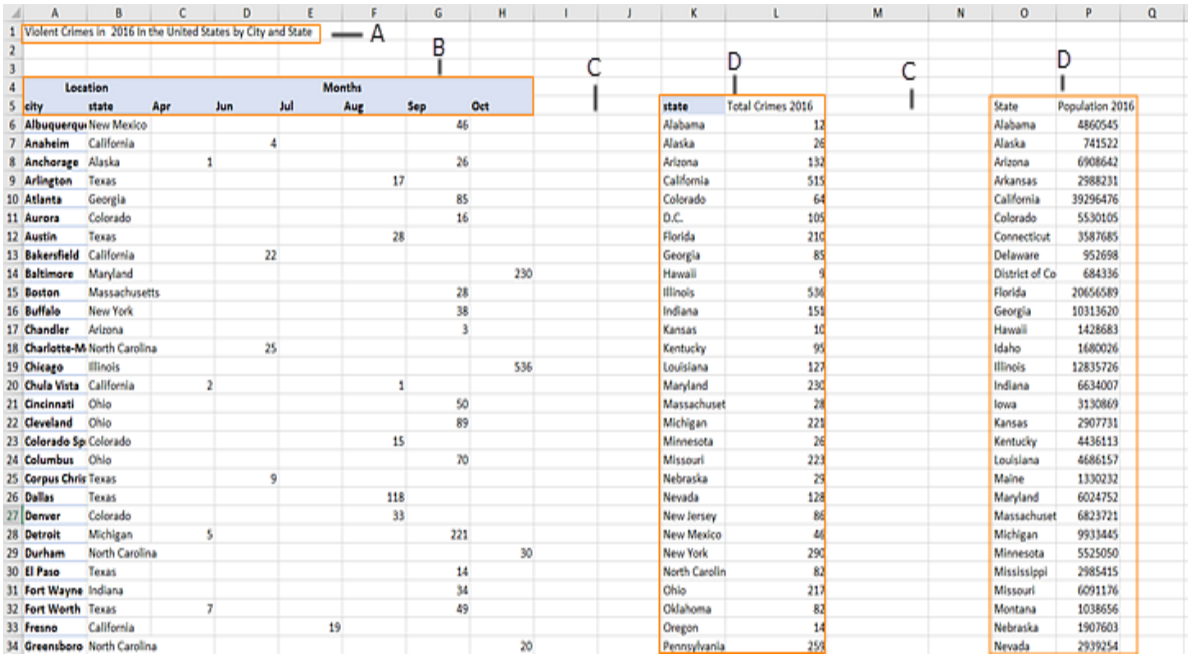
   If Data interpreter has misidentified the range of the found table, after you drag the found table to the canvas, click the drop-down arrow on that table, and then select **Edit Found Table** to adjust the corners of the found table (the top-left cell and bottom-right cell of the table).



6. After you have the data that you want to work with, you can apply any additional cleaning operations to your data so that you can analyze it.

Data Interpreter Example

In this example we are connecting to an Excel spreadsheet with violent crime data by city and state for the year 2016. This spreadsheet includes multiple tables on one sheet and some extra formatting.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Violent Crimes in 2016 In the United States by City and State																
2																	
3																	
4	Location		Months														
5	city	state	Apr	Jun	Jul	Aug	Sep	Oct			state	Total Crimes 2016			State	Population 2016	
6	Albuquerqu	New Mexico						46			Alabama	12			Alabama	4860545	
7	Anaheim	California			4						Alaska	26			Alaska	741522	
8	Anchorage	Alaska		1				26			Arizona	132			Arizona	6908642	
9	Arlington	Texas					17				California	515			Arkansas	2988231	
10	Atlanta	Georgia						85			Colorado	64			California	39296476	
11	Aurora	Colorado						16			D.C.	105			Colorado	5530105	
12	Austin	Texas					28				Florida	210			Connecticut	3587685	
13	Bakersfield	California			22						Georgia	85			Delaware	952698	
14	Baltimore	Maryland						230			Hawaii	9			District of Co	684336	
15	Boston	Massachusetts						28			Illinois	536			Florida	20656589	
16	Buffalo	New York						38			Indiana	151			Georgia	10313620	
17	Chandler	Arizona						3			Kansas	10			Hawaii	1428683	
18	Charlotte-M	North Carolina			25						Kentucky	99			Idaho	1680026	
19	Chicago	Illinois						536			Louisiana	127			Illinois	12835726	
20	Chula Vista	California		2			1				Maryland	230			Indiana	6634007	
21	Cincinnati	Ohio						50			Massachuset	28			Iowa	3130869	
22	Cleveland	Ohio						89			Michigan	221			Kansas	2907731	
23	Colorado Sp	Colorado					15				Minnesota	26			Kentucky	4436113	
24	Columbus	Ohio						70			Missouri	223			Louisiana	4686157	
25	Corpus Chris	Texas			9						Nebraska	29			Maine	1330232	
26	Dallas	Texas					118				Nevada	128			Maryland	6024752	
27	Denver	Colorado					33				New Jersey	86			Massachuset	6823721	
28	Detroit	Michigan		5				221			New Mexico	46			Michigan	9933445	
29	Durham	North Carolina							30		New York	290			Minnesota	5525050	
30	El Paso	Texas						14			North Carolin	82			Mississippi	2985415	
31	Fort Wayne	Indiana						34			Ohio	217			Missouri	6091176	
32	Fort Worth	Texas		7				49			Oklahoma	82			Montana	1038656	
33	Fresno	California				19					Oregon	14			Nebraska	1907603	
34	Greensboro	North Carolina							20		Pennsylvania	259			Nevada	2939254	

A. Title

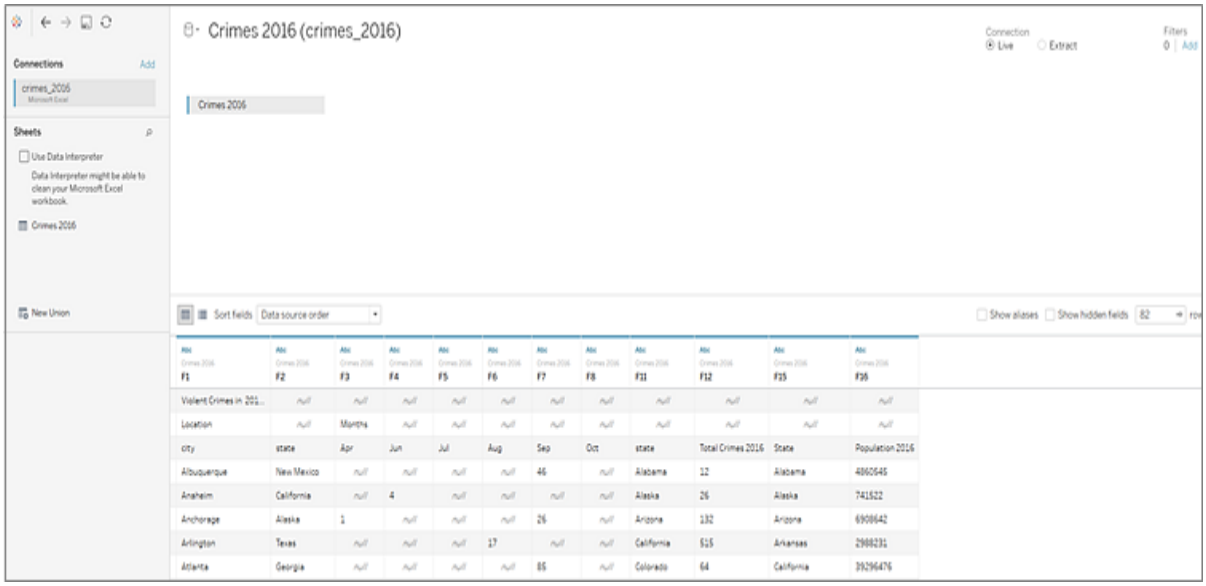
B. Merged header cells

C. Extra white space

D. Sub-tables

The extra formatting in this spreadsheet makes it difficult for Tableau to determine what the field headers and values are.

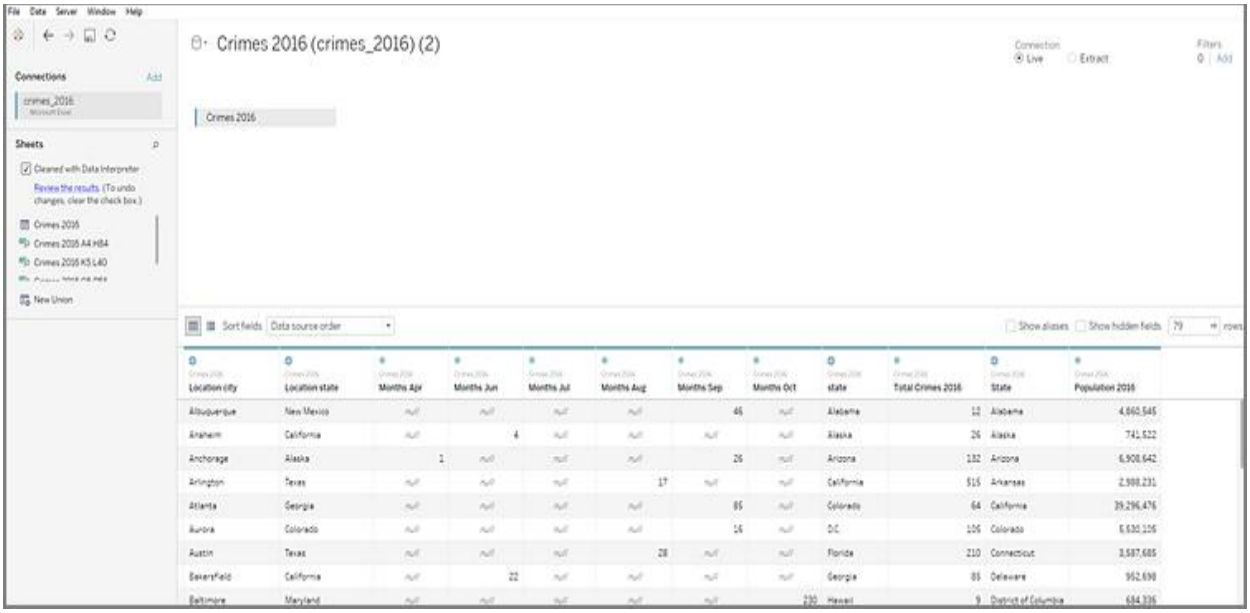
Instead, it reads the data vertically and assigns each column the default value F1, F2, F3 (Field 1, Field 2, Field 3) and so on. Blank cells are read as null values.



To see if Data Interpreter can help clean this data set, we select **Use Data Interpreter**.

Data Interpreter detected the proper headings for the fields, removed the extra formatting and found several sub-tables. The sub-tables are listed in the **Sheets** section in the Data pane and are named using the original sheet name and the cell ranges for each sub-table.

In this example there are three sub-tables: **Crimes 2016 A4:H84**, **Crimes 2016 K5:L40**, and **Crimes 2016 O5:P56**.

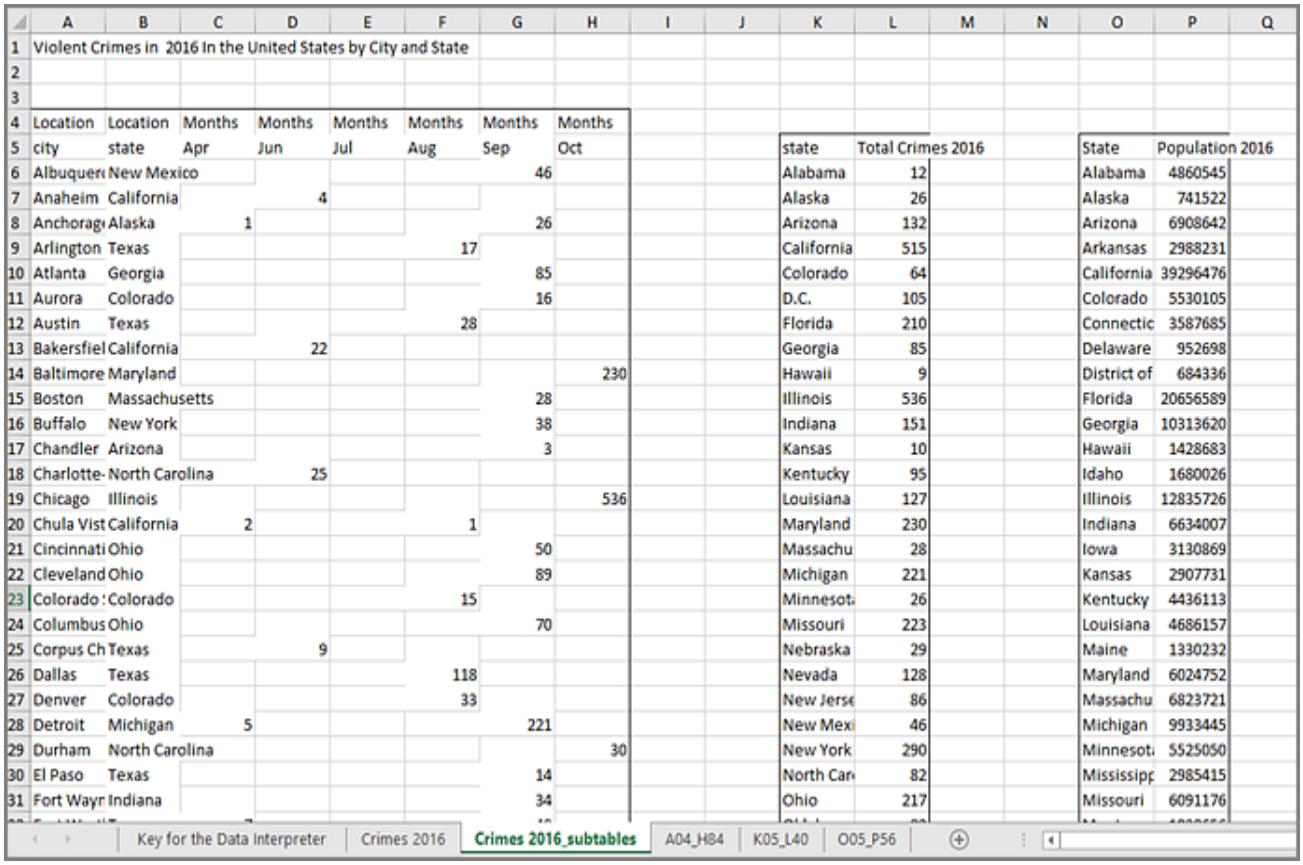


To examine the results of the Data Interpreter more closely, we click the **Review the results** link in the Data pane to view an annotated copy of the spreadsheet.

Here we see a copy of the original data, color coded to identify which data was identified as header data and which data was identified as field values.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Violent Crimes in 2016 In the United States by City and State																
2																	
3																	
4	Location	Location	Months	Months	Months	Months	Months	Months									Header
5	city	state	Apr	Jun	Jul	Aug	Sep	Oct			state	Total Crimes 2016			State	Populatio	Header
6	Albuquer	New Mexico						46			Alabama	12			Alabama	4860545	Data
7	Anaheim	California			4						Alaska	26			Alaska	741522	Data
8	Anchorag	Alaska		1				26			Arizona	132			Arizona	6908642	Data
9	Arlington	Texas					17				California	515			Arkansas	2988231	Data
10	Atlanta	Georgia						85			Colorado	64			California	39296476	Data
11	Aurora	Colorado						16			D.C.	105			Colorado	5530105	Data
12	Austin	Texas					28				Florida	210			Connectic	3587685	Data
13	Bakersfiel	California			22						Georgia	85			Delaware	952698	Data
14	Baltimore	Maryland							230		Hawaii	9			District of	684336	Data
15	Boston	Massachusetts						28			Illinois	536			Florida	20656589	Data
16	Buffalo	New York						38			Indiana	151			Georgia	10313620	Data
17	Chandler	Arizona						3			Kansas	10			Hawaii	1428683	Data

The next tab shows us the sub-tables that Data Interpreter found, outlined by the cell ranges.



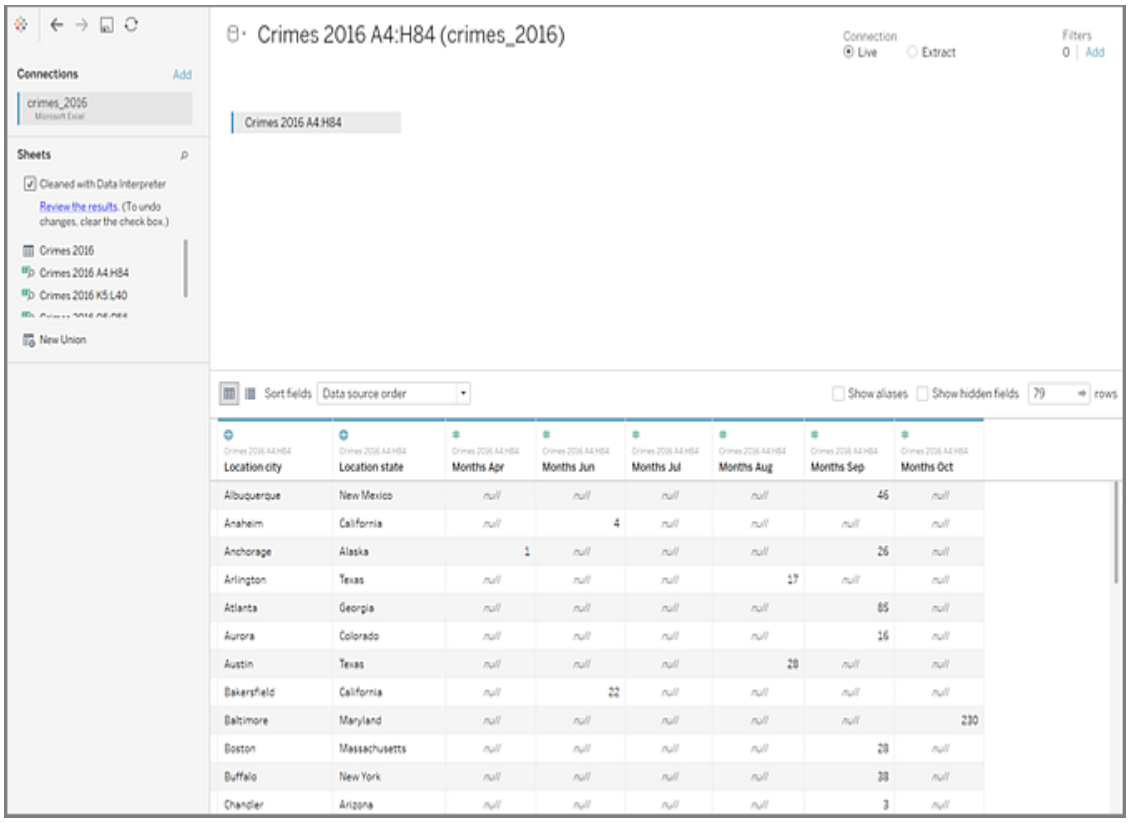
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Violent Crimes in 2016 In the United States by City and State																
2																	
3																	
4	Location	Location	Months	Months	Months	Months	Months	Months									
5	city	state	Apr	Jun	Jul	Aug	Sep	Oct									
6	Albuquer	New Mexico					46										
7	Anaheim	California		4													
8	Anchorag	Alaska	1				26										
9	Arlington	Texas				17											
10	Atlanta	Georgia					85										
11	Aurora	Colorado					16										
12	Austin	Texas				28											
13	Bakersfiel	California		22													
14	Baltimore	Maryland						230									
15	Boston	Massachusetts					28										
16	Buffalo	New York					38										
17	Chandler	Arizona					3										
18	Charlotte-	North Carolina		25													
19	Chicago	Illinois						536									
20	Chula Vist	California	2			1											
21	Cincinnati	Ohio					50										
22	Cleveland	Ohio					89										
23	Colorado :	Colorado				15											
24	Columbus	Ohio					70										
25	Corpus Ch	Texas		9													
26	Dallas	Texas				118											
27	Denver	Colorado				33											
28	Detroit	Michigan	5				221										
29	Durham	North Carolina						30									
30	El Paso	Texas					14										
31	Fort Wayr	Indiana					34										

state	Total Crimes 2016
Alabama	12
Alaska	26
Arizona	132
California	515
Colorado	64
D.C.	105
Florida	210
Georgia	85
Hawaii	9
Illinois	536
Indiana	151
Kansas	10
Kentucky	95
Louisiana	127
Maryland	230
Massachu	28
Michigan	221
Minnesot	26
Missouri	223
Nebraska	29
Nevada	128
New Jerse	86
New Mexi	46
New York	290
North Car	82
Ohio	217

State	Population 2016
Alabama	4860545
Alaska	741522
Arizona	6908642
Arkansas	2988231
California	39296476
Colorado	5530105
Connectic	3587685
Delaware	952698
District of	684336
Florida	20656589
Georgia	10313620
Hawaii	1428683
Idaho	1680026
Illinois	12835726
Indiana	6634007
Iowa	3130869
Kansas	2907731
Kentucky	4436113
Louisiana	4686157
Maine	1330232
Maryland	6024752
Massachu	6823721
Michigan	9933445
Minnesot	5525050
Mississipp	2985415
Missouri	6091176

Key for the Data Interpreter | Crimes 2016 | **Crimes 2016_subtables** | A04_H84 | K05_L40 | O05_P56

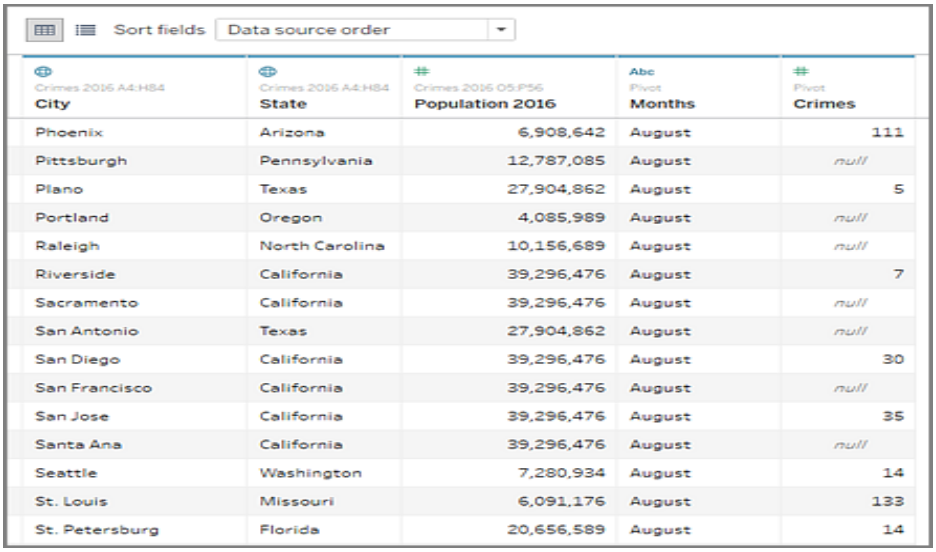
In this example the first sub-table, **Crimes 2016 A4:H84**, has the main data that we want to work with. To use this table as our data table, we can simply drag the original table off the canvas and then drag the new table to the canvas.



Once we have the data that we want to work with in the canvas, we can do some additional clean up on the data. For example we can:

- Change the field names so that they represent city, state, and month names.
- Pivot the months fields.
- Drag in the third sub-table **Crimes 2016 o5:P56** and join it to our first sub-table on the **State** field to include state populations for our analysis.
- Hide any duplicate fields that were added as a result of the join.

The results might look something like this:



Crimes 2016 A4:H84 City	Crimes 2016 A4:H84 State	Crimes 2016 O5:P56 Population 2016	Abc Pivot Months	Pivot Crimes
Phoenix	Arizona	6,908,642	August	111
Pittsburgh	Pennsylvania	12,787,085	August	*null*
Plano	Texas	27,904,862	August	5
Portland	Oregon	4,085,989	August	*null*
Raleigh	North Carolina	10,156,689	August	*null*
Riverside	California	39,296,476	August	7
Sacramento	California	39,296,476	August	*null*
San Antonio	Texas	27,904,862	August	*null*
San Diego	California	39,296,476	August	30
San Francisco	California	39,296,476	August	*null*
San Jose	California	39,296,476	August	35
Santa Ana	California	39,296,476	August	*null*
Seattle	Washington	7,280,934	August	14
St. Louis	Missouri	6,091,176	August	133
St. Petersburg	Florida	20,656,589	August	14

Now we are ready to start analyzing our data in Tableau.

When Data Interpreter is not available

The Data Interpreter option might not be available for the following reasons:

- **The data source is already in a format that Tableau can interpret:** If Tableau Desktop doesn't need extra help from Data Interpreter to handle unique formatting or extraneous information, the Data Interpreter option is not available.
- **Many rows or many columns:** The Data Interpreter option is not be available when your data has the following attributes:
  - Data contains more than 2000 columns.
  - Data contains more than 3000 rows and more than 150 columns.
- **The data source is not supported:** Data Interpreter is only available for Microsoft Excel, Text (.csv) files, PDF files and Google Sheets. For Excel, your data must be in the .xls or .xlsx format.