

Don't Be Scared of Relationships

The new data modeling capabilities create flexible data sources built around relationships.

Relationships combine data from different tables by looking at what columns (fields) those tables have in common and using that information to bring information from each table together in the analysis.

Unlike joins or unions, relationships form a data source without flattening multiple tables into a single table. Because of this, related data sources know which table each field is from. That means each field keeps its context, or level of detail. Related data sources can therefore handle tables with different granularity without issues of duplication or data loss.

In a related data source, the joins aren't fixed up front. Instead of merging all the data (and having to work with all the data regardless of what each viz requires), only the relevant data is combined—per sheet and as necessary. As you drag and drop, Tableau evaluates the relationships of the relevant fields and tables. Those relationships are used to write queries with the correct join types, aggregations, and null handling.

You can think about how the data fits together and what questions you want to answer, rather than how to combine the data or compensate for artifacts from the data source.

Relationships don't replace the previous ways of combining data: joins, unions, and blends. Rather, relationships are the new, flexible way to bring data together from multiple sources. The existing methods aren't going away, and are still useful in specific scenarios.

Note: For related information on how relationship queries work, see these Tableau blog posts:

- [Relationships, part 1: Introducing new data modeling in Tableau](#)
- [Relationships, part 2: Tips and tricks](#)
- [Relationships, part 3: Asking questions across multiple related tables](#)

Also see video podcasts on relationships from [Action Analytics](#) , such as [Why did Tableau Invent Relationships?](#) Click "Video Podcast" in the [Library](#) to see more.

Use Relationships

This topic walks through building a related data source and using one for analysis. If you would like to follow along, you can [download the Bookshop data set](#) . Choose Bookshop.xlsx for the raw data to start from scratch, or MinimalBookshop.tdsx to start with the related data source basics configured for you.

Note: Relationships are available in Tableau 2020.2 and later.

Video: Migrated Data

If you open an existing workbook from a previous version of Tableau (version 2020.1 or earlier) in Tableau 2020.2 or later, your data source tab might show a **Migrated Data** table.

There are multiple videos embedded in this topic. For your convenience, a transcript is provided in the expandable section underneath each video.

Migrated Data video transcript

If you open a workbook created before relationships were available (before version 2020.2), your data source may look different. If there were joined or unioned tables, they appear as a single logical table named Migrated Data. Double click to see the underlying tables—they're still there, and your analysis will still behave as expected.

Note that when you're only using the Migrated Data table, the Data pane behaves more like previous versions of Tableau Desktop.

You can even relate new tables to the directly to the Migrated Data table to extend the data model.

Once you've added relationships to the data model, the Data pane updates to the new layout, with dimensions and measures maintained within each table, and the Migrated Data table is displayed as a single, merged table.

Video: Relate tables

Note: The interface for editing relationships shown in this video differs slightly from the current release but has the same functionality.

Relate tables video transcript

To build a data source of related tables, connect to your data and drag a table to the canvas. Drag out a second table. A relationship will automatically form if it can. If Tableau cannot detect the related fields, you will be prompted to select them yourself.

The canvas you're seeing is a new layer of the data model where you can relate tables together. Dragging out additional tables adds them to the data model. The noodle shows which tables are related.

Relationships can be based on calculated fields. You can also specify how fields should be compared by using operators when you define the relationship.

Add joins

You can create a join by double clicking on a table to open the join canvas (also referred to as the physical layer), and then dragging out a table to create a join. Creating joins is the same experience as previous versions of Tableau—you can set a join clause, including a calculation if necessary, and configure the join type.

Once you close the join canvas, the table on the relationship canvas (also referred to as the logical layer) has a join icon to indicate the underlying join. This is because a join merges its tables together to form a single, new table. Related tables remain distinct.

Add unions

Add unions by dragging out New Union. Unions, like joins, actually merge the data, and show an icon on the logical layer. We recommend always renaming unions—or any tables with unclear names—to something more useful.

Video: Performance Options

Note: The interface for editing relationships shown in this video differs slightly from the current release but has the same functionality.

Performance Options video transcript

For any relationship, there are performance options as well as related fields. Tableau automatically chooses safe defaults for these options. However, if you are sure of the characteristics of your data, changing these options can give Tableau more information about how to best configure automatic joins during analysis. Setting the cardinality and referential integrity options gives Tableau more information about how to optimize queries.

Cardinality

Cardinality indicates whether or not the linking field records are unique. If multiple books can have the same author, but each author is listed only once in the author table, the cardinality for AuthID would be Many for Books and One for Author between those two tables. This setting impacts how Tableau handles aggregation before or after the join.

Referential integrity

Referential integrity indicates whether or not a record in one table is guaranteed to have a relationship in another table. If every book has an author, Book has referential integrity to Author. If not every author has a book, Author does not have referential integrity to Book. This setting impacts how Tableau chooses a join type and handles unmatched records.

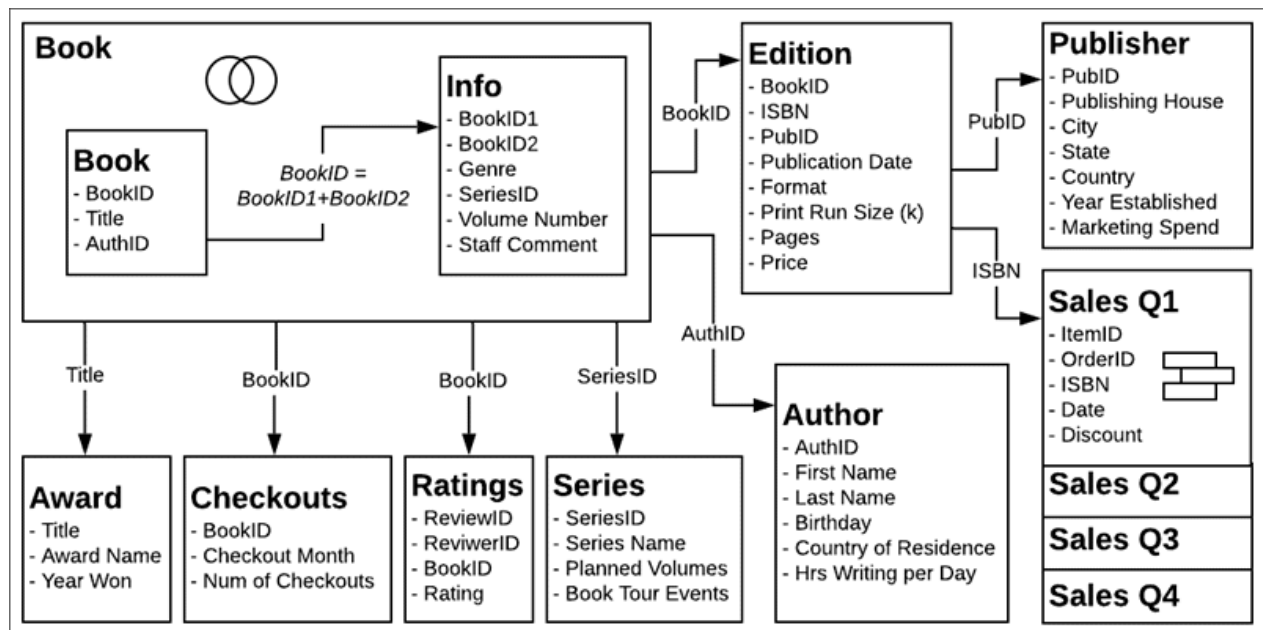
For more information, see [Cardinality and Referential Integrity \(cardinality_and_ri.htm\)](#).

Example: Bookshop Data

We'll do some analysis with this data source in a moment, so let's talk about the data.

The data set is around (fictitious) books, and it's important to consider the distinction between a book and an edition. A book is the conceptual work itself, with attributes such as the title, author, and genre. There are also editions of a book, with attributes such as a price and format (hardcover or paperback), that are identified with an ISBN. An edition of a book has a publisher and a page count, whereas a book may have won an award or be in a series.

You can follow along by downloading [MinimalBookshop.tdsx](#) (<https://help.tableau.com/samples/en-us/bookshop/MinimalBookshop.tdsx>), or be adventurous and build the model yourself from [Bookshop.xlsx](#) (<https://help.tableau.com/samples/en-us/bookshop/Bookshop.xlsx>). Note that you can hide many of the ID fields once the tables are combined.



Video: Work with multiple tables

Note: The interface for editing relationships shown in this video differs slightly from the current release but has the same functionality.

Work with multiple tables video transcript

The Data Source page

Before we go to a sheet tab, notice that the data grid view, below, shows only the fields from one table at a time. If we click to another table, we'll see the data it contains. There is no overall view because related data is not brought together until it's used in an analysis. For now, the data remains in each table and is therefore displayed by table.

If we click on a table with an underlying join or union, we do see that merged data, with any nulls that may involve.

The Data pane

When we click onto a sheet, you may notice that the Data pane looks different than in previous versions of Tableau. Instead of the primary organization of dimensions and measures, it's now focused on tables. Each table has its own breakdown for dimensions and measures, indicated by this line. It's possible to still group by folders instead, if you choose.

Number of Records and Count

There is also no **Number of Records** field. This is because the concept of Number of Records has changed. There isn't a single Number of Records for the data source. Each table has a **Count** field, which can be thought of as a local Number of Records for that table.

To get a sense of the number of records for multiple tables at the same time, you can use Measure Names and Measure Values or multi-select the count fields and use Show Me to create a viz.

Video: Basic calculations

Basic calculations video transcript

There are a few calculations we can do to make this data set a bit friendlier. First, we can create a calculation for the full author name that brings together first name and last name. We'll create a calculation called **Author Name** that is simply first name, a space, and last name. Note that this calculation stays in the Author table because it uses fields only from that table.

```
[First Name] + " " + [Last Name]
```

When we make a sale, the sales amount is a combination of the price—fixed by the publisher and specific to the edition, or the ISBN—and any discount at the time of the sale. So the calculation **Sales Amount** will be price times one minus the discount. Note that we're using the

ZN—or Zero Null—function to handle sales without discounts so we don't get nulls. This calculation goes to the shared area at the bottom of the Data pane as it uses fields from multiple tables. Aggregated calculations will also go to the bottom of the Data pane.

```
[Price] * (1-ZN([Discount]))
```

Video: Sets and groups

Sets and groups video transcript

We'll create a set for books to identify if they're in a series. In the Book table, right click on **Title** and select Create > Set. We'll name it **In a series**. On the **Condition** tab, choose **By field** with **Series Name, Count**, as ≥ 1 .

In the Edition table, the Format field is quite detailed, including specifying mass market and trade paperback. Sometimes that detail is useful, but sometimes we want to generalize between hardcover and paperback. Right click on **Format** field in the Data pane and select Create

> Group. We'll name it **Cover**. Control-click to select everything except Hardcover and then click **Group**. The name of the group is automatically highlighted, and we can call it **Paperback**.

Like with the Author Name calculation, these items we just created are relevant within a single table and therefore stay within those tables in the Data pane.

Analyze related data

Let's do some comparisons of how analysis looks between relationships and joins. We recommend that you download the companion workbooks for [Question 1](https://help.tableau.com/samples/en-us/bookshop/dontbescaresd1.twbx) (<https://help.tableau.com/samples/en-us/bookshop/dontbescaresd1.twbx>) and [Question 2](https://help.tableau.com/samples/en-us/bookshop/dontbescaresd2.twbx) (<https://help.tableau.com/samples/en-us/bookshop/dontbescaresd2.twbx>) rather than continue with your own data source from above. If you identify issues with the data source configuration or steps for Question 1, note that they are intentional for the purposes of the scenario.

Question 1: How many editions are there for each book?

Important: The video for Question 1 was recorded on beta software. The released version of 2020.2 correctly handles our attempt to introduce an error in the related data source. If you follow along, your workbook will not match the video—the relationships values will be correct. Please bear with us while we attempt to outsmart our product and force an error for educational purposes.

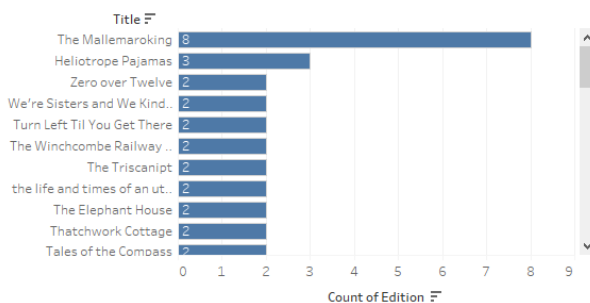
We'll set this up by doing some basic analysis with both data sources. We'll build a bar chart with each one, looking at the number of editions by title.

| Related | Joined |
|---------|--------|
|---------|--------|

Using the related data source:

1. Bring Title to Rows
2. Bring Edition (Count) to Columns
3. Click the Label icon in the toolbar to turn on labels
4. Click the Sort Descending icon in the toolbar

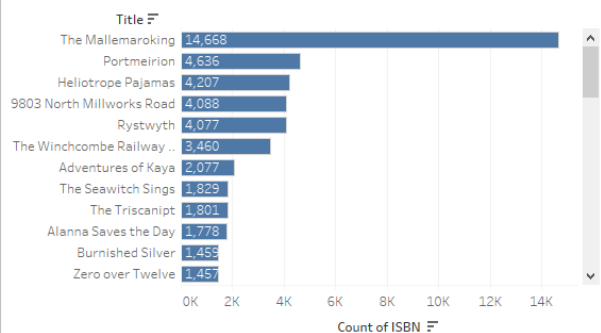
Relationships: How many editions are there for each book?



Using the joined data source:

1. Bring Title to Rows
2. Bring ISBN to Columns
3. Right Click on the pill and select Measure > Count
4. Click the Label icon in the toolbar to turn on labels
5. Click the Sort Descending icon in the toolbar

Joins: How many editions are there for each book?



It's immediately clear that Count is not the right aggregation for the joined data. However, something isn't right with the related data, either. Let's view the data and see what's going on and how to fix it. Watch the video (or read its transcript) to see how.

Video: Data validation and fixing the data model

Question 1 video transcript

Joined

Here we're looking at our joined data. On the joined viz, we'll take a look at a mark's underlying data.

This view data dialog has two tabs, Summary and Full Data.

- There are tons of rows for the same book—this is likely duplicated because of all the joins.

To handle the duplication, we'll change the aggregation on the ISBN field in the view to Count Distinct. Remember that we'll need to think about this duplication in almost every analytic scenario and address it with the correct aggregation, filtering, or how we structure the viz.

Related

Let's take a look at the structure of our related data source, and we'll take a peek at what's going on with our join as well. On the related viz, we'll take a look at a mark's underlying data.

And we see that the view data dialog again has two tabs, but this time it's the Edition tab instead of Full Data. Every table used in the viz will have its own tab.

- There's a doubling of rows for this book—and we know that joins can cause duplication, and that the Book table in this data source is really the Book table joined with Info and Award.

We can create a viz with Title and Award Name and scroll down to investigate. And sure enough, *The Mallemaroking* has won two awards. It would be better to relate these tables than to join them.

We'll go back to the Data Source tab and remove the Awards table from the join and relate it instead. And now we don't have to worry about the duplication issue any more for this data source.

Question 2: For authors with books in a series, who is has the most book tour events?

Video: Work with different levels of details

Question 2 video transcript

Let's look at which authors we'll be working with. Using the related data source, we'll bring **Author Name** and **Series Name** to Rows.

Because related data sources pare down unmatched data in tables that only contain dimensions, we're focusing on the data that is relevant to us. We can see that there are seven authors, two of whom have written for two series. If you don't want to see only matched values, you can restore those unmatched values (which is more like the legacy behavior) by going to the Analysis menu > Table Layout and checking Show Empty Rows.

| | |
|----------------|---------------|
| Related | Joined |
|----------------|---------------|

Using the related data source:

- 1. Bring the set **In a series** from the Book table to the filter shelf. The default is to filter only to members in the set.
- 2. Bring **Author Name** to Rows.
- 3. Bring **Book Tour Events** to Columns.

Using the joined data source:

- 1. Bring in getting the set to **In a series** to the filter i

l
t
e
r
s
h
e
l
f
.
2. B
r
i
n
g
A
u
t
h
o
r
N
a
m
e
t
o
R
o
w
s
.
3. B
r

i
n
g
**B
o
o
k
T
o
u
r
E
v
e
n
t
s**
t
o
C
o
l
u
m
n
s
.

Our
numb
ers
here

| |
|-----------------------|
| look a bit off. |
|-----------------------|

For the joined data, we know there's duplication because of the join, and we also know some authors span series. Because of this, we can't just change the aggregation to, say, MIN or MAX, because we lose information for authors with works in multiple series.

What we really want is the number of events per series, visualized by author. This is a classic case for Level of Detail (LOD) expressions. We'll create a calculation **Series Events**:

```
{FIXED [Series Name] : MIN ([Book Tour Events])}
```

Note that the MIN is to handle duplication of events for a single series.

Now if we bring this new field to Columns instead of the original events field, we'll get the correct values.

For the related data, we didn't need to do any of that. Relationships are smart enough to understand the native level of detail and the way the Author table relates through the Book and Info tables through to the Series table, and to correctly join and aggregate the Events measure back to Author Name—all without having to write LOD calculations.

So don't be scared of relating your data. See for yourself! You can [download the Bookshop data](#) or use some of your own. Try filtering, using table calculations, building a variety of chart types, configuring performance options, and pushing relationships as far as you can.

Tips when working with related data

- [View Underlying Data \(inspectdata_viewdata.htm\)](#) often to verify what data a mark represents.
- If you aren't required to join—and there are [reasons](#) (https://help.tableau.com/current/pro/desktop/en-us/relate_tables.htm#features-of-different-options-for-combining-data-relationships-joins-and-blends) you may need to—using a relationship provides more flexibility.

- If you don't want to see only matched values, you can restore those unmatched values by going to the Analysis menu > Table Layout > Show Empty Rows.

Related resources

Ready to tackle calculations with Relationships? Check out [Don't Be Scared of Calculations in Relationships \(datasource_dont_be_scared_calcs.htm\)](#).

Ready to keep exploring how to do complex analysis with Relationships? Check out [Don't Be Scared of Deeper Relationships \(datasource_dont_be_scared_deeper.htm\)](#).

For more information on the technical underpinnings of relationships straight from the Product Management team, check out the series on relationships on the Tableau Blog.

- [Relationships, part 1: Introducing new data modeling in Tableau](#)
- [Relationships, part 2: Tips and tricks](#)
- [Relationships, part 3: Asking questions across multiple related tables](#)

Also see video podcasts on relationships from [Action Analytics](#) , such as [Why did Tableau Invent Relationships?](#) Click "Video Podcast" in the [Library](#) to see more.

Other articles in this section

- [Don't Be Scared of Calculations in Relationships \(datasource_dont_be_scared_calcs.htm\)](#)
- [Don't Be Scared of Deeper Relationships \(datasource_dont_be_scared_deeper.htm\)](#)