Data Visualization Lab Introduction to Tableau

Nov. 14, 2022

Tableau Public

- *Tableau Public is a free platform to explore, create and publicly share data visualizations online"
 - It is free;
 - Everything you do in TP <u>must</u> be publicly accessible (can't save locally).
- → What does this imply?

Tableau Public

- → "Tableau Public is a free platform to explore, create and publicly share data visualizations online"
 - **!** It is free;
 - Everything you do in TP must be publicly accessible (can't save locally).
- → What does this imply?

TP can only be used in a <u>non-professional way</u> (as you know, businesses must keep data private)

Tableau Public

- → "Tableau Public is a free platform to explore, create and publicly share data visualizations online"
 - **!** It is free;
 - Everything you do in TP must be publicly accessible (can't save locally).
- → What does this imply?

 TP can only be used in a non-professional way (as you know, businesses must keep data private)
- → Does this mean that we are about to use *yet another useless platform* that you will not give a dఱn about for the rest of your lives?

Tableau Desktop

- → "Tableau Desktop is data visualization software that lets you see and understand data in minutes."
 - Not free, but free trial available;
 - Data can be saved locally and kept private.

It is designed to be used in professional environments.

→ As a student you can access it by applying for an <u>annual free license</u> for the Desktop version.

Know that it exists but <u>do not use it</u> for your exam project

What's next

What we will do

Create visualizations on Tableau

Learn to build dashboards and stories

Use Tableau Public

What we won't do

Preprocess data in Tableau

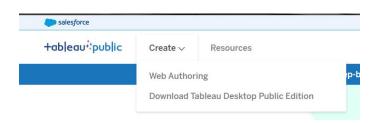
Work on private data

Use Tableau Desktop

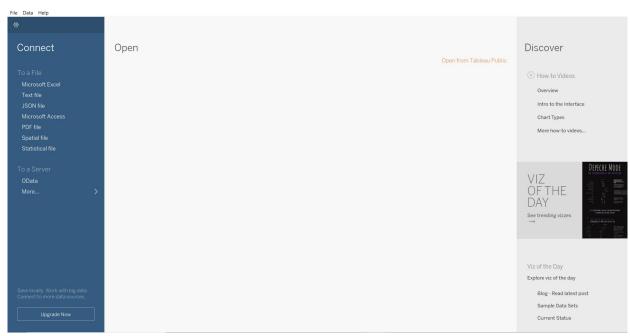
Introduction to Tableau

Installation and setup

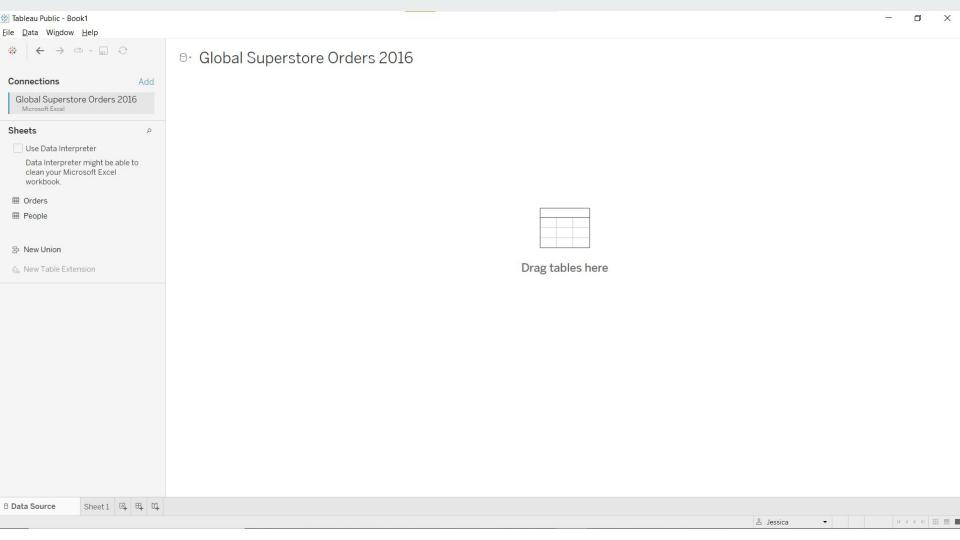
- → Head over to https://public.tableau.com/app/discover and create a new account for Tableau Public;
- → If you are not using your computer:
 - Create your account on the <u>TP homepage</u>;
 - ♦ Click on *Create>Web Authoring* to access the tool

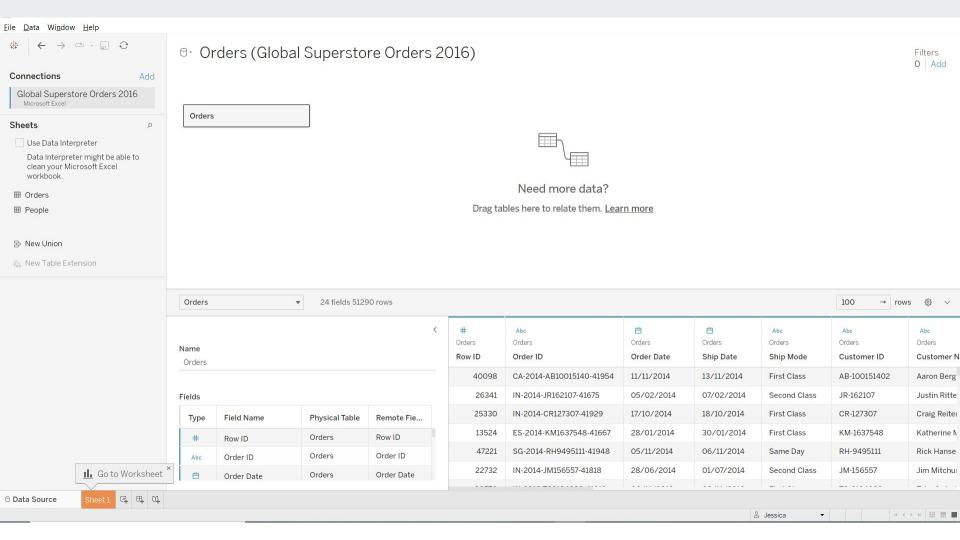


Connecting to data





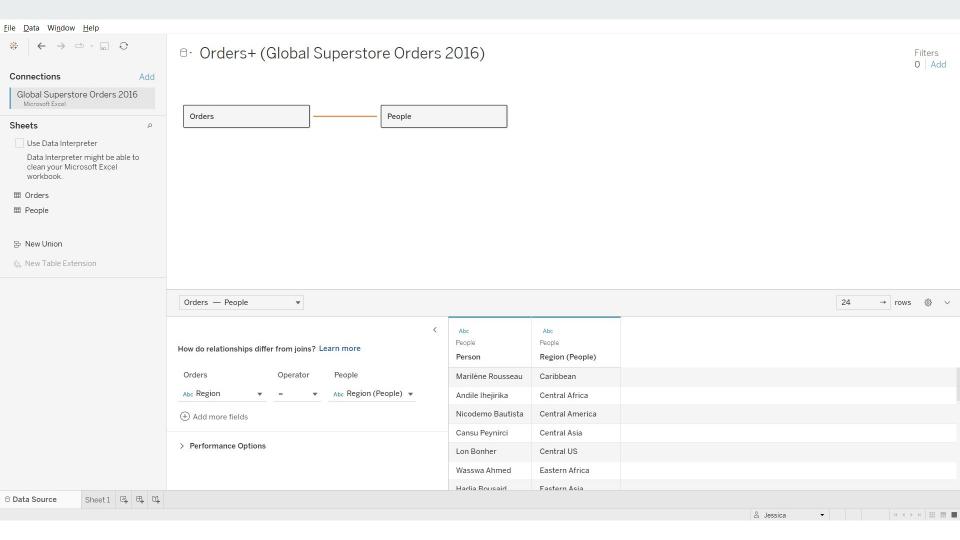




Joining multi-source data

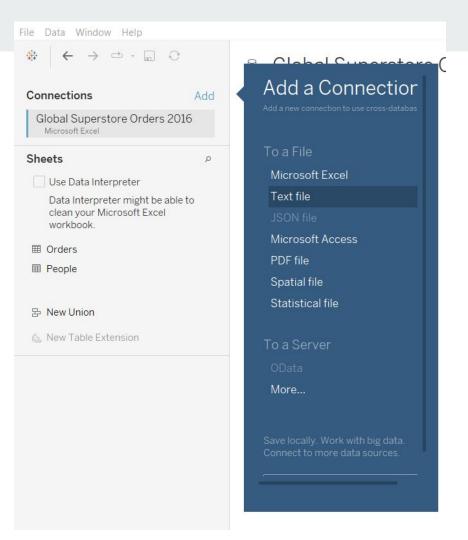
Connecting tables from the same file (Global Superstore Orders 2016)

- → Have a look at the file we just imported at the bottom of the page. We can see that for each transaction, a region is reported. Suppose we have a list of **vendors' names** so that we know the name of the vendor based on the region where the transaction has taken place.
- → It could be useful to integrate data from the two tables. To do so, we simply drag the **People** table onto the canvas.



Joining tables from different files (Global Superstore Returns 2016)

- → Otherwise, we can create an integrated data source by adding a connection;
- In our scenario, we might need to find out which orders, if any, were returned, but we have no returns listed in here. We need to resort to a different file.
- → Click on Add > Text File (or whatever type of file you might need). In this case we will add the Global Superstore Returns 2016.

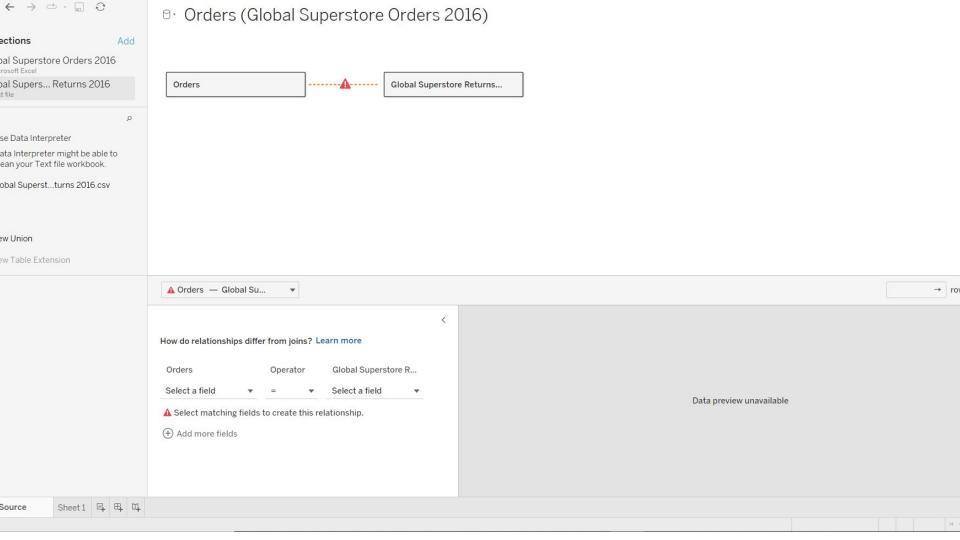


Joining data from different files

Once the other data source is imported we can do two different things:

1. Drag and drop the **Global Superstore Returns 2016.csv** table from the **Files** (once was Sheets) pane onto the canvas;

What happens? Why?

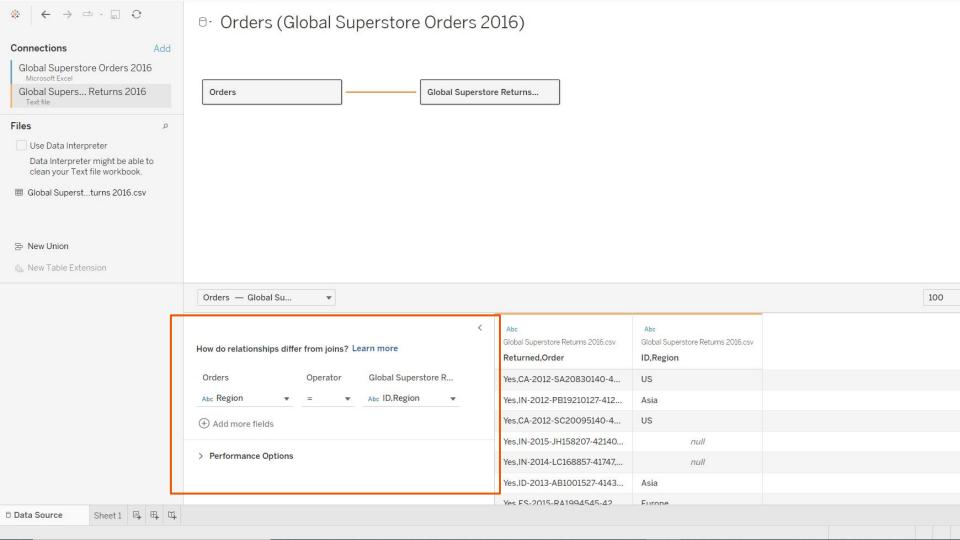


Joining data from different files

Once the other data source is imported we can do two different things:

1. Drag and drop the **Global Superstore Returns 2016.csv** table from the **Files** (once was Sheets) pane onto the canvas;

What happens? The relationship between tables is not created automatically Why? Column names in the two data sources did not match



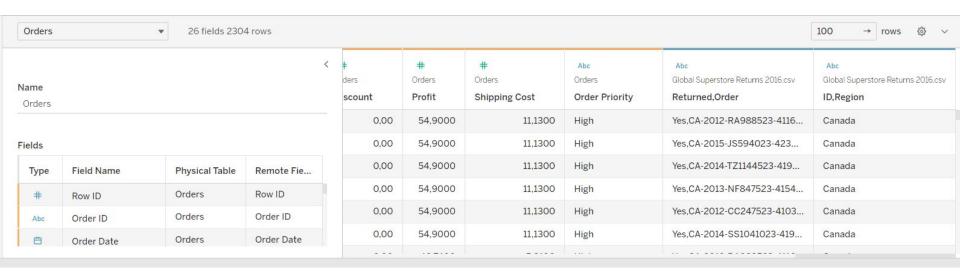
Joining data from different files

Once the other data source is imported we can do two different things:

- 1. Drag and drop the **Global Superstore Returns 2016.csv** table from the **Files** (once was Sheets) pane onto the canvas;
- 2. We can merge the two data sources by right-clicking on the **Orders** table on the canvas and then **Open** it. Once opened, drag and drop the **Global Superstore Returns 2016.csv** file onto the canvas, select **Join (Inner)** and again join by **Region** and **ID**, **Region** respectively.

What happens?

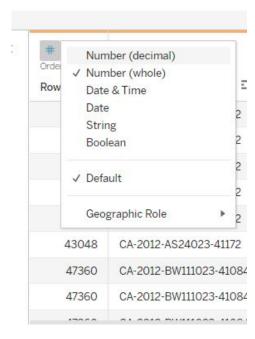
This second method will actually join the two data sources, meaning that you will be able to find information about returned orders in the Orders table. This might actually cause some loss of information, but we'll see this better in future lessons.



Data formatting

We will do a bit of data preparation by:

 Changing the Row ID type from Number (#) to String (Abc) – you can do this by simply clicking on the # icon in the column header;



Data formatting

We will do a bit of data preparation by:

- 1. Changing the **Row ID** type from Number (#) to String (Abc) you can do this by simply clicking on the # icon in the column header;
- 2. Splitting the **Order ID** to isolate the **Distribution Center** code you can do this by clicking on the dropdown menu caret on the column header and select
 - a. **Split**: selects on which character to split automatically;
 - b. **Custom split:** to customize the operation).

To rename the column, simply double click on the title and name it Distribution Center

| •Abc Calculation |
|------------------|
| Distribution c |
| CA |
| ~. |

Custom split with character, split only once at the first character. Column renamed as Distribution Center

| Abc Calculation | calculation | •Abc Calculation | *# Calculation |
|--------------------|--------------------|---------------------|--------------------|
| Order ID - Split 1 | Order ID - Split 2 | Order ID - Split 3 | Order ID - Split 4 |
| CA | 2012 | AS24023 | 41172 |
| CA | 2012 | AS24023 | 4117 |
| CA | 2012 | AS24023 | 4117 |
| CA | 2012 | AS24023 | 41177 |
| CA | 2012 | AS24023 | 4117 |
| CA | 2012 | AS24023 | 4117 |
| CA | 2012 | BM165023 | 4121 |
| CA | 2012 | BM165023 | 41216 |
| | 2040 | 01465000 | 4404 |

Split with no customization

| •Abc | |
|----------------|--|
| Calculation | |
| Distribution c | |
| CA | |
| | |

| •Abc Calculation | -# Calculation | Abc Calculation | *# Calculation |
|---------------------|--------------------|--------------------|--------------------|
| Order ID - Split 1 | Order ID - Split 2 | Order ID - Split 3 | Order ID - Split 4 |
| CA | 2012 | AS24023 | 41172 |
| CA | 2012 | AS24023 | 41172 |
| CA | 2012 | AS24023 | 41172 |
| CA | 2012 | AS24023 | 41172 |
| CA | 2012 | AS24023 | 41172 |
| CA | 2012 | AS24023 | 41172 |
| CA | 2012 | BM165023 | 41216 |
| CA | 2012 | BM165023 | 41216 |
| ~· | 2012 | DI MOTOGO | ***** |

Note that the newly created columns are put at the <u>end of the table</u> (if you think the Split operation has done nothing to your data, go look for the results there)

Try it yourself

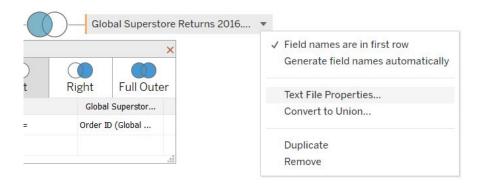
- → Import the Global Superstore Orders 2016 Excel file;
- → Join the Orders table from Global Superstore Orders 2016 to the Global Superstore Returns 2016 file;
- → Use a **left Join** on the **Order ID** column;
- → Change the **Row ID** type to String and isolate the **Distribution Center** from the **Order ID** column.

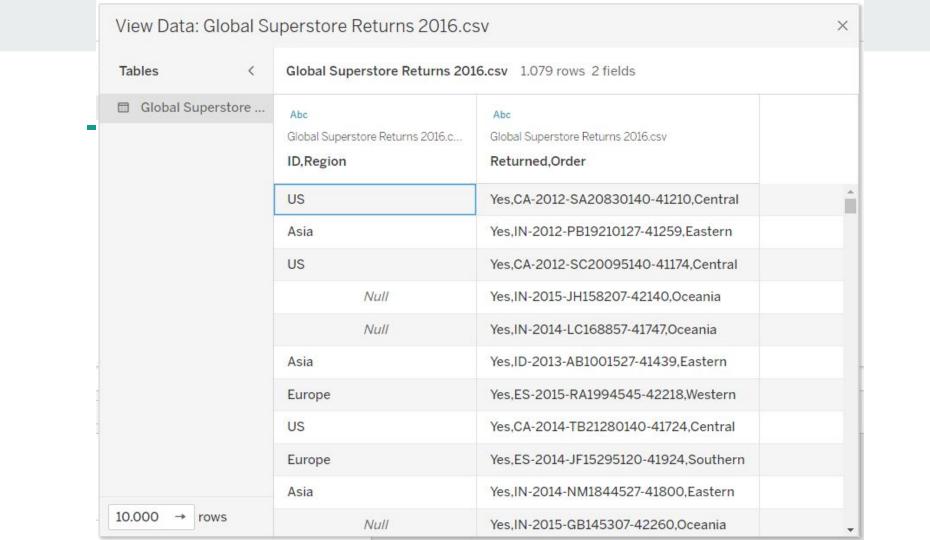
A little trick: sometimes TP is not capable of importing your files correctly; in these cases, you might need to correct the import (check the next slide for more information)

Try it yourself

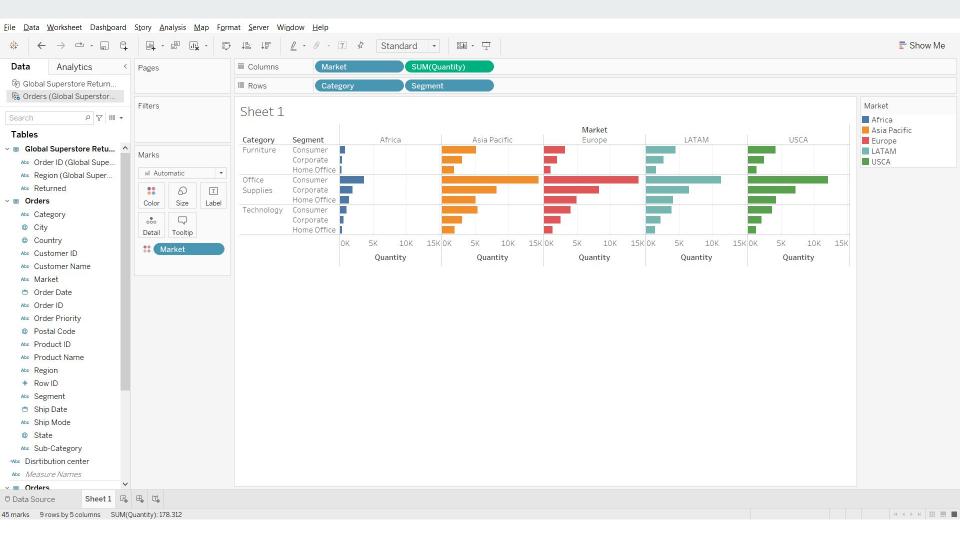
Right click on the Global Superstore Returns 2016 dropdown menu and open the Text File Properties menu. Update the Field separator.

Which separator do we need?





The Tableau Workspace Tableau Sheets



All you have to do is dragging the fields out from the pane on the left, called **Data Pane**.

The Data Pane is divided into Dimensions and Measures that represent the <u>column headers</u> in the Excel sheet:

- → Dimensions are categorical fields, often discrete and color coded blue in the data pane and in the view;
- → Measures are metrics, often continuous and color coded green in the data pane and in the view.

To add dimensions and measures to the view to create charts, simply drag and drop them from the Data Pane onto the **Rows** and **Columns** shelves.

Look at the Data Pane, which fields are categorical (and which are metrics instead)?

To create the viz, let's bring:

- → Category and Segment to Rows;
- → Quantity and Market to Columns.

To make the view clearer drag and drop the **Market** field onto the **Color** tab in the Marks pane.

What are we seeing here? What can we conclude about our data from this visualization?

To create the viz, let's bring:

- → Category and Segment to Rows;
- → Quantity and Market to Columns.

To make the view clearer drag and drop the **Market** field onto the **Color** tab in the Marks pane.

What are we seeing here? What can we conclude about our data from this visualization?

We are visualizing how sales are looking per category, customer segment and market, in terms of number of items sold. Africa seems to be an emerging market.

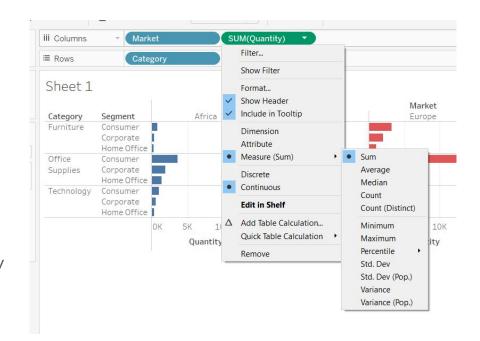
What happened to the Quantity field on the Columns shelf?



What happened to the Quantity field on the Columns shelf?

It displays the SUM of items sold per category, customer segment and market. TP automatically aggregates data with the supposed best measure.

Of course, you can change it as you need, either by selecting a predefined measure or by editing the formula "on shelf" (**Edit on shelf**).



Building views

Now suppose we want to analyze <u>sales over time</u> in our **Global Superstore Orders 2016** dataset.

Which measures and dimensions do we need?

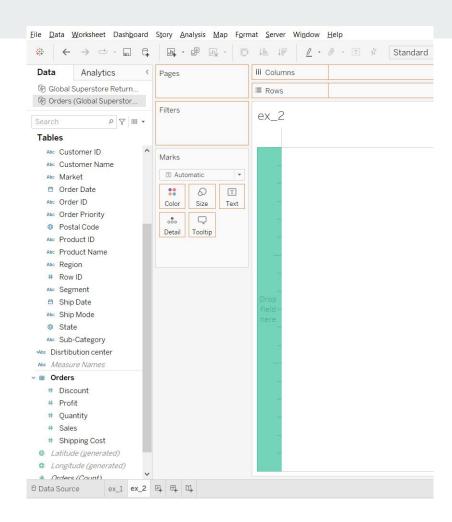
Building views

Now suppose we want to analyze <u>sales over time</u> in our **Global Superstore Orders 2016** dataset.

Which measures and dimensions do we need?

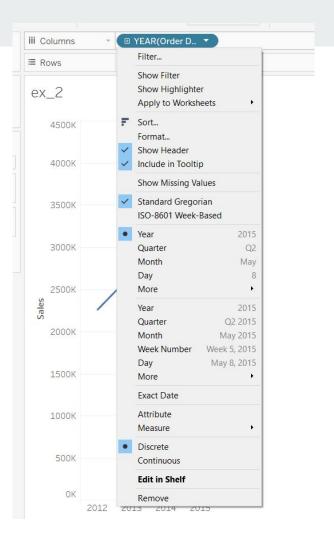
- **Sales** in Rows;
- Order date in Columns.

A little trick: you can drag and drop measures and dimensions directly onto the columns/rows areas in the view instead of onto the shelves. You can drop fields onto any areas surrounded with an orange border



TP automatically aggregates sales by YEAR (Order date), but as with measures we can change that.
But before doing so ...

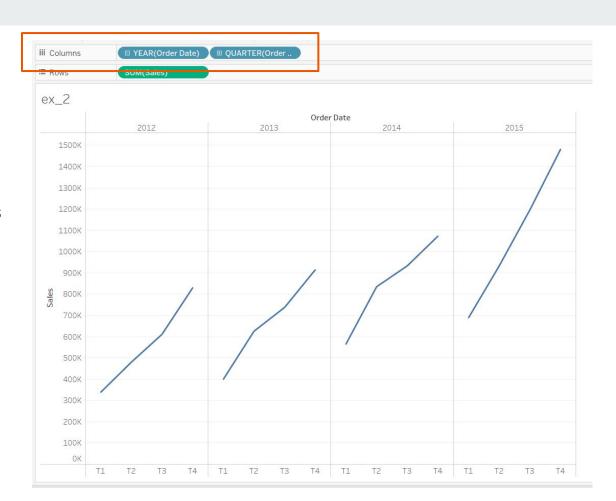
Do you notice anything about the Order date pill on the Columns shelf?



Suppose we want to visualize sales per quarters instead of years.

We can expand the YEAR (Order date) pill with the plus (+) symbol.

What if we wanted to compare sales in the same quarter over different years?



What if we wanted to compare sales in the same quarter over different years?

Simply swap the QUARTER (Order date) and YEAR (Order date) pills.



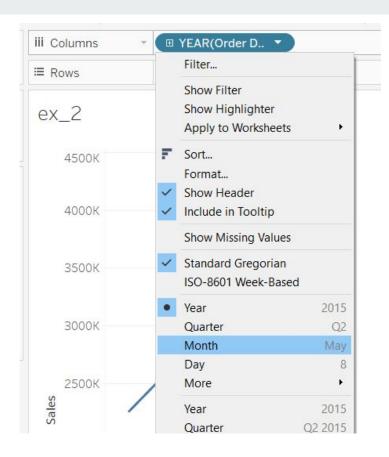
Now we want to analyze <u>sales over MONTHS</u> and compare YEARS in the same chart.

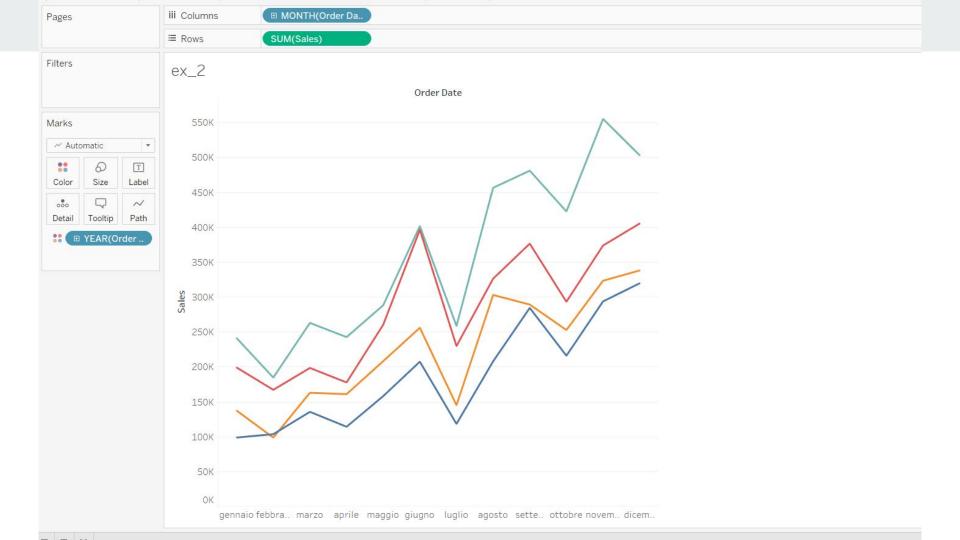
How can we do that? (Spoiler:

(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(Spoiler:
(

How can we do that?

First, change the aggregation level to MONTH and then we can drag and drop the **Order date** field from the Data pane onto the **Color** tab in the Marks pane.





Quick table calculations

Up until now, we have only displayed measures and dimensions that were <u>already present</u> as column headers in the imported dataset. Suppose we want to visualize <u>year over year growth</u> in sales.

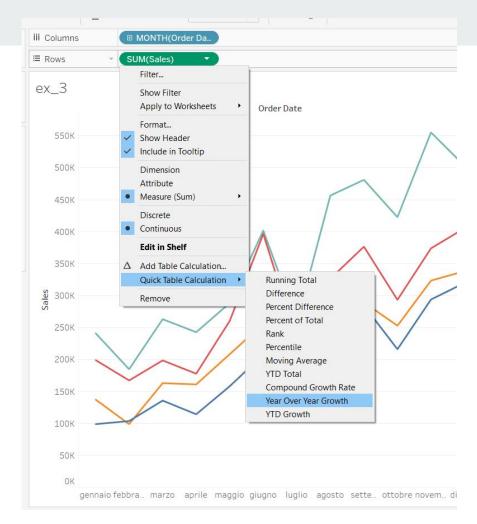
How can we compute that?

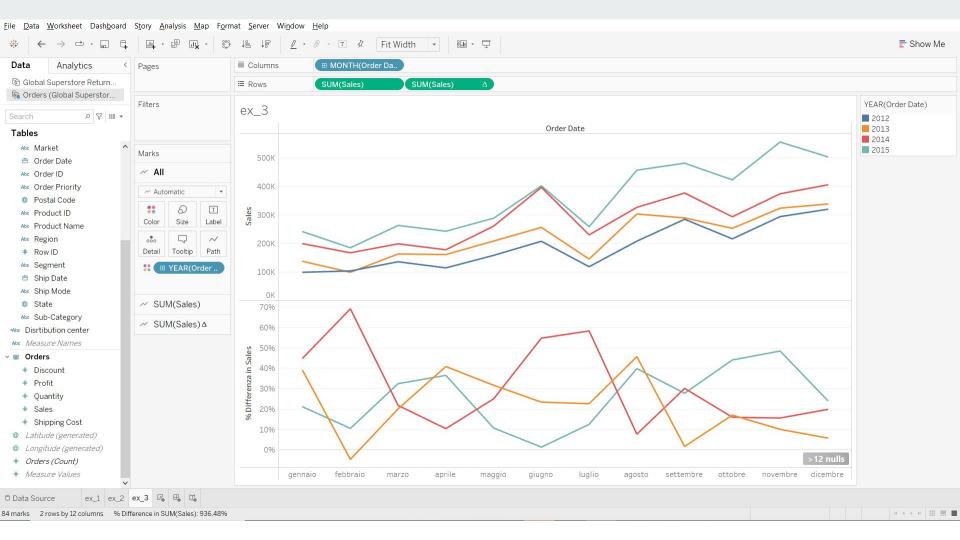
Quick table calculations

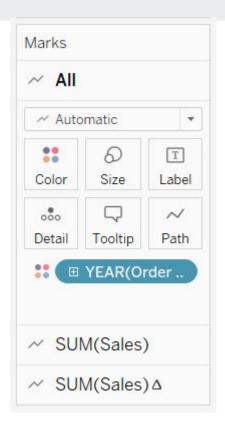
The **Quick table calculation** option allows you to add automatically computed measures to your visualization.

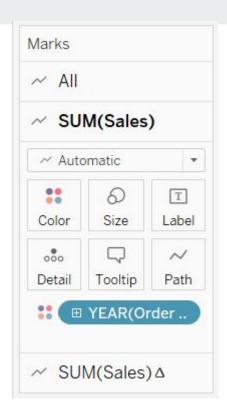
You won't need to work your way out of weird formulas, TP will do the hard work for you

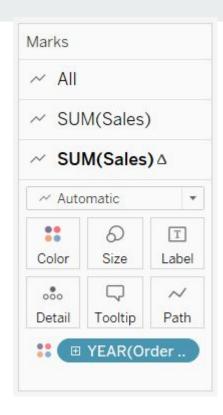
Now, to compare total sales and year over year growth, simply add the **Sales** field to the Rows shelf again.









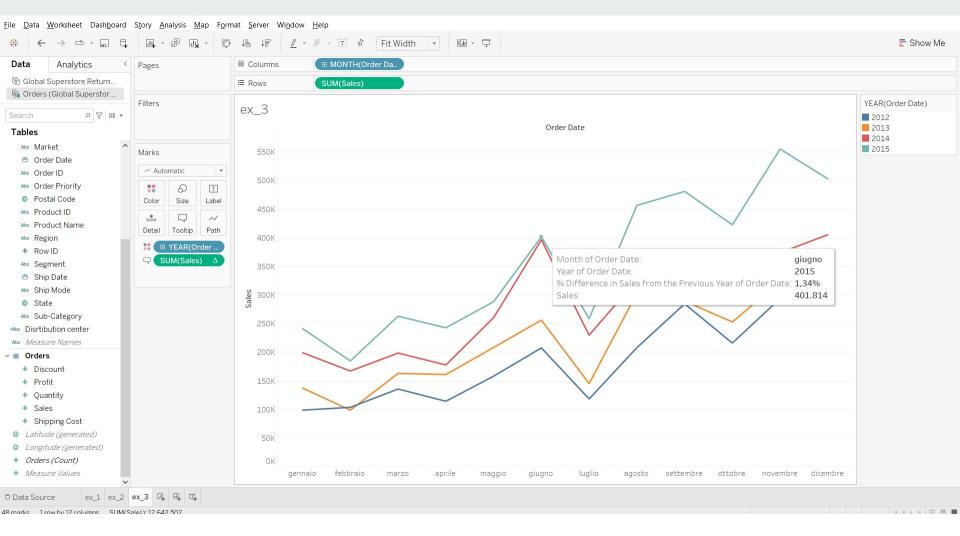


 \bigcirc A little trick: you can customize each measure (hence chart) independently in the Marks pane; in this case, you could adopt different color palettes for the lines depicting total sales and for those depicting year over year growth (SUM (Sales) \triangle)

The Mark Pane

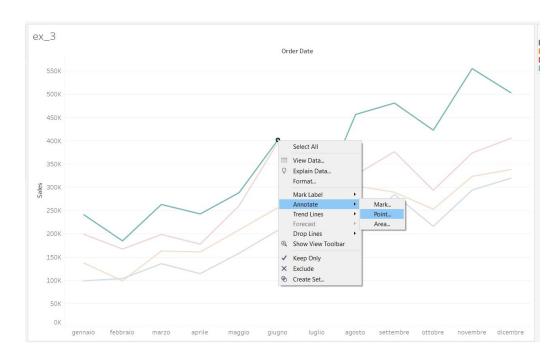
The Mark Pane allows you to customize your view with tooltips, colors and so on.

What if we want to display year over year growth in tooltips instead of having a separate chart? How can we do that?



Annotating data

If you see something you want to point out on your view, simply right click on that point and add an <u>annotation</u> (Annotate > Point).



Visualizing Quick Table Calculation data

Suppose you want to visualize how sales in each category changed over the course of each year.

Right now our chart has:

- 1. **Lines** displaying the sum of sales over all categories (aggregated) by month;
- 2. **Colors** representing the years in which the sales occurred;
- 3. Year over year growth displayed in **tooltips**.

Where do we need to add the Category dimension?

Visualizing Quick Table Calculation data

Suppose you want to visualize how sales in each category changed over the course of each year.

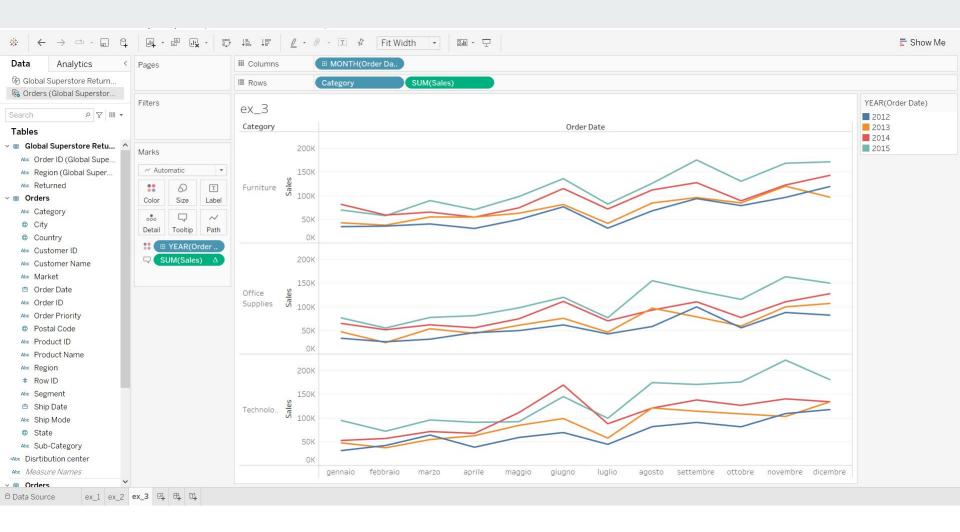
Right now our chart has:

- 1. **Lines** displaying the sum of sales over all categories (aggregated) by month;
- 2. **Colors** representing the years in which the sales occurred;
- 3. Year over year growth displayed in **tooltips**.

Where do we need to add the Category dimension?

Add Category to the Rows shelf.

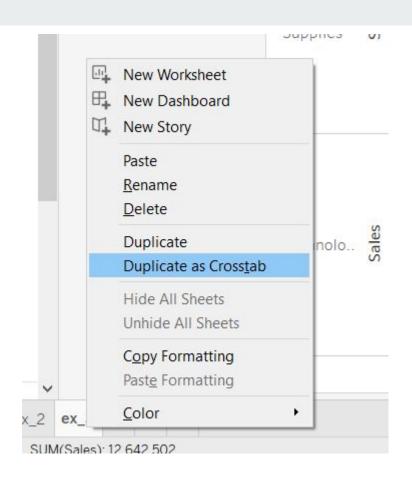
PNote that you can't swap Category and SUM (Sales) as we've done before



Visualizing Quick Table Calculation data

In TP you <u>cannot export data</u>, but you can visualize processed data (aka year over year growth) on a separate sheet.

To do that, simply right click on the **Sheet** tab and **Duplicate as Crosstab**.



| | | | Order Date | | | | | | | | | | | |
|------------|-----------|----------------------|------------|----------|---------|--------|---------|---------|--------|---------|---------|---------|-----------|----------|
| Category | Year of O | | gennaio | febbraio | marzo | aprile | maggio | giugno | luglio | agosto | settemb | ottobre | novembre | dicembre |
| Furniture | 2012 | % Difference in Sale | | | | | | | | | | | | |
| | | Sales | 34.464 | 35.799 | 40.277 | 30.690 | 49.769 | 76.585 | 31.383 | 68.000 | 93.934 | 79.094 | 96.558 | 119.531 |
| | 2013 | % Difference in Sale | 24,08% | 5,13% | 36,76% | 78,24% | 26,46% | 6,39% | 32,21% | 24,48% | 2,38% | 7,24% | 24,57% | -19,00% |
| | | Sales | 42.761 | 37.635 | 55.082 | 54.702 | 62.939 | 81.481 | 41.491 | 84.644 | 96.166 | 84.824 | 120.279 | 96.817 |
| | 2014 | % Difference in Sale | 91,31% | 57,08% | 18,59% | -0,21% | 18,16% | 41,45% | 73,18% | 32,67% | 32,79% | 5,30% | 5 2,10% | 47,91% |
| | | Sales | 81.805 | 59.118 | 65.323 | 54.587 | 74.371 | 115.251 | 71.854 | 112.296 | 127.701 | 89.319 | 122.803 | 143.203 |
| | 2015 | % Difference in Sale | -14,68% | -2,39% | 37,33% | 29,24% | 32,19% | 18,11% | 14,60% | 12,46% | 37,65% | 46,33% | 37,50% | 19,95% |
| | | Sales | 69.799 | 57.703 | 89.705 | 70.551 | 98.312 | 136.123 | 82.344 | 126.284 | 175.777 | 130.701 | 168.849 | 171.768 |
| Office | 2012 | % Difference in Sale | | | | | | | | | | | | |
| Supplies | | Sales | 33.527 | 26.135 | 31.579 | 45.563 | 49.731 | 61.793 | 42.807 | 58.390 | 99.987 | 55.700 | 080.88 | 82.424 |
| | 2013 | % Difference in Sale | 40,97% | -7,03% | 70,12% | -3,68% | 22,79% | 22,74% | 8,27% | 66,57% | -21,15% | 6,61% | 13,56% | 30,11% |
| | | Sales | 47.264 | 24.297 | 53.721 | 43.886 | 61.063 | 75.846 | 46.346 | 97.260 | 78.844 | 59.383 | 3 100.020 | 107.244 |
| | 2014 | % Difference in Sale | 37,49% | 112,17% | 15,43% | 27,07% | 22,42% | 46,99% | 51,66% | -4,23% | 40,48% | 30,33% | 10,71% | 19,28% |
| | | Sales | 64.984 | 51.553 | 62.008 | 55.766 | 74.756 | 111.489 | 70.291 | 93.150 | 110.762 | 77.393 | 3 110.736 | 127.926 |
| | 2015 | % Difference in Sale | 17,96% | 7,04% | 25,11% | 45,92% | 31,06% | 7,94% | 9,50% | 66,86% | 21,33% | 49,59% | 47,97% | 17,47% |
| | | Sales | 76.654 | 55.184 | 77.576 | 81.372 | 97.975 | 120.340 | 76.971 | 155.431 | 134.387 | 115.771 | 163.851 | 150.279 |
| Technology | 2012 | % Difference in Sale | | | | | | | | | | | | |
| | | Sales | 30.908 | 41.784 | 63.891 | 38.081 | 58.728 | 69.194 | 44.245 | 81.673 | 90.667 | 81.320 | 109.309 | 117.852 |
| | 2013 | % Difference in Sale | 53,39% | -11,63% | -15,05% | 64,03% | 43,65% | 42,86% | 29,73% | 48,44% | 26,15% | 33,71% | 5 -5,58% | 13,87% |
| | | Sales | 47.411 | 36.922 | 54.273 | 62.464 | 84.363 | 98.849 | 57.399 | 121.239 | 114.379 | 108.733 | 3 103.214 | 134.196 |
| | 2014 | % Difference in Sale | 10,52% | 53,21% | 31,30% | 8,01% | 32,02% | 71,76% | 52,94% | -0,16% | 20,79% | 16,52% | 36,08% | 0,10% |
| | | Sales | 52.398 | 56.569 | 71.263 | 67.469 | 111.372 | 169.780 | 87.784 | 121.043 | 138.157 | 126.695 | 140.451 | 134.325 |
| | 2015 | % Difference in Sale | 80,95% | 27,19% | 34,46% | 34,65% | -17,29% | -14,39% | 13,22% | 44,50% | | 39,15% | 58,47% | 34,82% |
| | | Sales | 94.815 | 71.951 | 95.819 | 90.849 | 92.114 | 145.351 | 99.390 | 174.905 | 170.994 | 176.295 | 222.579 | 181.097 |



g Swap dimensions to make the view a bit clearer

15 minute break

Answer the Wooclap when you come back:



Up next:

- → 10 minutes to answer the Woodlap
- → Lessons learned and recap
- → 15-minute Q&A at the end of the lesson

Play around with what we've seen today with different data sources (or start working on your projects)

Additional free data sources are available here

Recap and lessons learned

Recap

- → Data import:
 - Relationships and joins;
 - Data formatting.
- → Data visualization:
 - Measures and dimensions;
 - Expanding/aggregating dimensions;
 - Quick table calculations;
 - ♦ Intro to the Mark Pane;
 - Crosstab duplication.

Takeaway

- → Always double check for the relations between sheets/tables that Tableau automatically creates;
- → Relationships **do not merge** data, joins do;
- → Measures and dimensions' order matter (but now always) in the Columns/Rows shelves;

Link to the solution workbook for the lesson: PTBD