

PENS Project Learning Analytics

Ariadne Tsiakara
Computer Science Department
Faculty of Science and Technology
Middlesex University

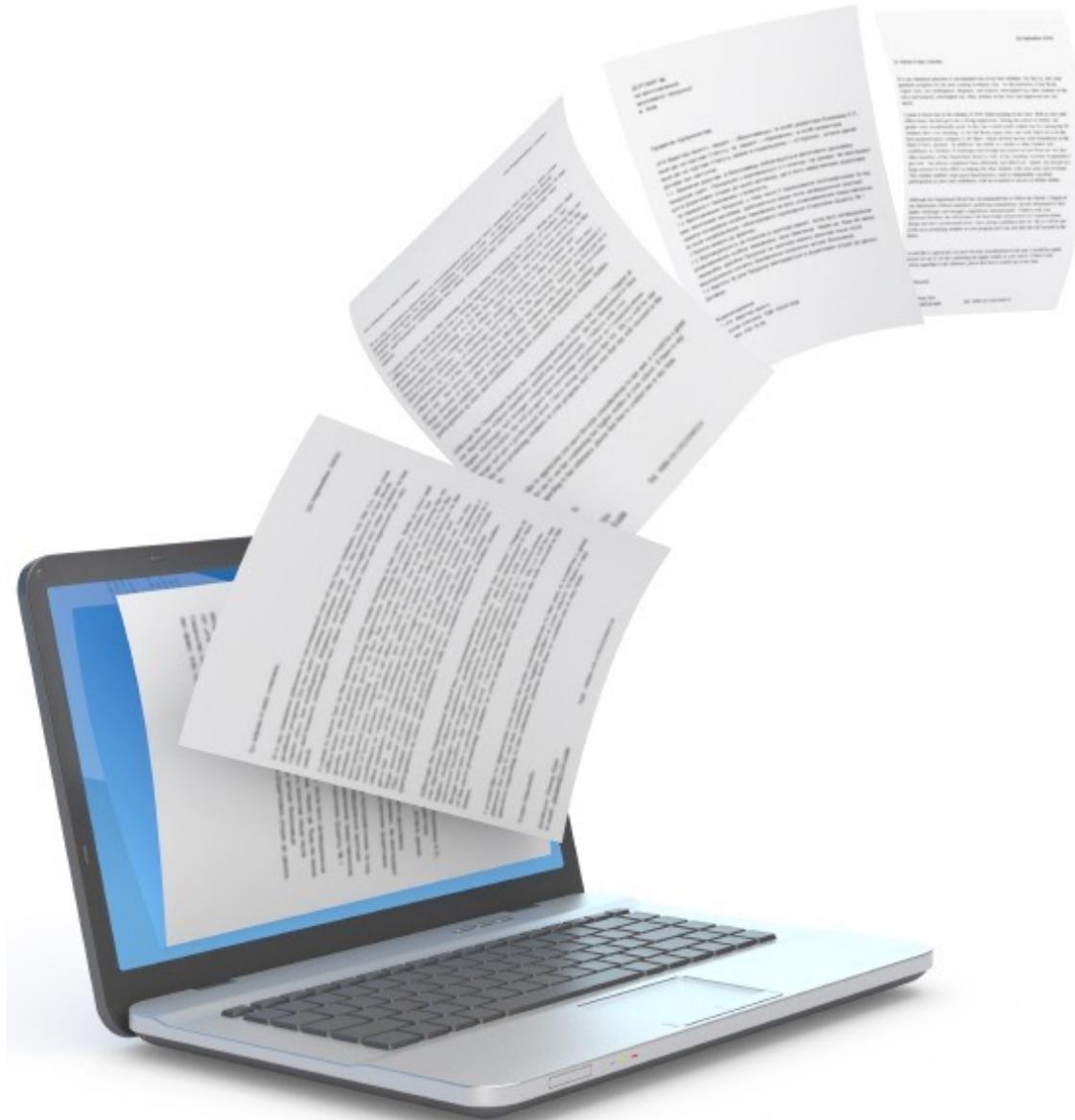
The training briefly describes aspects of the Tableau platform and how visual analytics can be used.

Emphasis is given on sharing good practice of how learning analytics are used to produce dashboards representing emerging patterns of behaviour, and activity.

Outline

- The need for visual analytics
- Using Tableau
- Examples of learning analytics

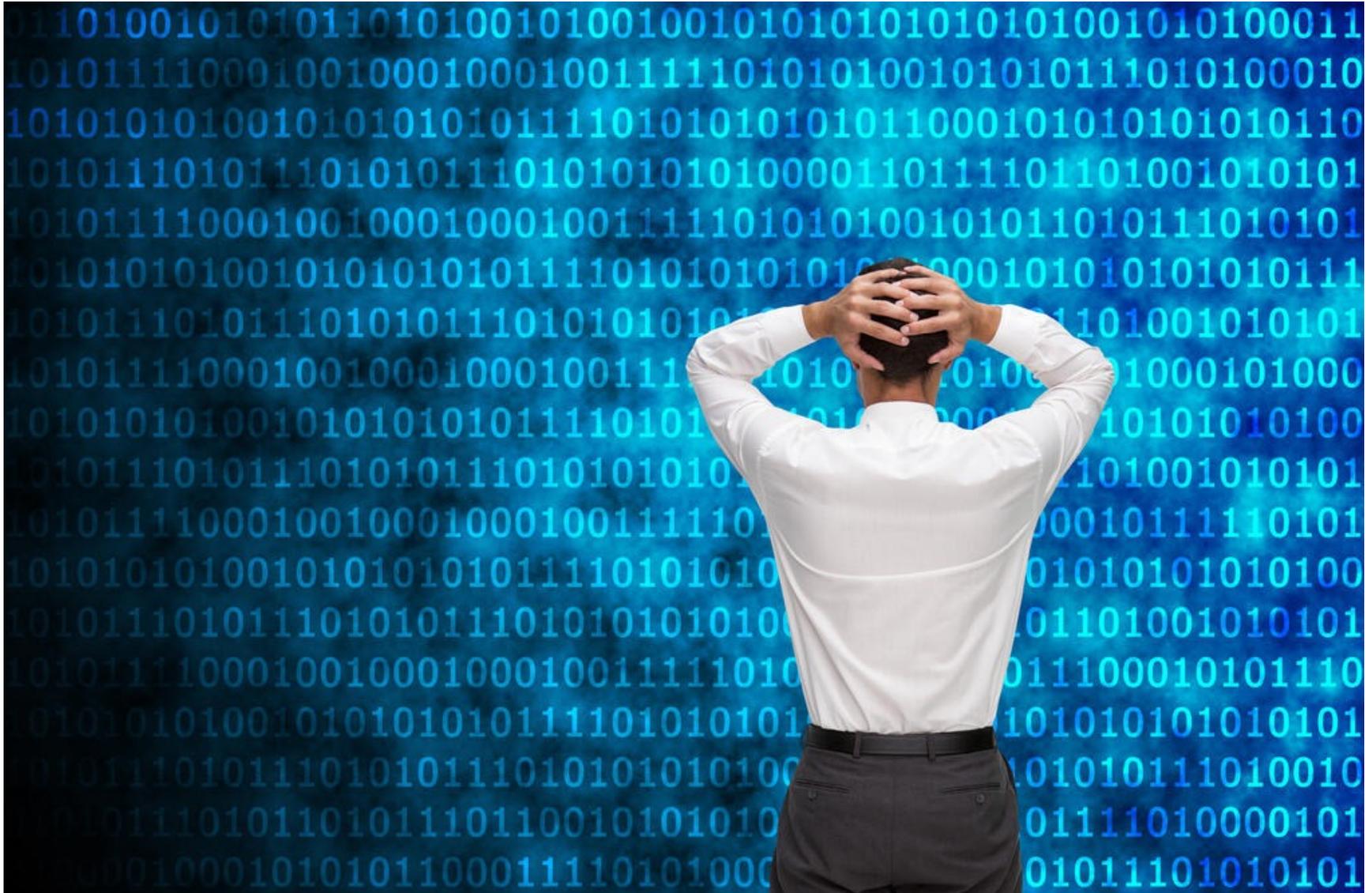
From papers to digital



Drowning in Data



Too much of Data



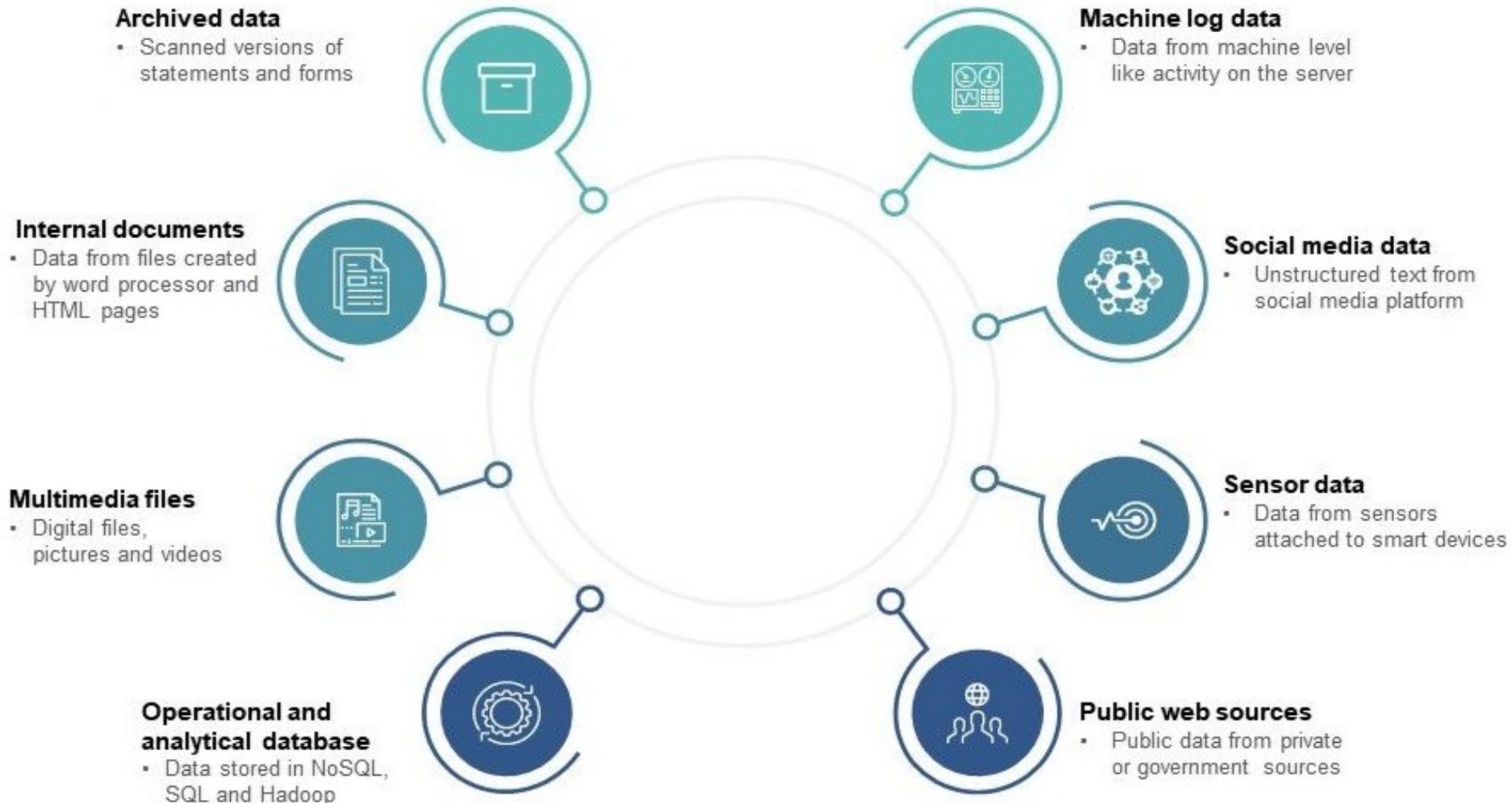
Traffic of Data



Social networks



Major Data Sources for Bigdata



Light in the end of tunnel



What to do with Data



Data Storage



Cloud Storage

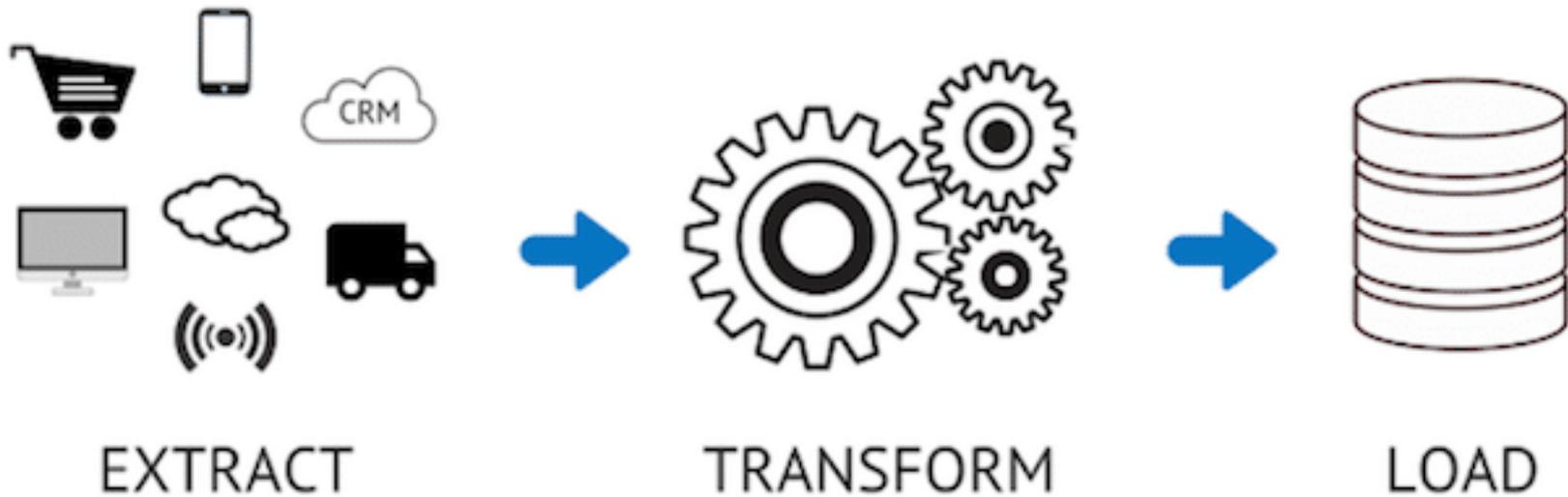




DATA ANALYSIS

- **Data analysis** is a process of inspecting, cleansing, transforming and modelling the data with the goal of discovering useful information, informing conclusion and supporting decision-making.
- Data analysis has multiple facets and approaches, including different techniques.
- **Data mining** is a particular data analysis technique that focuses on statistical modelling and knowledge discovery for predictive rather than purely descriptive purposes, while **Business Intelligence (BI)** covers data analysis that relies heavily on aggregation, focusing mainly on business information.
- **Data integration** is a precursor to data analysis and data analysis is closely linked to Data Visualization and Data Dissemination.

ETL Process



What is ETL Process

- **ETL** stands for **Extract, Transform, Load** and refers to the process of transferring data from one location to another. In addition to migrating data from one database to another, it also converts (transforms) databases into a single format that can be utilized in the final destination.
- The ETL Process became a popular concept in the 1970s and is often used in Data Warehousing.
- The ETL Process is preparing data for Analysis and Visualizations.

ETL



Data Cleaning



Data Cleansing (or Data Scrubbing) is the action of identifying and then removing or amending any data within a database that is:

- Incorrect
- Incomplete
- Duplicated

And under the [GDPR](#):

- Irrelevant
- Unnecessary

DATA CLEANING CHECKLIST

Up-to-date data



Data should be up-to-date in order to obtain maximum value from the data analysis.



Missing values



Count missing values and analyze where in the data they are missing. Missing values can disrupt some analyses and skew the results.



Duplicates



Duplicate IDs indicate multiple records for one person, e.g. someone holds multiple functions at the same time.



Numerical outliers



Numerical outliers are fairly easy to detect and remove. Define minimum and maximum to spot outliers easily.



Check IDs



Check data labels of all the fields to see whether some categorical values are mislabeled.



Define valid output



Define valid data labels for categorical data. Define data ranges for numerical variables. Non-matching data is presumably wrong.



DATA VISUALIZATION



Why is data visualization important

Because of the way the human brain processes information, using charts or graphs to visualize large amounts of complex data is easier than poring over spreadsheets or reports. Data visualization is a quick, easy way to convey concepts in a universal manner – and you can experiment with different scenarios by making slight adjustments.

Data visualization can also:

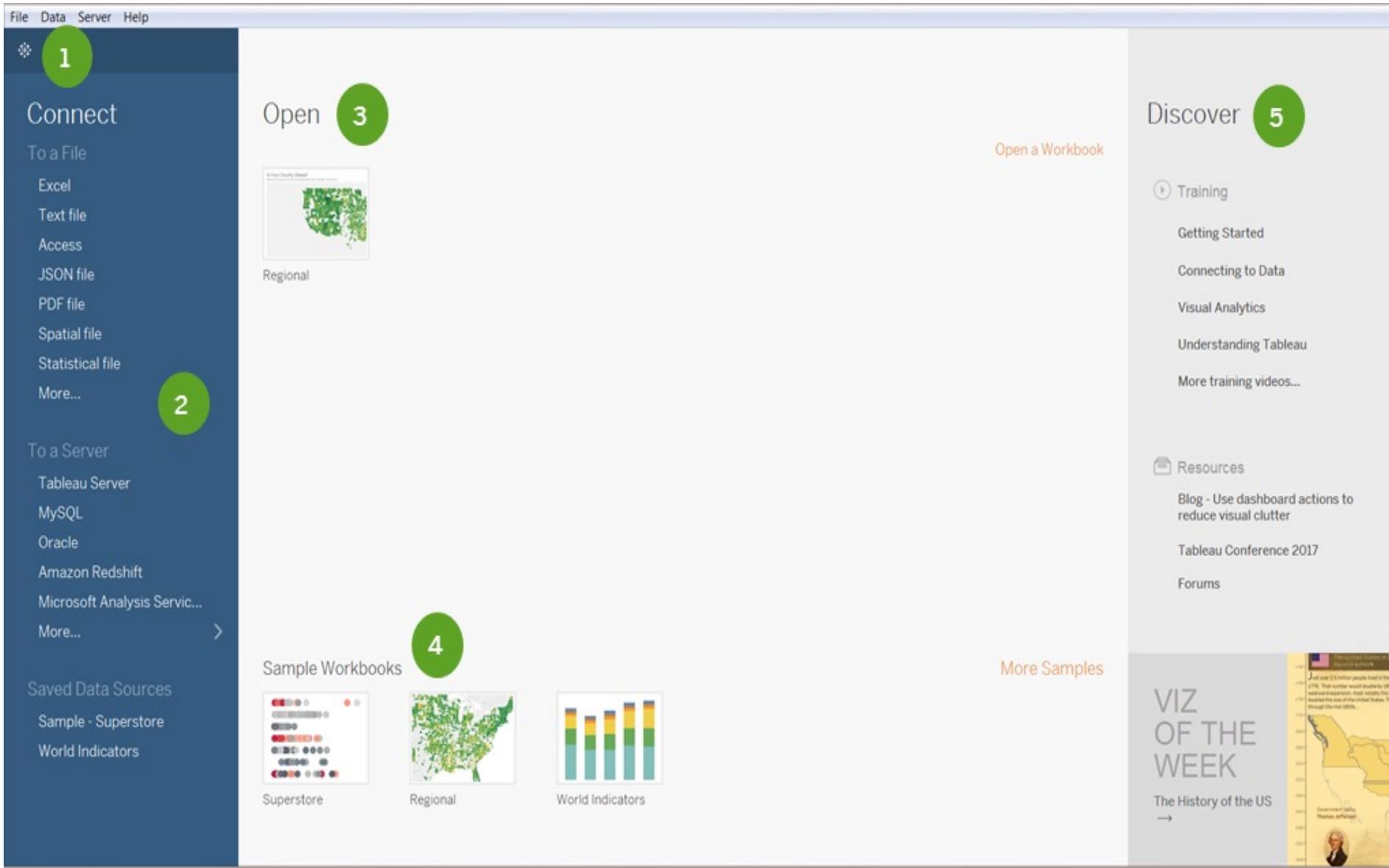
- Identify areas that need attention or improvement.
- Clarify which factors influence learner behaviour.
- Help you understand which learning activities to place where.
- Predict learning activity and behavioural patterns.

Why Tableau

- Tableau is a very effective tool to create interactive data visualizations very quickly. It is very simple and user-friendly.
- Tableau can create complex graphs giving a similar feel as the pivot table graphs in Excel. Moreover, it can handle a lot more data and quickly provide calculations on datasets.
- Users can create visualizations quickly and switch between types easily to find the model that best represents the message.
- It is extremely easy to integrate with multiple data sources and the user interface is well-organized.
- It can create visualizations for a large amount of data without crashing.

Introduction to Tableau

Tableau Home page



The screenshot shows the Tableau Home page interface with the following elements and callouts:

- 1**: A gear icon in the top-left corner of the sidebar.
- 2**: A circular callout next to the "More..." link in the "To a Server" section of the sidebar.
- 3**: A circular callout next to the "Open" button in the main content area.
- 4**: A circular callout next to the "Sample Workbooks" section in the main content area.
- 5**: A circular callout next to the "Discover" button in the right-hand sidebar.

The interface includes a top menu bar with "File", "Data", "Server", and "Help". The left sidebar contains sections for "Connect" (To a File, To a Server) and "Saved Data Sources". The main content area features "Open", "Sample Workbooks", and "Open a Workbook" options. The right sidebar contains "Discover" options for "Training" and "Resources". A "VIZ OF THE WEEK" banner is visible at the bottom right.

Tableau's Connections

Search

Action Matrix

Action Vector

Amazon Athena

Amazon Aurora

Amazon EMR Hadoop Hive

Amazon Redshift

Anaplan

Apache Drill

Aster Database

Azure SQL Data Warehouse

Box

Cloudera Hadoop

Databricks

Denodo

Dropbox

Exasol

Firebird

Google Ads

Google Analytics

Google BigQuery

Google Cloud SQL

Google Drive

Google Sheets

Hortonworks Hadoop Hive

IBM BigInsights

IBM DB2

IBM PDA (Netezza)

Intuit QuickBooks Online

Intuit QuickBooks Online (9.3-2018.1)

Kognitio

MapR Hadoop Hive

MariaDB

Marketo

MarkLogic

MemSQL

Microsoft Analysis Services

Microsoft PowerPivot

Microsoft SQL Server

MonetDB

MongoDB BI Connector

MySQL

OData

OneDrive

Oracle

Oracle Eloqua

Oracle Essbase

Pivotal Greenplum Database

PostgreSQL

Presto

Progress OpenEdge

Salesforce

SAP HANA

SAP NetWeaver Business Warehouse

SAP Sybase ASE

SAP Sybase IQ

ServiceNow ITSM

SharePoint Lists

Snowflake

Spark SQL

Splunk

Teradata

Teradata OLAP Connector

TIBCO Data Virtualisation

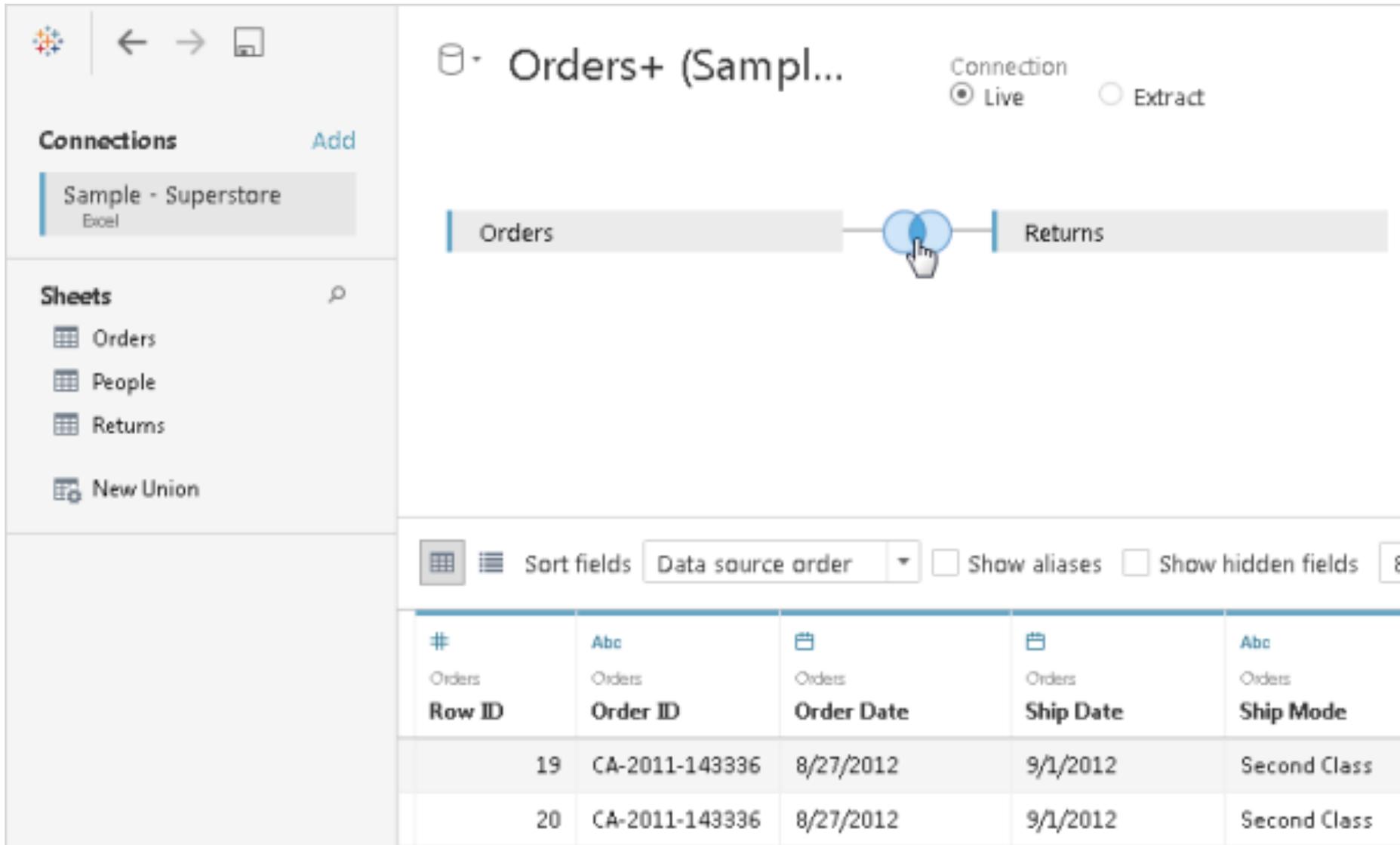
Vertica

Web Data Connector

Other Databases (JDBC)

Other Databases (ODBC)

Tableau Joins



The screenshot shows the Tableau interface for a data source named "Orders+ (Sampl...". The connection is set to "Live". In the center, two data sources, "Orders" and "Returns", are connected by a double-headed arrow, indicating a join. A hand cursor is pointing at the join symbol. On the left, the "Connections" pane shows "Sample - Superstore" (Excel) and the "Sheets" pane lists "Orders", "People", "Returns", and "New Union". At the bottom, there are controls for "Sort fields" (set to "Data source order"), "Show aliases", and "Show hidden fields". A data table is displayed at the bottom with columns: Row ID, Order ID, Order Date, Ship Date, and Ship Mode.

#	Orders	Orders	Orders	Orders
Row ID	Order ID	Order Date	Ship Date	Ship Mode
19	CA-2011-143336	8/27/2012	9/1/2012	Second Class
20	CA-2011-143336	8/27/2012	9/1/2012	Second Class

Tableau Joins

Patron



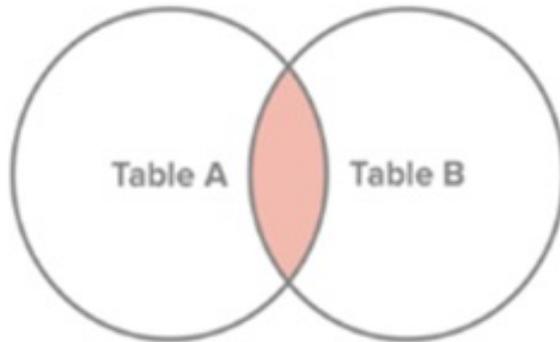
Contact

Join ✕

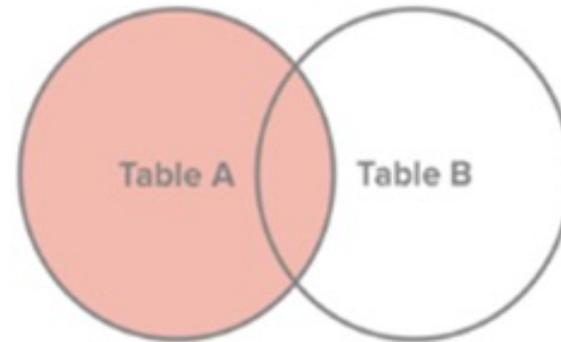
 Inner	 Left	 Right	 Full Outer
Data Source		Contact	
[First name]+ " "+[Last ...	=	Name	
<i>Add new join clause</i>			

SQL Joins

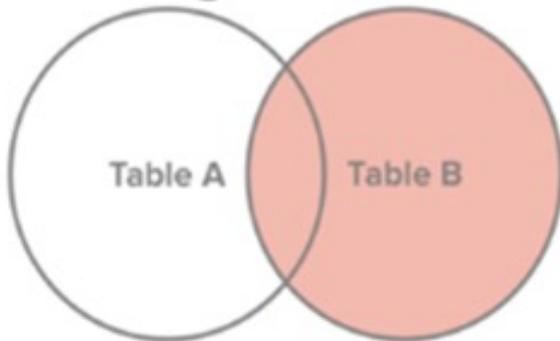
Inner Join



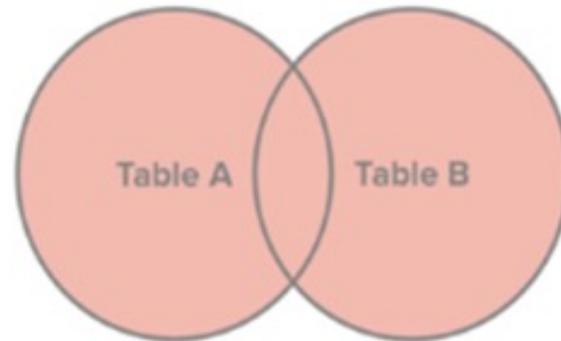
Left Join



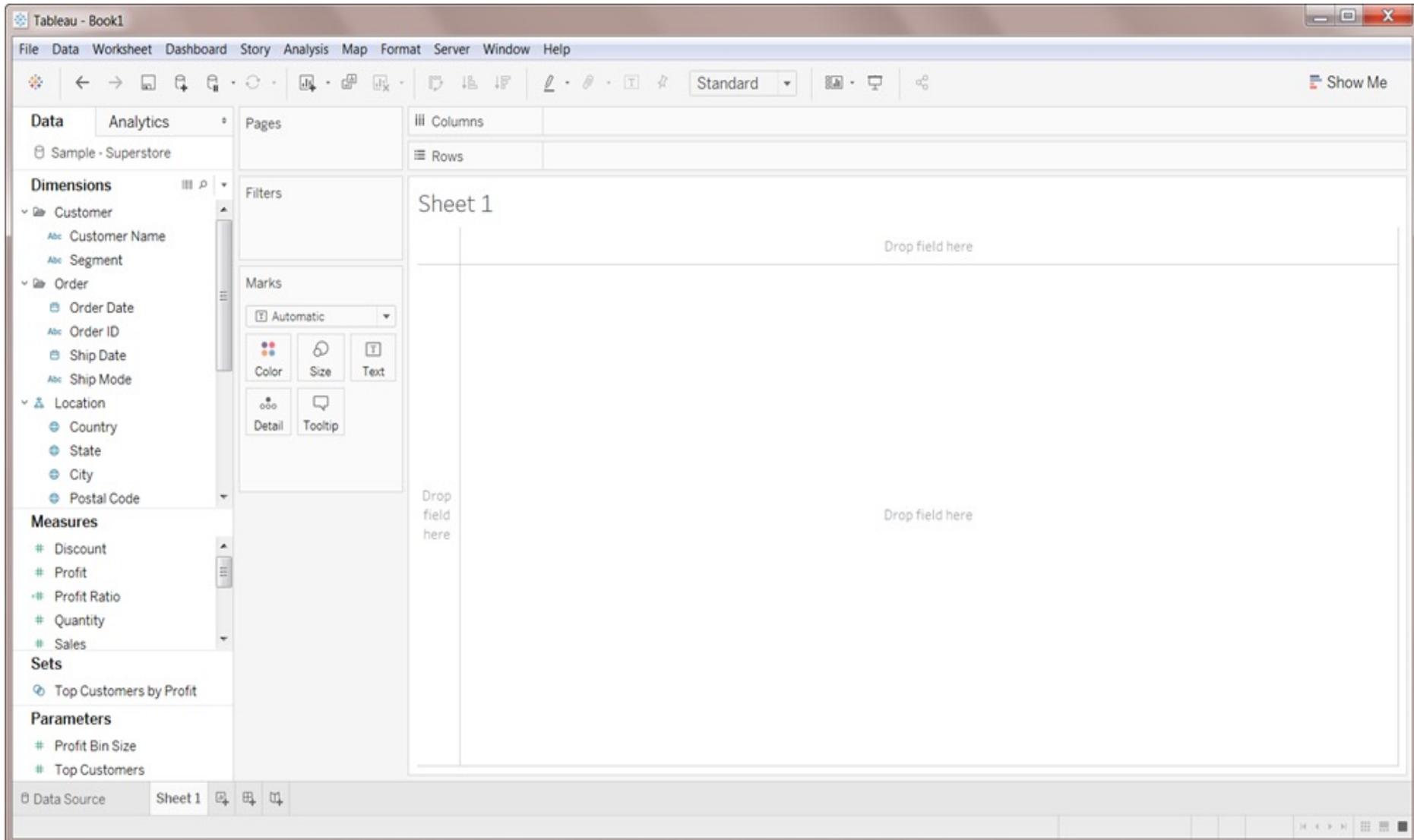
Right Join



Full Join



Dimensions & Measures



The screenshot displays the Tableau Desktop interface for a workbook named "Tableau - Book1". The main workspace is titled "Sheet 1" and is currently empty, with "Drop field here" prompts in the columns and rows areas. The left-hand pane is organized into several sections:

- Data:** Shows the data source as "Sample - Superstore".
- Dimensions:** A list of fields categorized by dimension:
 - Customer: Customer Name, Segment
 - Order: Order Date, Order ID, Ship Date, Ship Mode
 - Location: Country, State, City, Postal Code
- Measures:** A list of fields categorized by measure:
 - Discount, Profit, Profit Ratio, Quantity, Sales
- Sets:** Includes "Top Customers by Profit".
- Parameters:** Includes "Profit Bin Size" and "Top Customers".

The central "Marks" card is set to "Automatic" and includes options for Color, Size, Text, Detail, and Tooltip. The top menu bar includes File, Data, Worksheet, Dashboard, Story, Analysis, Map, Format, Server, Window, and Help. The top toolbar contains navigation and editing tools, and the bottom status bar shows the current data source and sheet name.

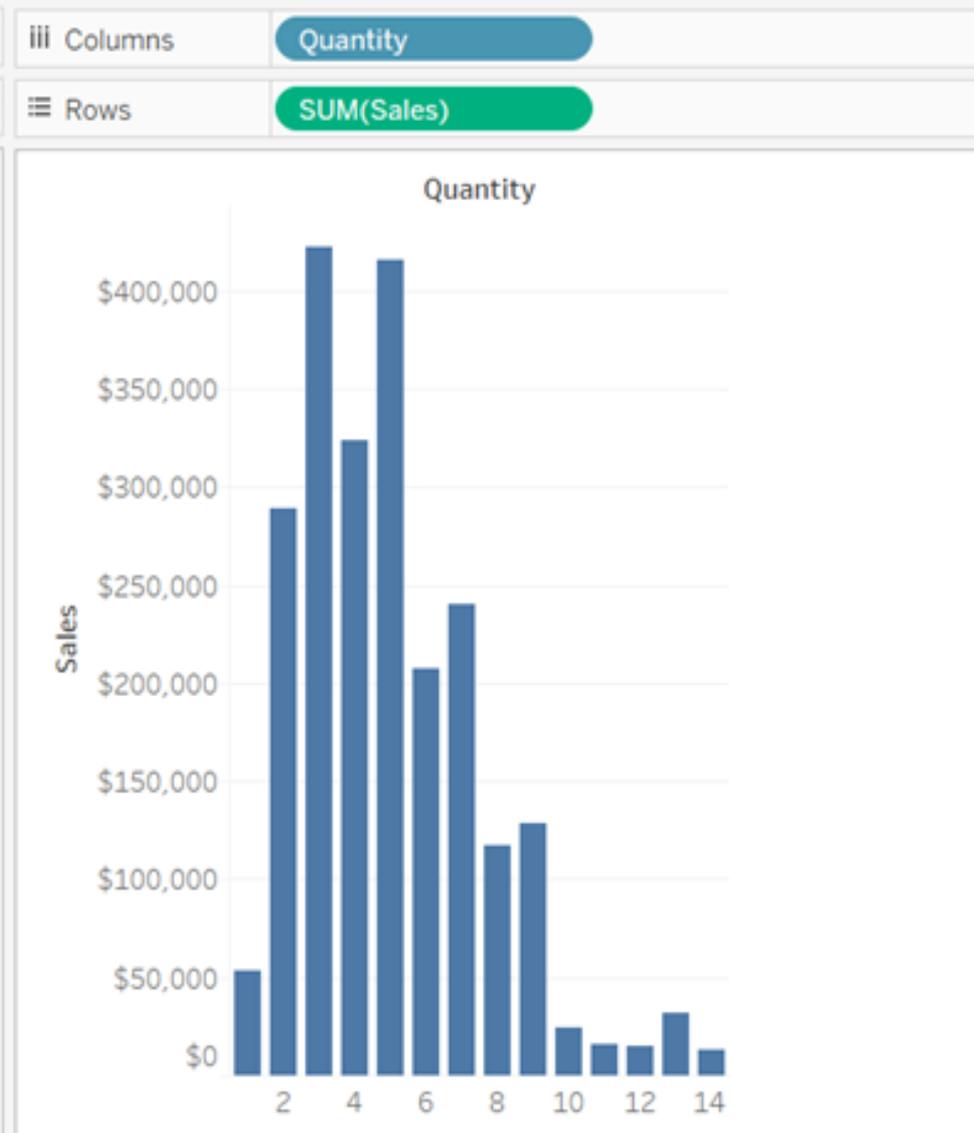
- **Dimensions** contain qualitative values (such as names, dates, or geographical data). Dimensions are used to categorize, segment, and reveal the details in our data. Dimensions affect the level of detail in the view.
- **Measures** contain numeric, quantitative values that you can measure. Measures can be aggregated. By dragging a measure into the view, Tableau by default applies an aggregation to that measure.

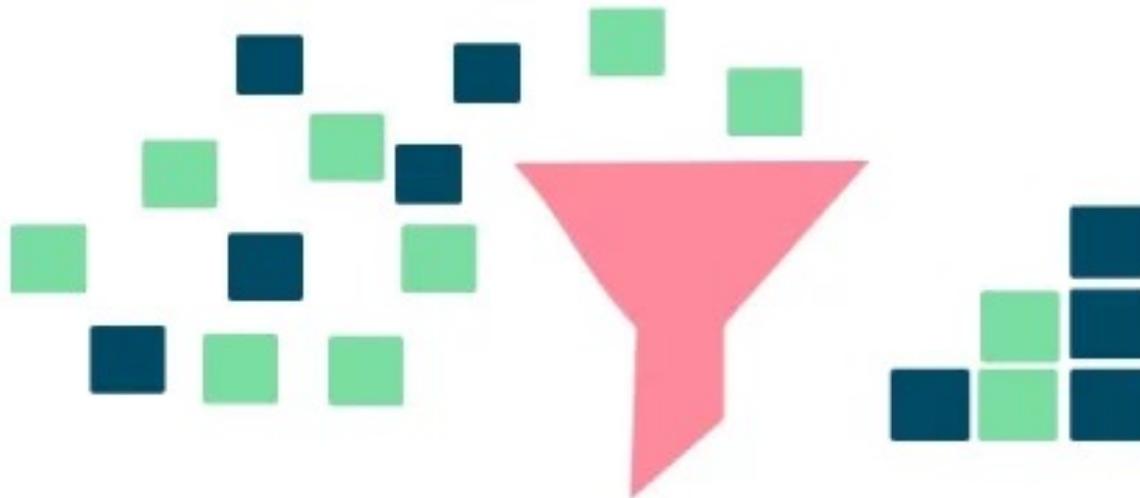
Tableau represents data differently in the view depending on whether the field is discrete (**blue**), or continuous (**green**).

Continuous and discrete are mathematical terms.

- **Continuous** means "forming an unbroken whole, without interruption".
- **Discrete** means "individually separate and distinct."

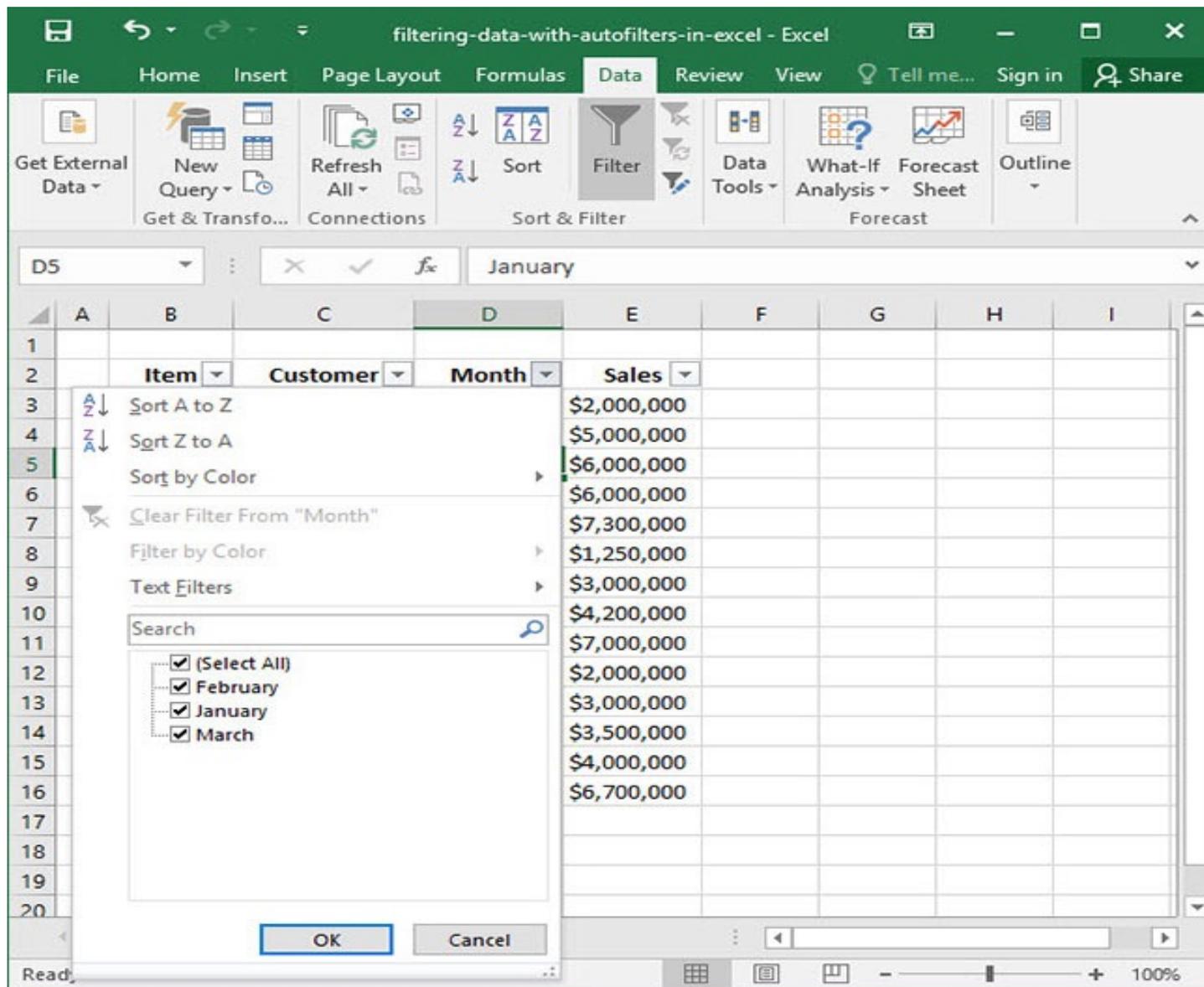
Continuous vs Discrete





FILTERING DATA

Data Filtering



The screenshot shows the Microsoft Excel interface with the 'Data' tab selected. The 'Filter' dropdown menu is open for the 'Month' column, displaying various options for sorting and filtering. The spreadsheet data is visible in the background, showing columns for Item, Customer, Month, and Sales.

Item	Customer	Month	Sales
		January	\$2,000,000
		January	\$5,000,000
		January	\$6,000,000
		January	\$6,000,000
		January	\$7,300,000
		January	\$1,250,000
		January	\$3,000,000
		January	\$4,200,000
		January	\$7,000,000
		February	\$2,000,000
		February	\$3,000,000
		January	\$3,500,000
		January	\$4,000,000
		January	\$6,700,000

Data Filtering

Contact [Add New Contact](#) [Hide Search Segment](#)

Contact Group *in group* Wolf-Runolfsson

Or

Tags *contains* Dokan

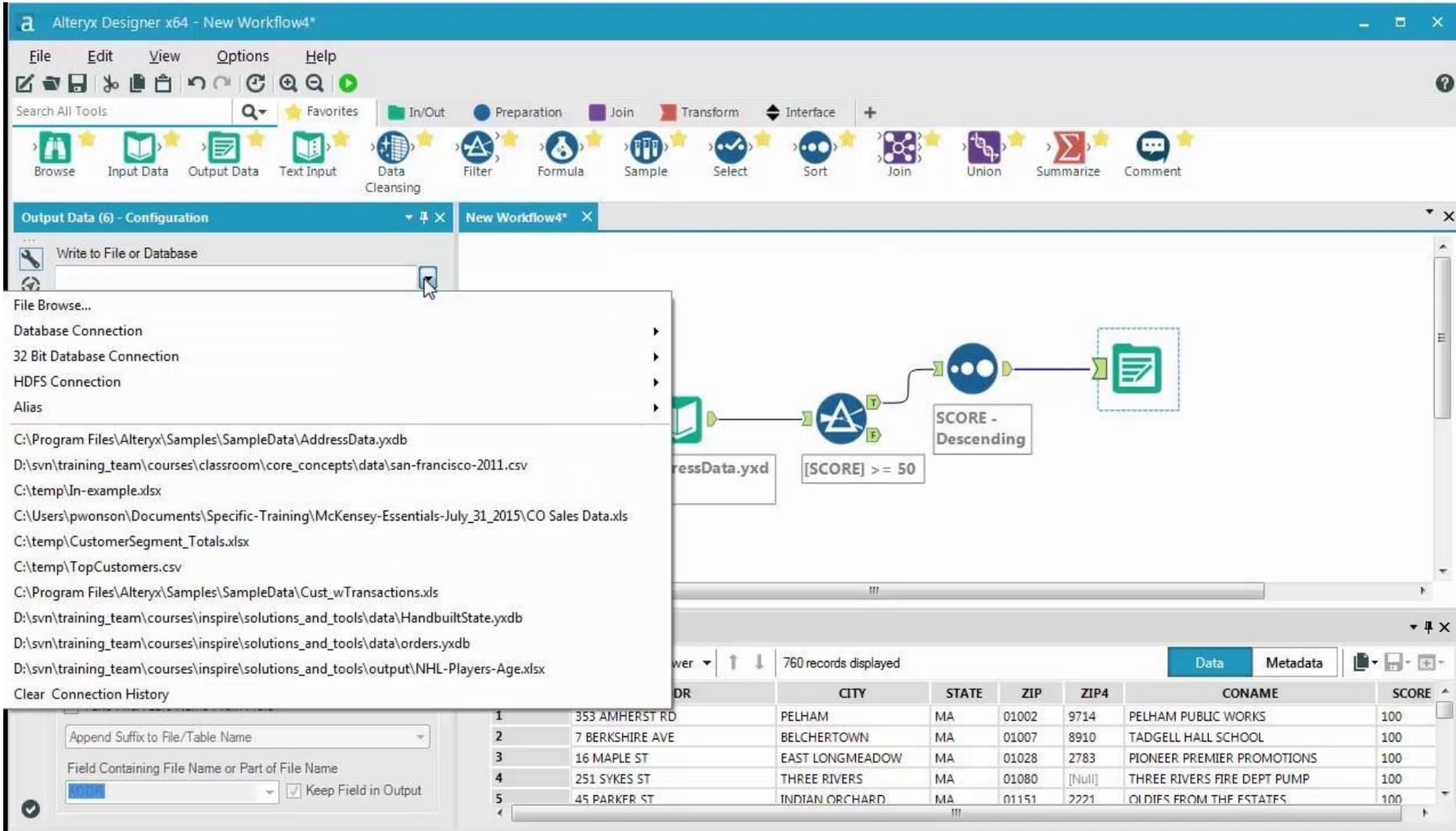
All (45) | [Customer](#) (17) | [Lead](#) (7) | [Opportunity](#) (13) | [Subscriber](#) (8) | [Trash](#) (0)

Bulk Actions [Shawon Chowdhury](#) Filter by Segment 1 item

<input type="checkbox"/>	Contact name	Email Address	Phone	Life stage	Owner	Created At
<input type="checkbox"/>	 Mertie Brown	bernhard.kassulke@harris.com	986.817.8244	Subscriber	Shawon Chowdhury	03/22/2018

Bulk Actions 1 item

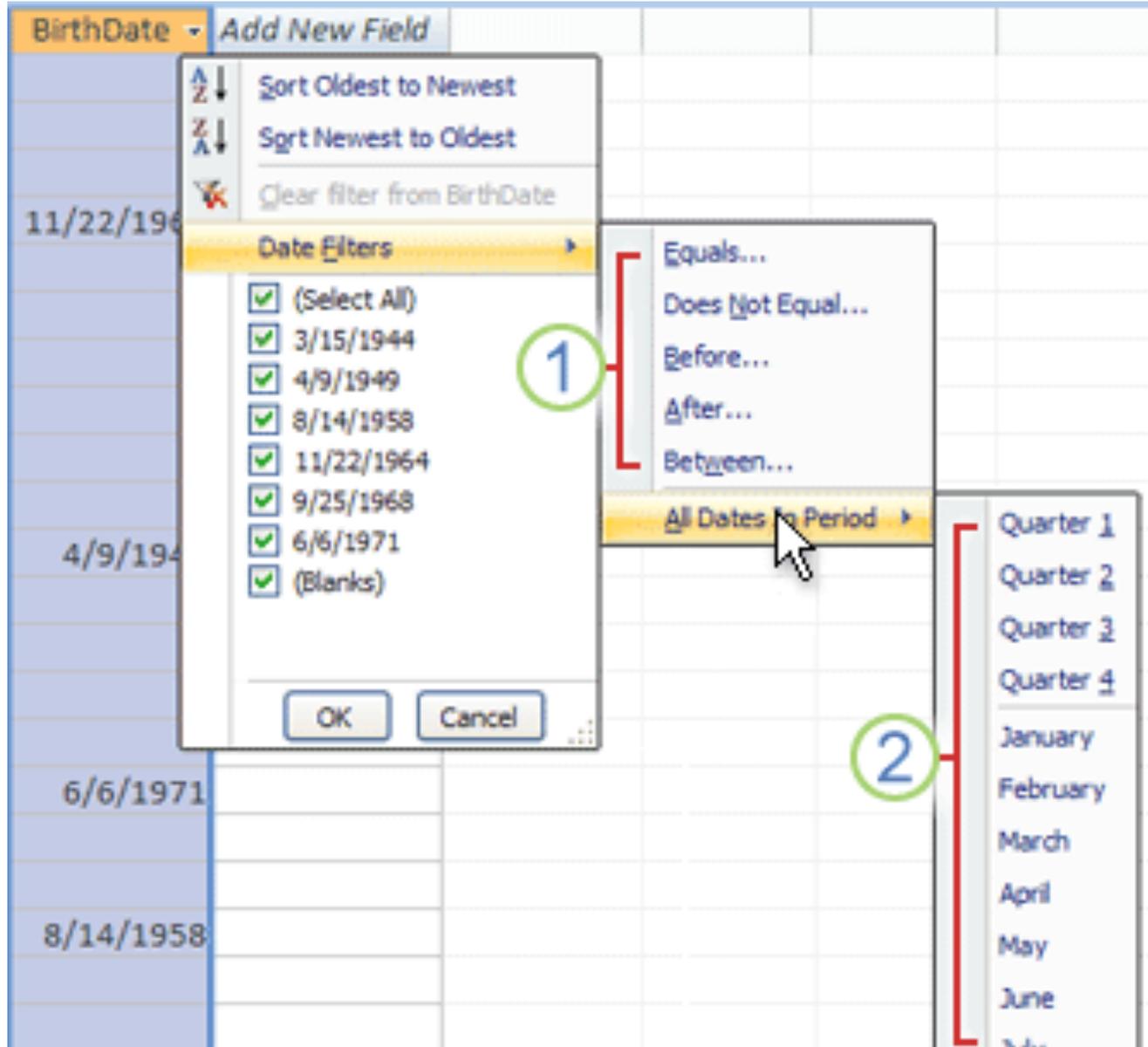
Data Filtering



The screenshot displays the Alteryx Designer x64 interface. The main window shows a workflow with the following steps: **Input Data** (green icon) -> **Filter** (blue icon) -> **Sort** (blue icon) -> **Output Data** (green icon). The **Filter** tool is configured with the expression `[SCORE] >= 50`. The **Sort** tool is configured with `SCORE - Descending`. A **File Browse...** dialog box is open over the **Input Data** tool, listing various file paths. The **Output Data** tool is currently selected, and its configuration panel is visible, showing options for **Write to File or Database**. At the bottom of the interface, a data table is displayed with 760 records. The table has columns: **IDR**, **CITY**, **STATE**, **ZIP**, **ZIP4**, **CONAME**, and **SCORE**. The first five rows of data are shown below.

IDR	CITY	STATE	ZIP	ZIP4	CONAME	SCORE	
1	353 AMHERST RD	PELHAM	MA	01002	9714	PELHAM PUBLIC WORKS	100
2	7 BERKSHIRE AVE	BELCHERTOWN	MA	01007	8910	TADGELL HALL SCHOOL	100
3	16 MAPLE ST	EAST LONGMEADOW	MA	01028	2783	PIONEER PREMIER PROMOTIONS	100
4	251 SYKES ST	THREE RIVERS	MA	01080	[Null]	THREE RIVERS FIRE DEPT PUMP	100
5	45 PARKER ST	INDIAN ORCHARD	MA	01151	2221	OLD DIES FROM THE PSTATES	100

Data Filtering



BirthDate ▾ Add New Field

- Sort Oldest to Newest
- Sort Newest to Oldest
- Clear filter from BirthDate

Date Filters ▸

- (Select All)
- 3/15/1944
- 4/9/1949
- 8/14/1958
- 11/22/1964
- 9/25/1968
- 6/6/1971
- (Blanks)

OK Cancel

Equals...

Does Not Equal...

Before...

After...

Between...

All Dates in Period ▸

- Quarter 1
- Quarter 2
- Quarter 3
- Quarter 4
- January
- February
- March
- April
- May
- June

1

2

Data Filtering

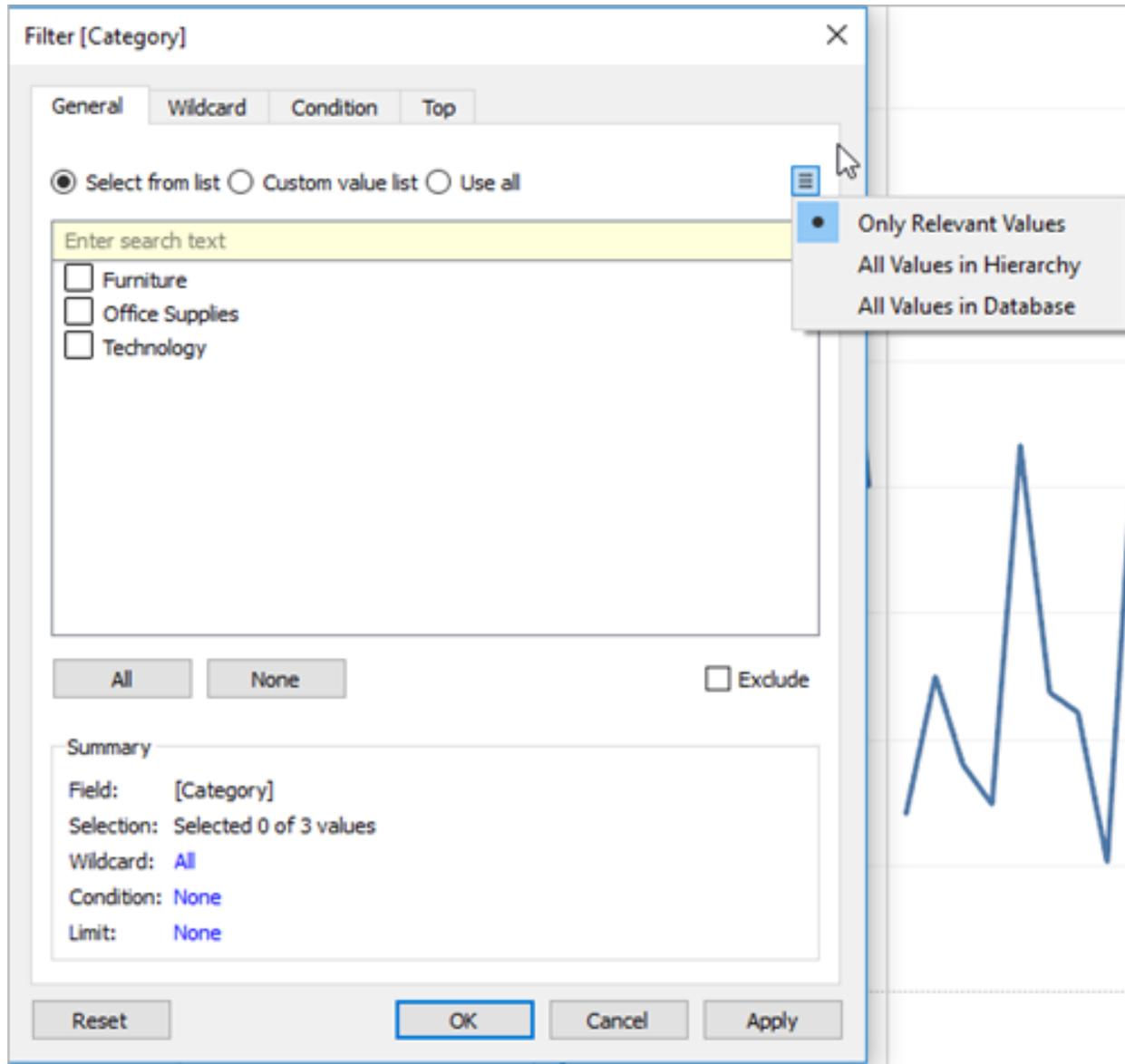
 **Filter Data** ? ✕

Filter Matching: Match ANY of the following +
 Match ALL of the following

Library Complex Name	▼	Starts With	▼	SL8500	▼	✕
Library Name	▼	Contains	▼	12	▼	✕
Library Top Level Indicator	▼	Is	▼	DEGRADED	▼	✕

Apply Reset Cancel

Data Filtering



The image shows a 'Filter [Category]' dialog box with the following elements:

- General** tab selected.
- Radio buttons: Select from list, Custom value list, Use all.
- Search input: 'Enter search text' (highlighted in yellow).
- List of categories: Furniture, Office Supplies, Technology.
- Buttons: 'All', 'None', and 'Exclude' (checkbox).
- Summary section: Field: [Category], Selection: Selected 0 of 3 values, Wildcard: All, Condition: None, Limit: None.
- Buttons: 'Reset', 'OK', 'Cancel', 'Apply'.

A tooltip is visible over the 'Select from list' radio button, containing the following options:

- Only Relevant Values
- All Values in Hierarchy
- All Values in Database

In the background, a line chart is visible on a grid, showing a blue line with several peaks and troughs.

- **Filtering data** means to set conditions so that only certain **data** is displayed. It is done to make it easier to focus on specific information in a large dataset or in a table of data.
- **Filtering** does not remove or modify data; it simply changes which rows or columns appear.

- Tableau performs a numbers of filters on the view in a very specific order; this is called the Order of Operations. Filters are executed in the following order:
 1. Extract filters
 2. Data source filters
 3. Context filters
 4. Filters on dimensions (whether on the Filters shelf or in filter cards in the view)
 5. Filters on measures (whether on the Filters shelf or in filter cards in the view)

- Extract filters
 - Extracts are saved subsets of data that are used to improve performance or to take advantage of Tableau functionality that is not available or supported in your original data.
 - After data extraction, the total amount of data can be reduced by using filters and configuring other limits.

- Extract filters
 - After the creation of an extract, the data can be refreshed from the original data source.
 - By refreshing the data, there are 2 options:
 - either do a full refresh, which replaces all of the contents in the extract,
 - or an incremental refresh, which only adds rows that are new since the previous refresh.

Data Filtering

Extract filters

- The primary method to create an extract of the data

Connection

Live

Extract

| [Edit](#)

[Refresh](#)

Filters

0 | [Add](#)

Extract will include all data.

Data Filtering

Extract Data ✕

Specify how to store data in the extract:

Data Storage

Single table Multiple tables

Store data in your extract together using a single table. [Learn more](#)

Use this option if you need to use extract filters, aggregation, top N, etc.

Specify how much data to extract:

Filters (optional)

Filter	Details
--------	---------

Aggregation

Aggregate data for visible dimensions

Roll up dates to

Number of Rows

All rows

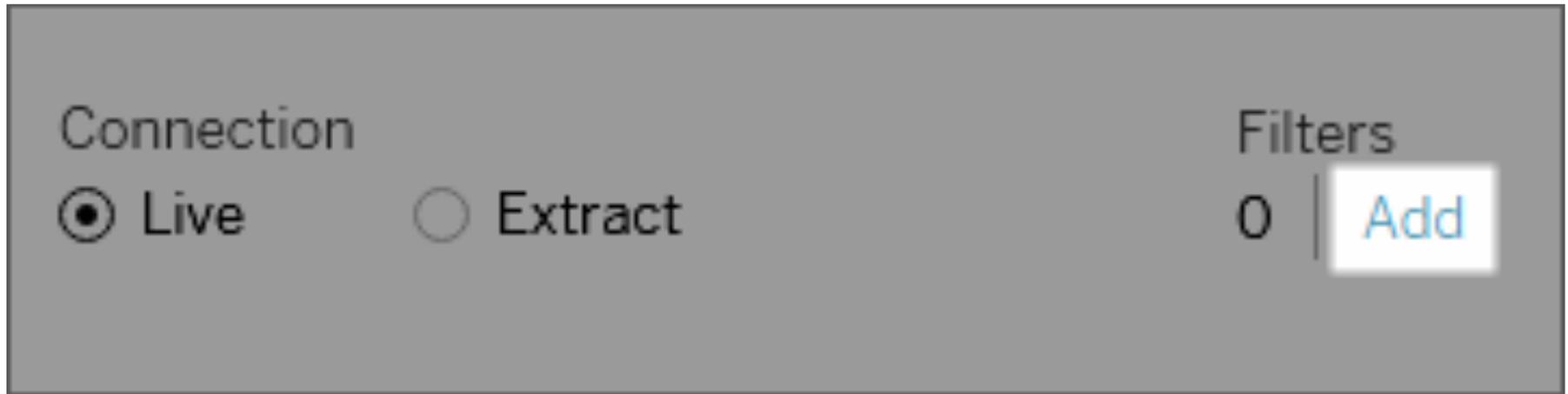
Incremental refresh

Top: rows

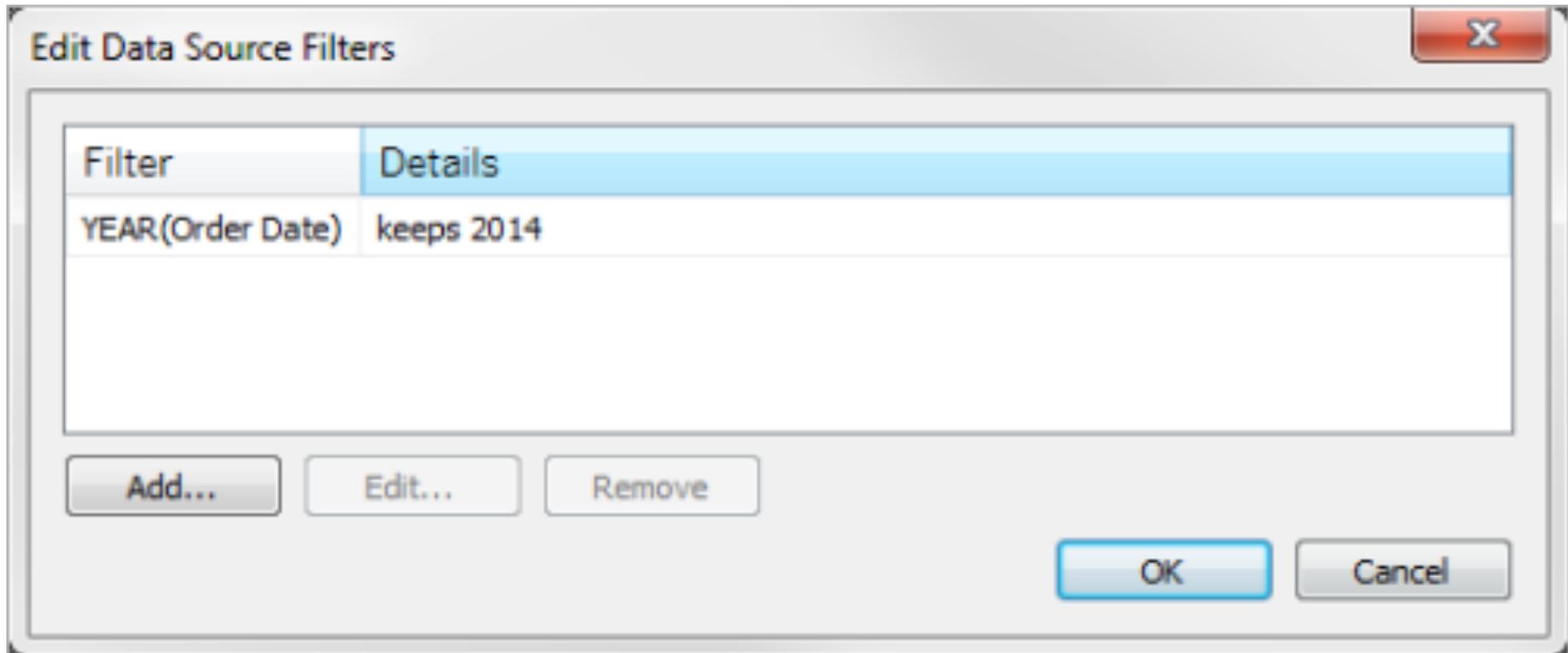
- Data source filters
 - You create filters on a data source, in order to reduce the amount of data in the data source.
 - Are really useful for restricting the data that users can see when you publish a workbook or a data source.

Data Source filters

- The primary method to create a data source filter

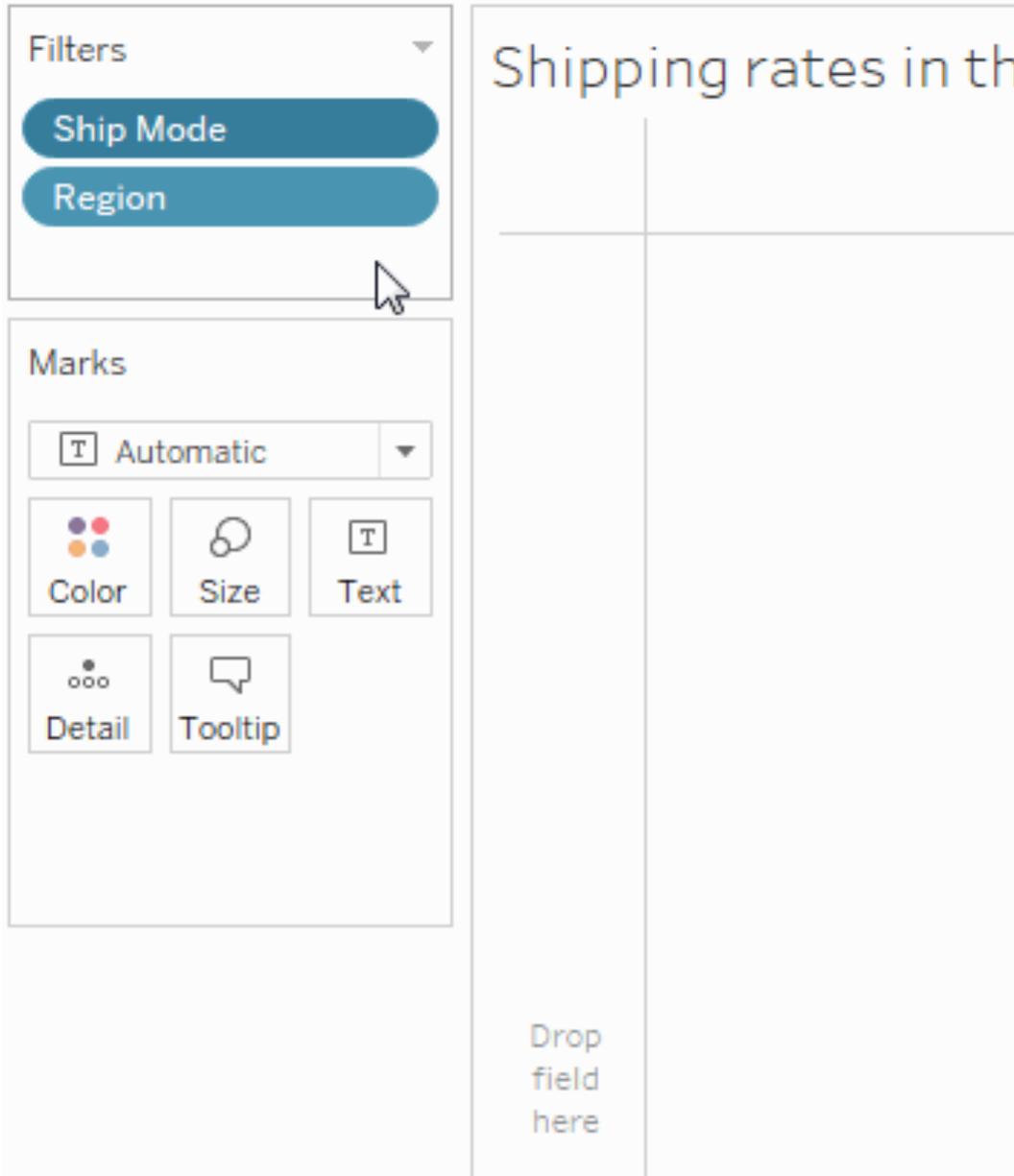


Data Filtering



- Context filters
 - **Improve performance** – If you set a lot of filters or have a large data source, the queries can be slow. You can set one or more context filters to improve performance.
 - **Create a dependent numerical or top N filter** – You can set a context filter to include only the data of interest, and then set a numerical or a top N filter.

Data Filtering



The image shows a user interface for data filtering and visualization. On the left, there are two panels: 'Filters' and 'Marks'. The 'Filters' panel contains two blue buttons labeled 'Ship Mode' and 'Region'. The 'Marks' panel contains a dropdown menu set to 'Automatic' and five icons for 'Color', 'Size', 'Text', 'Detail', and 'Tooltip'. On the right, a large white area is titled 'Shipping rates in th' and contains a grid. At the bottom of this area, the text 'Drop field here' is visible.

Filters

Ship Mode

Region

Marks

Automatic

Color

Size

Text

Detail

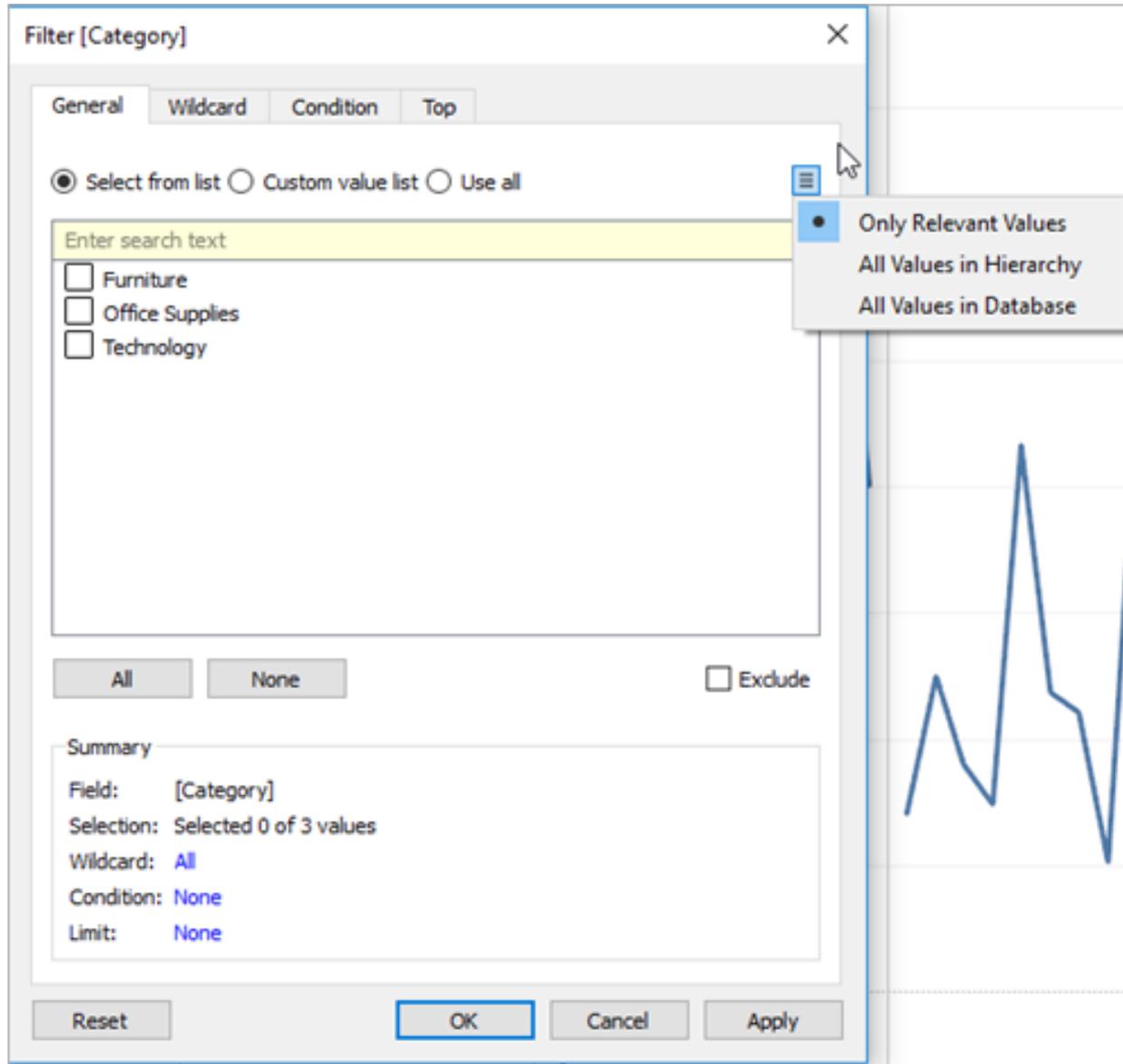
Tooltip

Shipping rates in th

Drop field here

- Filters on dimensions (whether on the Filters shelf or in filter cards in the view)
 - **Dimensions** contain discrete categorical data, so filtering this type of field generally involves selecting the values to include or exclude.

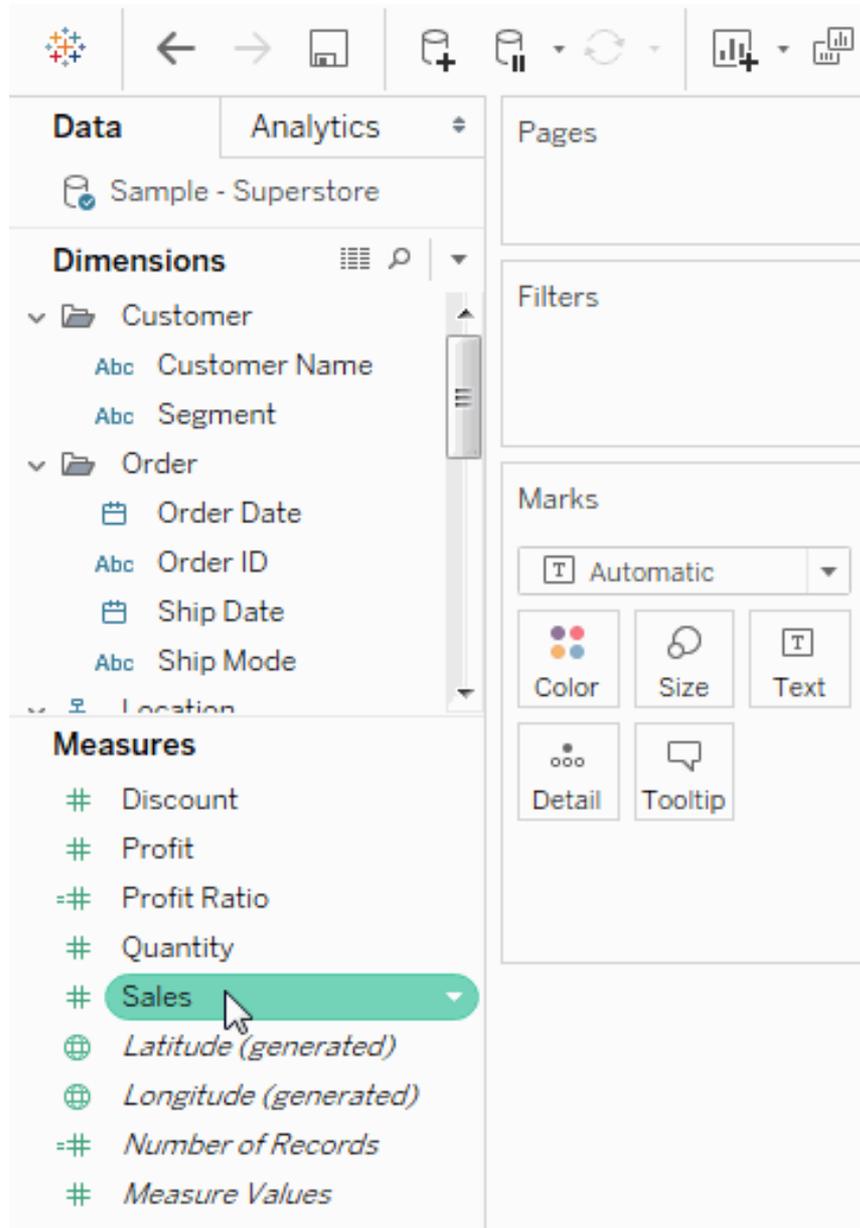
Data Filtering



The screenshot shows a 'Filter [Category]' dialog box with a dropdown menu open. The dialog has tabs for 'General', 'Wildcard', 'Condition', and 'Top'. Under 'General', there are radio buttons for 'Select from list' (selected), 'Custom value list', and 'Use all'. A search box contains 'Enter search text'. Below it are three checkboxes: 'Furniture', 'Office Supplies', and 'Technology', all of which are unchecked. At the bottom of the dialog are buttons for 'All', 'None', and 'Exclude' (unchecked). A 'Summary' section shows: Field: [Category], Selection: Selected 0 of 3 values, Wildcard: All, Condition: None, and Limit: None. At the very bottom are 'Reset', 'OK', 'Cancel', and 'Apply' buttons. The dropdown menu is open, showing three options: 'Only Relevant Values' (selected with a blue dot), 'All Values in Hierarchy', and 'All Values in Database'. The background shows a line graph with a blue line on a grid.

- Filters on measures (whether on the Filters shelf or in filter cards in the view)
 - **Measures** contain quantitative data, so filtering this type of field generally involves selecting a range of values that you want to include.

Data Filtering



The screenshot displays the Tableau Desktop interface. At the top, there is a toolbar with navigation and analysis icons. Below the toolbar, the 'Data' pane shows the 'Sample - Superstore' data source. The 'Dimensions' pane is expanded to show a tree view of fields: 'Customer' (Customer Name, Segment) and 'Order' (Order Date, Order ID, Ship Date, Ship Mode). The 'Measures' pane is visible at the bottom, listing various metrics. The 'Sales' measure is highlighted with a green background and a mouse cursor. The 'Marks' card is set to 'Automatic' and includes options for Color, Size, Text, Detail, and Tooltip.

Data Analytics

Sample - Superstore

Dimensions

- Customer
 - Customer Name
 - Segment
- Order
 - Order Date
 - Order ID
 - Ship Date
 - Ship Mode

Measures

- Discount
- Profit
- Profit Ratio
- Quantity
- Sales**
- Latitude (generated)
- Longitude (generated)
- Number of Records
- Measure Values

Marks

Automatic

Color Size Text

Detail Tooltip

Data Filtering

Filter Field [Sales]

How do you want to filter on [Sales]?

- # **All values**
- # Sum
- # Average
- # Median
- # Count
- # Count (Distinct)
- # Minimum
- # Maximum
- # Standard deviation
- # Standard deviation (Population)
- # Variance
- # Variance (Population)
- # Attribute

Next > Cancel

Data Filtering

Filter [Sales] ✕

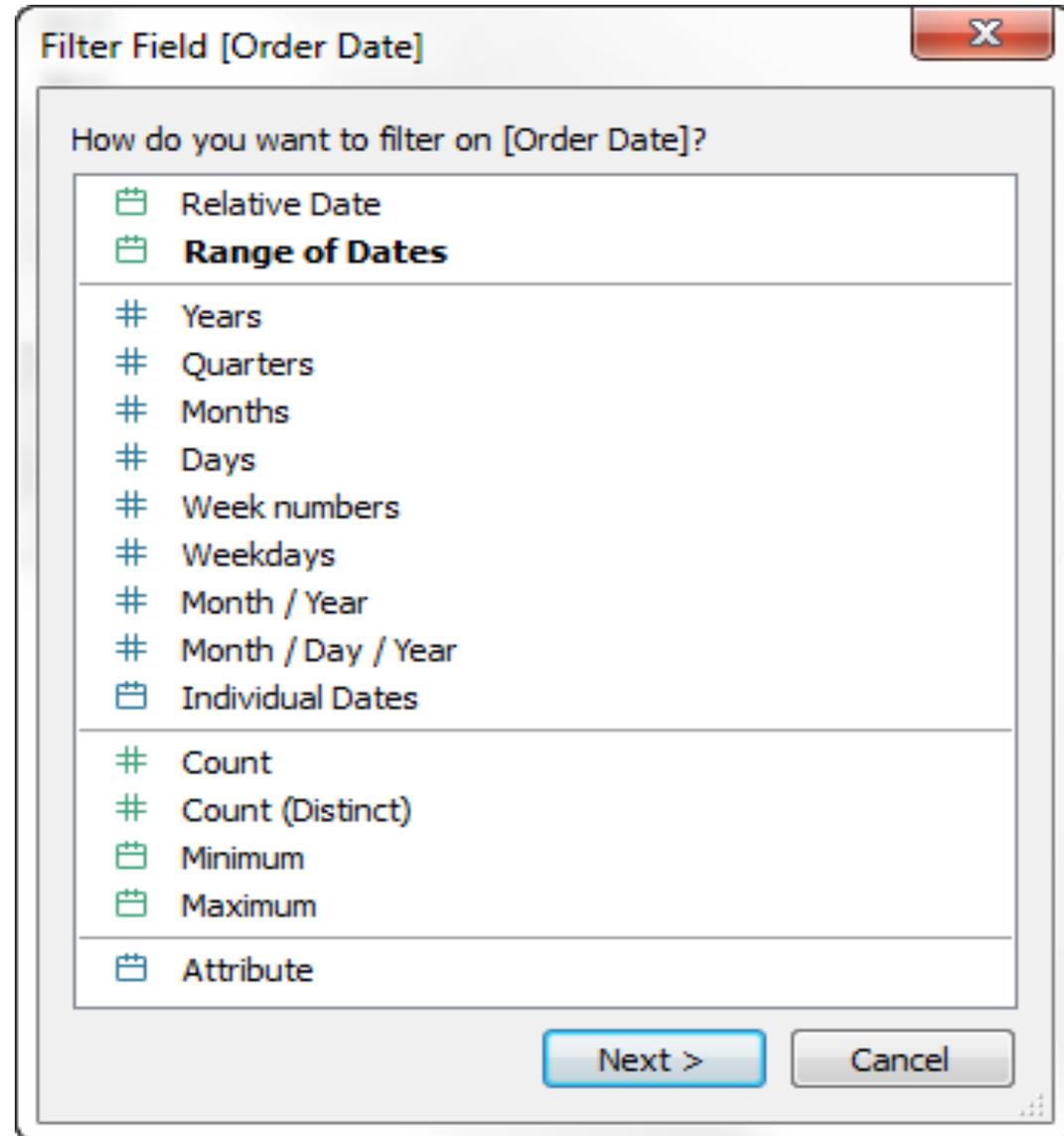
Range of values

1,090.448 210,681.032

Show: Include Null Values

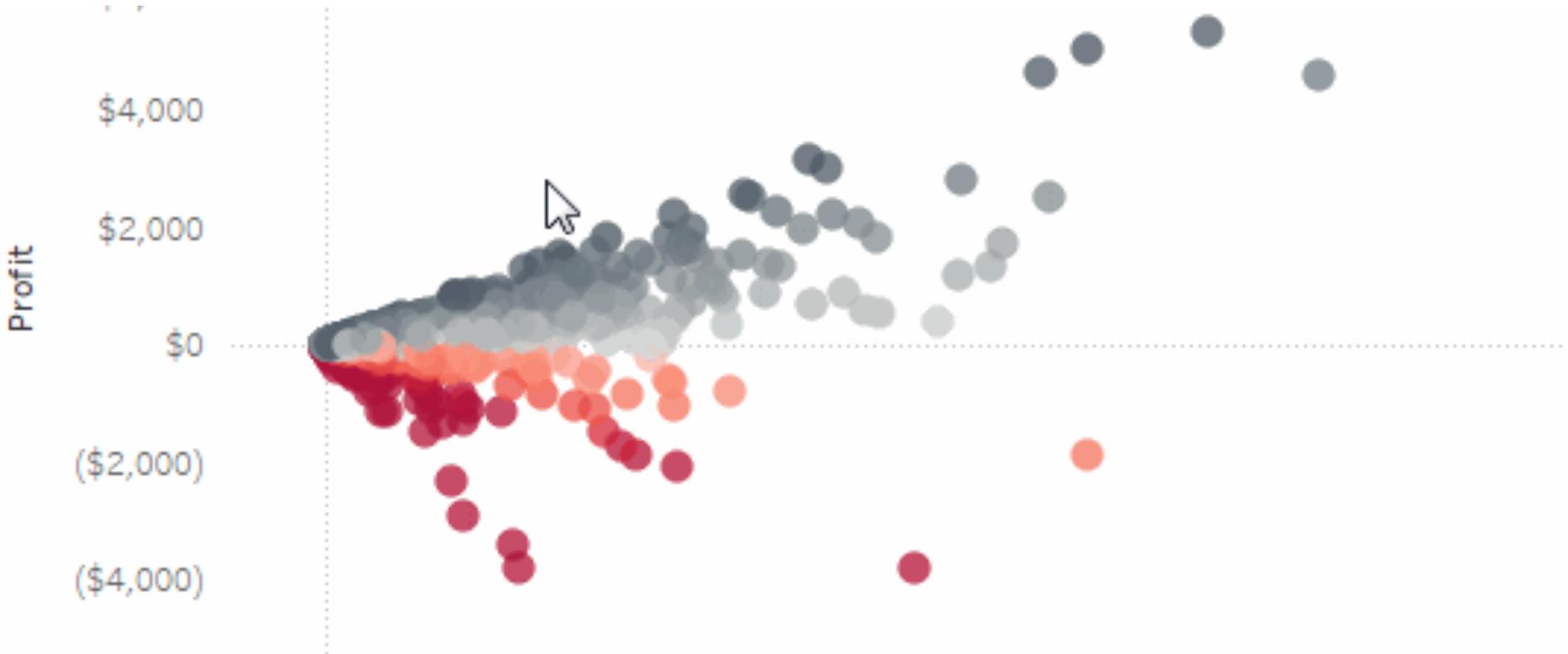
Additional Filtering

- Filter dates

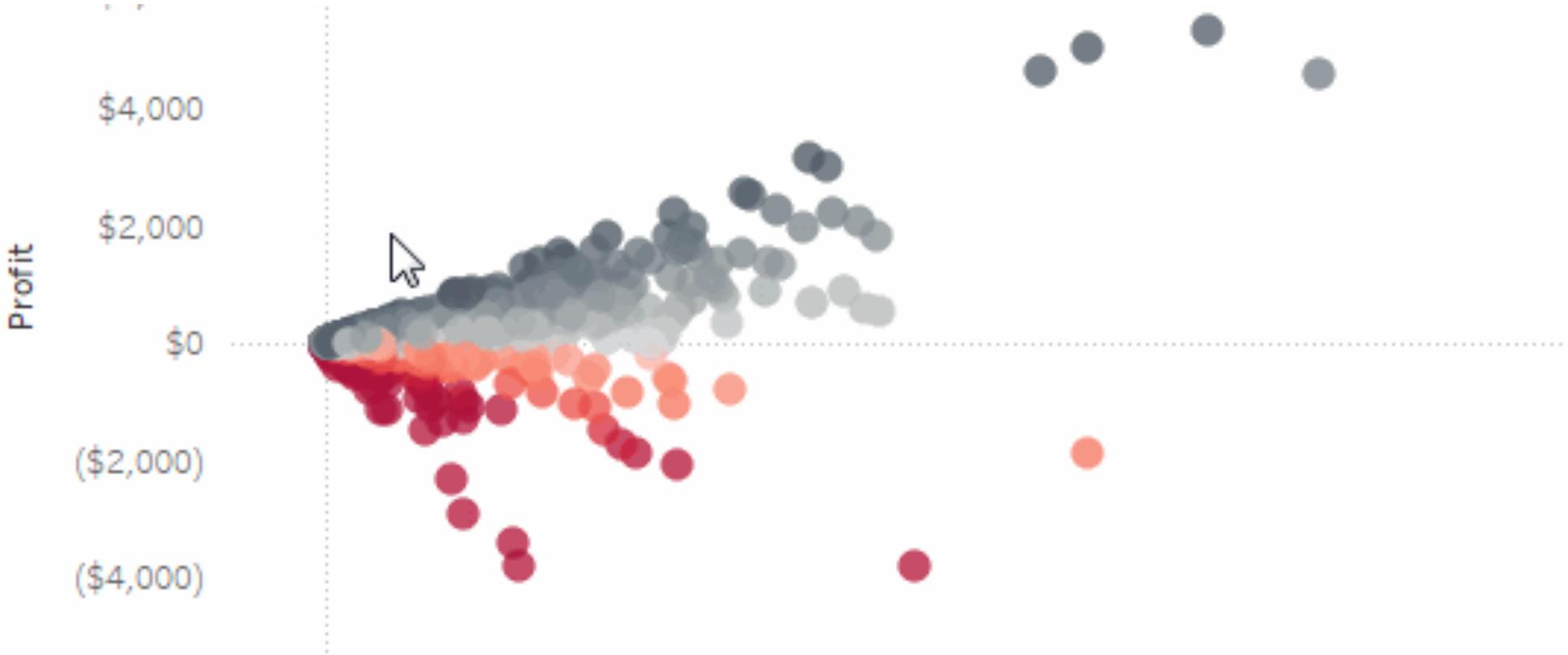


- **Select to Keep or Exclude Data in your View**
 - You can filter individual data points (marks), or a selection of data points from your view.
 - To filter marks from the view, select a single mark (data point) or click and drag in the view to select several marks.

Data Filtering



Data Filtering



- Apply filters to Multiple worksheets
 - When a filter is added to a worksheet, by default it filter applies only to the current worksheet. Sometimes, however, this filter have to be applied to other worksheets in the workbook.
 - You can select specific worksheets to apply the filter to or apply it globally to all worksheets that use the same data source or related data sources.

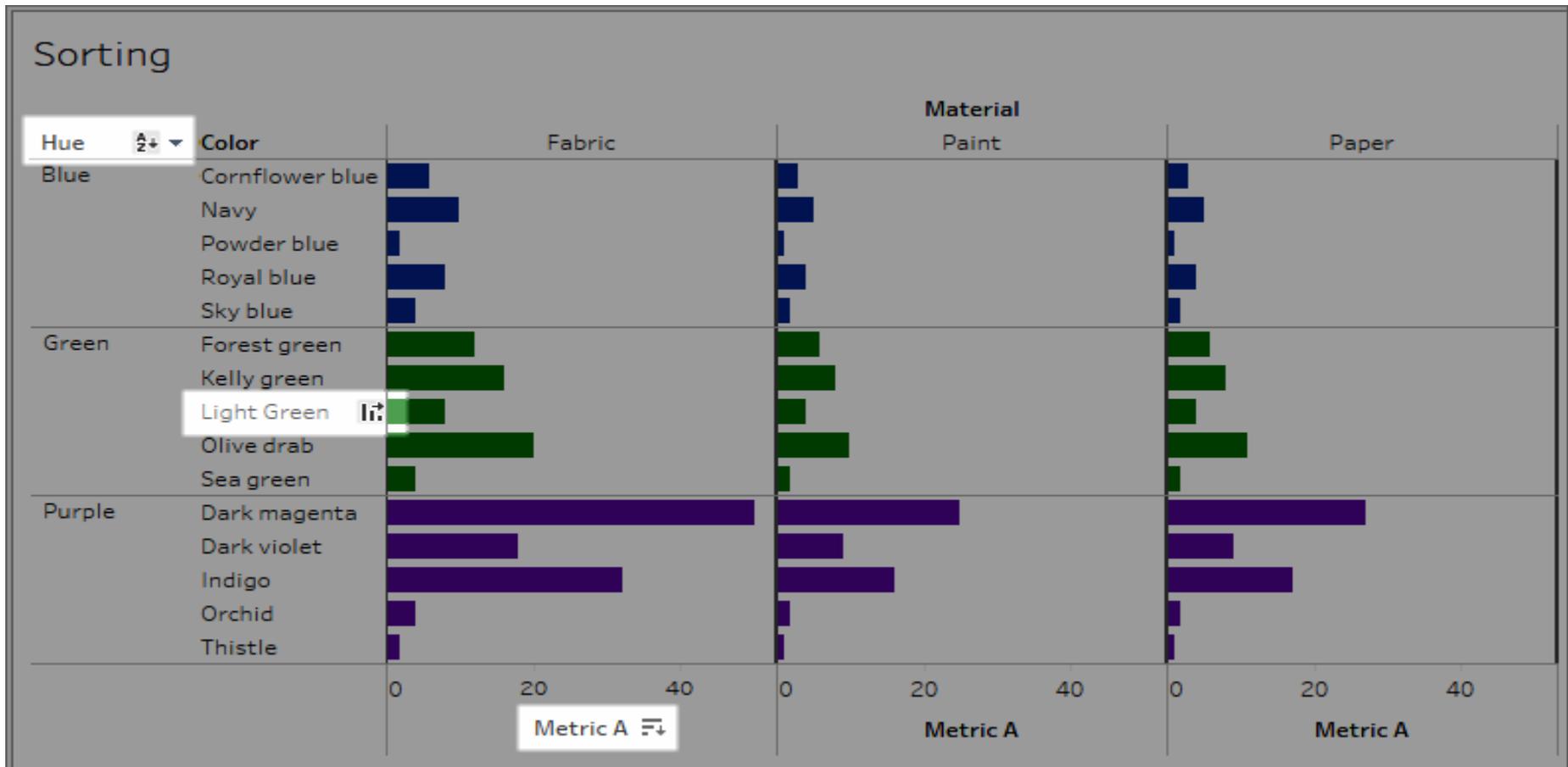
Sort Data



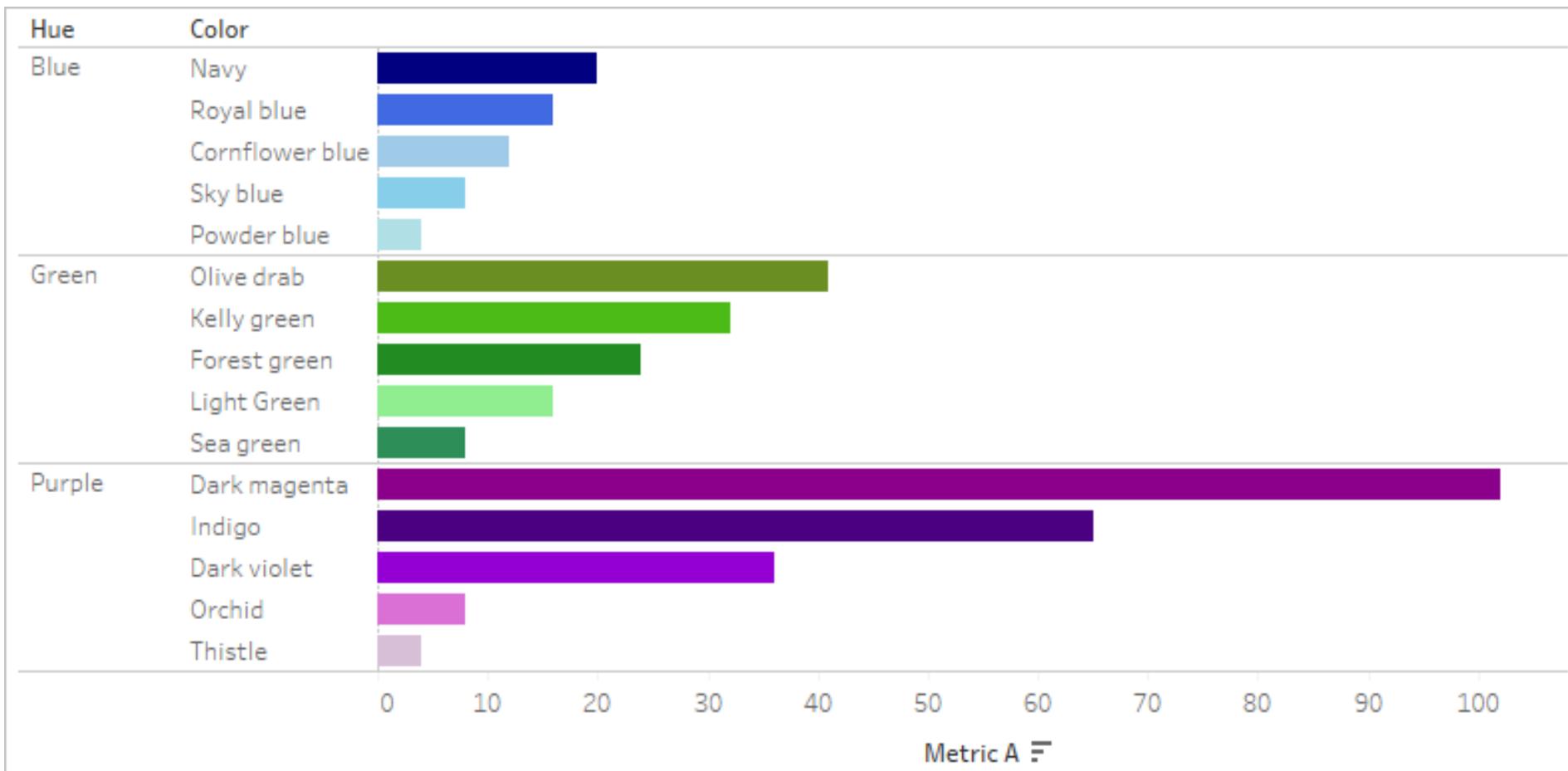
- Why Data Sorting is important?
 - **Data sorting** is the process that involves arranging our **data** into some meaningful order to make it easier to understand, analyse or visualize them.
 - When working with research **data**, **sorting** is a common method used for visualizing **data** in a form that makes it easier to comprehend the story the **data** is telling.

- There are many ways to sort data in a visualization:
 - Sort data on an axis
 - Sort specific fields in the visualization
 - Sort data using the toolbar
 - Sort data using headers or legends
 - Create a nested sort

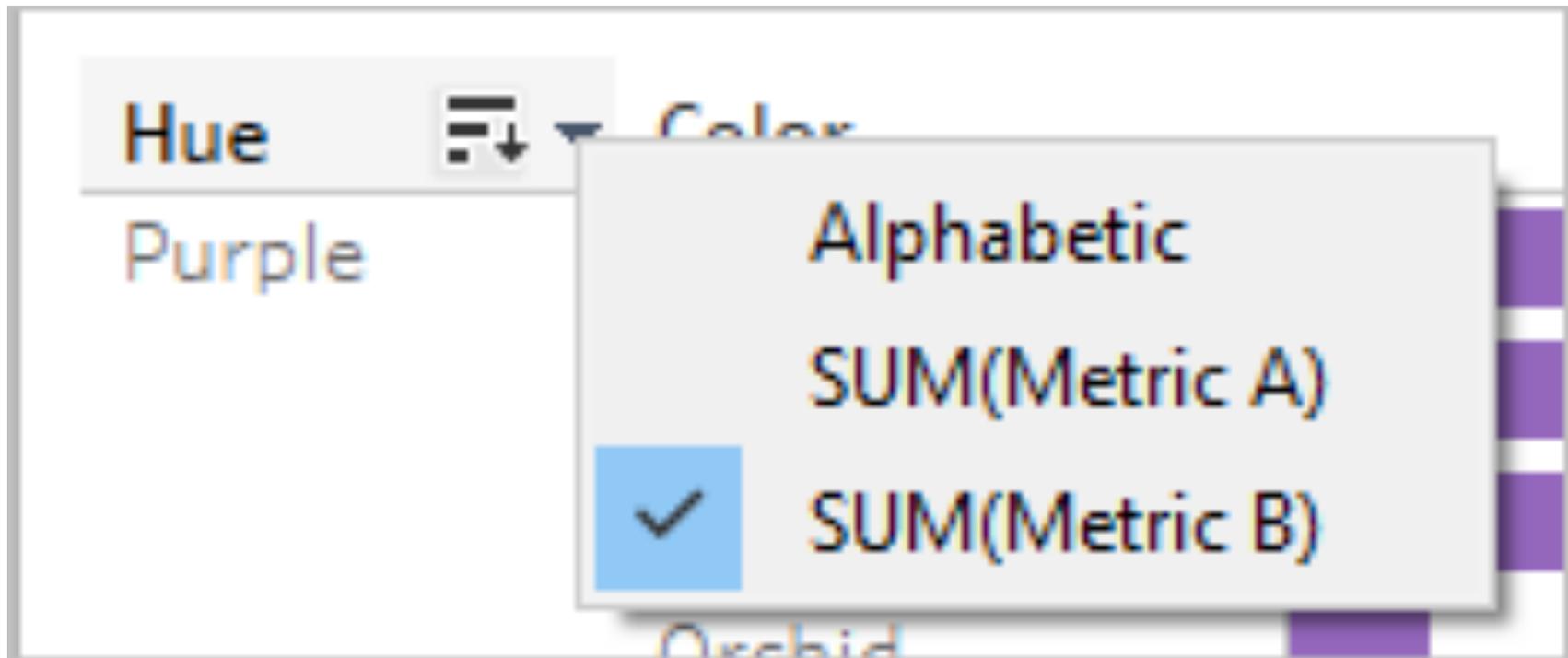
- Sort from an Axis



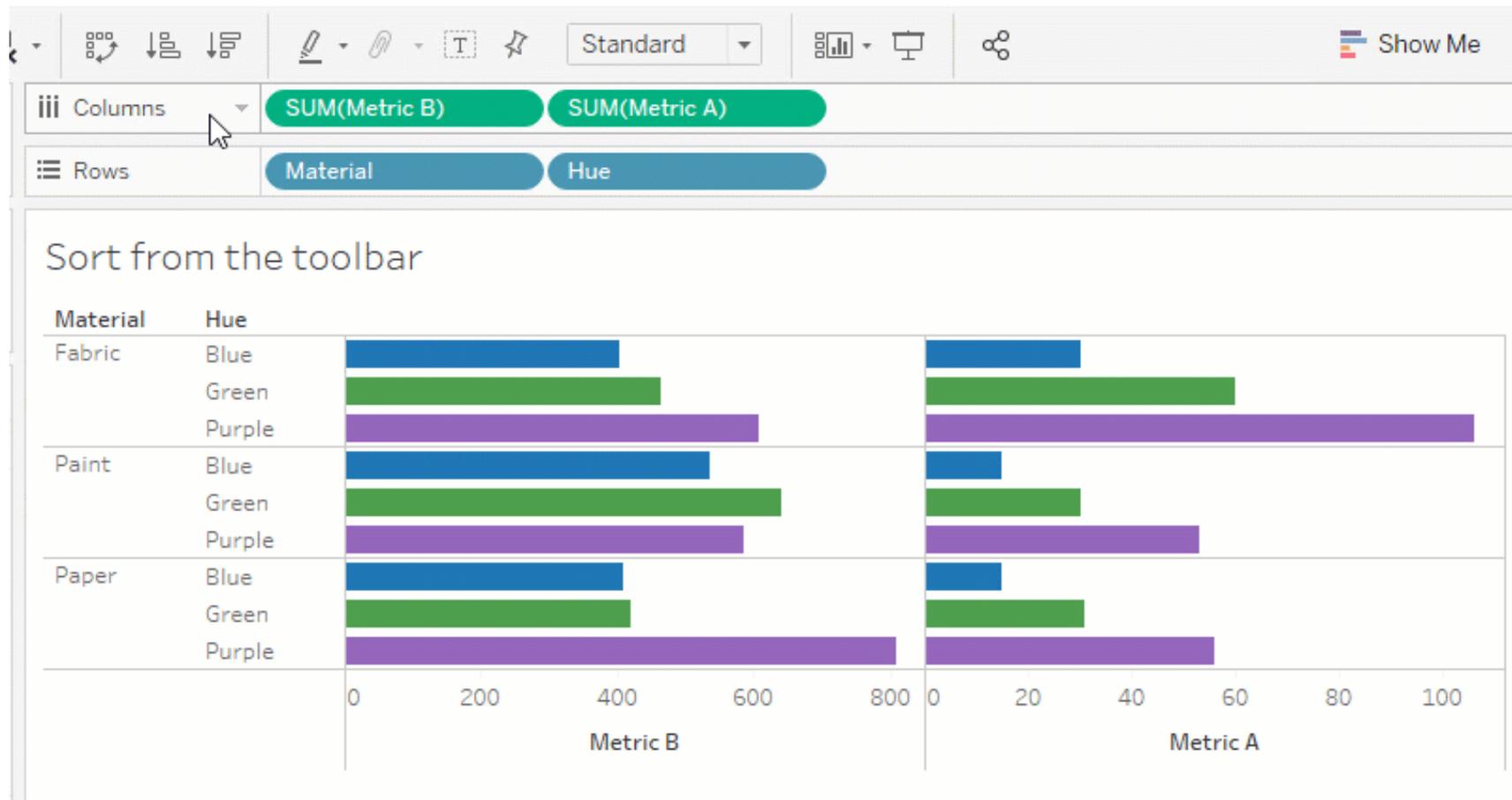
- Sort from an Axis



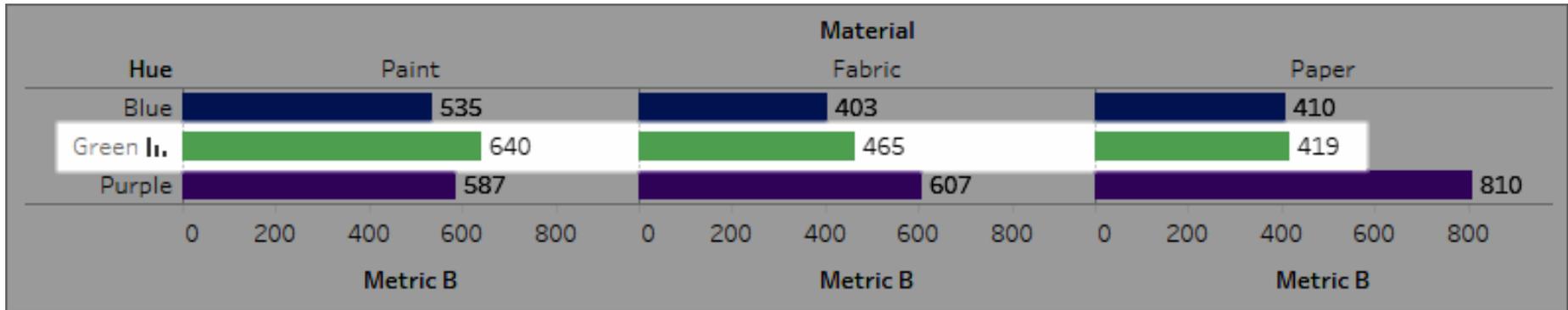
- Sort from specific fields in the visualization



- Sort Data using Toolbar

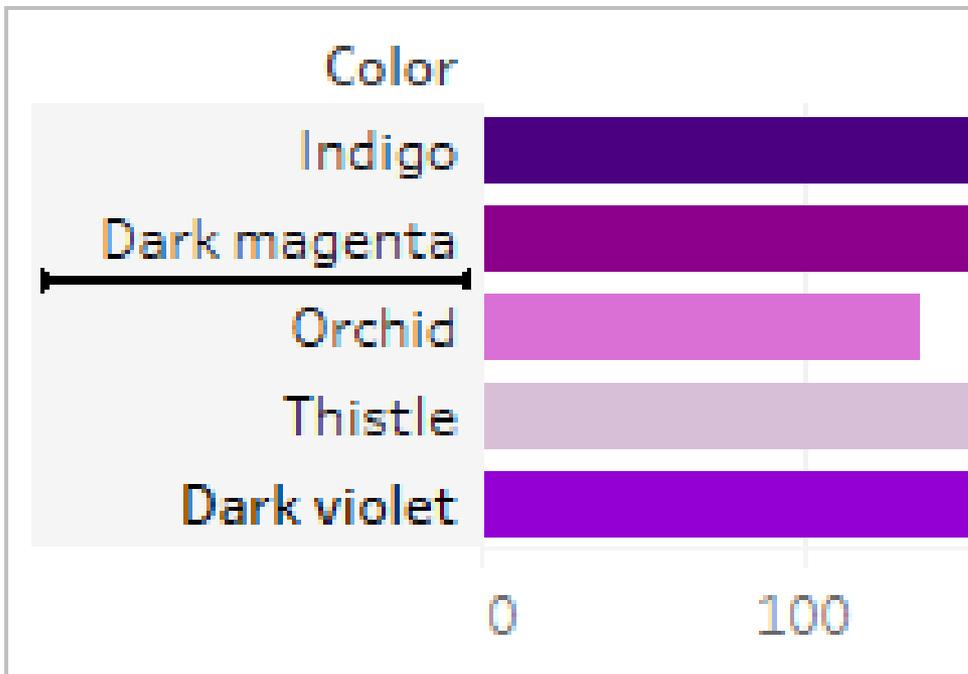


- Sort Data using headers or legends

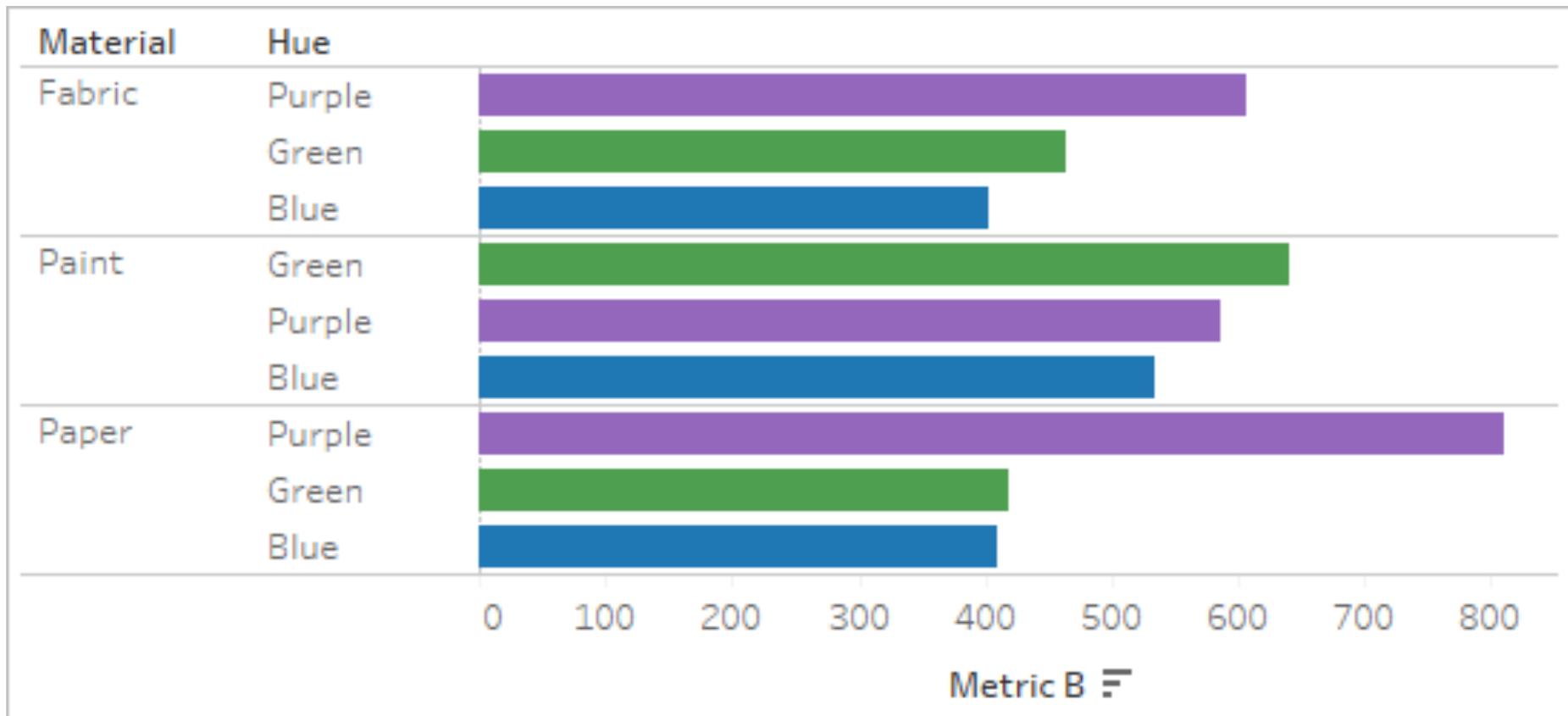


Sorting Data

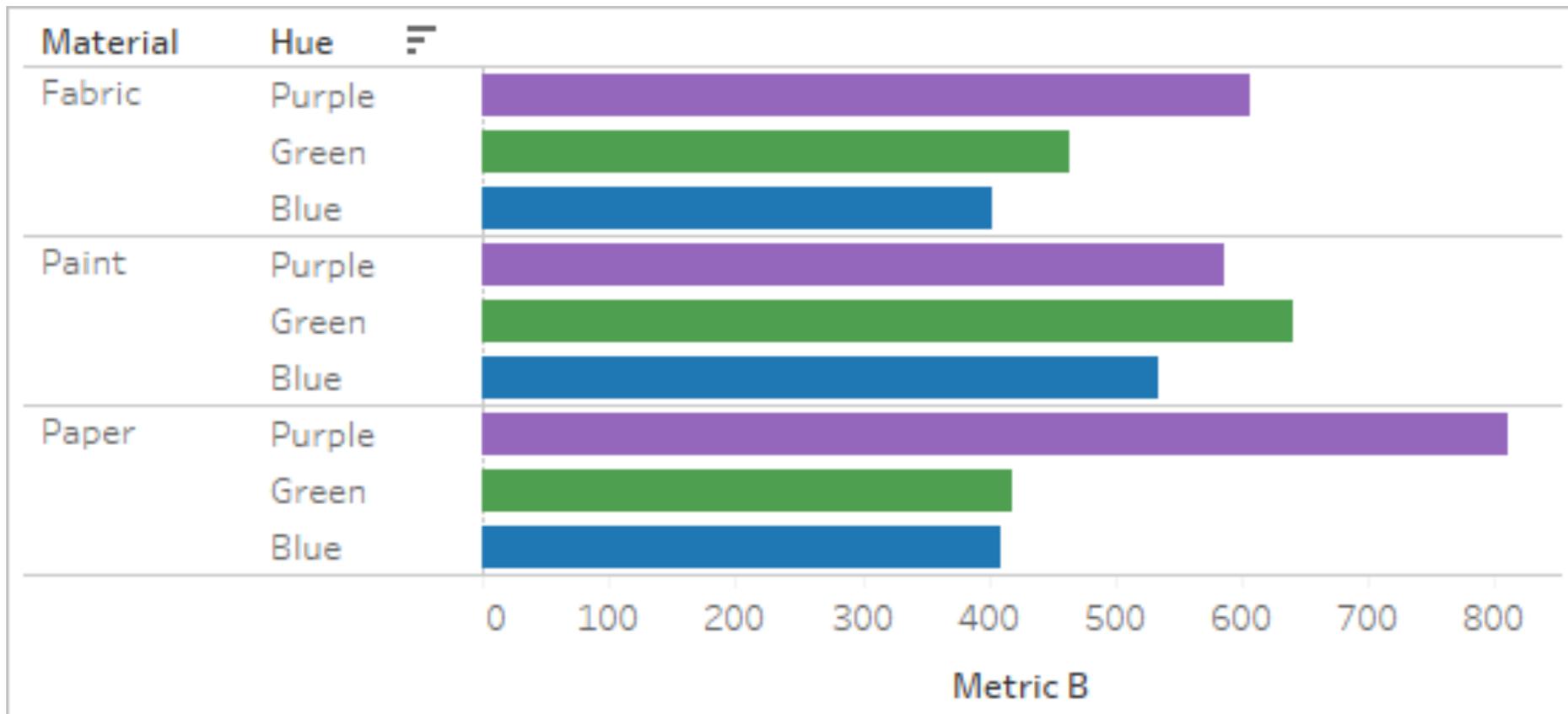
- Sort by Drag and Drop



- Create an Nested Sort



- Create an Non-Nested Sort

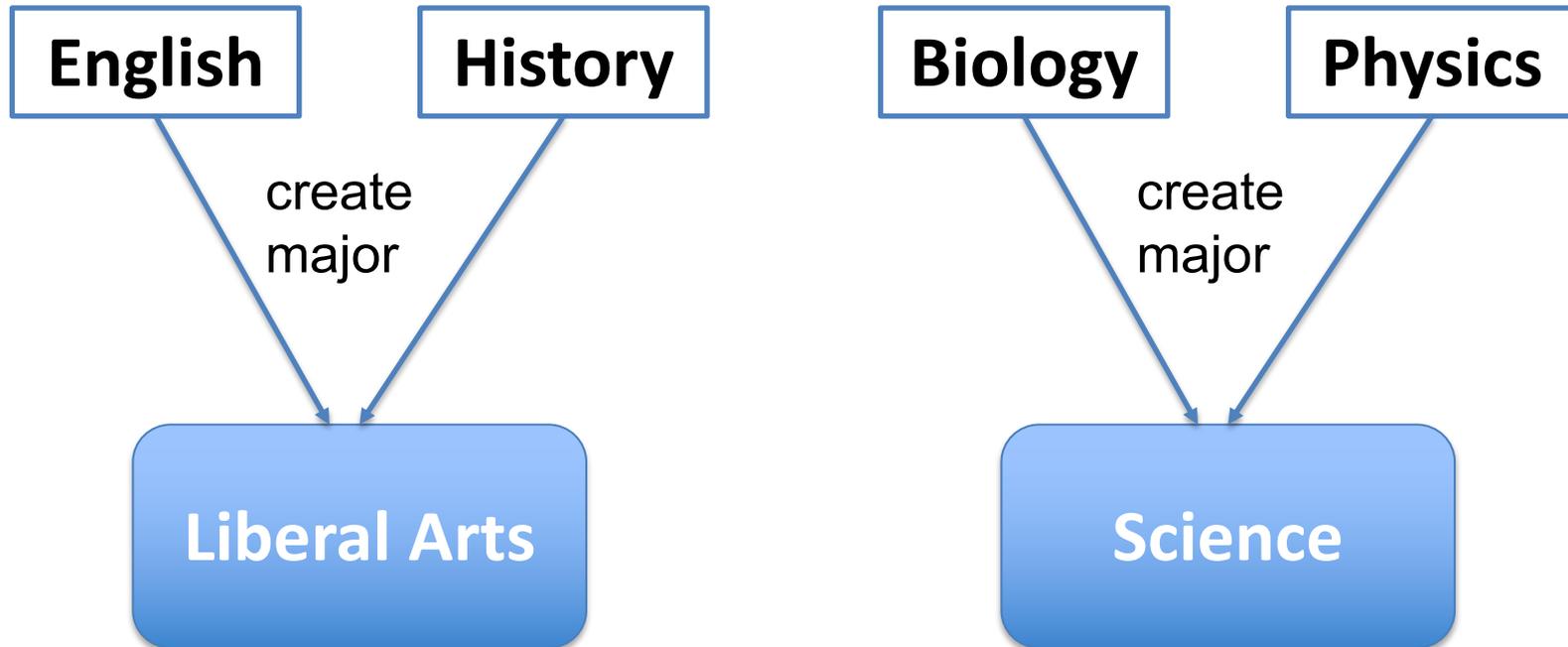


Group your Data

- On Tableau 'Groups' are used for:
 - to combine related members in a field
 - correcting data errors
 - answering "what if" type questions

Group your Data

- Combine related members in a field

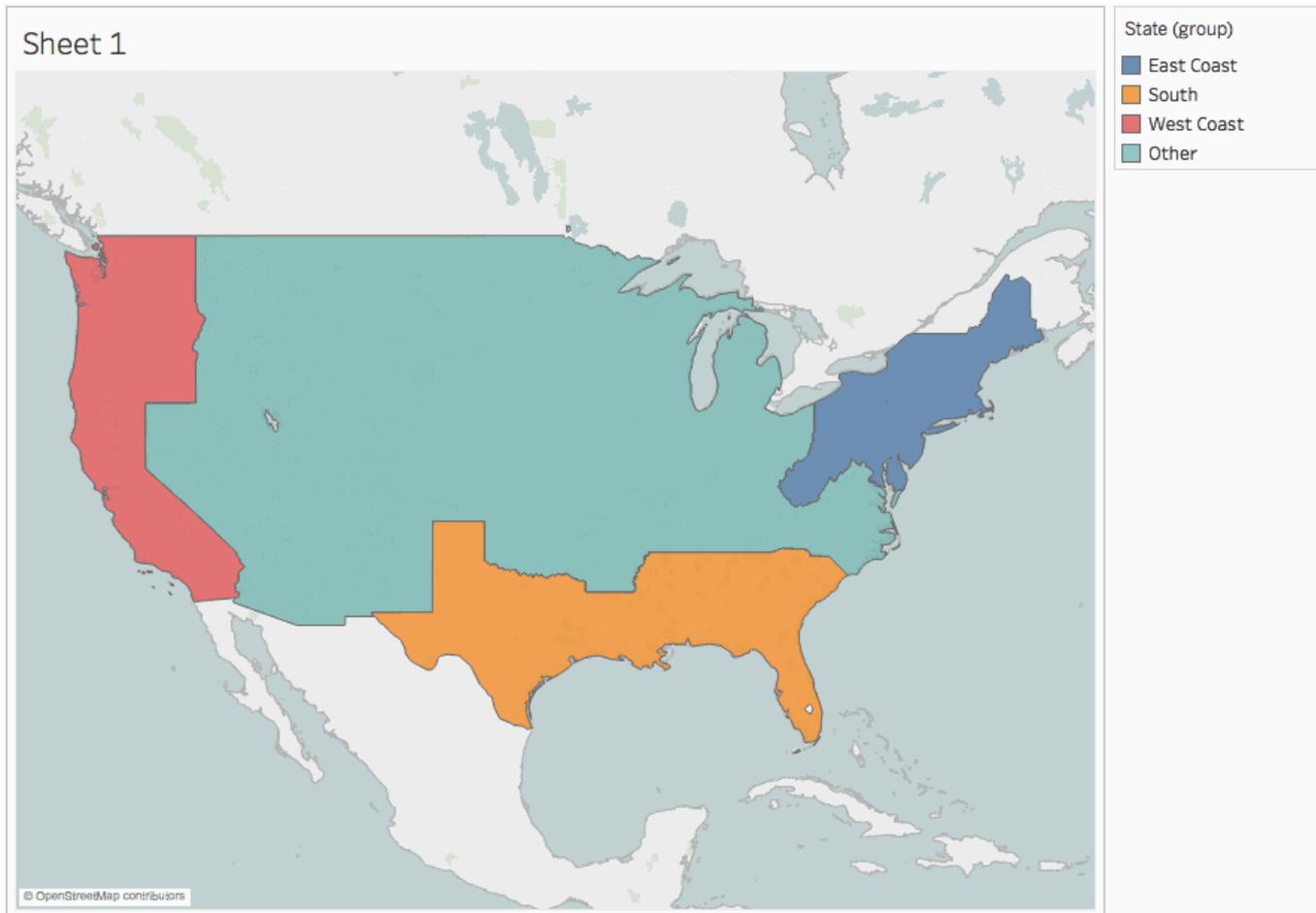


- Correcting data errors

EXAMPLE 1	EXAMPLE 2
combining 'CA' , 'Calif.' and 'California'	combining 'Cst 3340' , 'cst-3340' , and 'CST 3340'
into one data point 'CA'	into one data point 'CST3340'

Group your Data

- ‘What if’ type questions



- How to create a Group

There are various ways to create a group.

- By creating a group from a field in the **Data** pane.
- By selecting data in the view and then clicking the group icon.

Group your Data

- Group by selecting data in the view
- In the view, select one or more data points and then, on the tooltip that appears, click the group icon.

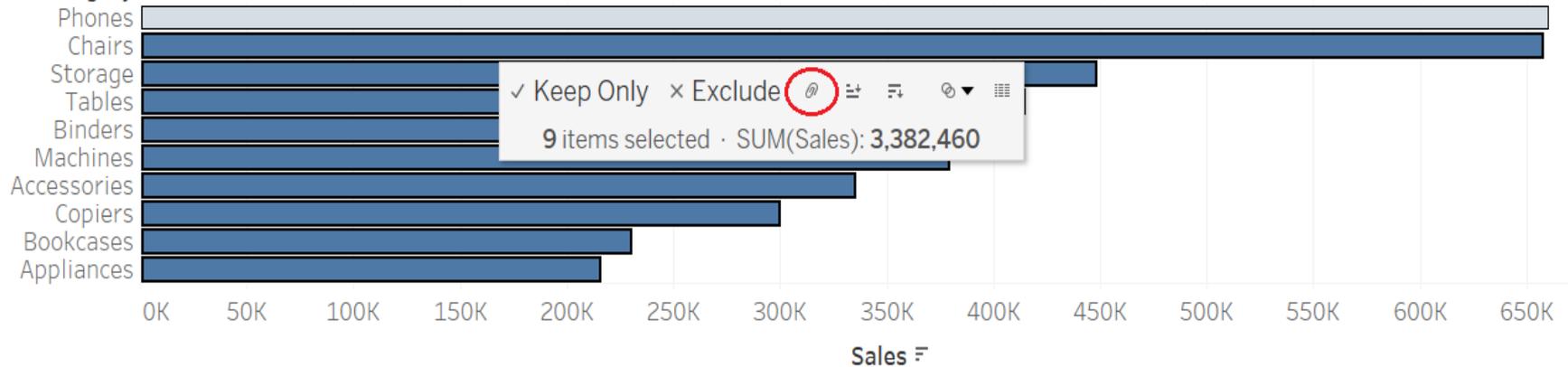


Group your Data

- Group by selecting data in the view

Sheet 2

Sub-Category

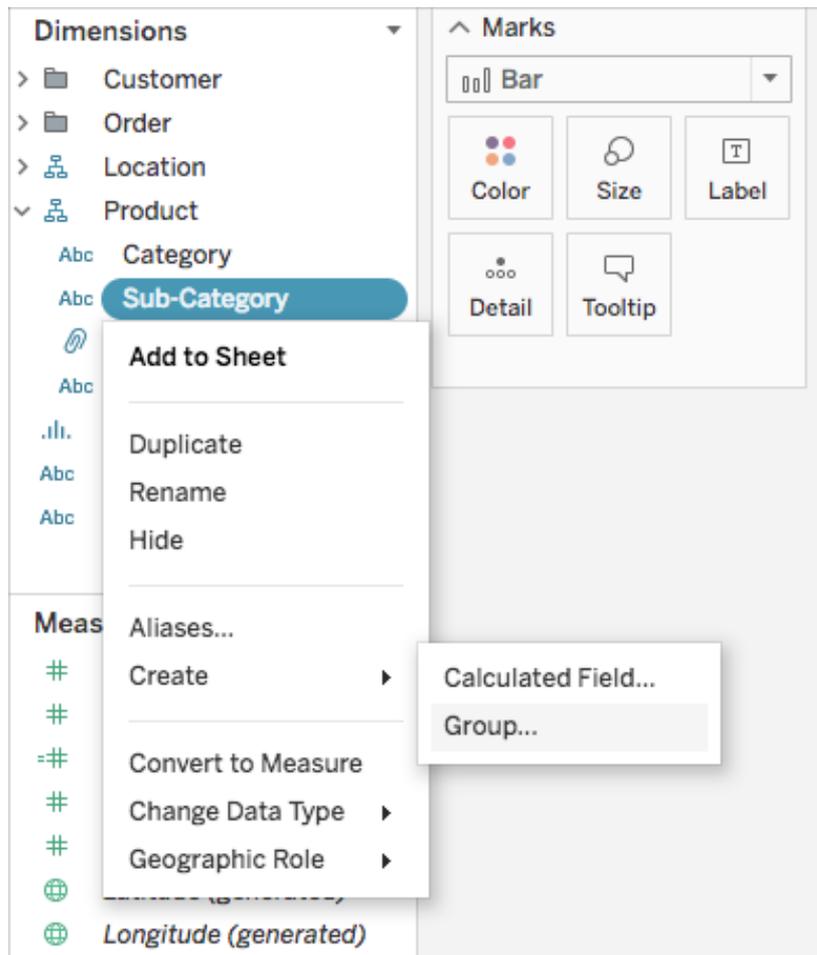


Group your Data

- Group from a field in the Data pane
- In the **Data** pane, right-click a field and select **Create** → **Group**.

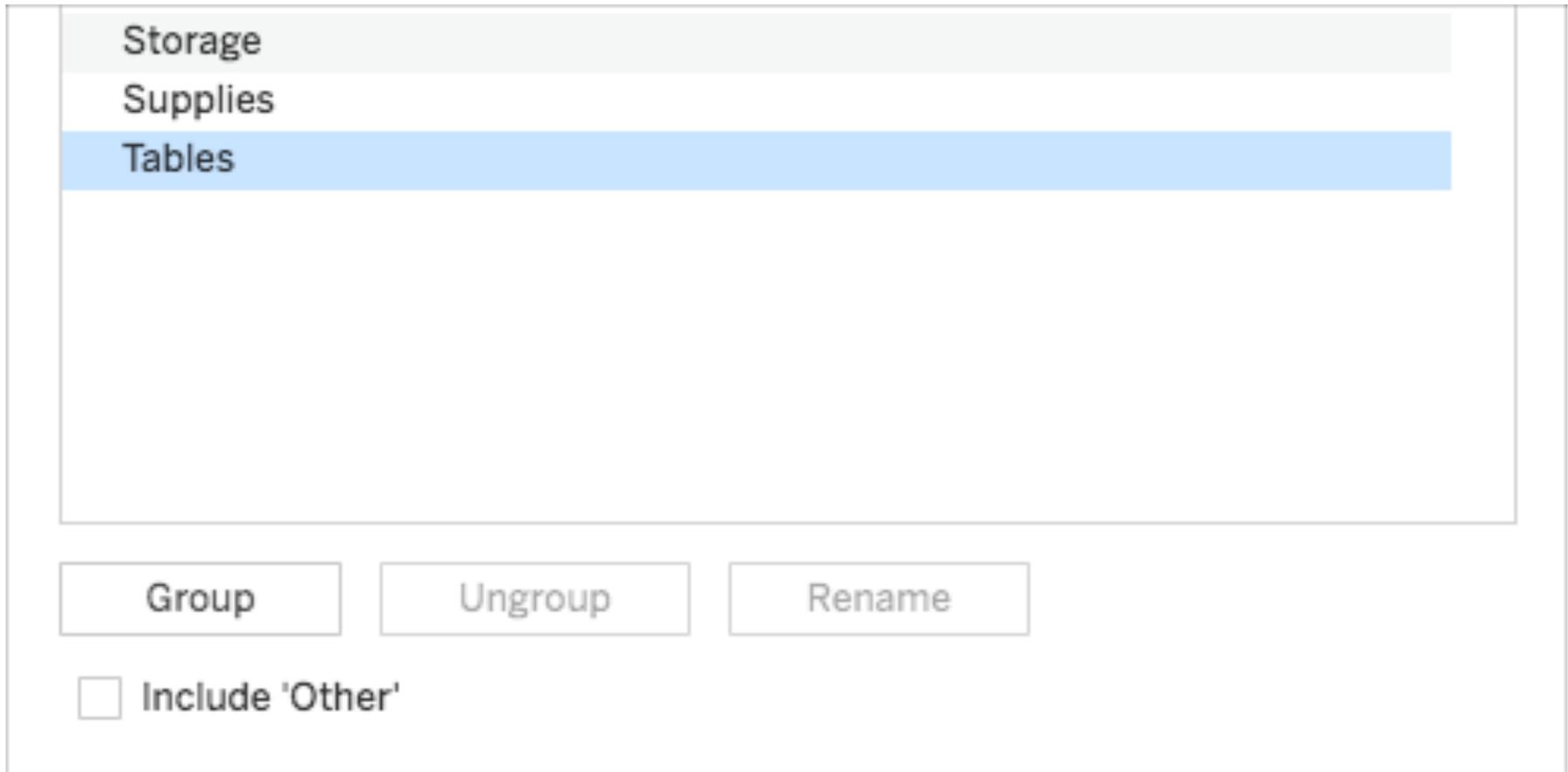
Group your Data

- Group from a field in the Data pane



Group your Data

- Group from a field in the Data pane



The screenshot shows a data pane with a list of items: Storage, Supplies, and Tables. The 'Tables' item is selected, highlighted in blue. Below the list are three buttons: 'Group', 'Ungroup', and 'Rename'. At the bottom left, there is a checkbox labeled 'Include Other'.

Storage
Supplies
Tables

Group Ungroup Rename

Include 'Other'

Group your Data

Edit Group [Sub-Category (group)]

- Accessories
- Appliances
- Art
- Binders
- ▼ Bookcases & Tables
 - Bookcases
 - Tables
- Chairs
- Copiers
- Envelopes
- Fasteners
- Furnishings
- Labels
- Machines
- Paper
- Phones
- Storage
- Supplies

Include 'Other'

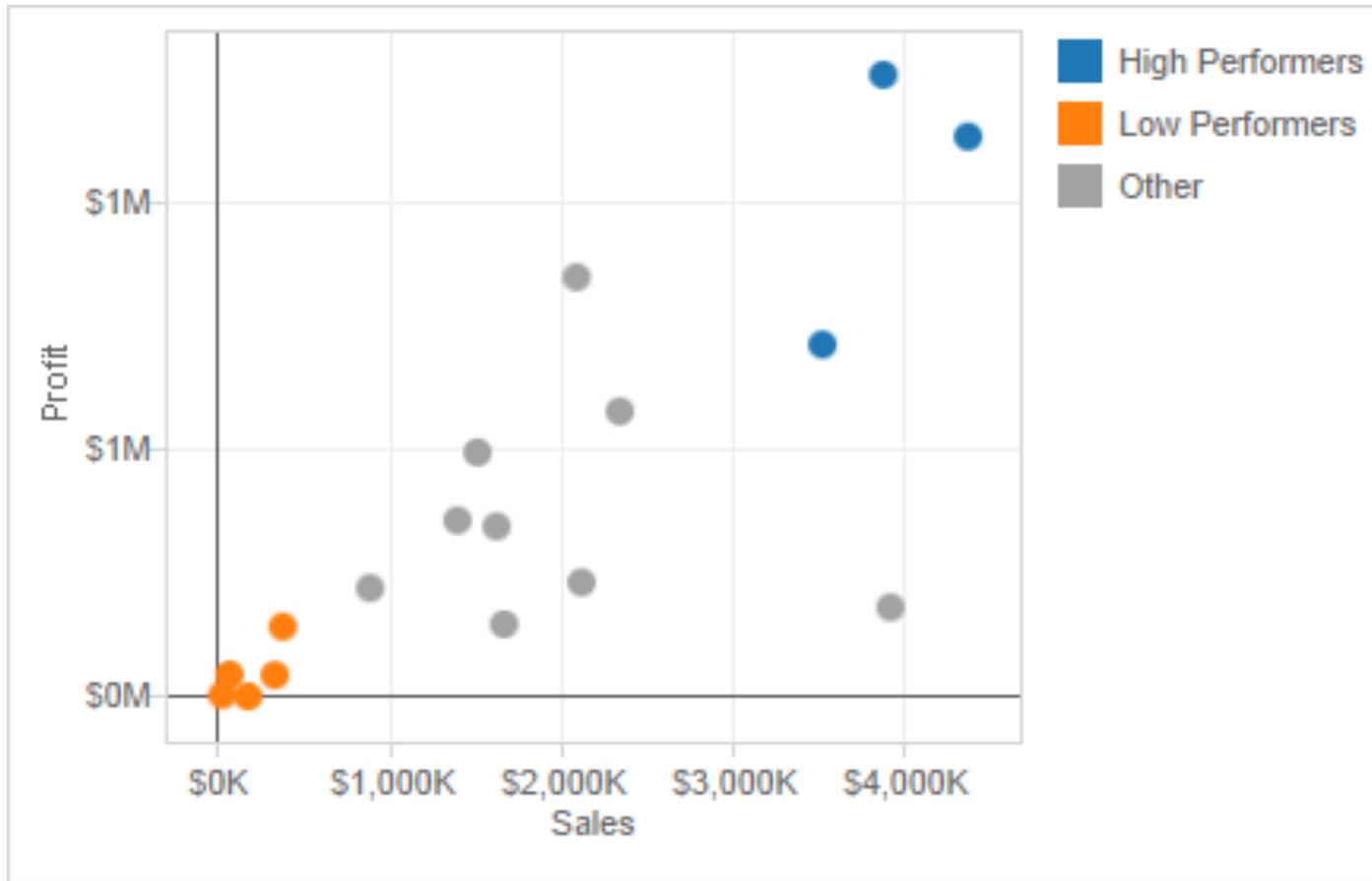
- Include an Other Group

The Include Other option is useful for

- highlighting certain groups
- comparing specific groups against everything else.

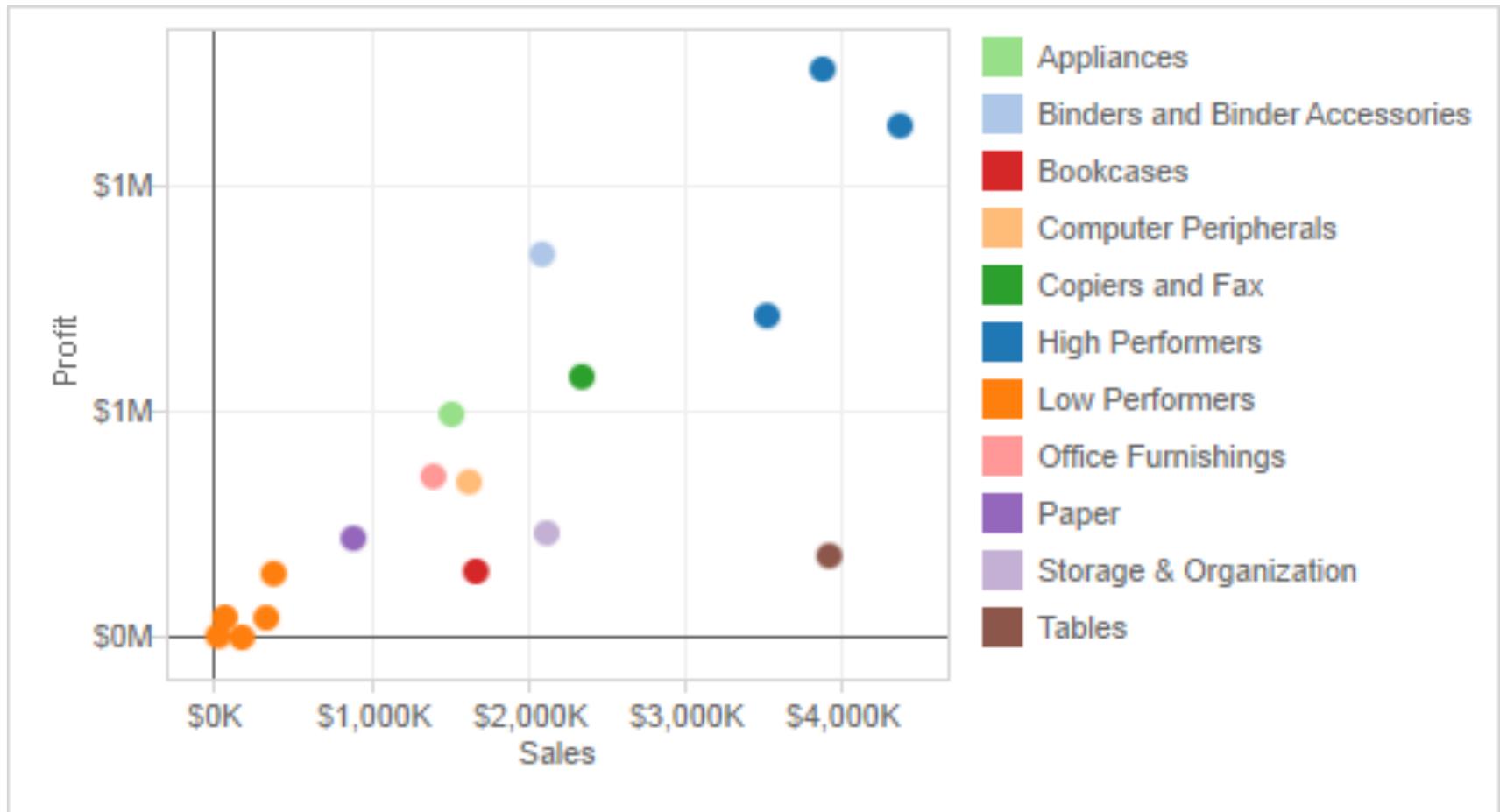
Group your Data

- Include others



Group your Data

- Does not include others



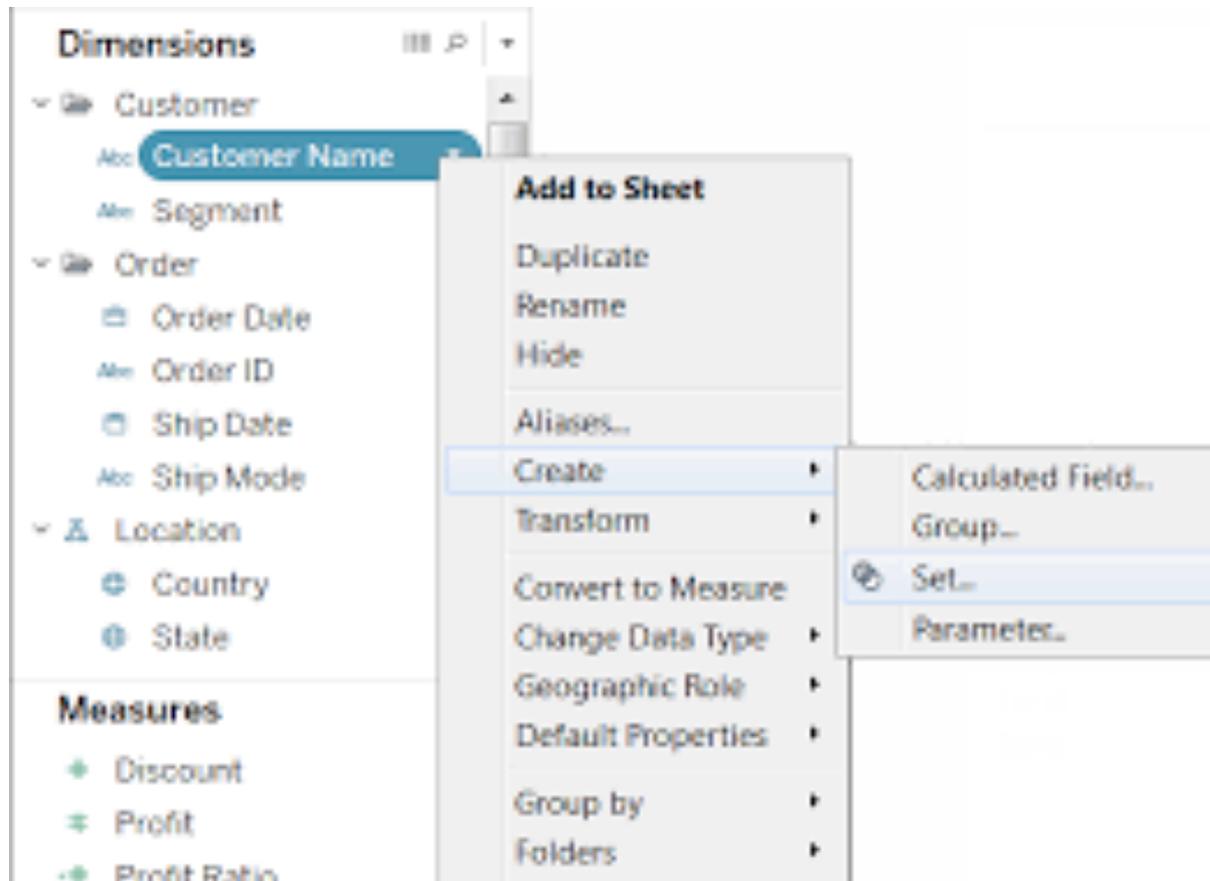
Set your Data



- Sets are used to compare and ask questions about a subset of data. Sets are custom fields that define a subset of data based on some conditions.
- Two types of sets:
 - Dynamic
 - Interactive

Set your Data

- Create a dynamic set



Set your Data

Create Set ✕

Name:

General **Condition** Top

Select from list Custom value list Use all

Enter search text

- "While you Were Out" Message Book, One Form per Page
- #10 Gummed Flap White Envelopes, 100/Box
- #10 Self-Seal White Envelopes
- #10 White Business Envelopes, 4 1/8 x 9 1/2
- #10- 4 1/8" x 9 1/2" Recycled Envelopes
- #10- 4 1/8" x 9 1/2" Security-Tint Envelopes
- #10-4 1/8" x 9 1/2" Premium Diagonal Seam Envelopes
- #6 3/4 Gummed Flap White Envelopes
- 1.7 Cubic Foot Compact "Cube" Office Refrigerators
- 1/4 Fold Party Design Invitations & White Envelopes, 24 8-...
- 12 Colored Short Pencils

Exclude

Summary

Field: [Product Name]
Selection: Selected 0 of 1850 values
Wildcard: All
Condition: None
Limit: None

Set your Data

Create Set ✕

Name:

General Condition Top

None

By field:

Sales

Range of Values

Min:

Max:

By formula:

Set your Data

Create Set ✕

Name:

General Condition **Top**

None

By field:

by

By formula:

by

Set your Data

Data Analytics

Orders+ (Sample - Supers...)

Dimensions

- Product Name
- Region
- Row ID
- Segment
- Ship Date
- Ship Mode
- State
- Sub-Category

People

Measures

- Model.txt**
 - Path
- Orders**
 - Discount
 - Profit
 - Quantity
 - Sales
- Index
- TC Date Part

Sets

- High Sales & Profit
- Top Customers

Sets

 High Sales & Profit

 Top Customers

- Create a fixed set
- The members of a fixed set do not change.
- A fixed set can be based on
 - a Single Dimension or
 - Multiple Dimensions.

Set your Data



Set your Data

Create Set ✕

Name:

Members (2 total): Exclude

Category	Sub-Category	Year of Order Date
Technology	Machines	2016
Technology	Phones	2016

Separate members by Technology, Machines, 2016 Add to Filters shelf

- Show In/Out members in a set

In most cases, when you drag a set to the visualization, the set will be displayed using the In/Out mode.

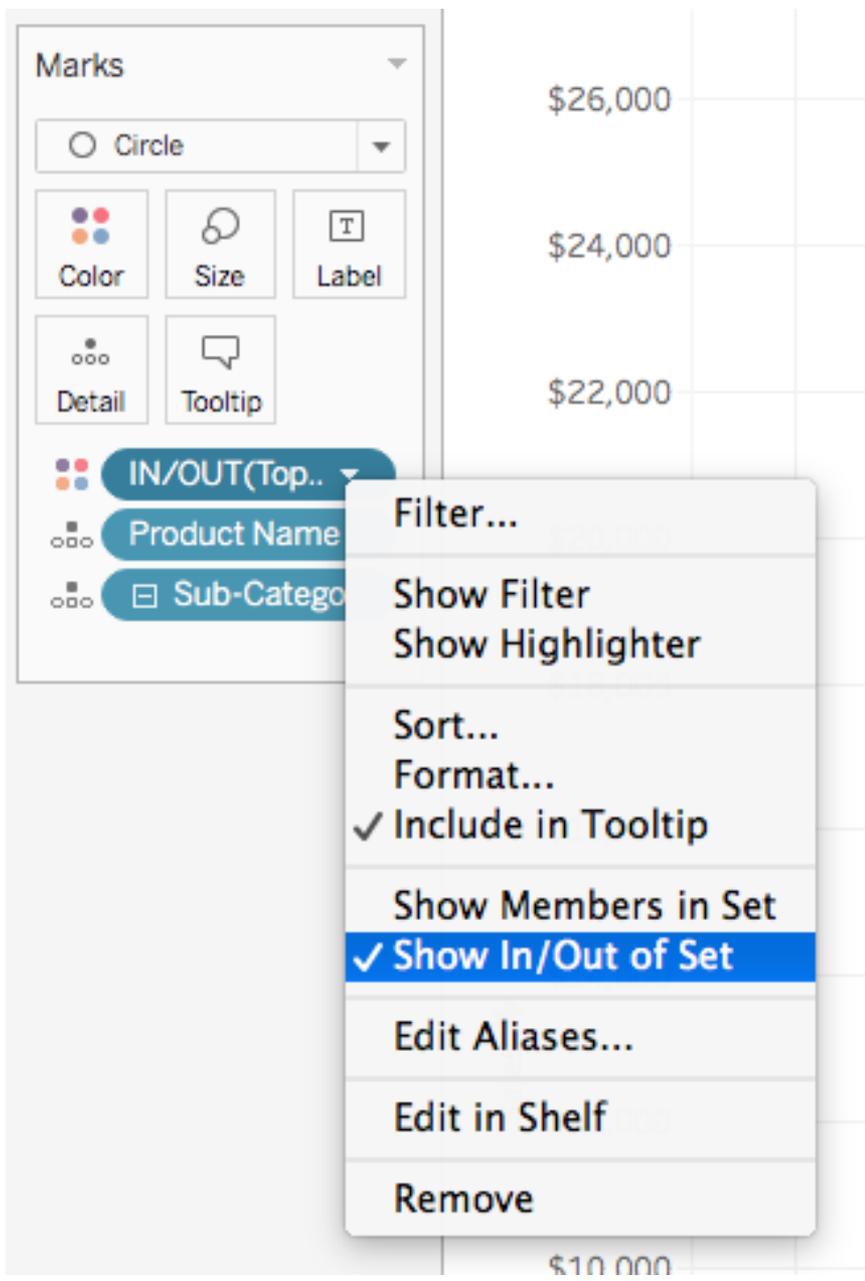
This mode separates the set into two categories:

- **In** - The members in the set.
- **Out** - Any members that are not part of the set.

- Show Members in a set

Showing the members in the Set automatically adds a filter to the view that includes only the members of the set.

Set your Data



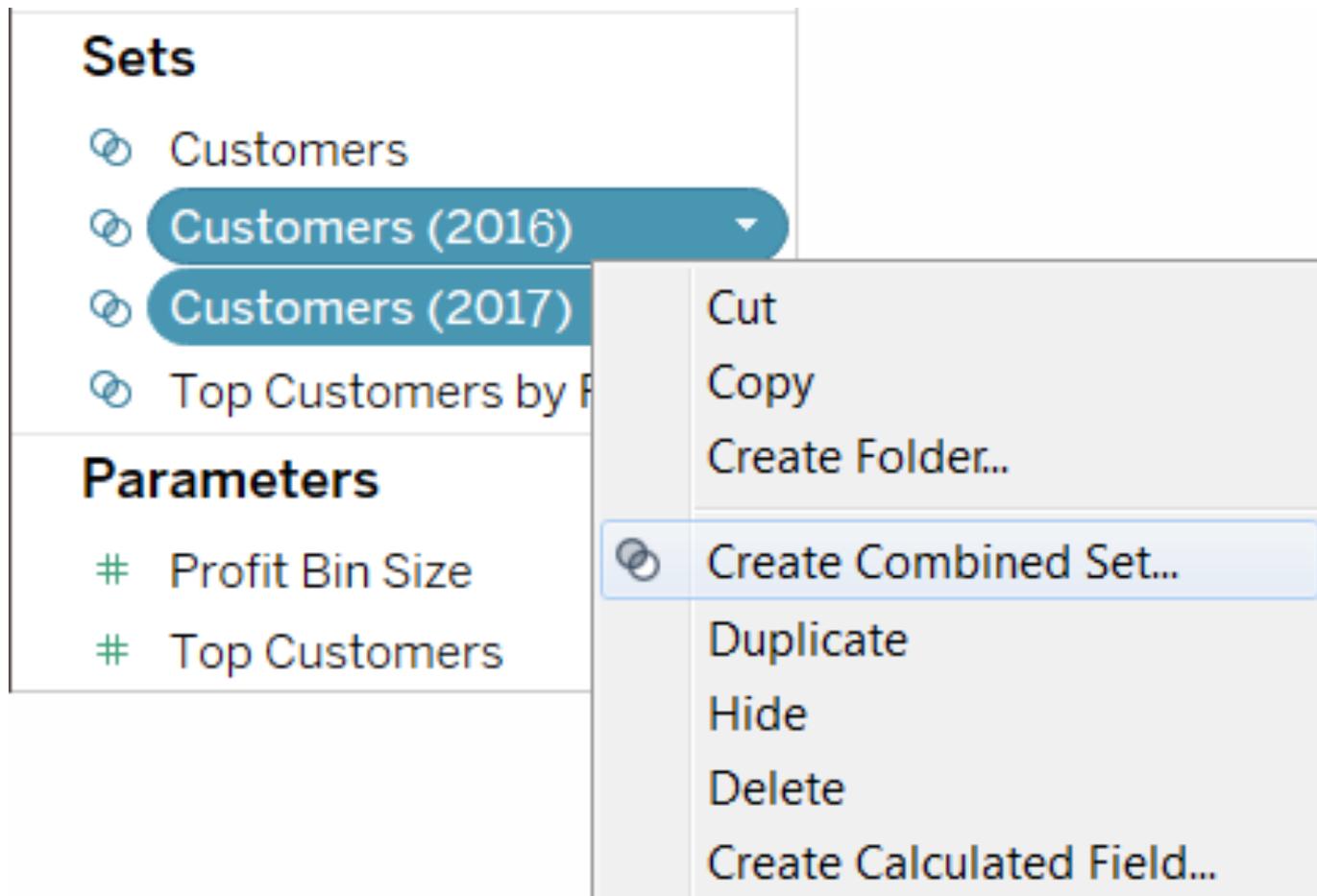
- ## Combine Sets

You can combine two sets to compare the members.

When you combine sets you create a new set containing:

- either the combination of all members, just the members that exist in both,
- or members that exist in one set but not the other.

- Combine Sets



The screenshot shows a software interface with a 'Sets' panel on the left and a context menu on the right. The 'Sets' panel has a title 'Sets' and a list of items: 'Customers', 'Customers (2016)', 'Customers (2017)', and 'Top Customers by P'. The 'Customers (2016)' and 'Customers (2017)' items are highlighted with blue bars. Below the 'Sets' panel is a 'Parameters' section with a title 'Parameters' and a list of items: '# Profit Bin Size' and '# Top Customers'. The context menu is open over the 'Customers (2016)' item and contains the following options: 'Cut', 'Copy', 'Create Folder...', 'Create Combined Set...', 'Duplicate', 'Hide', 'Delete', and 'Create Calculated Field...'. The 'Create Combined Set...' option is highlighted with a blue border.

Sets

- Customers
- Customers (2016)
- Customers (2017)
- Top Customers by P

Parameters

- # Profit Bin Size
- # Top Customers

Context Menu:

- Cut
- Copy
- Create Folder...
- Create Combined Set...
- Duplicate
- Hide
- Delete
- Create Calculated Field...

- Combine Sets

Create Set [Set 3] ✕

Name:

How would you like to combine the two sets?

Sets: 

 All members in both sets

 Shared members in both sets

 "Customers (2016)" except shared members

 "Customers (2017)" except shared members

Separate members by East, Green Tea, 2012

Group vs Set



VS



- The differences between Groups & Sets
 - The most significant difference is that sets are dynamic while groups are not.
 - Sets offer greater flexibility as they are linked to a condition.
 - Groups have only one dimension while with sets you can group across multiple dimensions
 - Sets can be combined

- The differences between Groups & Sets

- Sets are more complex but offer greater flexibility.

VS

- However, many times **a group will do the job** if the flexibility offered by the set **is not really necessary** for what you are doing.

- The differences between Groups & Sets
 - A great advantage of sets is that the data within can be exported very easily and used from the business.

VS

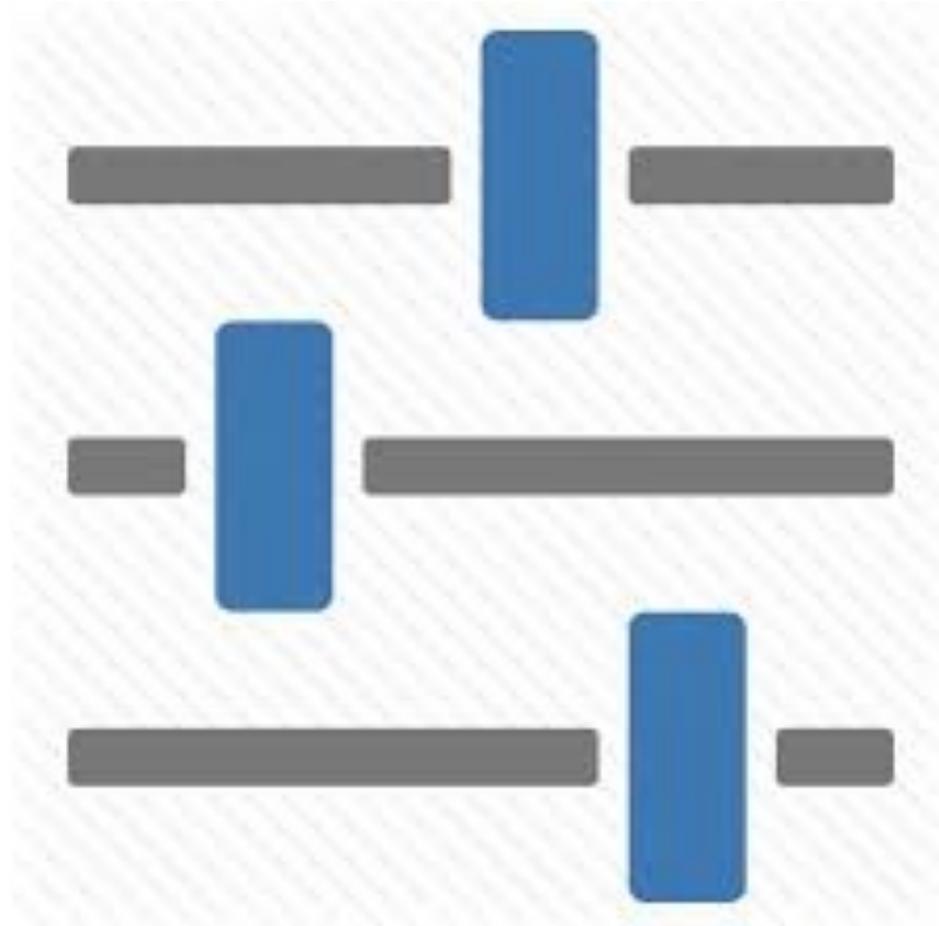
- Groups do not offer this advantage.

- The differences between Groups & Sets
 - With sets you can choose **"IN/OUT"** or **"Show Members in Set"**.
- VS
- Groups the only option is **group/ungroup**.

- Sets are extremely useful when comparing one group of things against another because of their flexibility.
- For example: A dimension member can only belong to a single category in a group. Using sets, that same dimension value ("Cats") could exist in many sets like: "Mammals", "Furry Things", "Pets".

- Sets can be referenced directly in a calculation, and since they imply a filter, they can be quite useful in a hierarchy.
- For example: You might drop a set into a hierarchy so that when someone opens that level of the hierarchy up, the values are pre-filtered by what the set does.

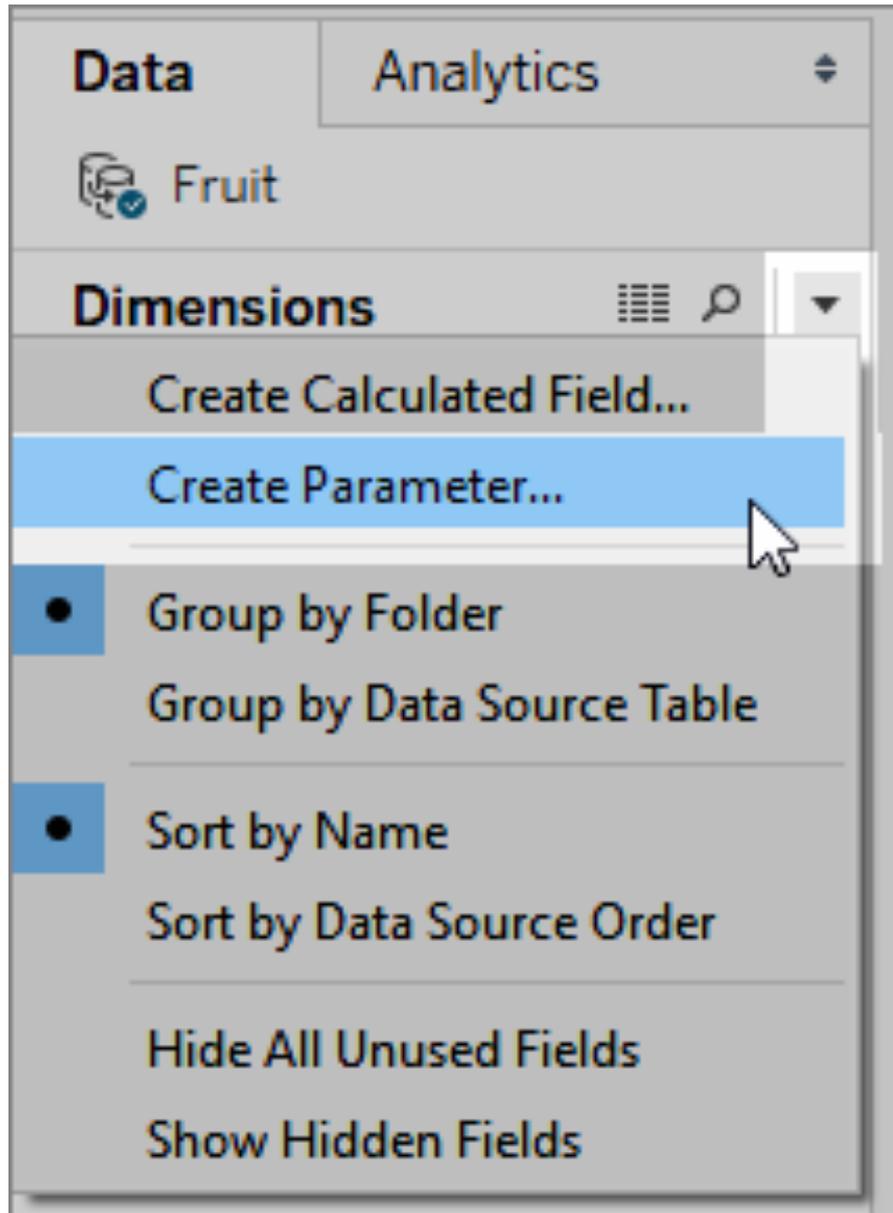
Parameters



- What Parameters are?
- Parameters are dynamic values that can replace constant values in Calculations, Filters, and Reference Lines.
- Parameters can be more dynamic and interactive by using them in Parameter Actions.

- Parameters are used:
 - In Filters
 - In Reference lines
 - in a Calculation

Parameters



Parameters

Create Parameter ✕

Name:

Properties

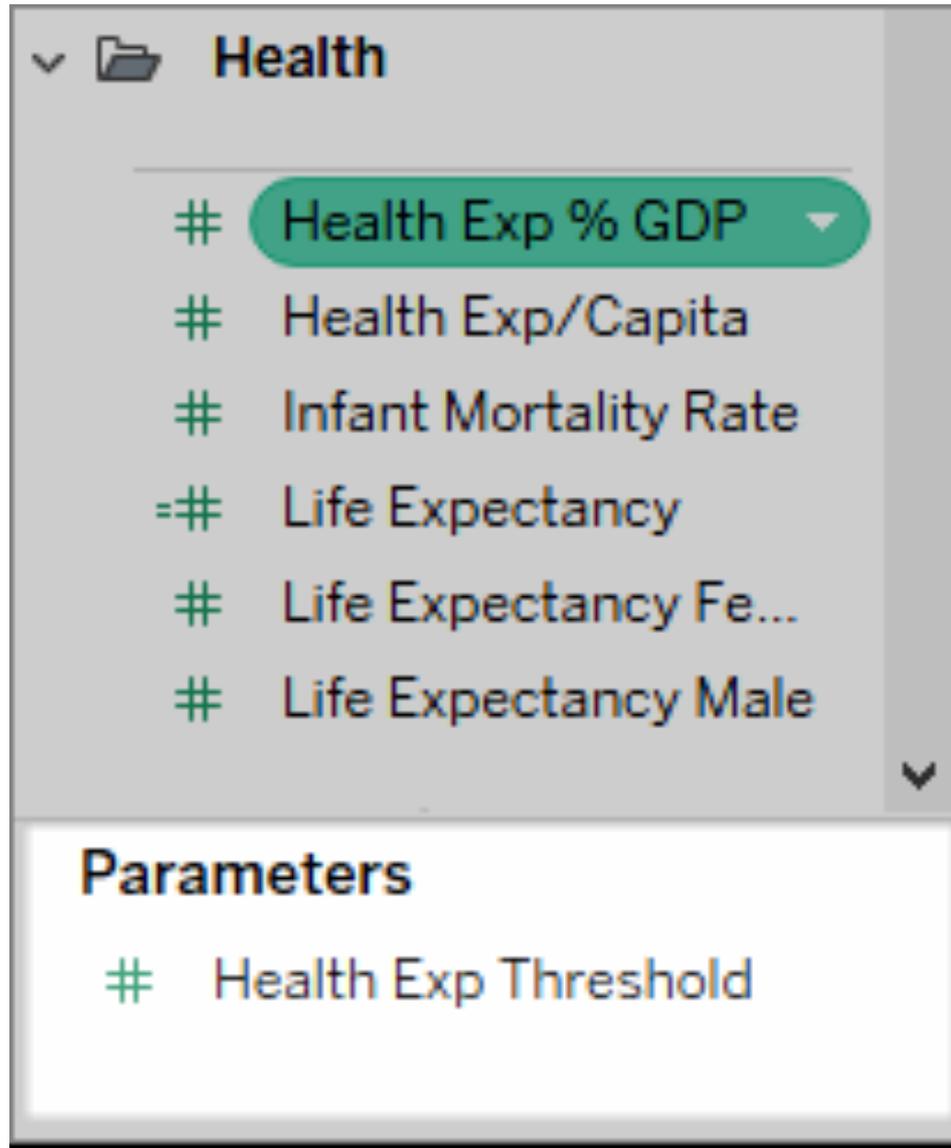
Data type:

Current value:

Display format:

Allowable values: All List Range

Parameters



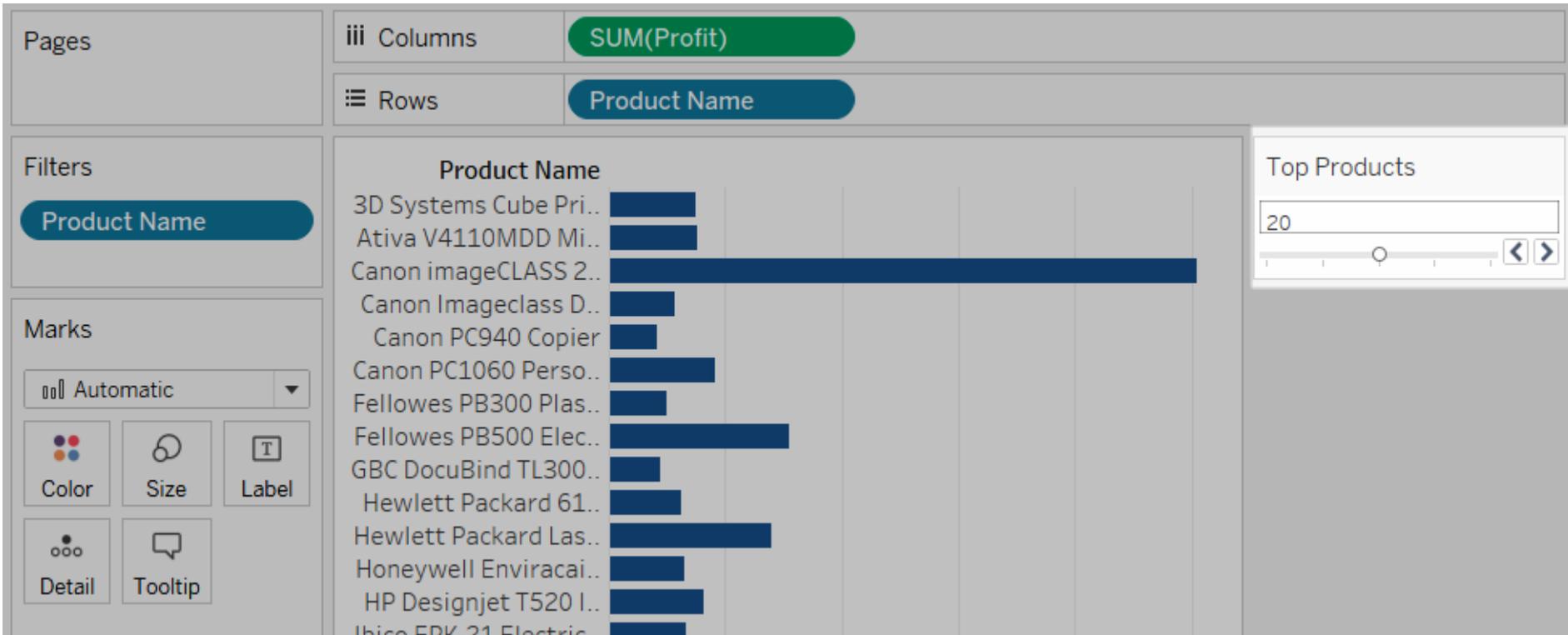
Health

- # Health Exp % GDP
- # Health Exp/Capita
- # Infant Mortality Rate
- =# Life Expectancy
- # Life Expectancy Fe...
- # Life Expectancy Male

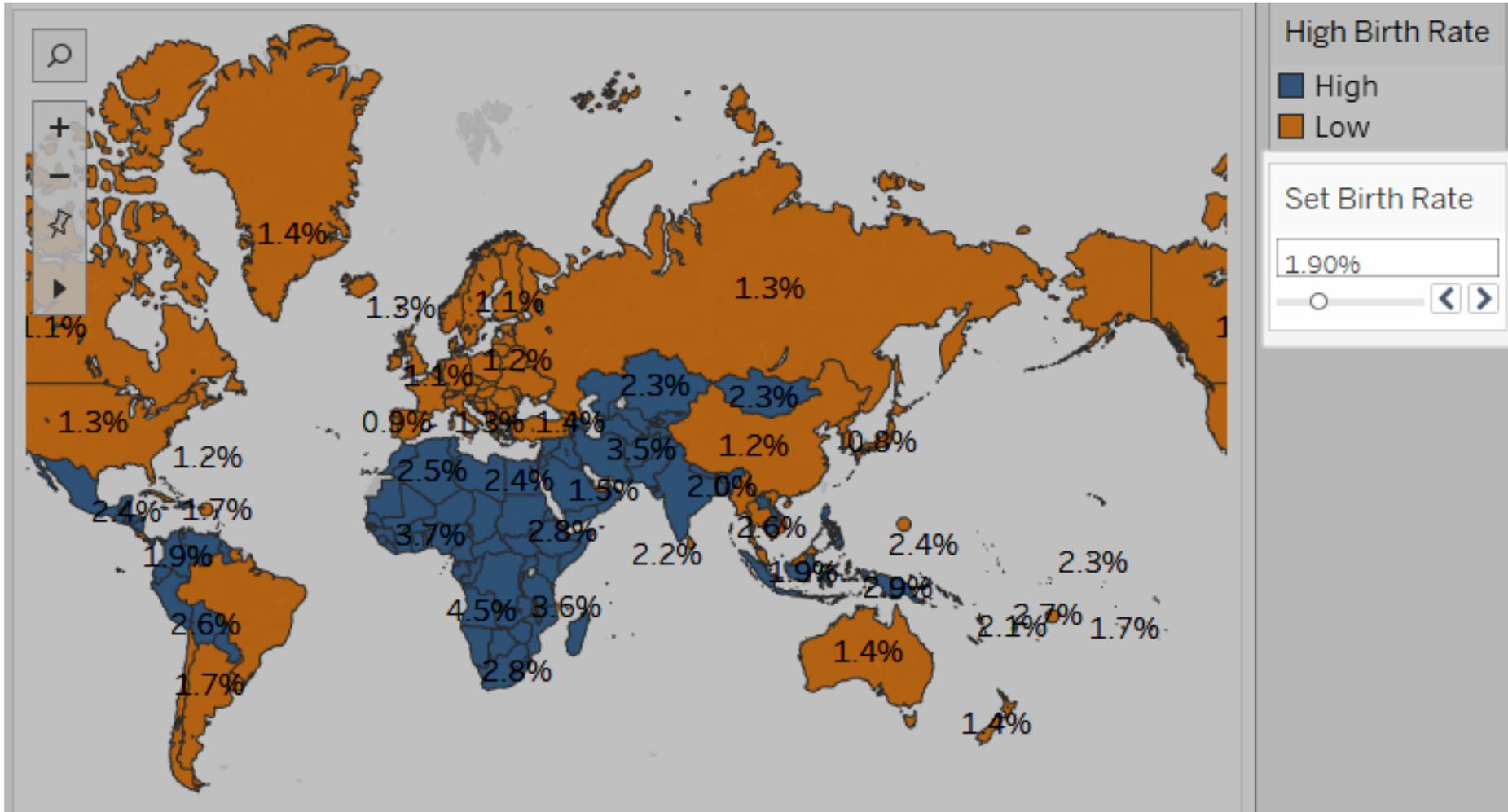
Parameters

- # Health Exp Threshold

Parameters



Parameters



Calculations



- When a calculated field is created, a new field (or column) in the data source is created, the values or members of which are determined by a calculation that you control.
- This new calculated field is saved to the data source in Tableau, and can be used to create more robust visualizations.

- Calculations are used:
 - To Segment data
 - To Convert the data type of a field, such as converting a string to a date.
 - To Aggregate data
 - To Filter results
 - To Calculate Ratios

- Types of Calculations

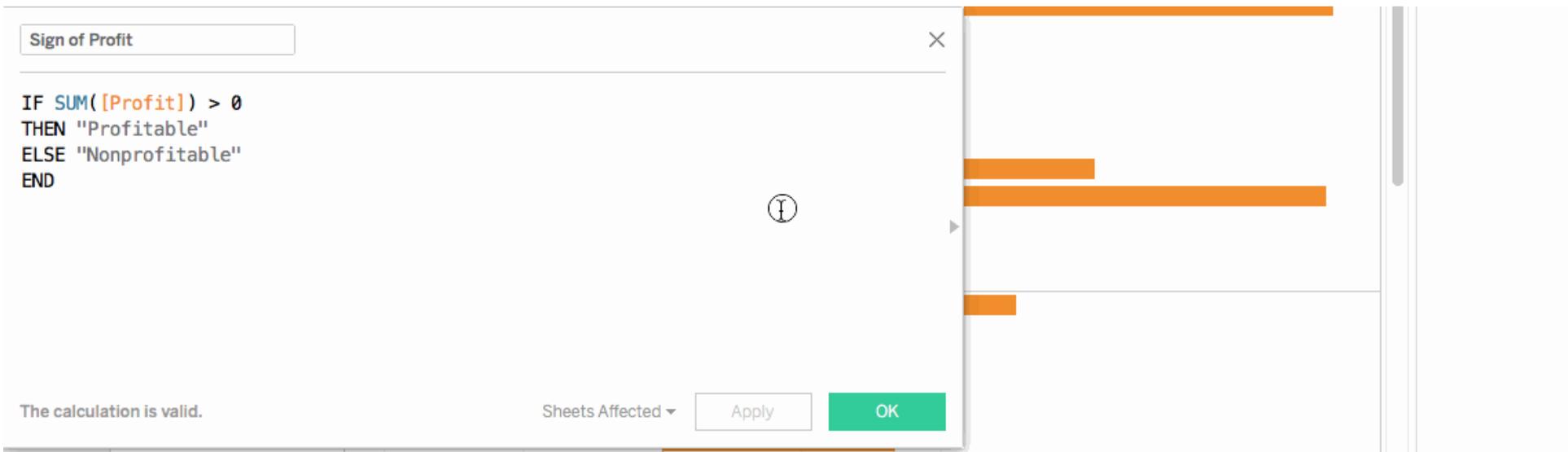
There are three main types of calculations:

- Basic calculations
- Table calculations
- Level of Detail (LoD) expressions

- **Basic Calculations**

Basic calculations allow you to transform values or members at the data source level of detail (a row-level calculation) or at the visualization level of detail (an aggregate calculation).

- Basic Calculations



Sign of Profit

```
IF SUM([Profit]) > 0  
THEN "Profitable"  
ELSE "Nonprofitable"  
END
```

The calculation is valid.

Sheets Affected ▾ Apply OK

- Basic Calculations

×

`[Sales] - [Profit]`

The calculation is valid. 1 Dependency ▾

Calculations

Pages

Filters

Segment

Columns: SUM(Profit) | SUM(Cost)

Rows: Segment | Sub-Category

Filters

Segment

Marks

All

Automatic

Colour | Size | Label

Detail | Tooltip

SUM(Profit)

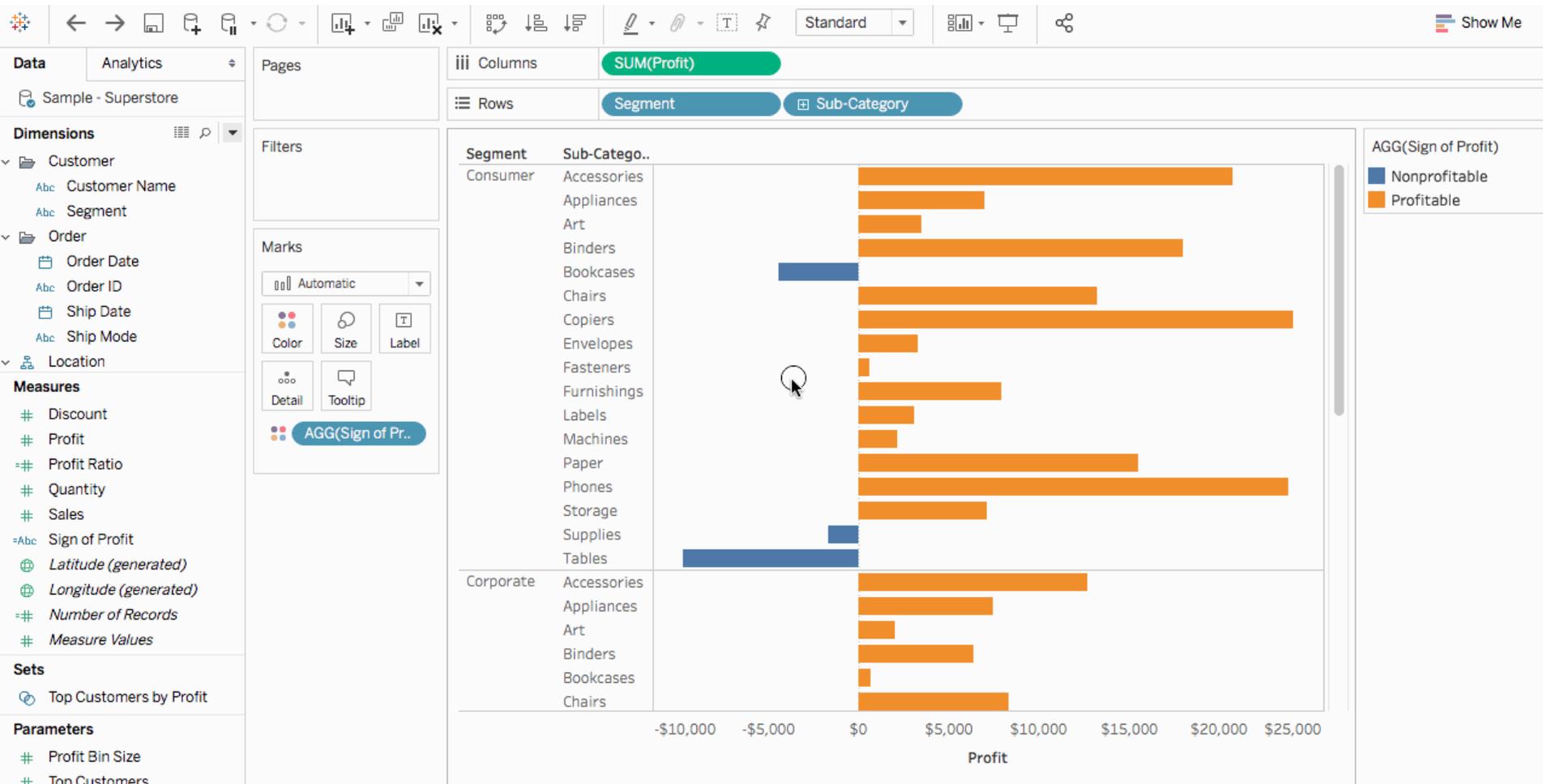
SUM(Profit) | SUM(Cost)

SUM(Profit)

-19,456 | 48,167



Calculations



- **Table Calculations**

Table calculations allow you to transform values at the level of detail of the visualization only.

Table calculations are a special type of calculated field that computes on the local data in Tableau.

They are calculated based on what is currently in the visualization and do not consider any measures or dimensions that are filtered out of the visualization.

- Table Calculations

Table calculations are used for a variety of purposes, including:

- Transforming values to rankings
- Transforming values to show running totals
- Transforming values to show percent of total
- Addressing and Partitioning

Calculations

Table Calculation ×
Difference in Sales

Calculation Type
Difference From ▼

Compute Using

- Table (across)
- Table (down)
- Table (across then down)
- Table (down then across)
- Pane (down)
- Pane (across then down)
- Pane (down then across)
- Cell

Specific Dimensions

- Year of Order Date
- Month of Order Date
- Quarter of Order Date

At the level ▼

Relative to Previous ▼

Show calculation assistance

Calculations

Pages

Columns

YEAR(Order Date)

Rows

QUARTER(Order D..

MONTH(Order Dat..

Filters

Marks

Automatic

Color

Size

Text

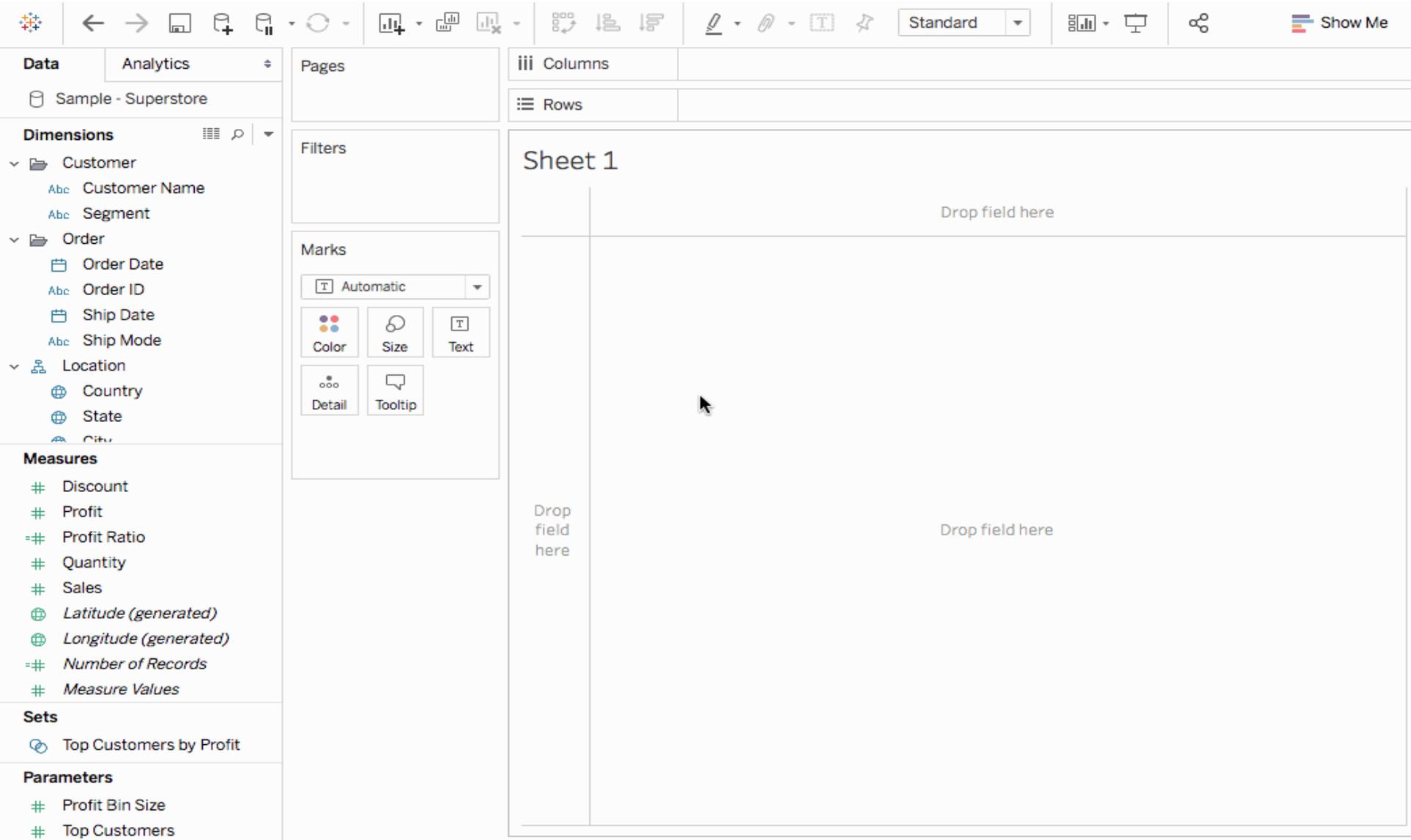
Detail

Tooltip

SUM(Sales)

		Order Date			
Quarter of Orde..	Month of Order ..	2011	2012	2013	2014
Q1	January		\$4,238	\$268	\$26,111
	February		\$7,400	\$10,657	-\$2,584
	March		-\$17,224	\$12,719	\$2,723
Q2	April		\$5,900	\$5,053	\$864
	May		\$6,483	\$26,559	-\$11,040
	June		-\$9,798	\$14,633	\$8,829
Q3	July		-\$5,181	\$9,675	\$9,988
	August		\$8,989	-\$3,633	\$28,251
	September		-\$17,181	\$8,312	\$17,581
Q4	October		-\$48	\$25,058	\$21,331
	November		-\$2,656	\$6,220	\$30,134
	December		\$5,374	\$22,318	-\$6,763

Calculations



Data | Analytics | Sample - Superstore

Dimensions

- Customer
 - Customer Name
 - Segment
- Order
 - Order Date
 - Order ID
 - Ship Date
 - Ship Mode
- Location
 - Country
 - State
 - City

Measures

- Discount
- Profit
- Profit Ratio
- Quantity
- Sales
- Latitude (generated)
- Longitude (generated)
- Number of Records
- Measure Values

Sets

- Top Customers by Profit

Parameters

- Profit Bin Size
- Top Customers

Columns

Rows

Filters

Marks

Automatic

Color | Size | Text

Detail | Tooltip

Sheet 1

Drop field here

Drop field here

Drop field here

Basic Calculations vs Table Calculations

- Table Calculations and Calculated Fields are similar in the sense that they both use functions to compute the results.
- The difference is how and where the computing takes place, where the result is saved and if it can be reused in more worksheets.

- Table Calculations are simpler and their scope is more limited compared to Calculated Fields.
- Calculated fields are much more diverse enabling deeper analysis.

- Level of Detail (LoD) expressions

LOD Calculations allows to compute values at the data source level and the visualization level (like basic calculations).

However, LOD calculations give more control on the level of granularity.

- Level of Detail (LoD) expressions
They can be performed at:
 - a more granular level (**INCLUDE**),
 - a less granular level (**EXCLUDE**),
 - or an entirely independent level (**FIXED**)
with respect to the granularity of the
visualization.

- **INCLUDE**

INCLUDE level of detail expressions compute values using the specified dimensions in addition to whatever dimensions are in the view.

INCLUDE can be useful when you want to calculate at a fine level of detail in the database and then re-aggregate and show at a coarser level of detail in your view.

- **INCLUDE**

Sales by State ×

```
{ INCLUDE [State] : SUM (Sales) }
```

The calculation is valid. 2 Dependencies ▾ Apply OK

Calculations

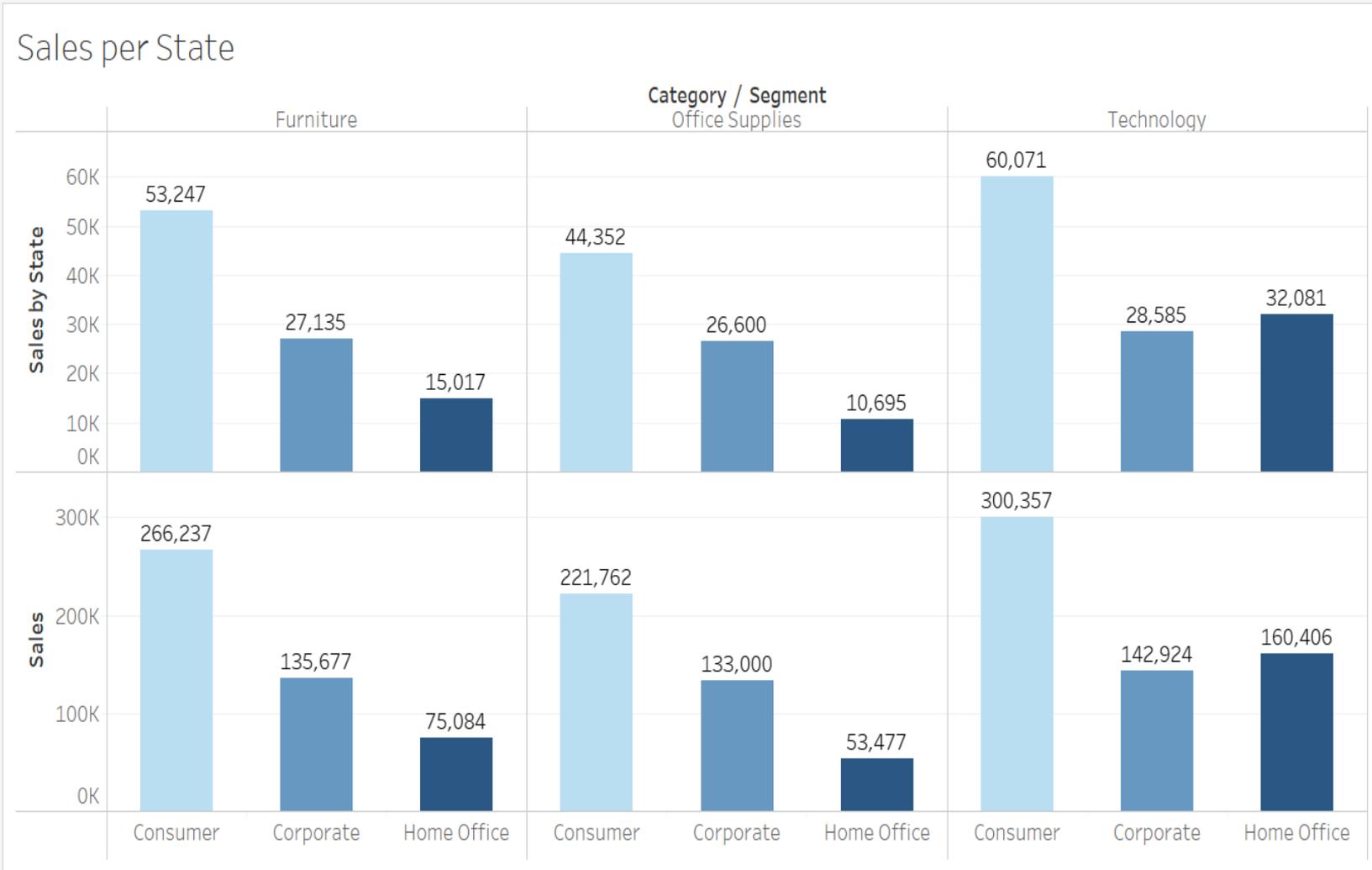
Pages

Columns: Category, Segment

Rows: AVG(Sales by State), SUM(Sales)

Filters: State

Marks: All, Automatic, Colour, Size, Label, Detail, Tooltip, Segment, SUM(Sales), AVG(Sales by...), SUM(Sales)



- EXCLUDE

EXCLUDE level of detail expressions declare dimensions to ignore from the view level of detail.

EXCLUDE can be useful for ‘percent of total’ or ‘difference from overall average’ scenarios. They are comparable to such features as Totals and Reference Lines.

- EXCLUDE

Exclude Region ×

```
{EXCLUDE [Region] :SUM ([Sales]) }
```

The calculation is valid. 1 Dependency ▾ Apply OK

Calculations

Pages

Columns: Region, SUM(Sales)

Rows: MONTH(Order D..)

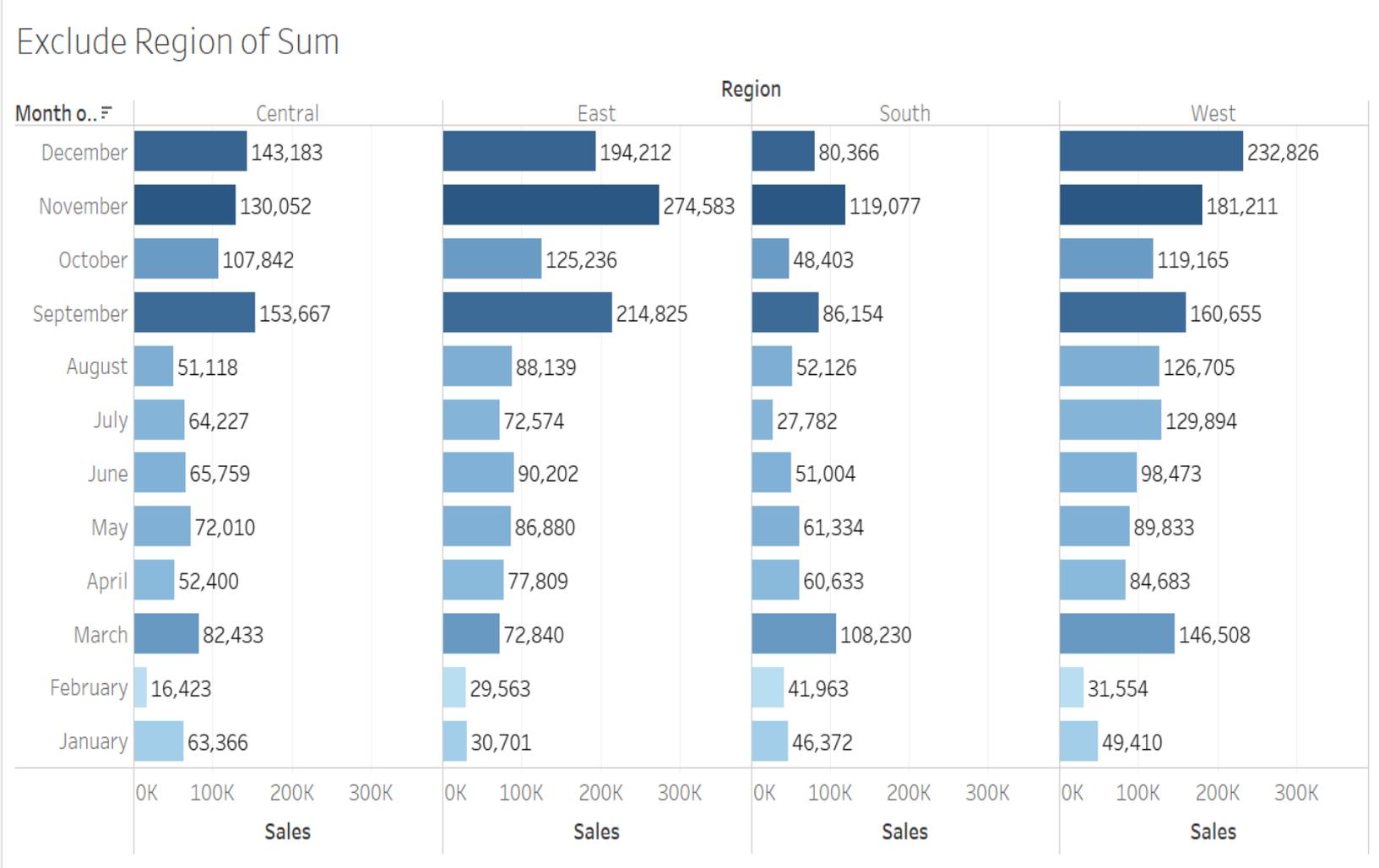
Filters

Marks

- Bar
- Colour
- Size
- Label
- Detail
- Tooltip
- ATTR(Exclude ..)
- SUM(Sales)

ATTR(Exclude Region)

119,503 704,922



- **FIXED**

FIXED level of detail expressions compute a value using the specified dimensions, without reference to the dimensions in the view.

FIXED level of detail expressions do not consider the view level of detail, the calculation only uses the dimension referenced in the calculation.

- **FIXED**

Sales by Region ×

```
{FIXED[Region]:SUM([Sales])}
```

The calculation is valid. 1 Dependency ▾ Apply OK

Calculations

Pages

Columns

Rows

Region State

Filters

Sheet 5

Region State

Region	State	
Central	Illinois	501,240
	Indiana	501,240
	Iowa	501,240
	Kansas	501,240
	Michigan	501,240
	Minnesota	501,240
	Missouri	501,240
	Nebraska	501,240
	North Dakota	501,240
	Oklahoma	501,240
East	South Dakota	501,240
	Texas	501,240
	Wisconsin	501,240
	Connecticut	678,781
	Delaware	678,781
	District of Columbia	678,781
	Maine	678,781
	Maryland	678,781
	Massachusetts	678,781
	New Hampshire	678,781
New Jersey	678,781	

Marks

Text

Color Size Text

Detail Tooltip

Region

SUM(Sales By ..)

Region

- Central
- East
- South
- West

- LOD expression syntax

A level of detail expression has the following structure:

{[FIXED | INCLUDE | EXCLUDE] <*dimension declaration*> : <*aggregate expression*>}

- Ad-Hoc Calculations

Ad-hoc calculations are calculations that you can create and update as you work with a field on a shelf in the view.

- Ad-Hoc Calculations
 - Ad-hoc calculations are supported on the **Rows**, **Columns**, **Marks**, and **Measure Values** shelves;
 - they are not supported on the **Filters** or **Pages** shelves.

- Ad-Hoc Calculations

iii Columns	<input type="text" value="YEAR(Order Date)"/>
☰ Rows	<input type="text"/>

- Ad-Hoc Calculations

iii Columns	SUM(Sales)
☰ Rows	Sub-Category $\frac{\text{SUM}([\text{Profit}])}{\text{SUM}([\text{Sales}])}$
Apply (Ctrl+Enter)	

- Ad-Hoc Calculations

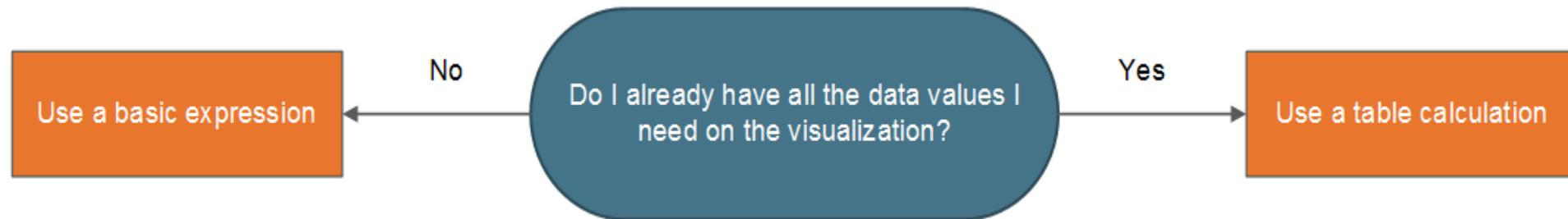
iii Columns	⊕ YEAR(Order Date)
☰ Rows	AGG(SUM([Profit])/S..

Tips for Calculations

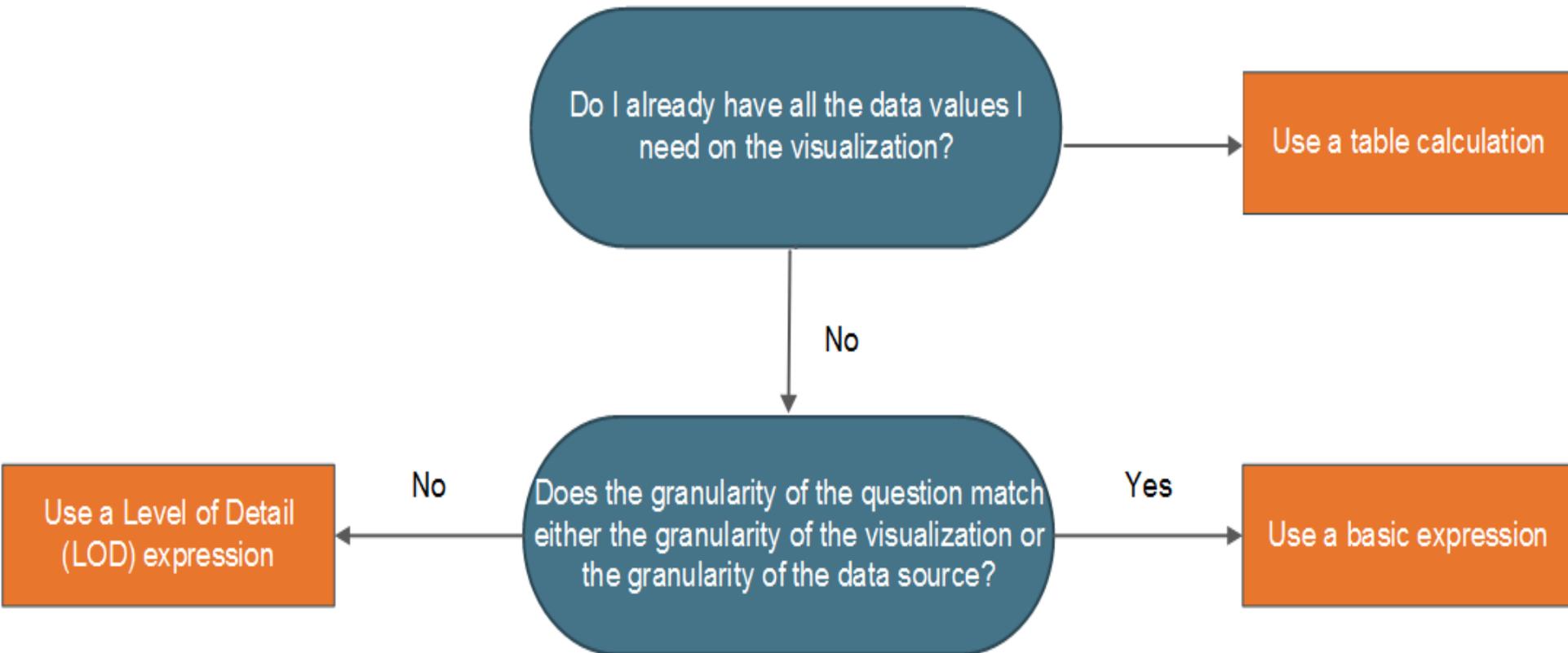


- Which calculation is right for my analysis?
 - Depends on the needs of your analysis and the question you need to answer.
 - When trying to decide, consider the following questions

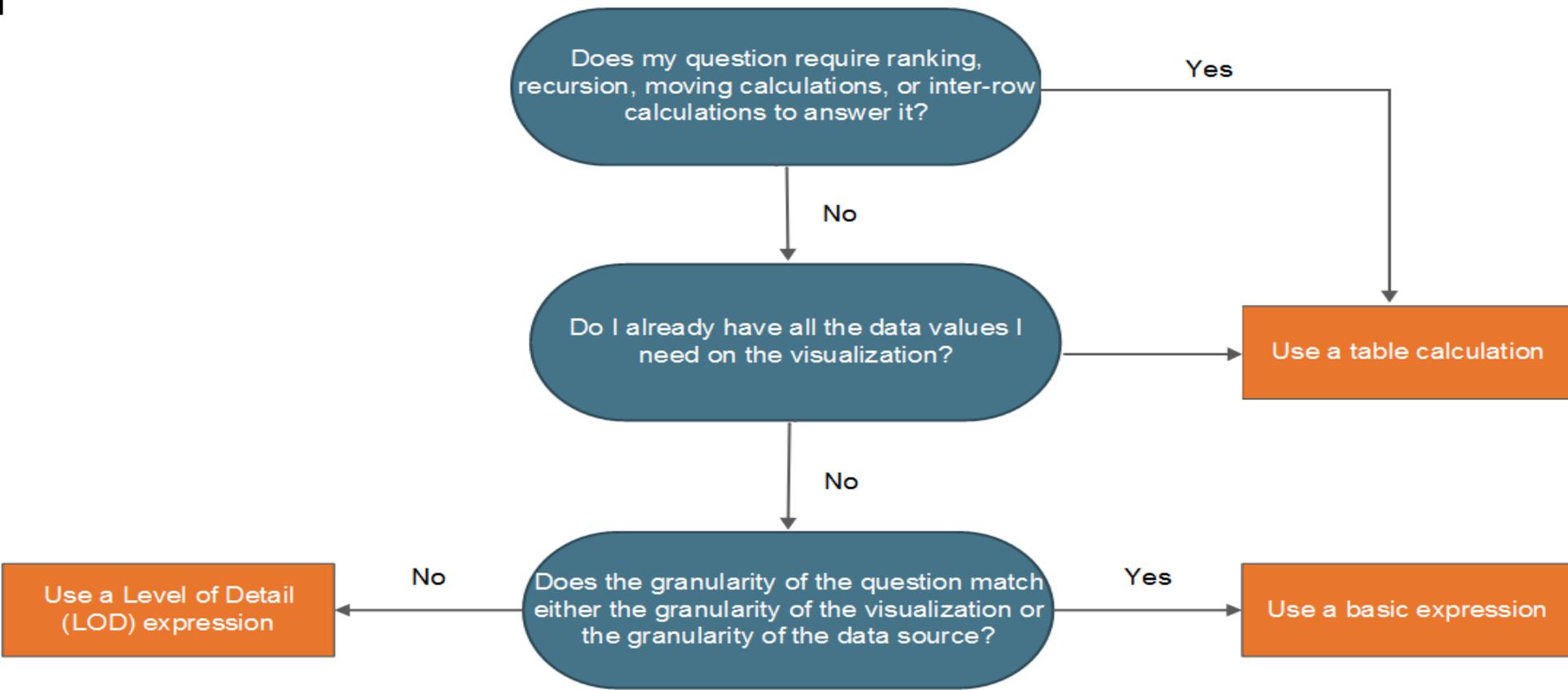
- Q1: Do you already have all the data values you need on the visualization?



- Q2: Does the granularity of the question match the granularity of the visualization or the granularity of the data source?



- Q3: Do I need Ranking, Inter-row calculations, Moving calculations or Recursion?



- Tips for Learning How to Create Calculations

There is no easy way to know exactly how to create the perfect formula; it takes **practice and research**.

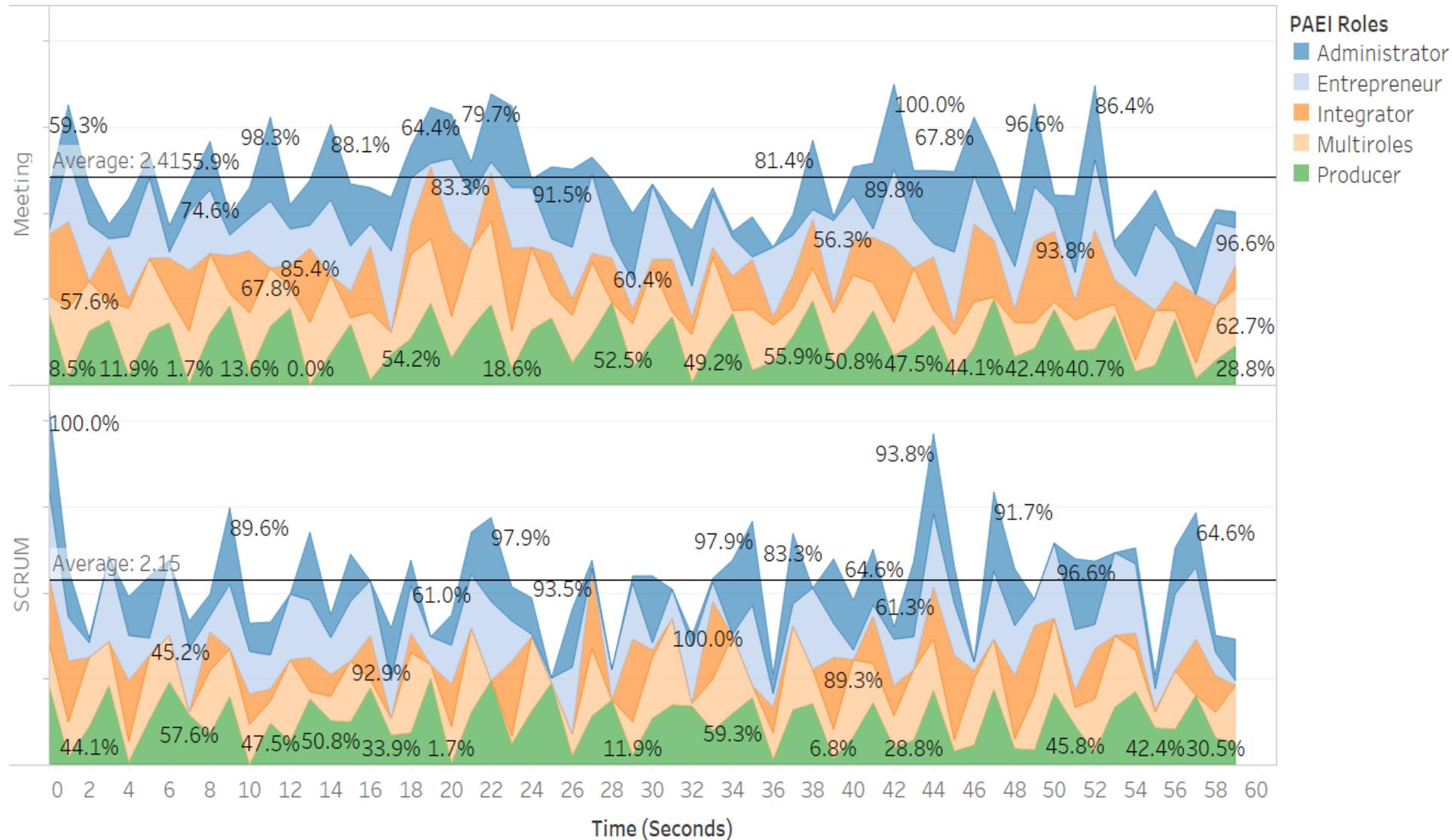
- Tips for Learning How to Create Calculations
 - **Know your question or purpose.** If you know the type of data you need, this can help you choose the correct function, as well as format your formula properly.
 - **Learn Tableau functions.** There are many different functions available. Each type serves a different purpose.

- Tips for Learning How to Create Calculations
 - **Learn how to format calculations.** Once you are familiar with the different types of Tableau functions and their purpose, make sure to learn how to format calculations using the proper syntax.
 - **Learn from other examples.**

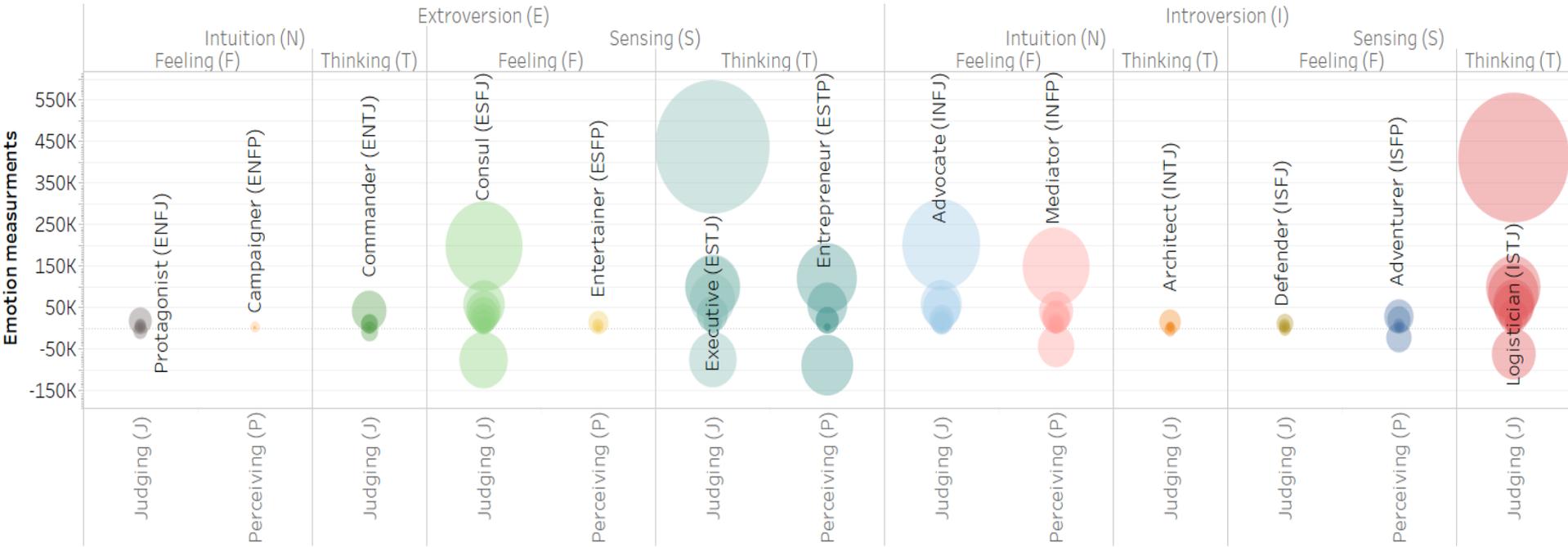
Examples of data analytics in Intelligent Environments

- Smart sensor data
- Augmented reality (Google Glass) data
- Personalised assessment data
- Student presentation data

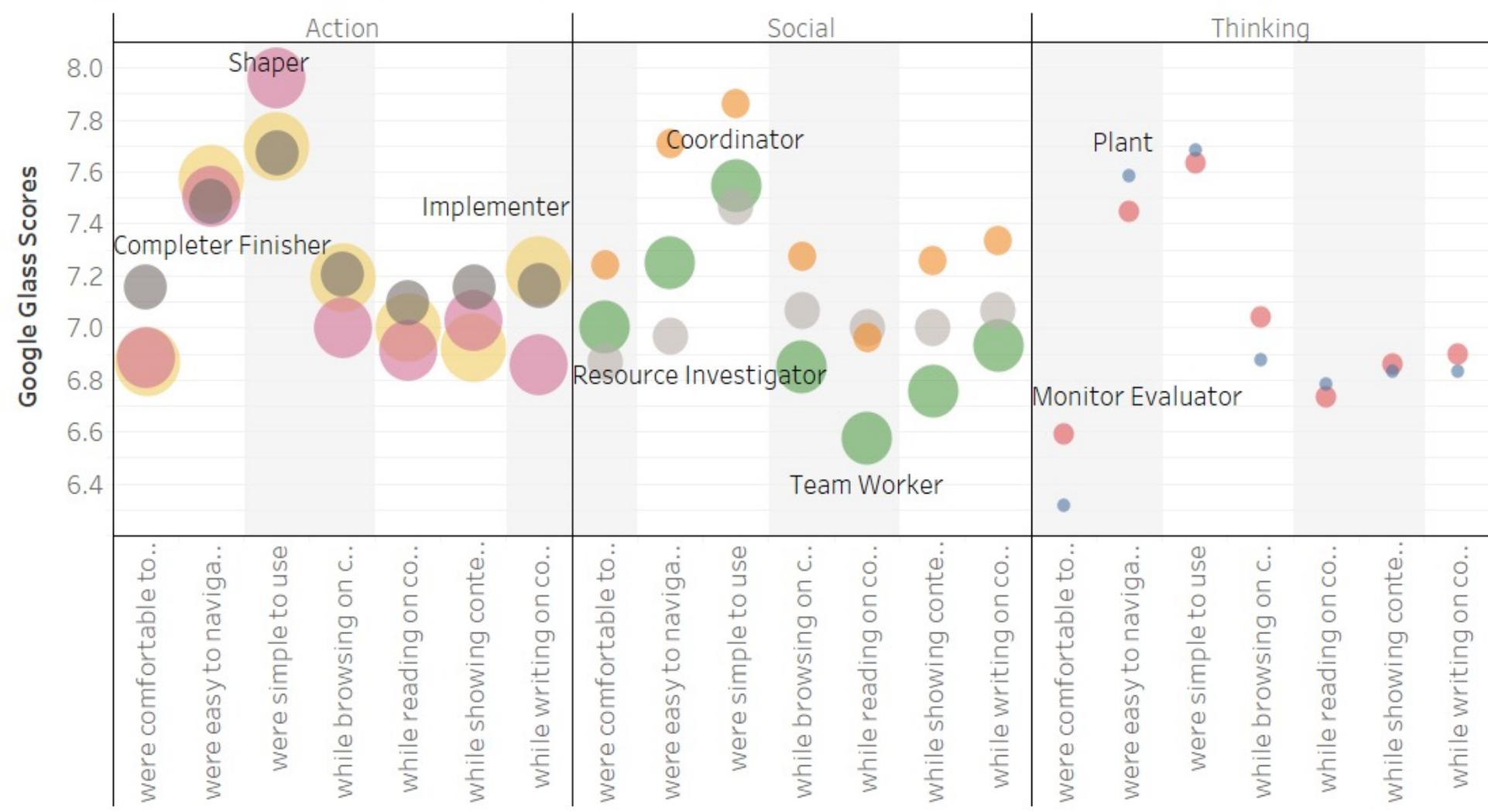
Smart Sensors



Smart Sensors

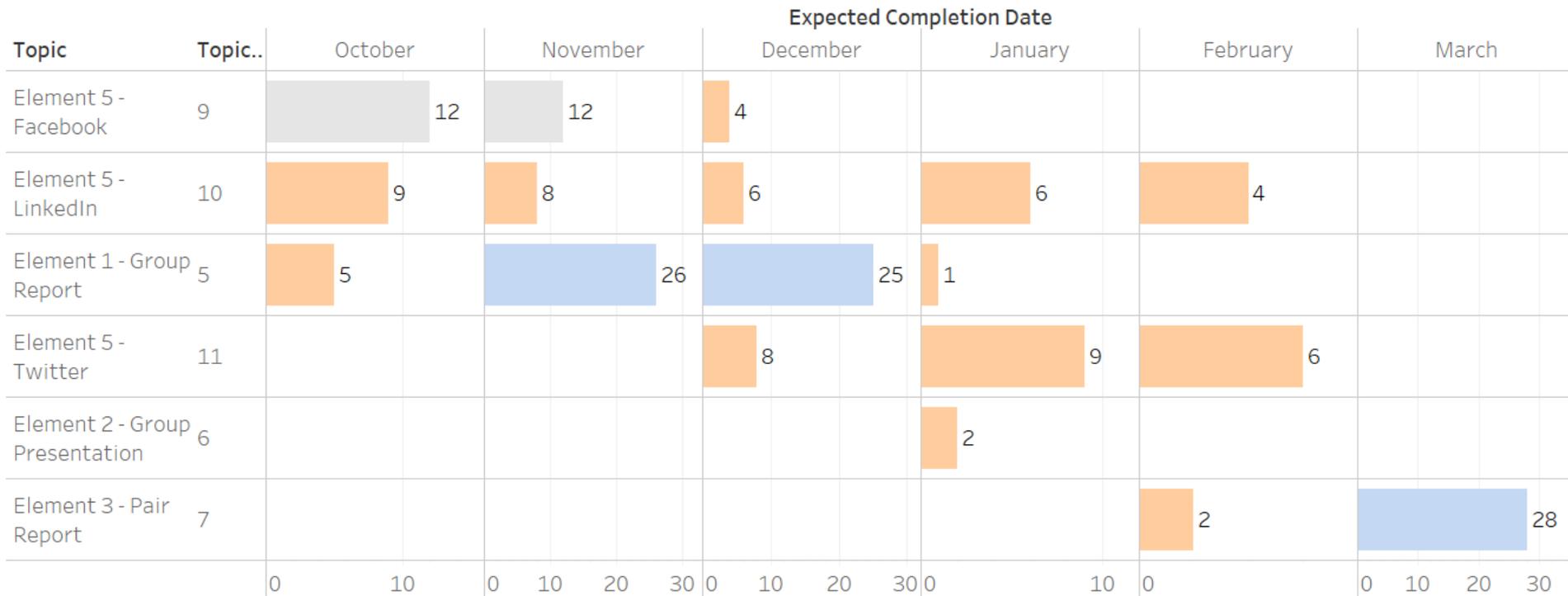


Google Glasses responses by Role



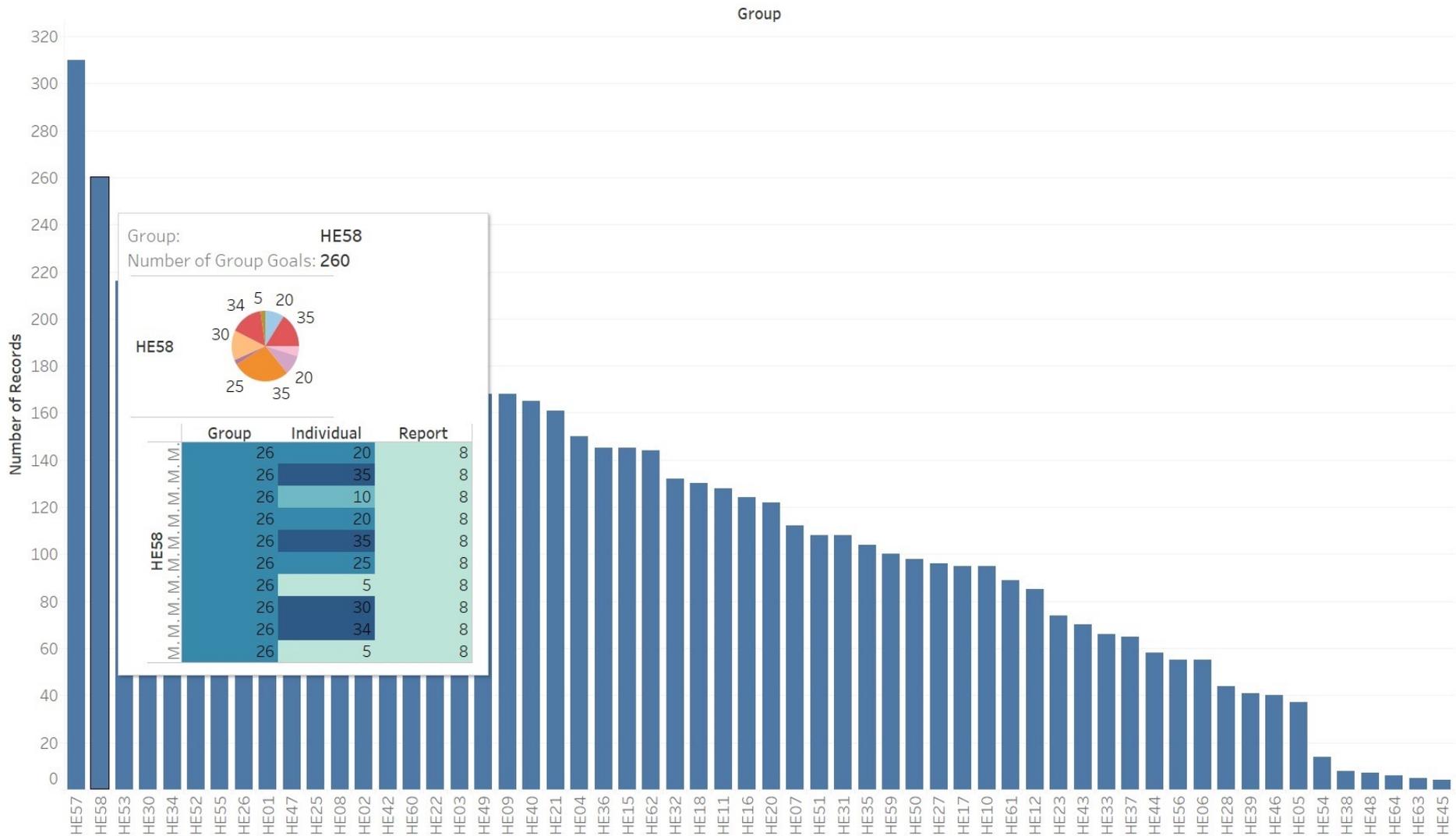
SOBs

SOBs per Topic



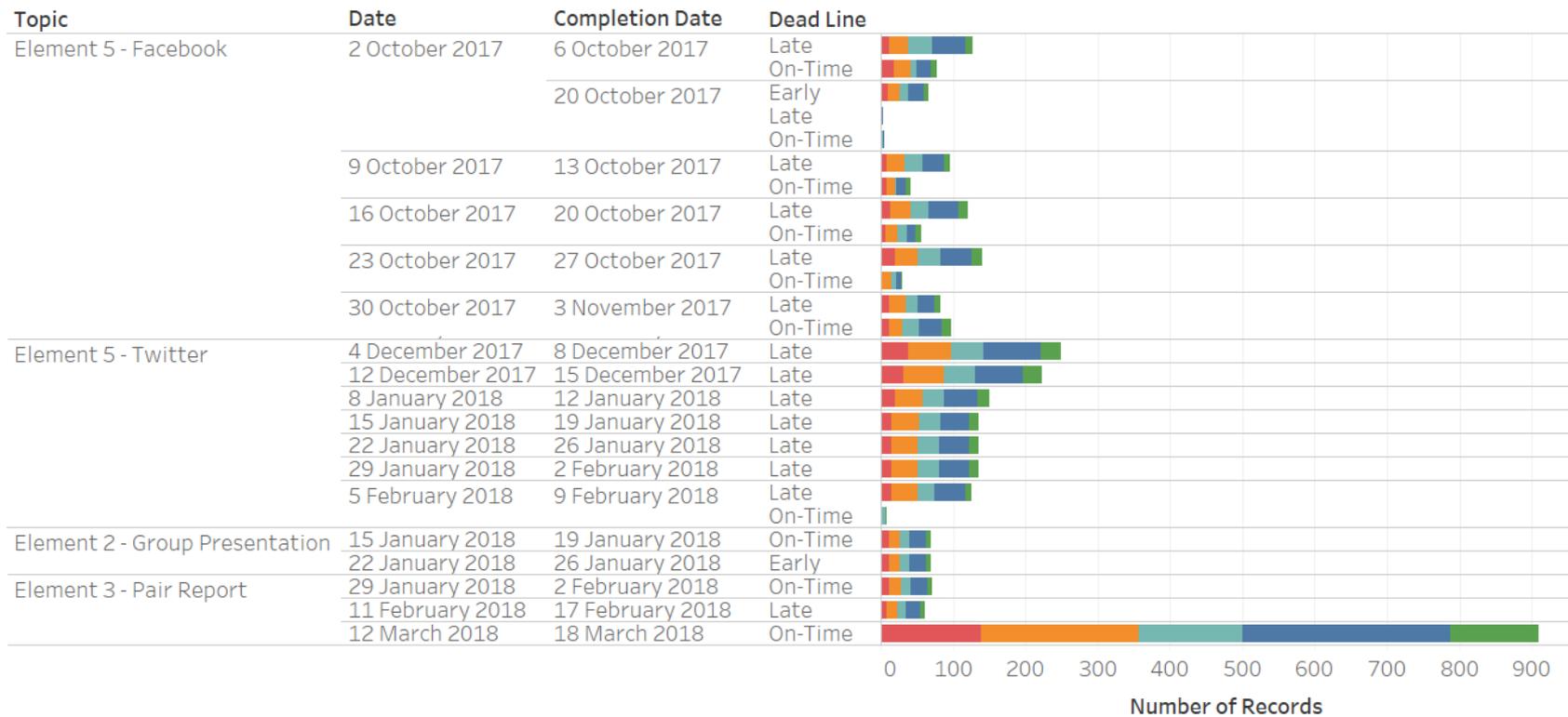
GOALS

Group Goals



SOBs per VARK

Observations per Topic & VARK

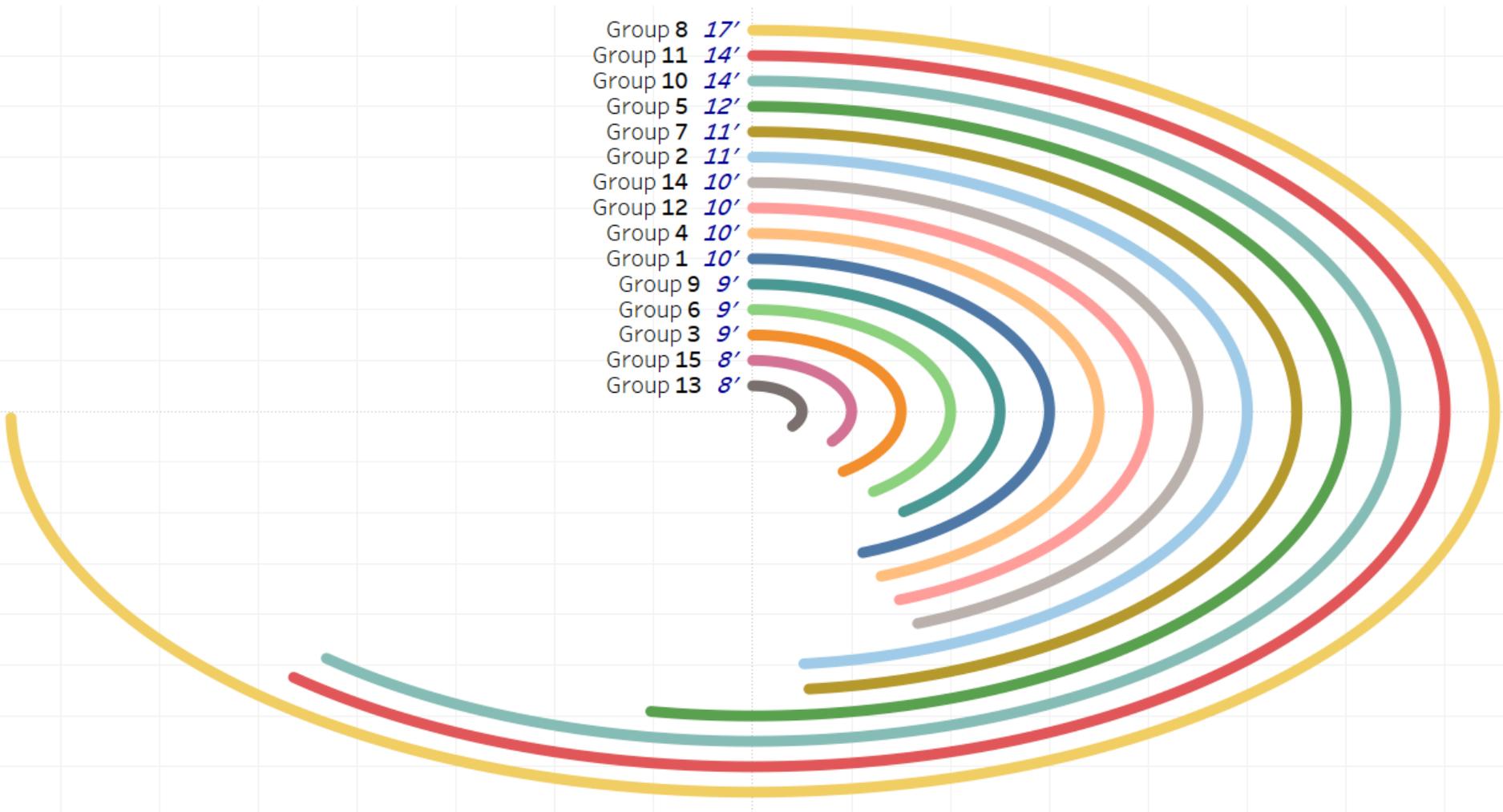


VARK Modalities

- Visual
- Aural
- Read\Write
- Kinaesthetic
- Multimodal

Group Presentations

Group 8 17'
Group 11 14'
Group 10 14'
Group 5 12'
Group 7 11'
Group 2 11'
Group 14 10'
Group 12 10'
Group 4 10'
Group 1 10'
Group 9 9'
Group 6 9'
Group 3 9'
Group 15 8'
Group 13 8'



Questions



a.tsiakara@mdx.ac.uk