

# Information Visualization

Course Module IN6221

**Information Visualization Tool**  
**Tableau**

WKW School of Communication and Information,  
NTU

# Tableau Installation

- <https://www.tableau.com/academic/students>

The screenshot shows the Tableau website's academic section. At the top, there's a navigation bar with the Tableau logo, PRICING, SIGN IN, and a search icon. Below the navigation, a sub-menu for 'ACADEMIC PROGRAMS' is visible with options like 'Why Tableau', 'Products', 'Solutions', 'Resources', and 'Partners'. A prominent blue button labeled 'FREE ACADEMIC LICENSE' is located on the right. The main content area features a large, stylized graphic of a forest with green trees and blue squares. The text 'Tableau for Students' is centered, with a subtext 'Download Tableau Desktop and Tableau Prep Builder to gain data skills'. A blue button below the subtext says 'GET TABLEAU FOR FREE'. At the bottom, a paragraph explains the program: 'We offer free one-year Tableau licenses to students at accredited academic institutions through our Tableau for Students program. Receive access to our entire eLearning suite once verified.'

PRICING SIGN IN

Why Tableau ▾ Products ▾ Solutions ▾ Resources ▾ Partners ▾

FREE ACADEMIC LICENSE

ACADEMIC PROGRAMS

## Tableau for Students

Download Tableau Desktop and Tableau Prep Builder to gain data skills

GET TABLEAU FOR FREE

We offer free one-year Tableau licenses to students at accredited academic institutions through our Tableau for Students program. Receive access to our entire eLearning suite once verified.

# Tableau Installation

You're almost there!

Are you an instructor? Visit [tableau.com/teaching](https://tableau.com/teaching) to request your license.

Students at accredited academic institutions worldwide are eligible for a free one-year license to activate Tableau Desktop and Tableau Prep. Complete the form below to confirm your eligibility and unlock your new free license. You must be 16 years of age or older to request a license.  
How does verifying work?

Country/Region

Country/Region (of school)

Select One

Personal information

Legal First Name\*

Legal First Name

Legal Last Name\*

Legal Last Name

Email\*

Email

I use Tableau

Confirm Email

Date of birth\*

At home

Month Day Year

School Information

School Name\*

School Name

How will you be using your Tableau license?\*

Select one

VERIFY STUDENT STATUS

Are you an instructor? Visit [tableau.com/teaching](https://tableau.com/teaching) to request your license.

Information entered into this web form will be used for verification purposes. The information will be shared with Tableau. All use of Tableau products must comply with United States export control and economic sanction laws. [Tableau Privacy Policy](#)

Verification services powered by SheerID. [SheerID FAQs](#)

Tableau will send you a download link for Tableau Desktop Including the activation key.

# Introduction to Tableau

- Tableau is a widely used business intelligence (BI) and analytics software trusted by companies like Amazon, Lenovo, and Honeywell to **explore**, **visualize**, and **securely share** data in the form of **Worksheets**, **Dashboards**, and **Stories**.



## The Pros of Tableau Software

- Remarkable Visualization Capabilities
- Ease of Use
- High Performance
- Multiple Data Source Connections
- Thriving Community and Forum
- Mobile-Friendliness

Commercial Alternatives



# Interface Overview

Tableau - Book1

File Data Server Help

Connect

Search for Data

Tableau Server

To a File

- Microsoft Excel
- Text file
- JSON file
- Microsoft Access
- PDF file
- Spatial file
- Statistical file
- More...

To a Server

- Microsoft SQL Server
- MySQL
- Oracle
- Amazon Redshift
- More...

Save Data Sources

Sample - Superstore

World Indicators

Open

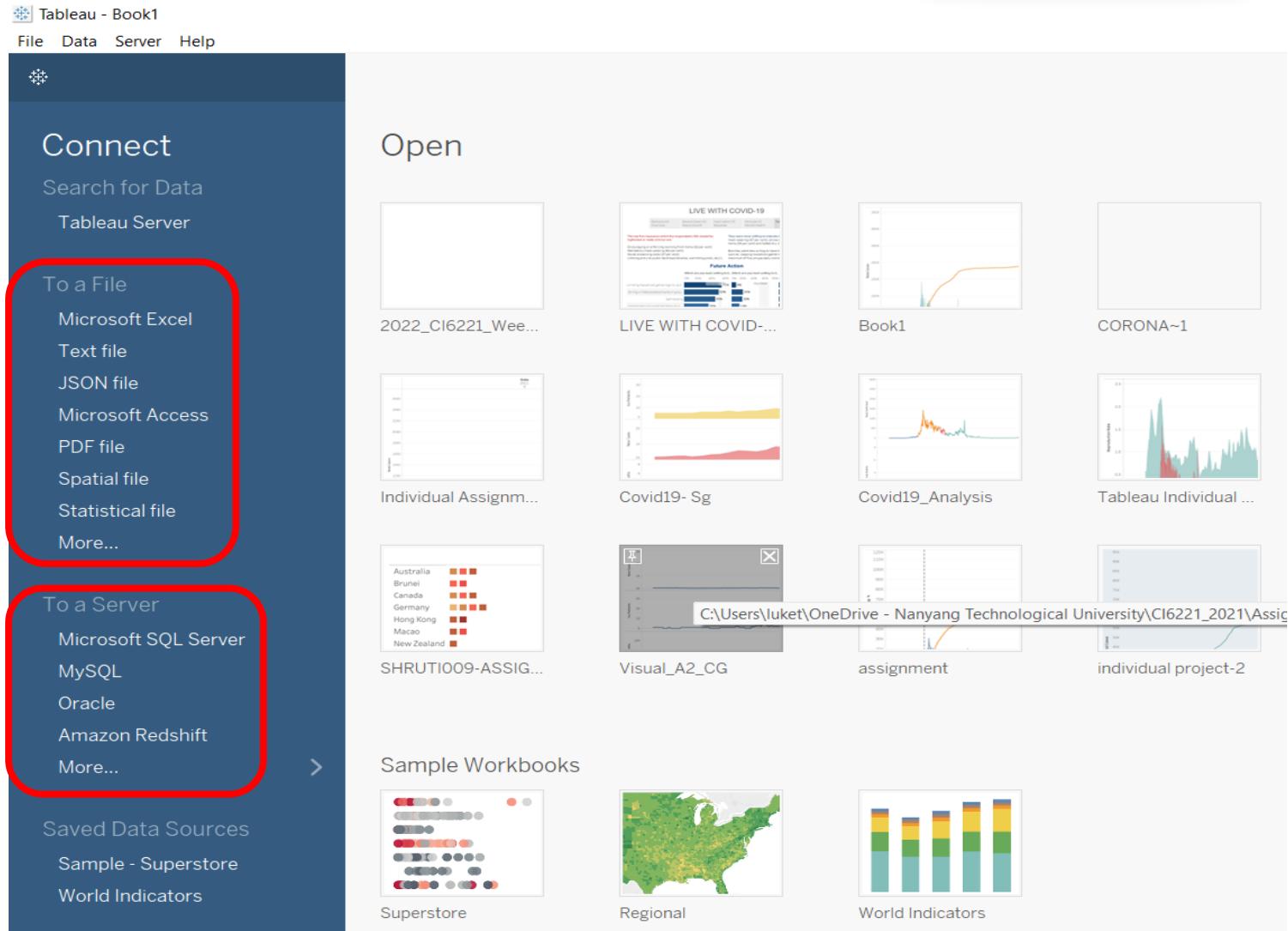
2022\_CI6221\_Wee...  
LIVE WITH COVID-19  
Book1  
CORONA~1

Individual Assignm...  
Covid19- Sg  
Covid19\_Analysis  
Tableau Individual ...

Australia  
Brunei  
Canada  
Germany  
Hong Kong  
Macao  
New Zealand  
SHRUTI009-ASSIG...  
Visual\_A2(CG)  
assignment  
individual project-2

Sample Workbooks

Superstore  
Regional  
World Indicators



# Understanding the Sample Data

## Sample Superstore.xls

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name
32298	CA-2012-124	31/7/2012	31/7/2012	Same Day	RH-19495	Rick Hansen
26341	IN-2013-778	5/2/2013	7/2/2013	Second Class	JR-16210	Justin Ritter
25330	IN-2013-712	17/10/2013	18/10/2013	First Class	CR-12730	Craig Reiter
13524	ES-2013-157	28/1/2013	30/1/2013	First Class	KM-16375	Katherine Murray
47221	SG-2013-432	5/11/2013	6/11/2013	Same Day	RH-9495	Rick Hansen
22732	IN-2013-423	28/6/2013	1/7/2013	Second Class	JM-15655	Jim Mitchum
30570	IN-2011-818	7/11/2011	9/11/2011	First Class	TS-21340	Toby Swindell
31192	IN-2012-863	14/4/2012	18/4/2012	Standard Class	MB-18085	Mick Brown
40155	CA-2014-135	14/10/2014	21/10/2014	Standard Class	JW-15220	Jane Waco
40936	CA-2012-116	28/1/2012	31/1/2012	Second Class	JH-15985	Joseph Holt
34577	CA-2011-102	5/4/2011	9/4/2011	Second Class	GM-14695	Greg Maxwell
28879	ID-2012-284	19/4/2012	22/4/2012	First Class	AJ-10780	Anthony Jacobs
45794	SA-2011-183	27/12/2011	29/12/2011	Second Class	MM-7260	Magdalene Morse
4132	MX-2012-13	13/11/2012	13/11/2012	Same Day	VF-21715	Vicky Freymann
27704	IN-2013-739	6/6/2013	8/6/2013	Second Class	PF-19120	Peter Fuller
13779	ES-2014-509	31/7/2014	3/8/2014	Second Class	BP-11185	Ben Peterman
36178	CA-2014-143	3/11/2014	6/11/2014	Second Class	TB-21175	Thomas Boland
12069	ES-2014-165	8/9/2014	14/9/2014	Standard Class	PJ-18835	Patrick Jones
22096	IN-2014-117	31/1/2014	1/2/2014	First Class	JS-15685	Jim Sink
49463	TZ-2014-819	5/12/2014	7/12/2014	Second Class	RH-9555	Ritsa Hightower
46630	PL-2012-782	8/8/2012	10/8/2012	First Class	AB-600	Ann Blume
31784	CA-2011-154	29/10/2011	31/10/2011	First Class	SA-20830	Sue Ann Reed
21586	IN-2011-448	2/5/2011	3/5/2011	First Class	JK-15325	Jason Klamczynski
13528	ES-2013-286	27/2/2013	1/3/2013	Second Class	LB-16795	Laurel Beltran
1570	US-2014-133	31/7/2014	1/8/2014	First Class	NP-18325	Naresj Patel
3484	MX-2014-165	5/9/2014	8/9/2014	First Class	VD-21670	Valerie Dominguez
30191	IN-2011-102	17/12/2011	20/12/2011	First Class	PB-19210	Phillip Breyer
11645	ES-2011-469	14/3/2011	17/3/2011	Second Class	EB-14110	Eugene Barchas
37311	CA-2013-159	11/3/2013	12/3/2013	First Class	KF-16285	Karen Ferguson
22999	IN-2012-448	25/2/2012	25/2/2012	Same Day	BP-11230	Benjamin Patterson
32011	US-2011-138	16/7/2011	20/7/2011	Second Class	DD-10525	Rick Reed

Tableau - Book1

File Data Server Window Help

Connections Add

Sample - Superstore Microsoft Excel

Sheets

Orders

Use Data Interpreter  
Data Interpreter might be able to clean your Microsoft Excel workbook.

Orders (highlighted with a red circle)

People

Returns

Orders

People

Returns

NewUnion

Orders 21 fields 9994 rows

Name

Orders

Fields

Type Field Name Physical Table Remote File...

# Row ID Orders Row ID

Abc Order ID Orders Order ID

Order Date Orders Order Date

Need more data?

Drag tables here to relate them. [Learn more](#)

Connection  
 Live  Extract

#	Abc	Orders	Orders
Row ID	Orders	Order Date	Ship Date
1	CA-2021-152156	8/11/2021	11/11/2021
2	CA-2021-152156	8/11/2021	11/11/2021
3	CA-2021-138688	12/6/2021	16/6/2021
4	US-2020-108966	11/10/2020	18/10/2020
5	US-2020-108966	11/10/2020	18/10/2020
6	CA-2019-115812	9/6/2019	14/6/2019

# Data Connection View

Click on “Sample Superstore” and drag “Orders” sheet into Data Connection Canvas.

**Connect**

Search for Data  
Tableau Server

To a File  
Microsoft Excel  
Text file  
JSON file  
Microsoft Access  
PDF file  
Spatial file  
Statistical file  
More...

To a Server  
Microsoft SQL Server  
MySQL  
Oracle  
Amazon Redshift  
More...

Saved Data Sources  
**Sample - Superstore**  
World Indicators

File Data Server Window Help

Connections Add

Sample - Superstore Microsoft Excel

Sheets

Use Data Interpreter  
Data Interpreter might be able to clean your Microsoft Excel workbook.

Orders People Returns

People Returns

New Union

New Table Extension

**Data Connection Canvas**

**Sample - Superstore**

Connection  Live  Extract

Filters 0 | Add

Tables: Right-click for more options, e.g., rename

Orders People Returns

Datatype: Right-click to change

Fields

20 fields 10194 rows

Name Orders

Type Field Name Physical Table Remote Fie...

Type	Field Name	Physical Table	Remote Fie...
Abc	Order ID	Orders	Order ID
	Order Date	Orders	Order Date
	Ship Date	Orders	Ship Date

Columns: Right-click for more options, e.g., rename

Abc	Order ID	Orders	Order ID
US-2020-103800	3/1/2020		
US-2020-112326	4/1/2020		
US-2020-112326	4/1/2020		
US-2020-112326	4/1/2020		

# Live versus Extract

Decide whether to connect **live** or **extract**.

- Connecting **live** leaves the data in the database or source file => get up-to-the-second changes in data visualized in Tableau.
- Connecting **live** can result in a **slow experience**, depending on the database.

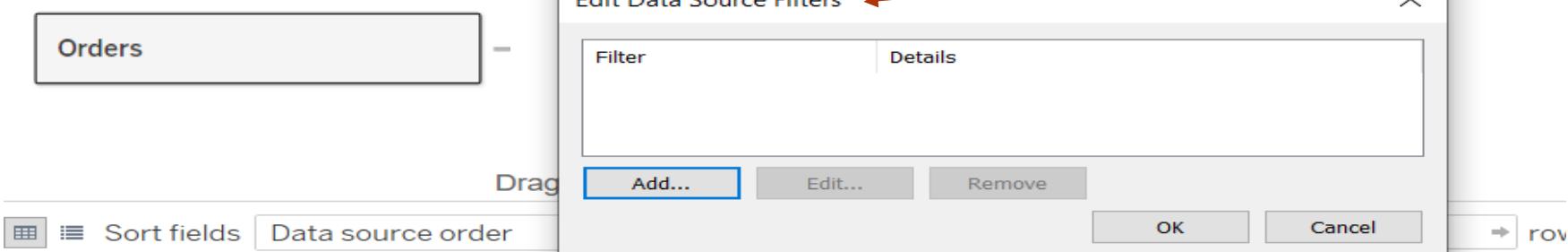
The other option is to **extract** the data into Tableau's **in-memory** data engine => help when connecting to slow database or to **take query load off critical systems**.

- Possible to import only **some** of the data and bring in **specific elements** (click add Filters)

For now, we'll **connect live**

□ - (Sample - Superstore)

Select the Orders data table



# Sheet View

## Menus & Toolbar

## Data Window

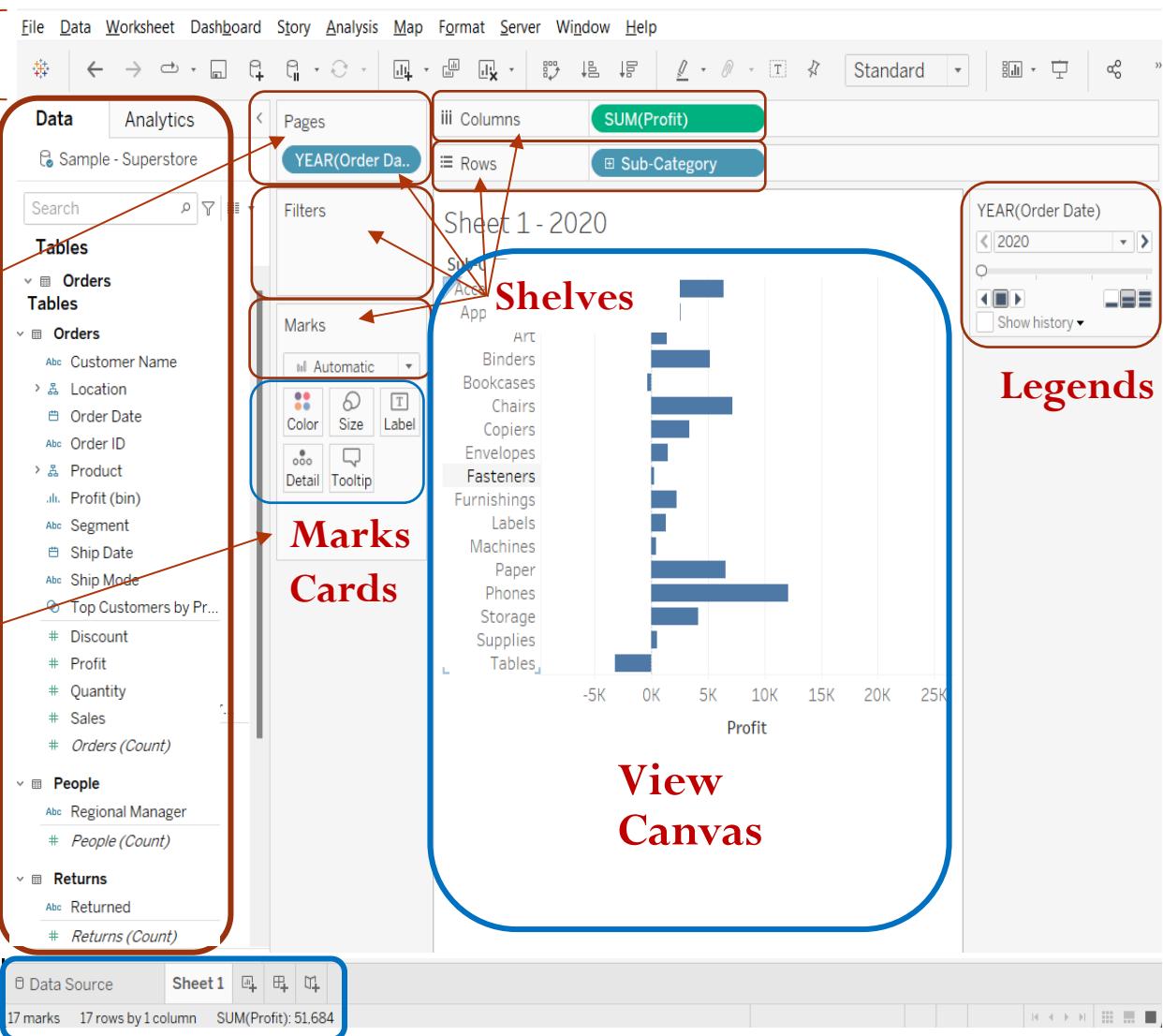
**Pages shelf** break a view into a series of pages to better analyze how a specific field affects the rest of the data in a view

页面架将视图分解为一系列页面，以更好地分析特定字段如何影响视图中的其余数据

**Mark cards** – allows adding of features details to the charts

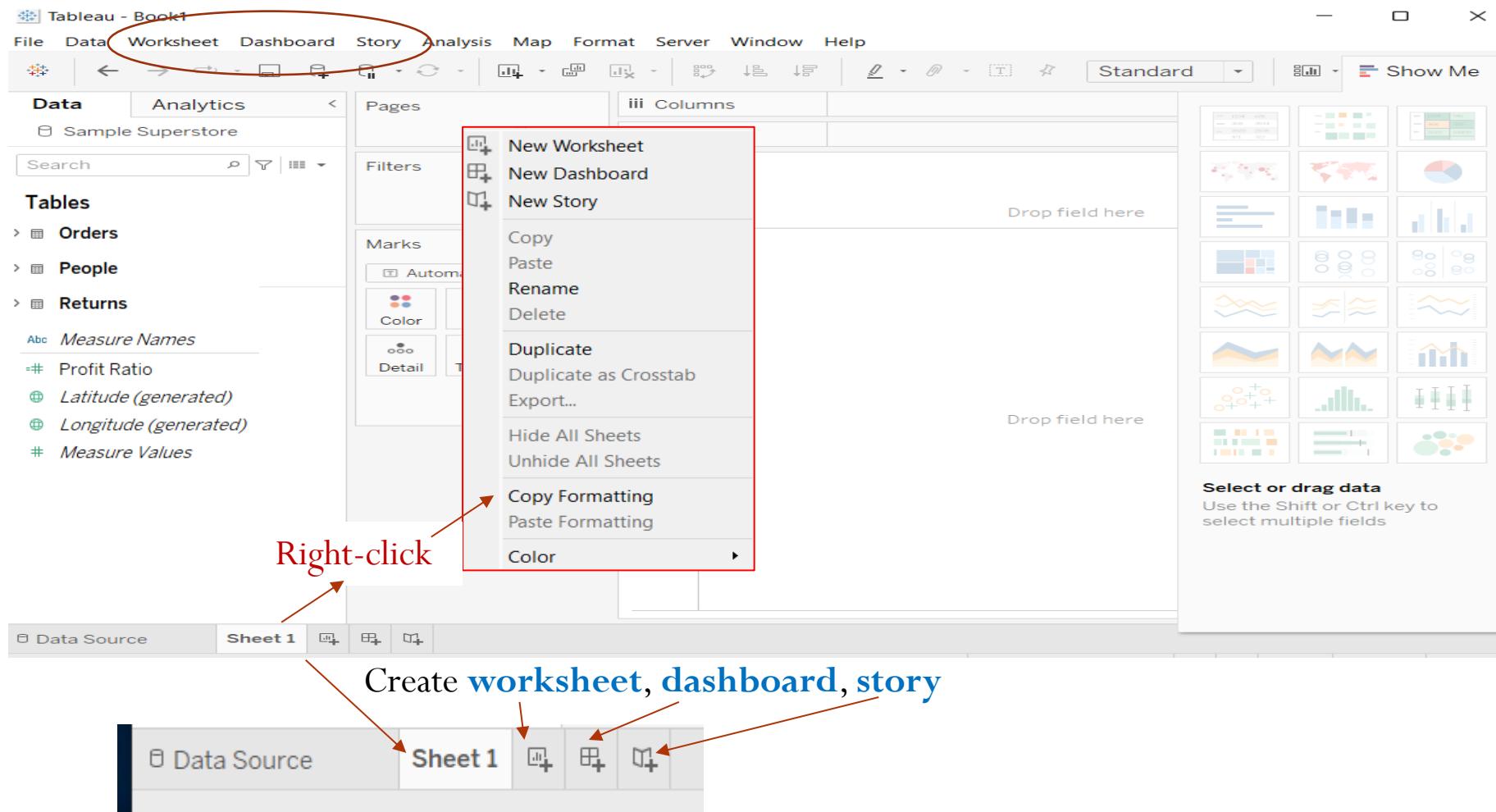
允许将特征详细信息添加到图表中

## Access Tabs



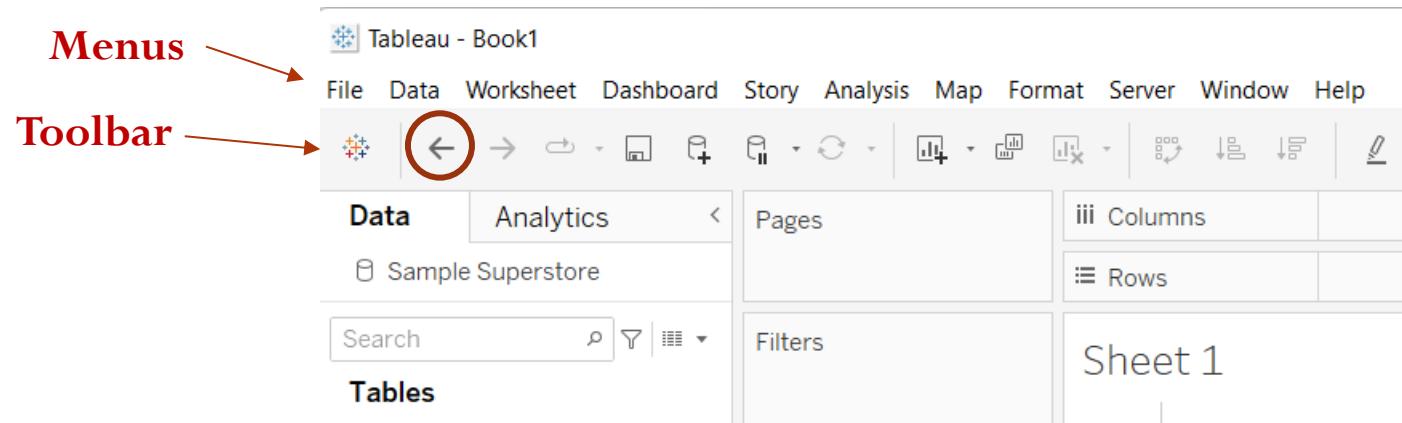
# Menus & Toolbar

- Create **worksheets**, **dashboards**, and **stories** with top or bottom tabs.
- **Right-click on “Sheet”** to do things like rename the sheets, duplicate sheets, copy formatting, and many other things.



# Menus & Toolbar

- At the top are the **menus**. (layout may look slightly different on a Mac) which contain controls functions
- Below is the **toolbar**, with buttons like **undo – no limit to undo** - allows you to explore! There's **no automatic save** in Tableau, so make sure to save your work periodically.



# Data Window

The screenshot shows the Tableau Data window. At the top, there are tabs for 'Data' (which is selected) and 'Analytics'. Below the tabs, it says 'Sample - Superstore'. There is a search bar and a toolbar with various icons. Under the 'Tables' section, there is a list of tables: 'Orders' (selected), 'Customer Name', 'Location', 'Order Date', and 'Order ID'. A red bracket on the left side groups 'Table: Orders' and 'Fields: dimensions' under the 'Tables' heading.

Table: Orders

Fields: dimensions

Dimensions contain **qualitative** values  
(such as names, dates, or geographical data).

Measures contain numeric, **quantitative** values that can have **calculation** performed on it. Measures can be **aggregated** (count of the measure)

测量包含可以进行计算的数值、定量值。  
措施可以汇总（措施的计数）

= equal sign indicated  
**aggregated** value

Fields: measures

- abc Measure Names
- # Profit Ratio
- ⊕ Latitude (generated)
- ⊕ Longitude (generated)
- # Measure Values

On the left of the screen is the **data window**.

On the **data tab**, the top **lists all open data connections**, and depending on which one is selected, the **fields** from that data source are listed below, broken out into **dimensions** and **measures**.

## Dimensions

Abc	Orders
Order ID	Order Date
US-2020-103800	3/1/2020
US-2020-112326	4/1/2020
US-2020-112326	4/1/2020
US-2020-112326	4/1/2020
US-2020-141817	5/1/2020

## Measures

#	Orders
Profit	5.55
	-5.49
	4.27
	-64.77
	4.88

# Connecting to Multiple Tables

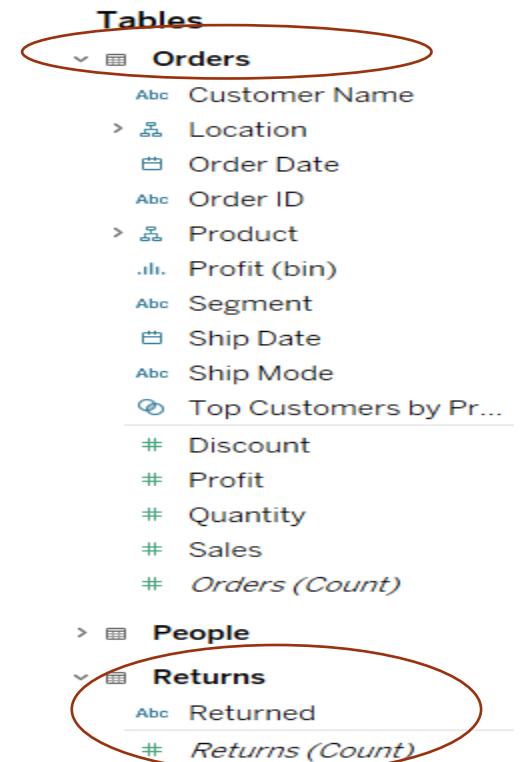
- To add columns from other tables in the **same data source** (i.e. other worksheets in the Excel file), **click on the Data Source tab** (at bottom left).
- Join the **Returns** table to **Orders** table. Double click or **drag out Orders** to the canvas.

Select **Order ID** as the **common field (foreign key)** between the Returns and Orders tables.

We can edit the join clause, or even create a new one (drag New Union to canvas)

The screenshot shows the Power BI Data Editor interface. On the left, the 'Connections' pane displays a connection to 'Sample - Superstore Microsoft Excel'. The 'Sheets' pane lists several tables: 'Orders', 'People', 'Returns', and 'New Union'. The 'Data Source' tab is selected at the bottom. In the main workspace, a diagram shows the 'Orders' table connected to the 'Returns' table via the 'Order ID' column. A preview window below shows the relationship with the 'Order ID' column selected. The 'Connections' pane also includes a 'Data Interpreter' section and a 'New Table Extension' option.

We now have columns from both **Orders** and **Returns** in our data window.

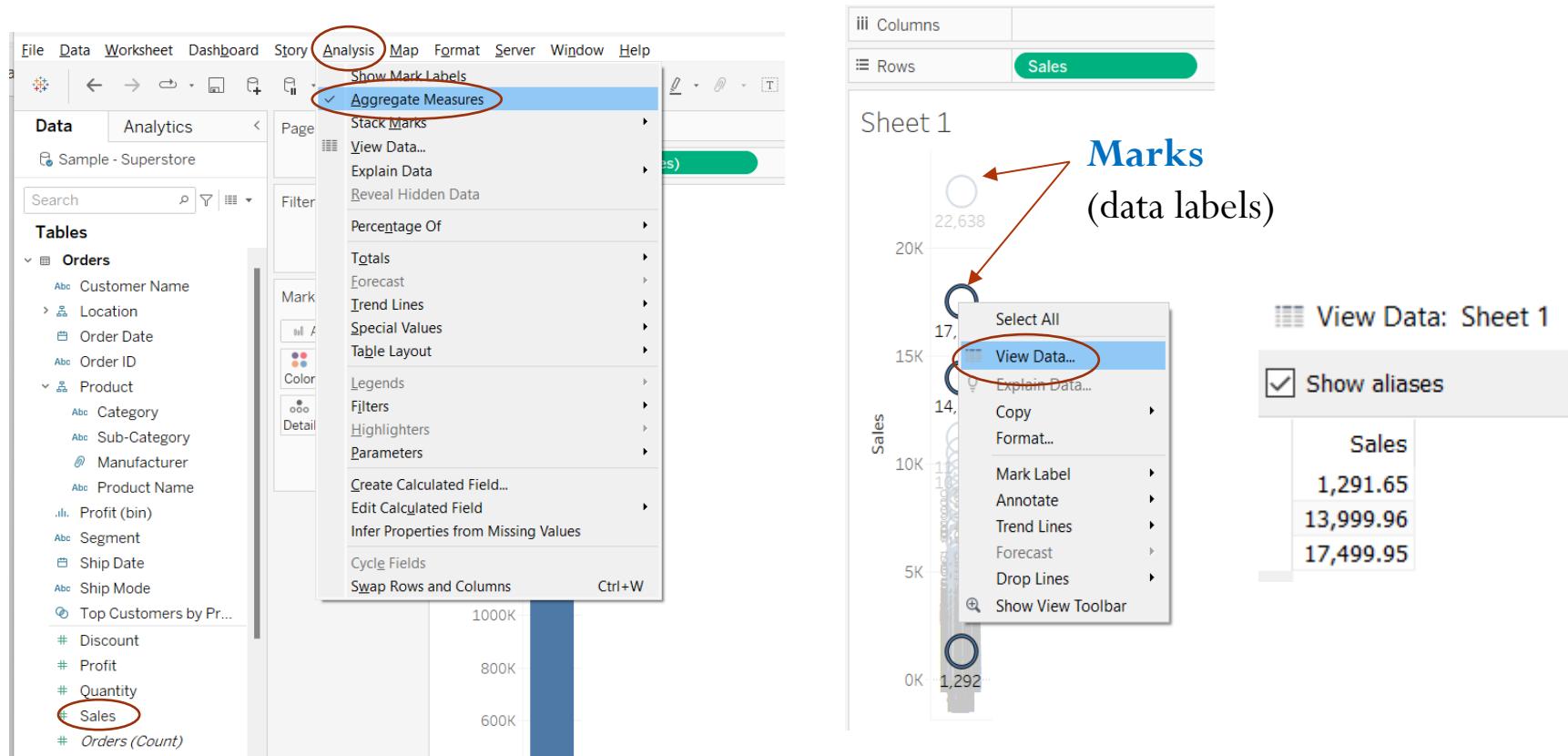


Columns from **Returns** table

# Viewing Fields Data

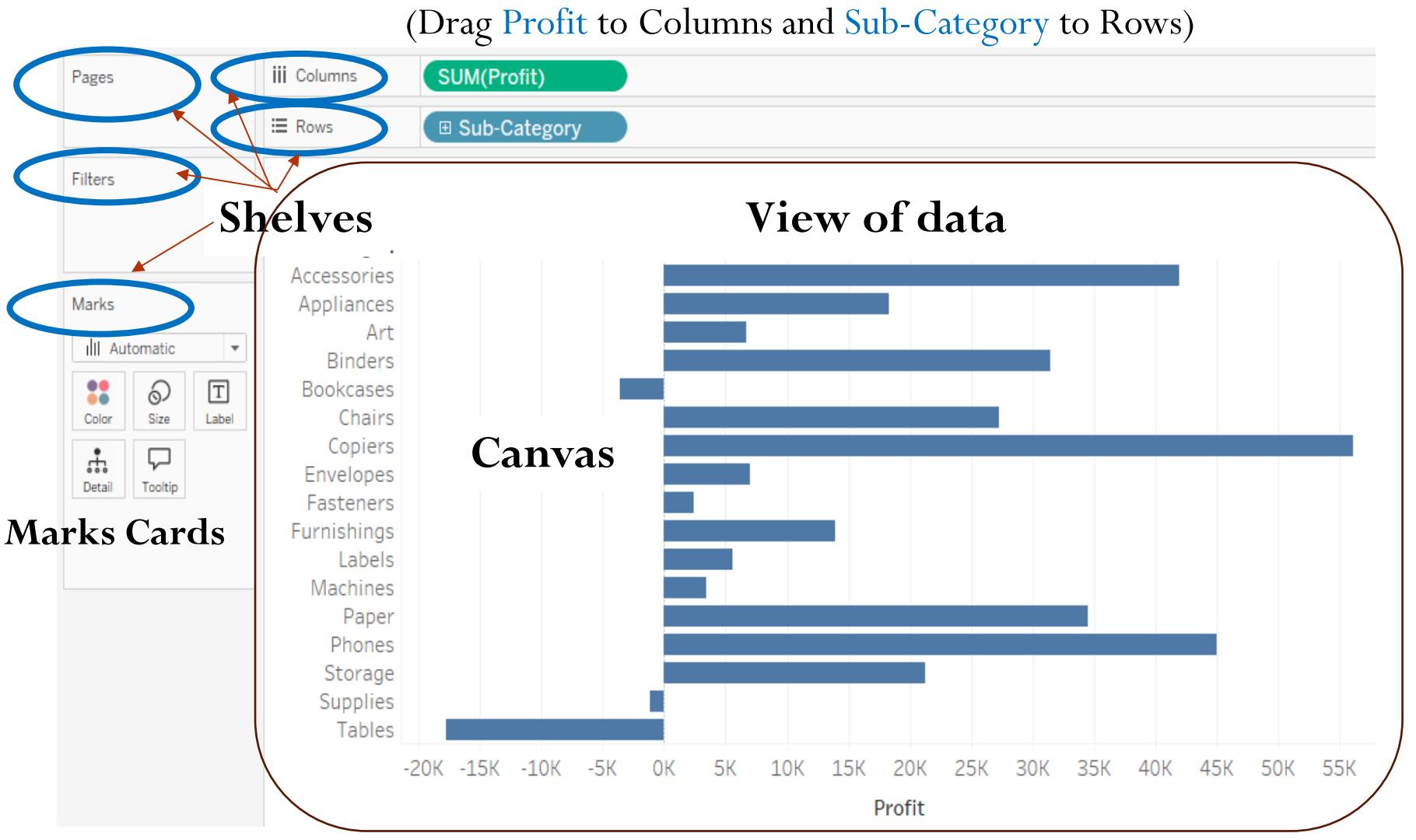
Double-click **Sales** – this gives a single bar. Tableau has aggregated the data. What if you want to see Sales at a more granular level? click **Analysis**, uncheck **Aggregate Measures**. Right-click a **mark** (data label) or several marks (ctrl+select) => **View Data** => see **actual data**.

双击销售-这会给出一个单条。Tableau已经汇总了数据。如果你想在更精细的层面上看到销售呢？单击分析，取消选中汇总措施。右键单击一个标记（数据标签）或几个标记（ctrl+select）=>查看数据=>查看实际数据。



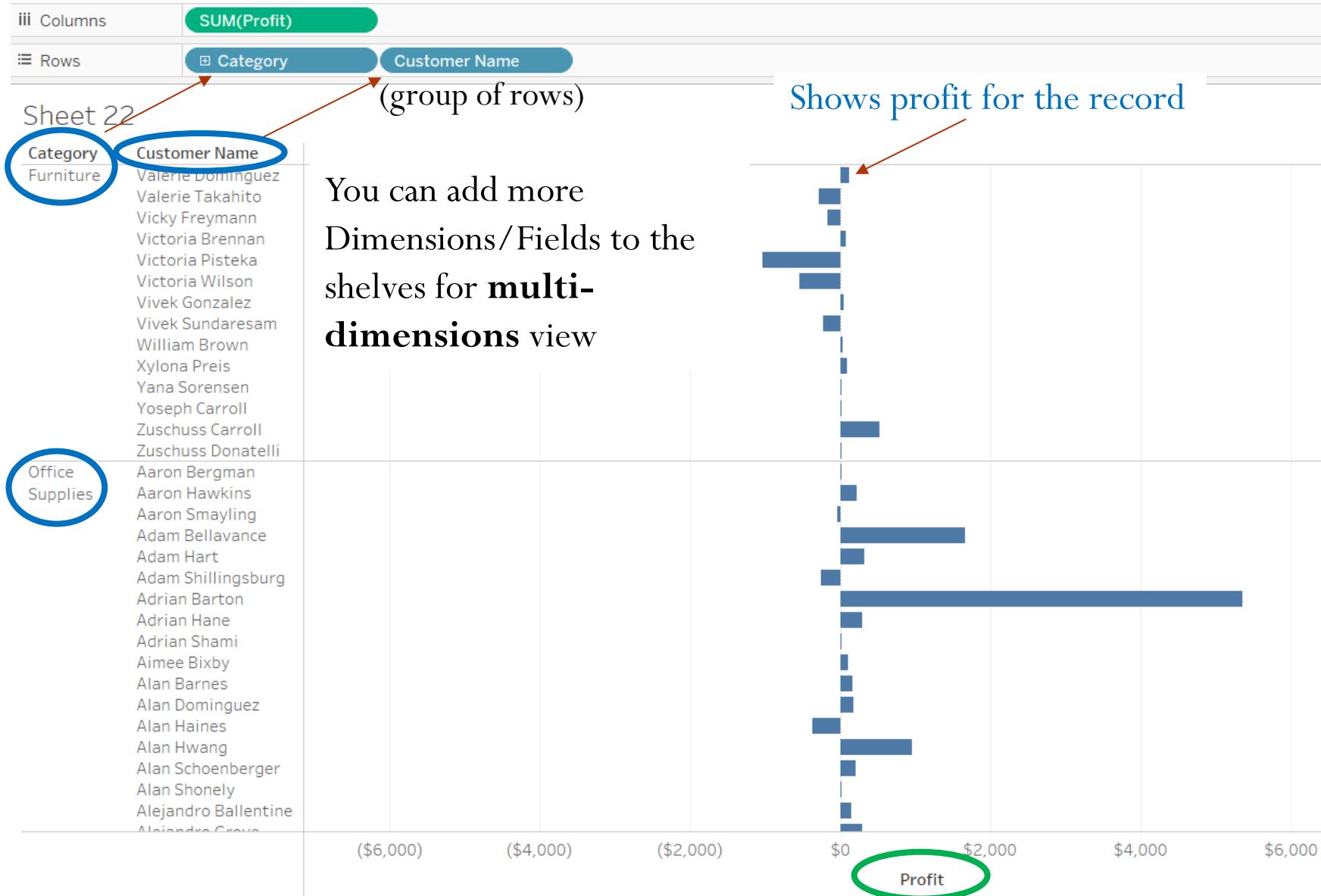
# Shelves & Cards

A **view** can be built by dragging and dropping fields from the data window into the **canvas** directly, or onto the **shelves**.



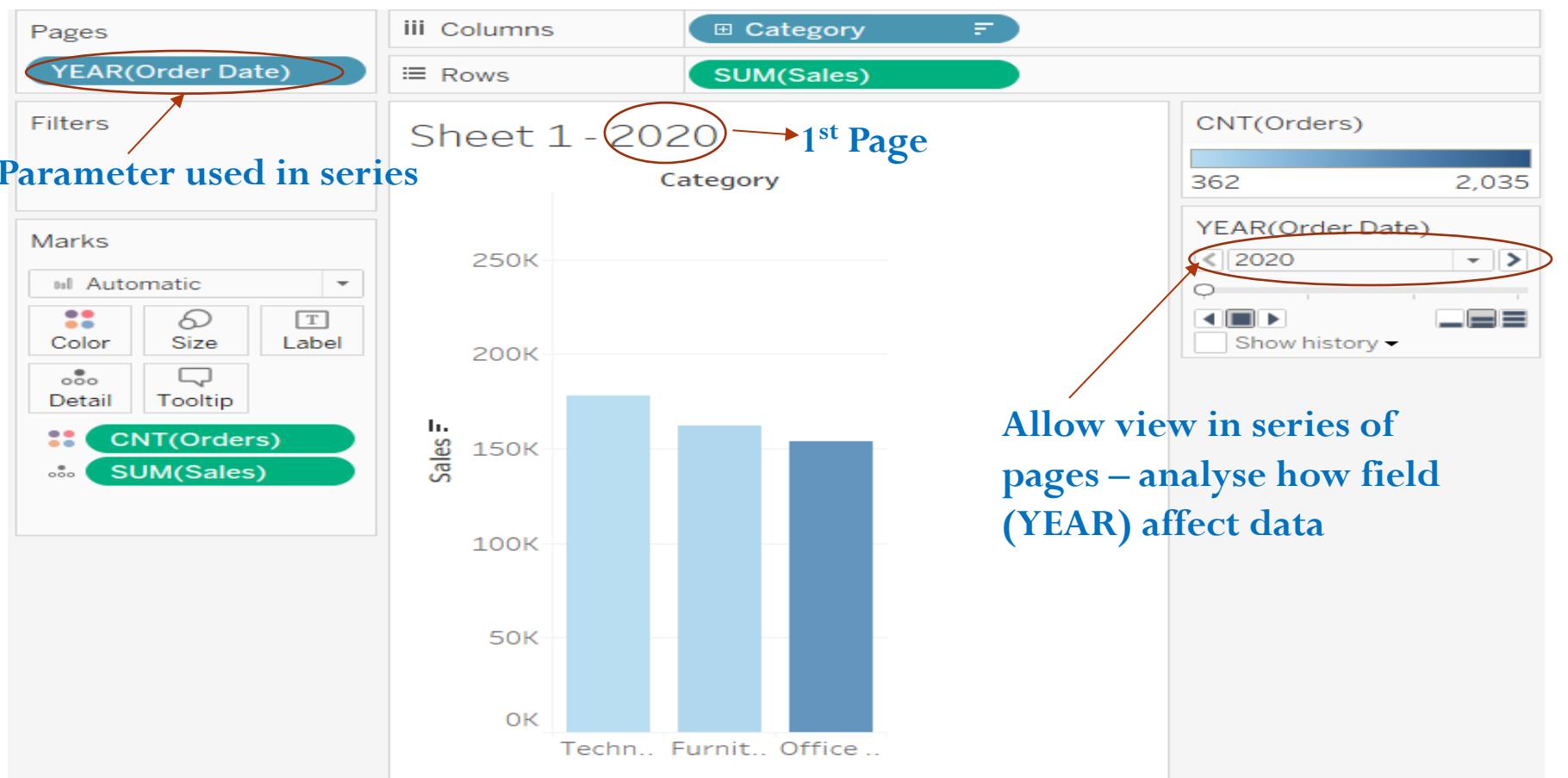
# Shelves & Cards

Tableau displays data using **marks**, each corresponding to **a row (or a group of rows)** in the data source.



# Pages Shelf

- The Pages shelf breaks a view into a **series of pages** to better analyze **how a specific field affects the rest of the data** in a view.
- Placing a **dimension** (Order Date) on the Pages shelf will add a **new (Order Date) field for each member (Year)** in the dimension. Placing a **measure** on the Pages shelf will convert the measure to **discrete – parameter will loop through the series**.

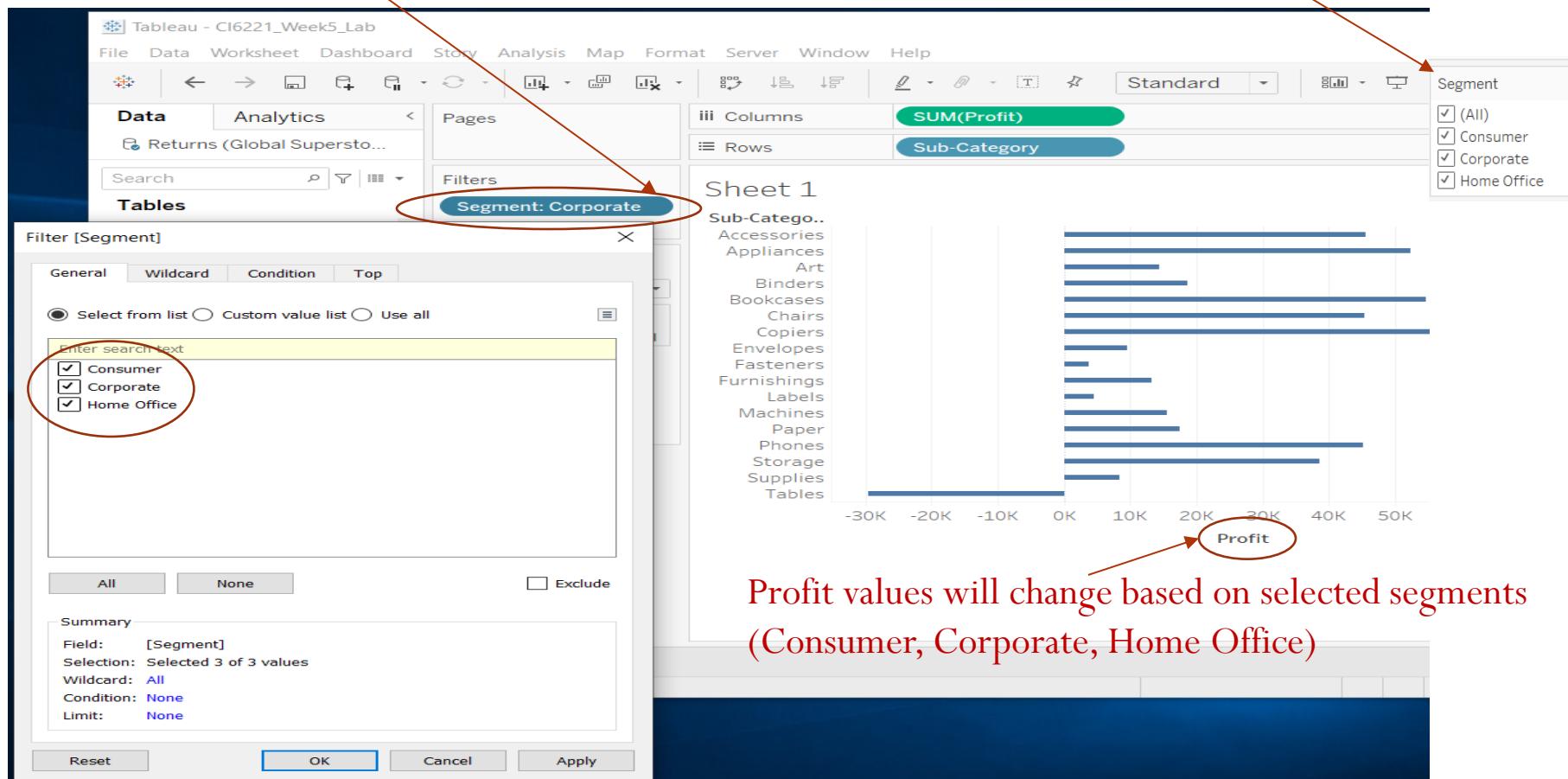


# Filter Shelf

**Filter shelf** – select data to view based on values.

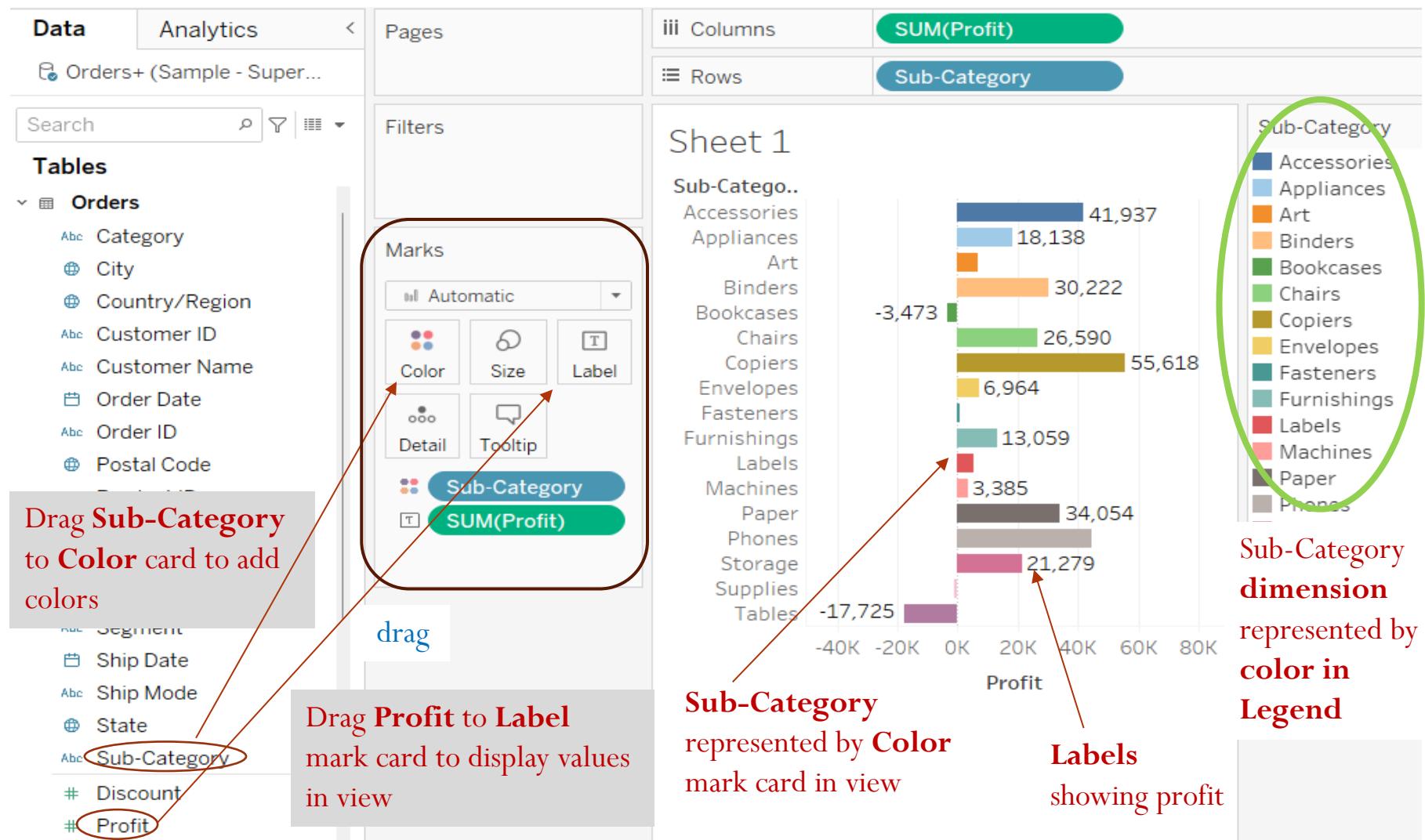
Drag “**Segment**” to **Filters** to show **Filter options**. Select required “Segments” (e.g., Consumer, Corporate, Home Office) to view filtered data (Profit values will change based on selected segments)

Click “Segment” dropdown icon => “**Show Filter**” => display filter on right panel for easy selection.



# Marks Card

- The Marks Card is made up of several shelves, each of which can have **fields placed on them** and can be clicked on to **edit their characteristics**



# Marks Card

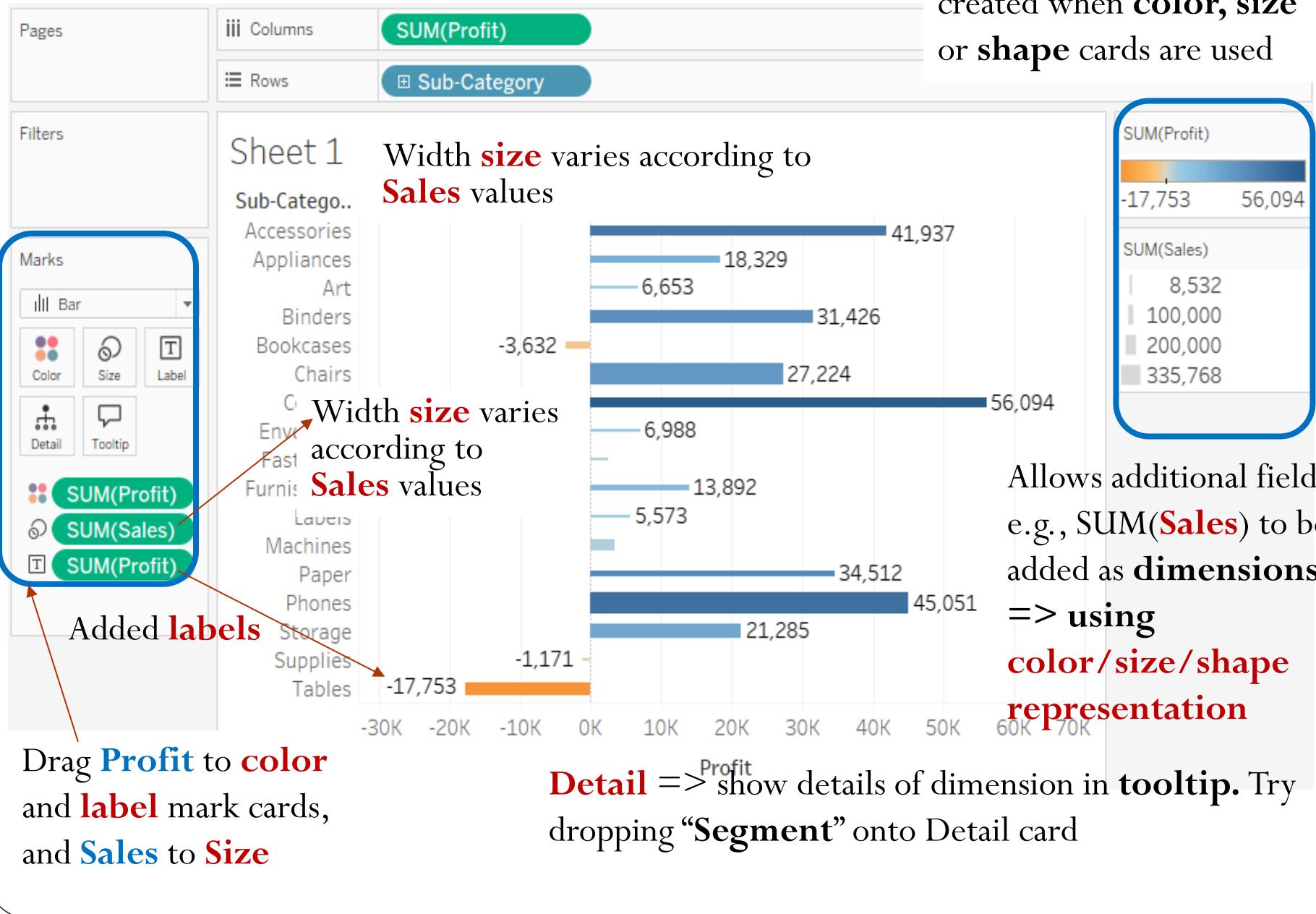
Select different **chart types** at list down box

Click on **SUM(Profit)** to change display from Sum to Average values

Tableau desktop interface showing a bar chart of Profit by Category. The chart displays the following data:

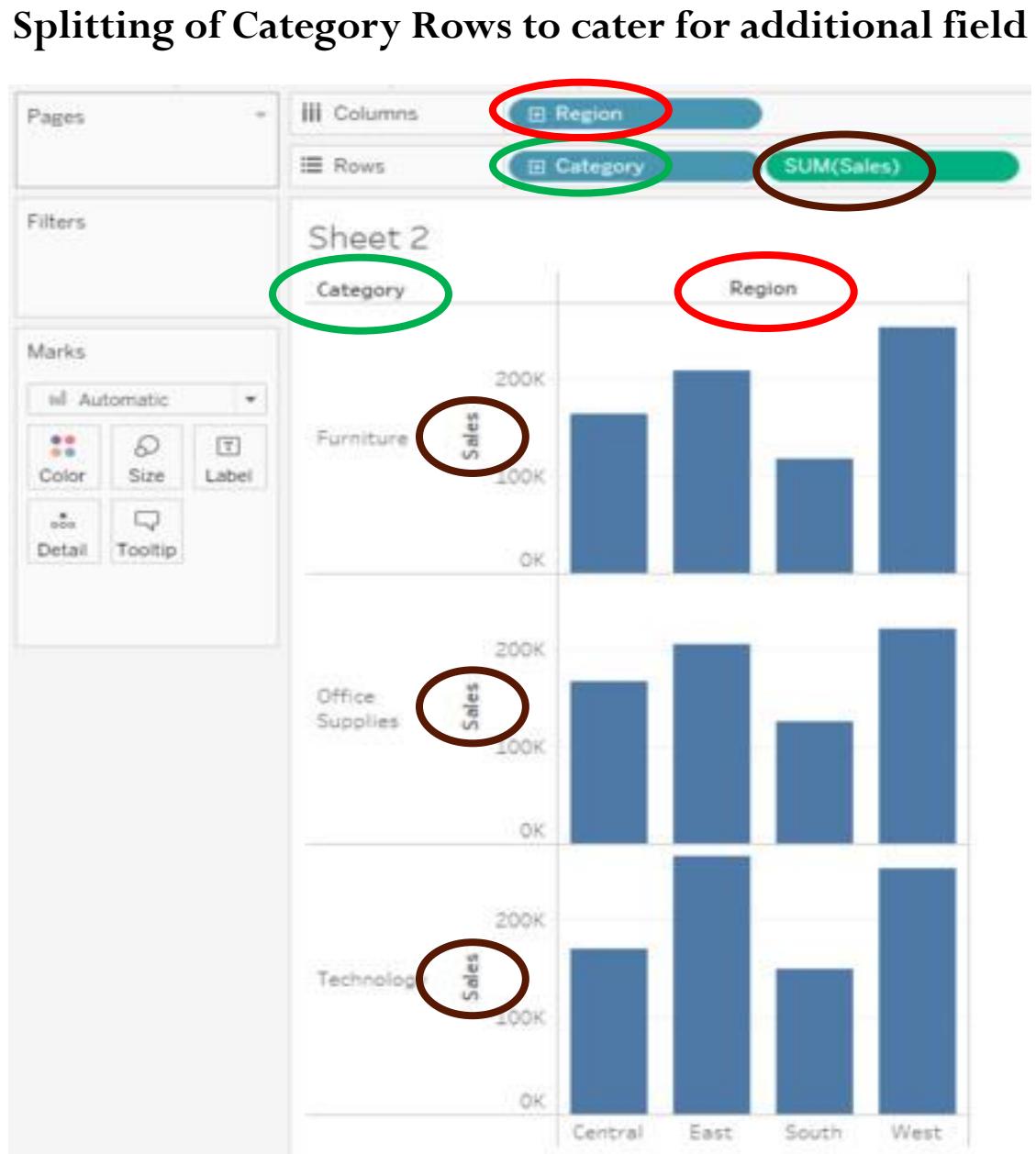
Category	Profit
Accessories	-39,354
Appliances	20,397
Art	31,213
Binders	60,803
Bookcases	12,953
Chairs	
Copiers	
Envelopes	
Fasteners	
Furnishings	
Labels	
Machines	
Paper	
Phones	
Storage	
Tables	
Technology	

# Marks Card



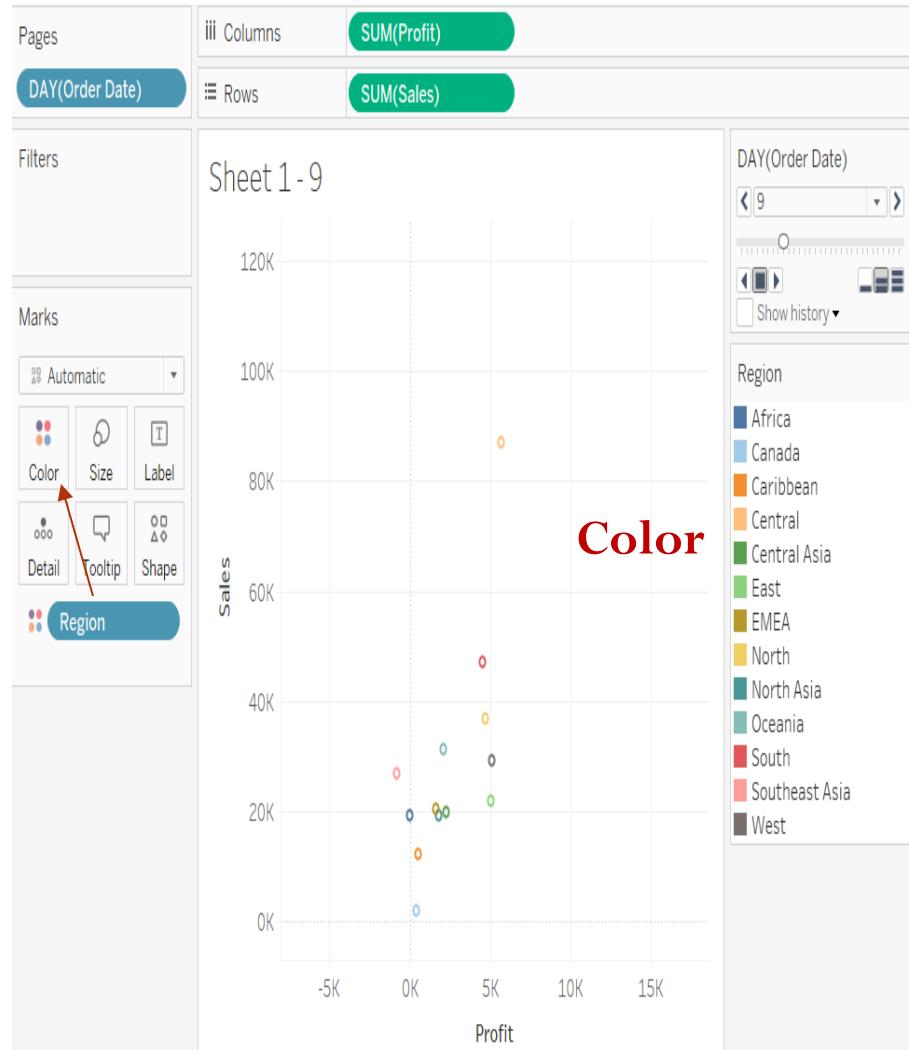
# Marks Card

- Tableau **does not support 3D Charts**
  - 3D layout **not ideal** – blockage, difficult to view, tendency to give misleading values (slanted lines)
  - Use **sub-rows** instead for additional dimension

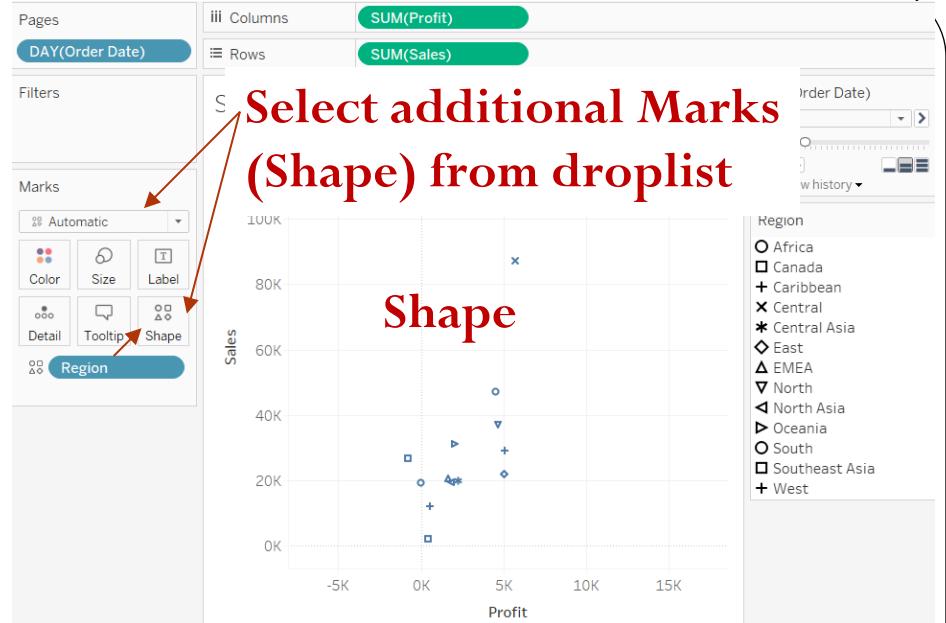


# Marks Shelf

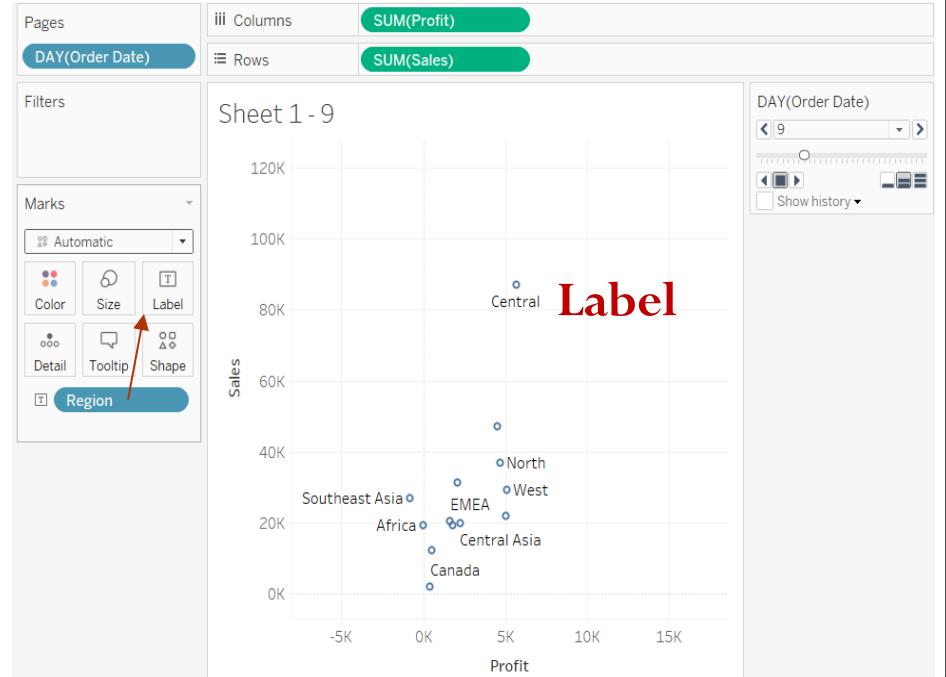
## Choice of different Marks



Color



Shape



Label

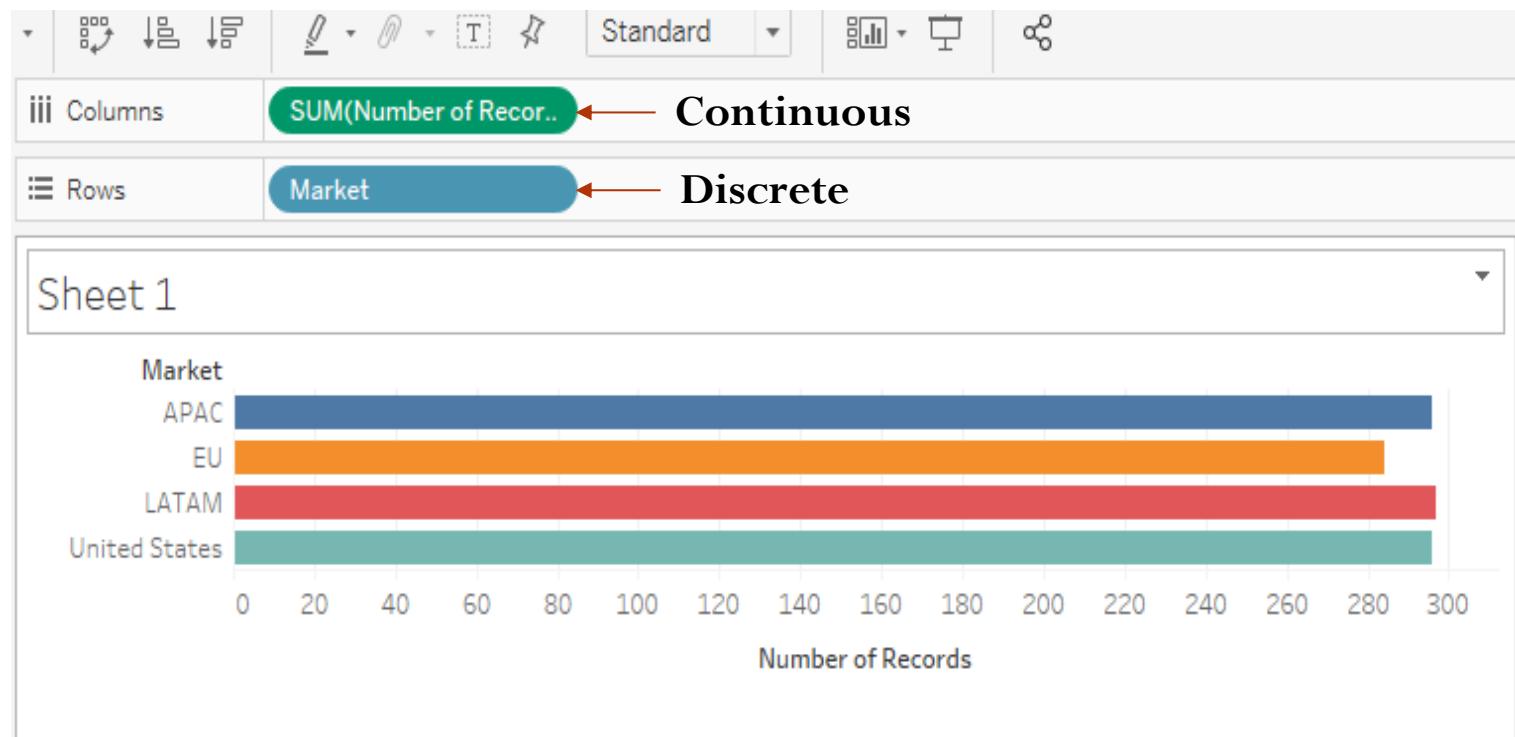
Select additional Marks (Shape) from dropdown

# Paying Attention to Pill Color (Continuous vs Discrete)

When we bring a field into the view from the Data Window, Tableau creates a pill. The pill color indicates whether the pill is continuous or discrete. **Discrete** pills are **BLUE**, **Continuous** pills are **GREEN**.

当我们从数据窗口将字段带入视图时，Tableau会创建一个药丸。药片的颜色表示药片是连续的还是离散的。离散药片是蓝色的，连续药片是绿色的。

**Text** and **categories** are inherently **discrete**. **Numbers** can also be discrete if they can only take one of a **limited set** of **distinct, separate** values. On the other hand, numbers are **continuous** if they can take on any value in a **range**.



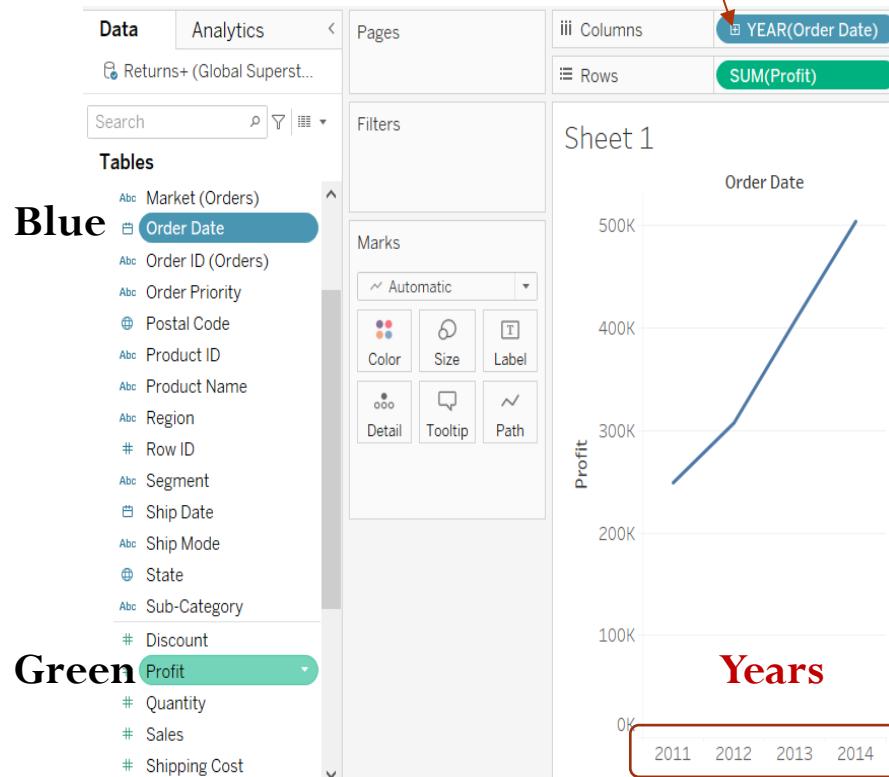
# Continuous vs. Discrete Dates

Dates can be both continuous and discrete. If date icon is **blue** in the data window, then the default when we bring out that pill will be **discrete**, otherwise it will be **continuous**.

日期可以是连续的，也可以是离散的。如果数据窗口中的日期图标是蓝色的，那么当我们拿出那颗药片时，默认值将是离散的，否则它将是连续的。

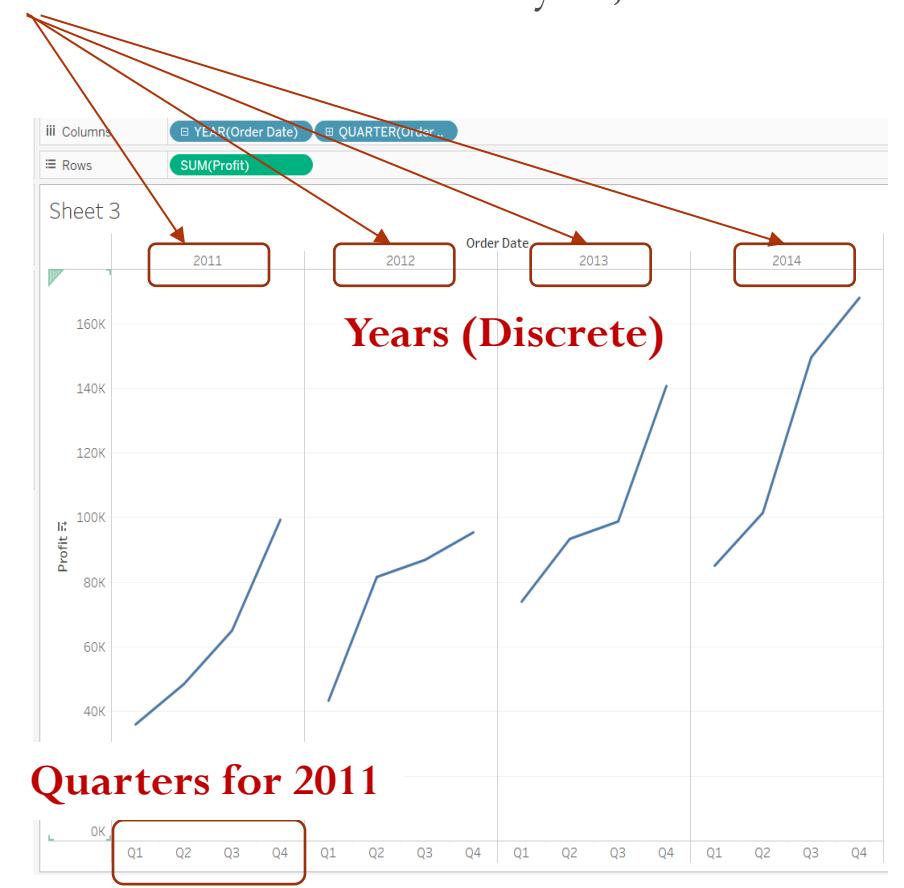
**Create New sheet.** Double-click **Profit**, then bring **Order Date** to columns => **blue pill** => **discrete date**. Each part of the date is treated like a category => **click the +** => expand to years and quarters) => separate section of the viz for each year, rather than one continuous line.

**click the +**



**Green**

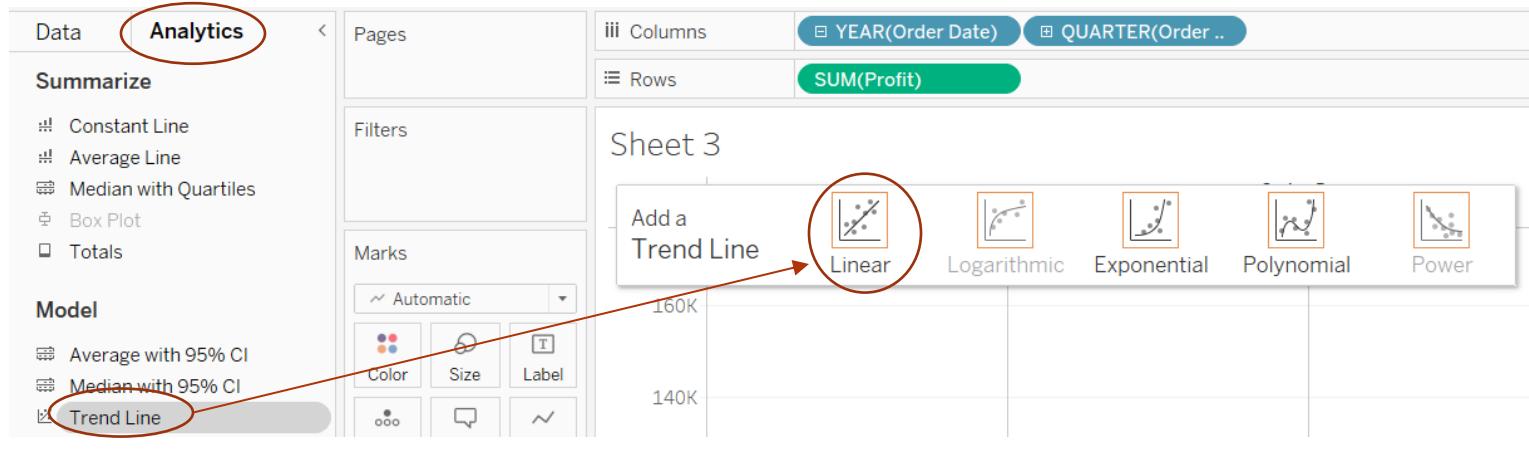
- # Profit
- # Quantity
- # Sales
- # Shipping Cost



# Continuous vs. Discrete Dates

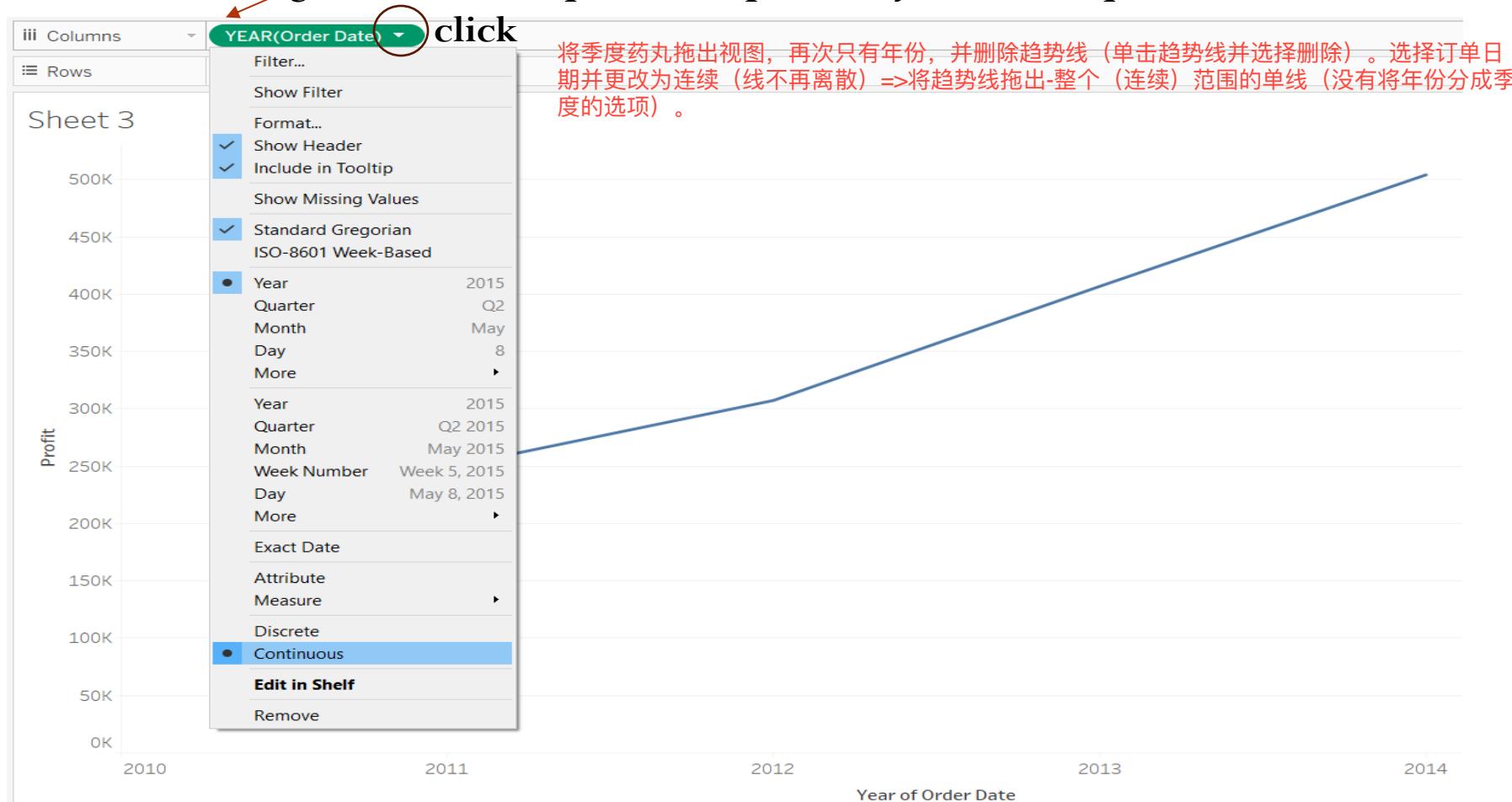
Go to the **Analytics** tab and drag out a **trendline** – each year has a separate line. Discrete dates allow you to treat each “date part” (year, month, etc.) as a separate variable. For example, to filter sales by month (e.g. show the sales in July), a discrete date works best.

转到“分析”选项卡，然后拖出一条趋势线——每年都有一条单独的线。离散日期允许您将每个“日期部分”（年份、月份等）视为单独的变量。例如，要按月筛选销售额（例如显示7月份的销售额），离散日期效果最好。



# Continuous vs. Discrete Dates

Drag the Quarter pill out of the view to have only Year again and remove the trendline (click on trendline and select remove). Select Order Date and change to Continuous (line no longer discrete) => drag a trendline out – single line for entire (continuous) range (there is no option to split the years into quarters).



# Dimensions & Measures

- Dimensions (categorical fields) set the **granularity**, or the level of detail in the view => things you **group by** or **drill down by**.
- It does not make sense to think about computing an average city or average region => **slice or group** the data using these fields to view the data by some **combination** of these categories (e.g., group cities by countries, regions, etc)  
• 考虑计算平均城市或平均地区=>使用这些字段对数据进行切片或分组，通过这些类别的某些组合（例如，按国家、地区等对城市进行分组）来查看数据是没有意义的。
- What **dimensions** we use to build the view will **determine how many marks** we have – Region (under Location) has **4 categories**, so it would give us **4 marks**.
- Measures (numerical data) => they're **aggregated up to the granularity set by the dimensions** in the view.

测量 (数字数据) =>它们被聚合到视图中维度设置的粒度。

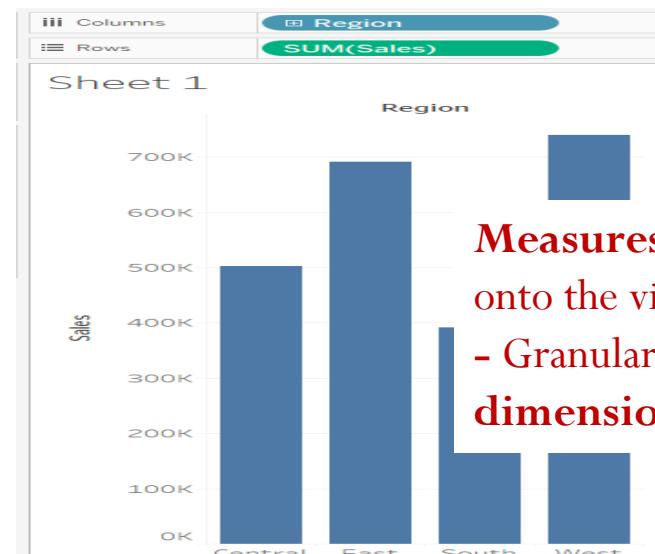
iii Columns      Region

Rows

Sheet 1

Region				
Central	East	South	West	
Abc	Abc	Abc	Abc	

4 marks = 4 dimensions



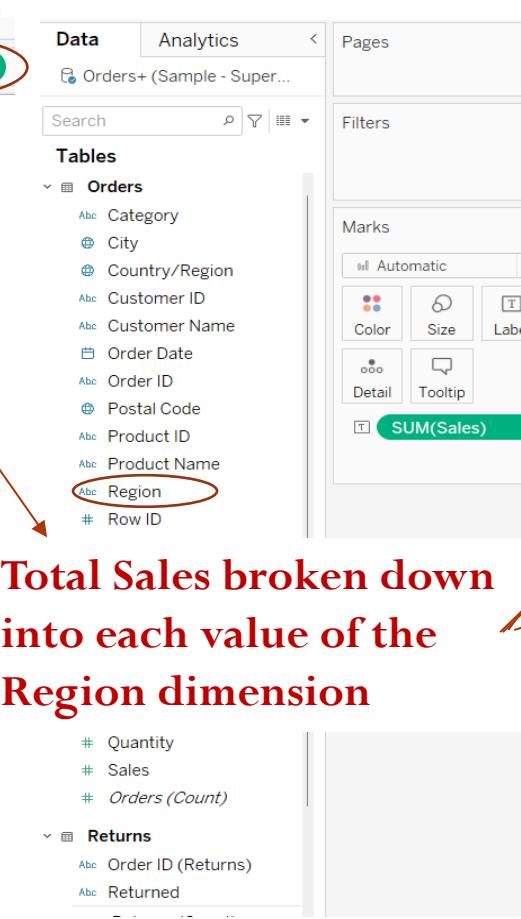
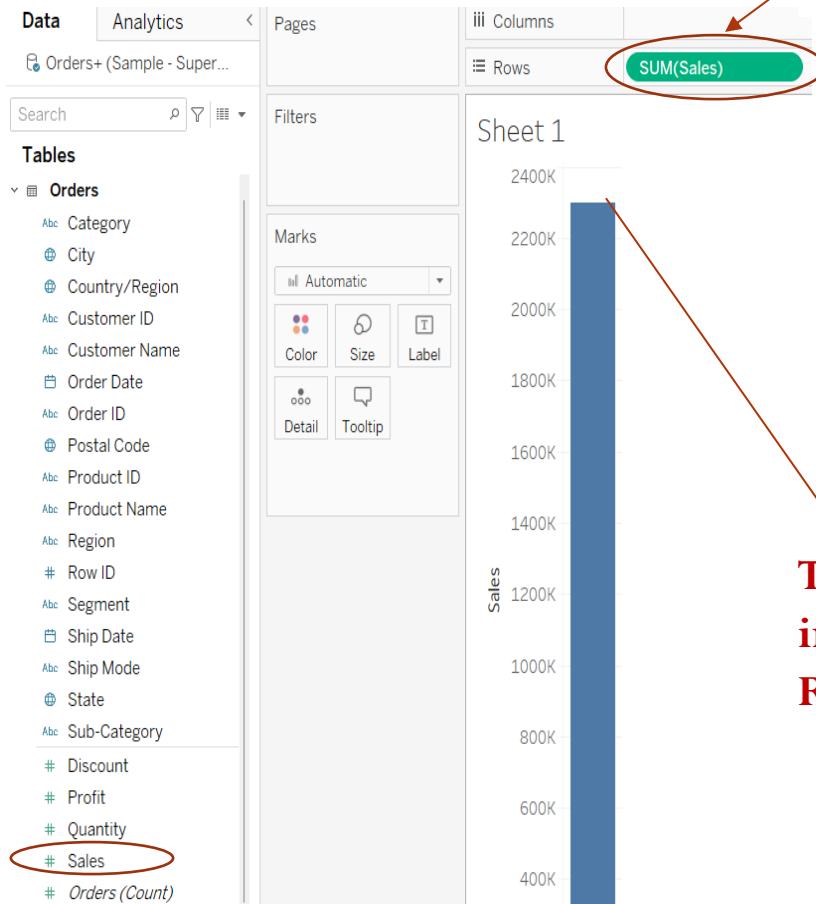
Measures (Sales) come out onto the view as **aggregates**  
- Granularity set by **Region dimension**

# Dimensions & Measures

The **value of a measure** depends on the **context of the dimensions**. For example, the sum of **Sales** (double click it) is different if there is no dimension in the view (just a single overall sum) => When we add **Region** (double click it) as a dimension – now we have a **Sum of Sales** for each **Region**.

测量值取决于维度的上下文。例如，如果视图中没有维度（只有一个整体总和）=>当我们添加区域（双击）作为维度时，销售总和（双击）是不同的——现在我们每个区域都有销售总和。

No dimension added



Total Sales broken down  
into each value of the  
Region dimension

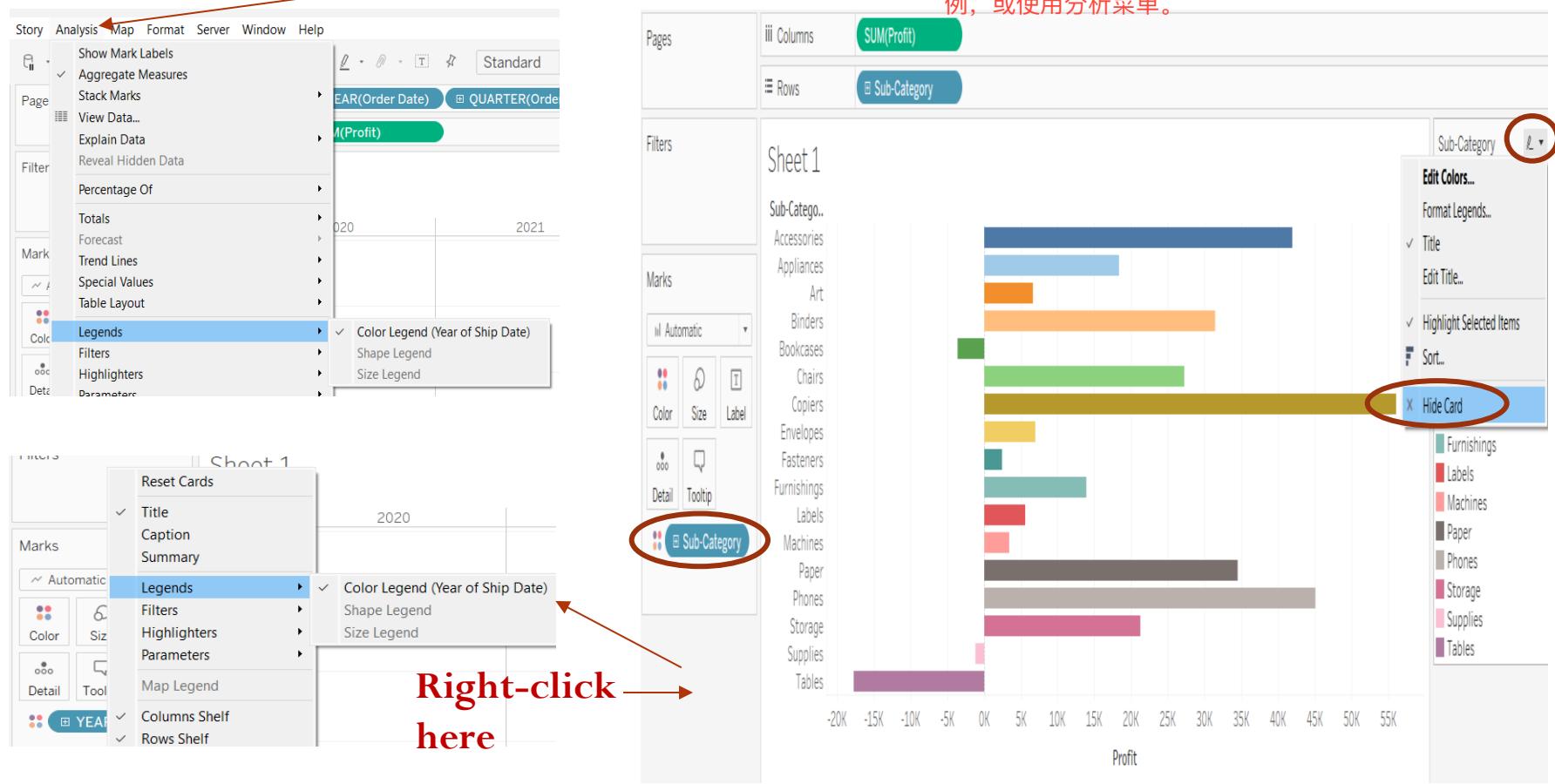


# Legends

A legend will automatically be created when a field is placed on the **color**, **size** or **shape** card.

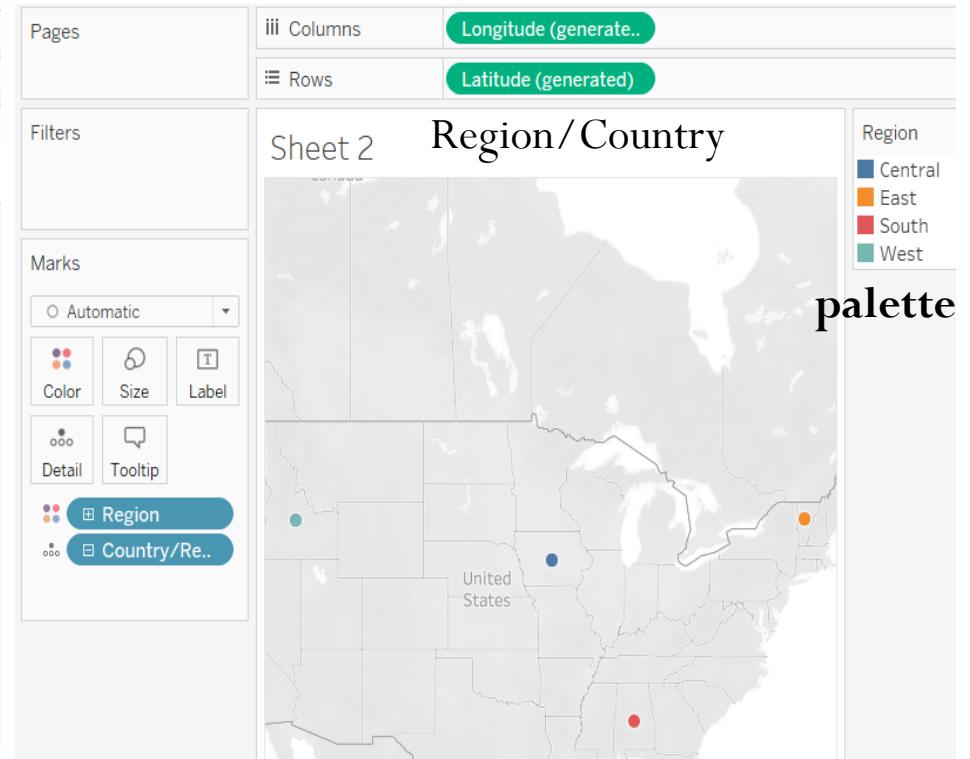
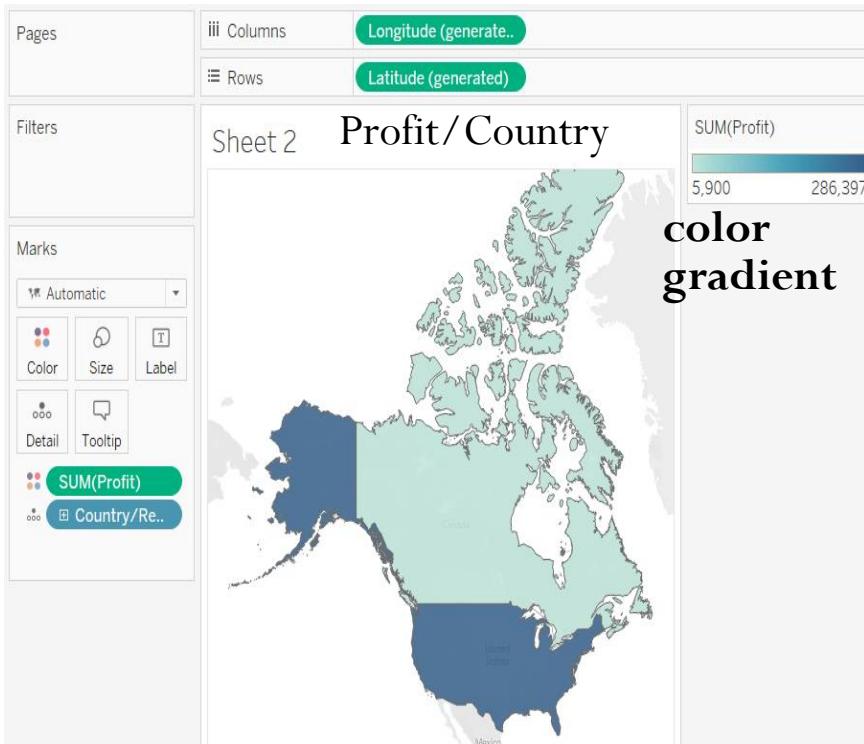
Legends can be removed by clicking on the Legend menu, and selecting **hide card**. To bring a legend back, either right click in the space under the marks card, select Legend and choose which legend to bring back, or use the **Analysis** menu.

当字段放在颜色、尺寸或形状卡片上时，将自动创建图例。可以通过单击“图例”菜单并选择隐藏卡片来删除图例。要带回图例，请在标记卡下的空格中右键单击，选择图例并选择要带回的图例，或使用分析菜单。



# Color & Maps

- A **measure on color** defaults to a **filled map** - show a **color gradient**. A **dimension on color** defaults to a **symbol map** - show a **palette of distinct colors**.  
颜色测量默认为填充地图-显示颜色渐变。颜色的维度默认为符号映射-显示不同颜色的调色板。
- Create new sheet, double click **country** (under location) to get a map
- Drag a **numeric measure (Profit)** to **color** => **filled map** with gradient shading.  
拖动一个数字度量 (利润) 到颜色=>用渐变阴影填充地图。
- Undo, and drag a **discrete dimension (Region)** to **color** => **unique color** for each region.

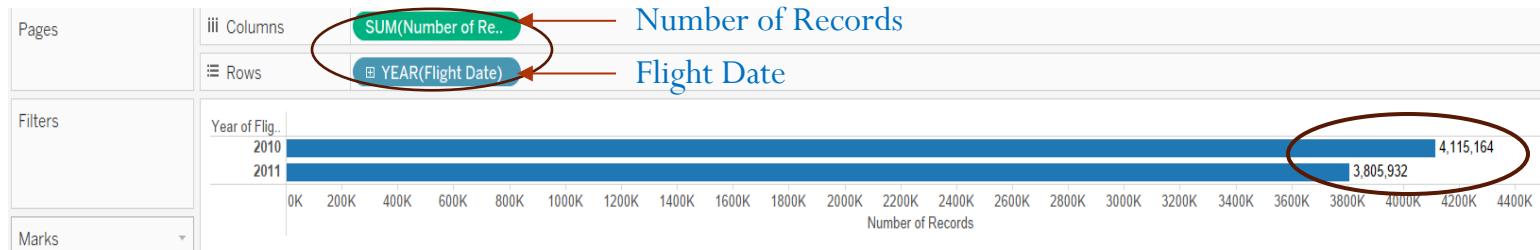


# In-class Exercise

The **Flights\_data.twbx** file contains information on all U.S. airline flights from 2010 and 2011.

Create a Tableau worksheet to answer each of the following questions:

- A quick viz of flights over time shows a drop of more than 300,000 flights from 2010 to 2011:



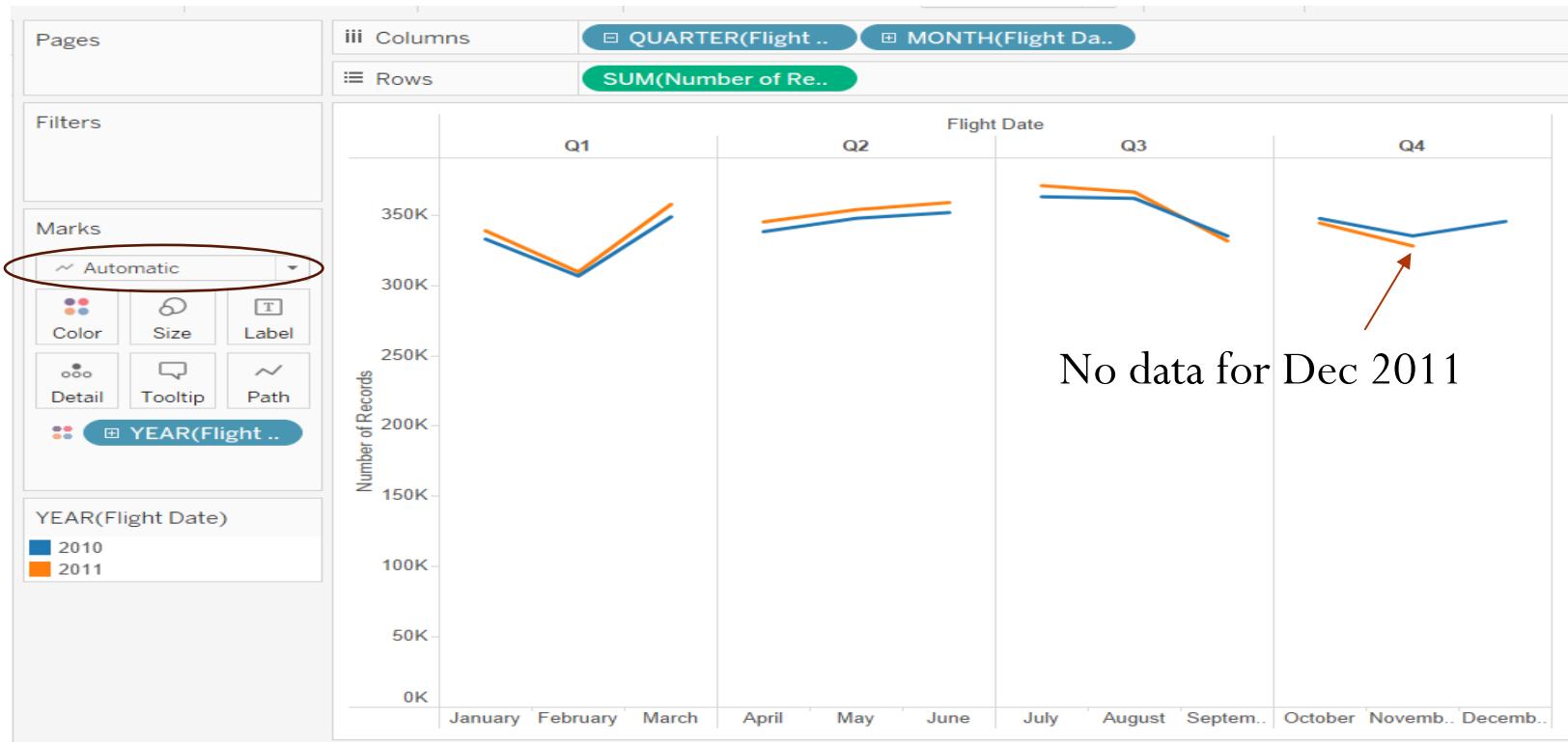
From your experience, you know that flight activity **did not drop** so significantly between 2010 and 2011. Why does this viz seem to indicate otherwise?

- 1) Investigate flights over time in more detail - drill down to quarters/months - determine why the yearly numbers show a large drop.
- 2) Identify days that are particularly light in terms of travel each year. Are there particular holidays that seem to experience substantially lower numbers of flights?
- 3) Determine who operates the most flights. Is there one airline that operated more flights than any other single airline for each week in this dataset?
- 4) Which states experience the longest departure delays. Create a map that visualizes the departure delay that you can expect in each origin state. Allow the user to filter by airline. Colors should range from dark blue for states with short departure delays to dark red for states with long departure delays.

# Exercise Solution

1) Investigate flights over time in more detail to determine why these yearly numbers show such a large drop.

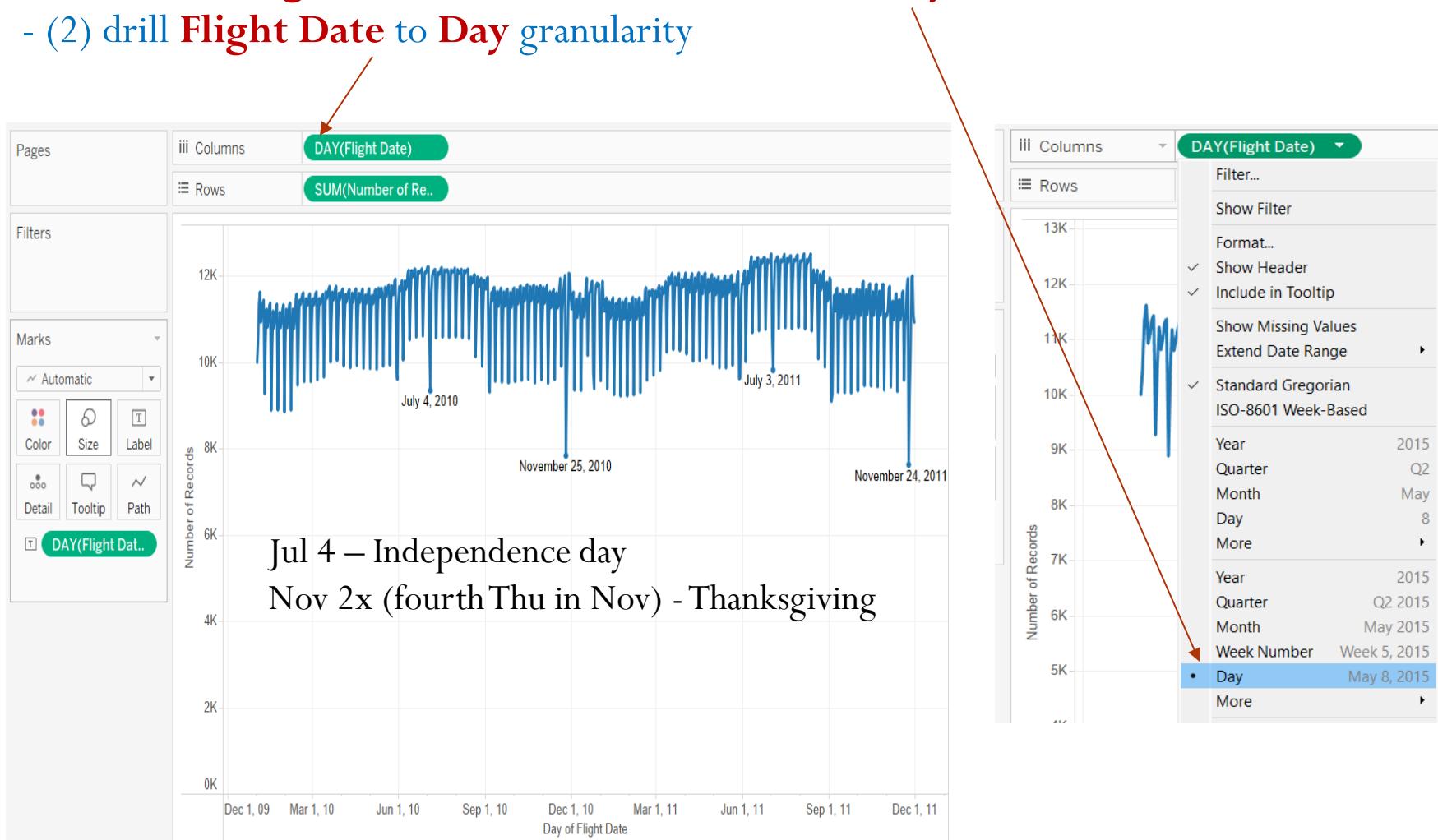
- drag **number of records** to rows
- drag **flight date** to column => drill down to **quarters/months**.
- move **year** to **color mark card** to compare the different years (2010 and 2011)
- change to line graph if need to.



# Exercise Solution

2) Are there particular holidays that seem to experience substantially lower numbers of flights?

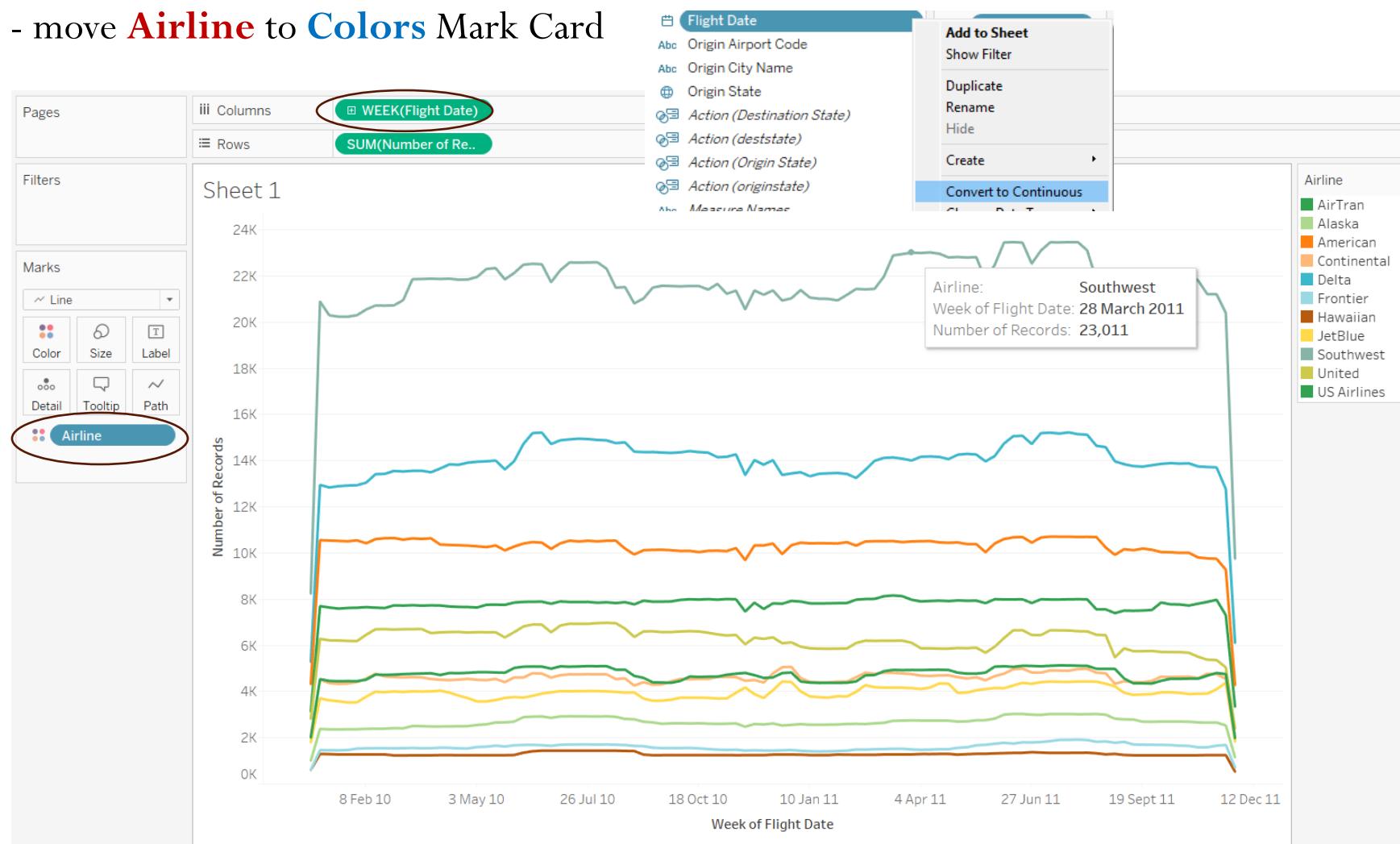
- (1) convert **Flight Date** to continuous – make sure **Day format** is selected
- (2) drill **Flight Date** to **Day** granularity



# Exercise Solution

3) Is there one airline that operated more flights than any other single airline for **each week** in this dataset?

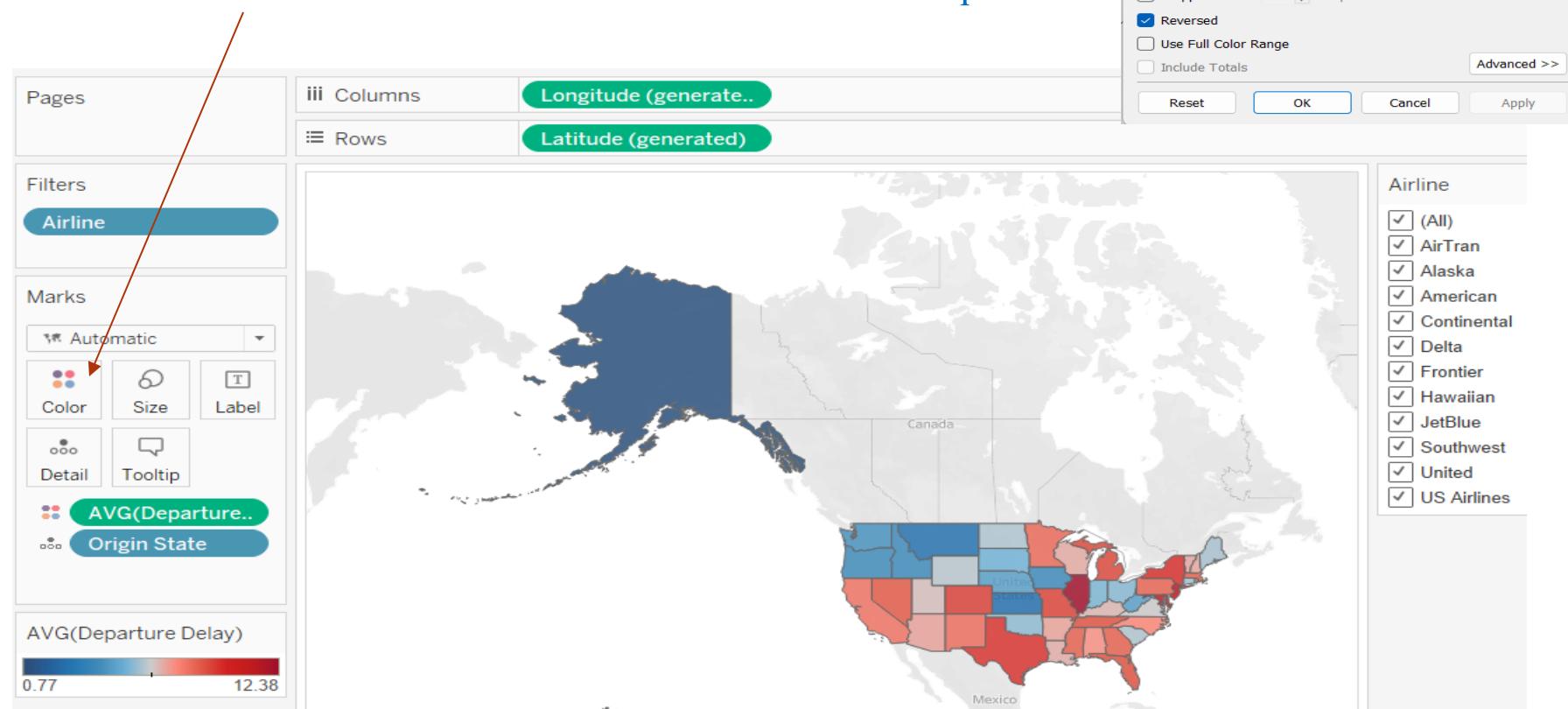
- drill down **Flight Date (convert to continuous)** to Week
- move **Airline** to **Colors** Mark Card



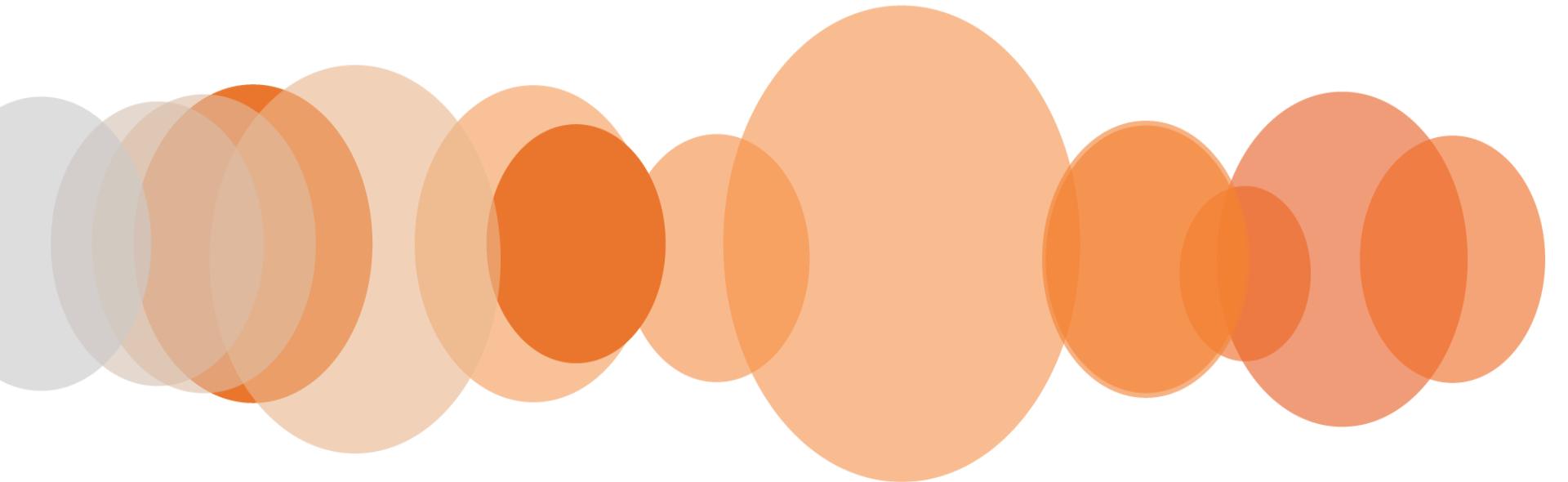
# Exercise Solution

4) Create a map that visualizes the **departure delay** that you can expect in each **origin state**. Allow the user to filter by airline. Colors should range from dark blue for states with short departure delays to dark red for states with long departure delays.

- double-click **Origin State** – bring out the map
- change **Departure Delay** to **AVG**
- click on **Color Mark Card** to edit => select red-blue palette



# Handling Poorly Formatted Data



Open “**Data Prep – Flights.xlsx**” Excel file - shows number of resolved incidents per Employee/month.

A	B	C	D	E	F	G	H	I	J	K	L	M	N
Flights Data Summary													
This report was generated on 1-1-15													
Employee	1/1/2014	1/2/2014	1/3/2014	1/4/2014	1/5/2014	1/6/2014	1/7/2014	1/8/2014	1/9/2014	1/10/2014	1/11/2014	1/12/2014	
B-002	4	1	5	2	3	0	3	1	2	0	2	5	
E-055	1	2	1	3	4	1	4	0	2	1	4	0	
E-075-II	14	17	16	15	18	16	14	17	12	13	14	12	
B-066	4	4	5	2	5	0	0	2	0	1	0	3	
C-025-II	17	13	17	18	17	17	12	15	17	17	14	15	
E-030	2	2	1	1	0	3	5	5	0	2	4	1	
C-001-II	14	14	14	14	13	18	17	14	13	18	15	14	
E-038	4	1	0	4	0	2	5	0	2	2	2	2	
C-054	2	5	4	4	2	3	0	5	5	5	3	5	
A-081	3	2	4	5	2	2	2	4	1	4	2	0	
B-031-II	14	14	14	14	15	13	15	14	12	16	12	18	
D-019	2	3	0	0	4	4	1	2	5	0	5	5	
E-096	2	0	4	4	5	3	3	0	5	4	2	0	
D-026	0	2	0	2	5	3	1	0	0	2	5	4	
E-022	3	3	4	3	4	2	0	3	2	3	3	1	
C-015	1	5	3	5	2	1	3	3	1	1	5	2	
B-062-II	14	12	16	16	16	18	12	12	18	16	12	17	
E-029	5	1	2	4	0	3	5	4	5	3	4	5	
A-037	2	2	0	2	3	4	2	0	2	1	2	2	
E-087-II	14	17	13	17	18	13	13	12	13	16	13	16	
C-040	5	0	4	5	3	5	2	1	1	4	2	1	
A-077	3	5	3	5	2	3	5	3	4	5	4	4	
C-041-II	18	18	15	15	15	17	17	12	18	17	16	16	
D-005	4	0	5	3	2	3	5	1	0	2	3	3	
E-046-II	12	13	14	17	16	14	14	18	13	16	15	17	
C-053	2	5	5	0	2	1	0	4	5	1	1	3	

# Data Interpreter

NEED Check

Open “Data Prep – Flights.xlsx” Excel file using Tableau

There are some issues - **no column names**, the **headers** from Excel have **a lot of nulls**, and so on.

Use **Data Interpreter** (Tableau’s built-in tool for data preparation).

Click to turn on

Drag “Resolved Incidents” to data connection canvas

Need more data?

Drag tables here to relate them. [Learn more](#)

No column names

Null values

(Due to report headers found in excel file)

Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc
Resolved Incidents	Resolved I...	Resolved In...	Resolved In...	Resolved In...	Resolved In...							
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	
Flights Data Sum...	Null	Null	Null	Null								
This report was ...	Null	Null	Null	Null								
Employee	1/1/2014	1/2/2014	1/3/2014	1/4/2014	1/5/2014	1/6/2014	1/7/2014	1/8/2014	1/9/2014	1/10/2014	1/11/2014	
B-002	4	1	5	2	3	0	3	1	2	0	2	
E-055	1	2	1	3	4	1	4	0	2	1	4	
E-075	14	17	16	15	18	16	14	17	12	13	14	

Output from  
Data Interpreter  
- report headers  
deleted

Abc	#	#	#	#	#	#	#	#	#	#	:
Resolved Incidents	Resolved Incide...										
Employee	1/1/2014	1/2/2014	1/3/2014	1/4/2014	1/5/2014	1/6/2014	1/7/2014	1/8/2014	1/9/2014	1/10/2014	1/11/2014
B-002	4	1	5	2	3	0	3	1	2	0	2
E-055	1	2	1	3	4	1	4	0	2	1	4
E-075	14	17	16	15	18	16	14	17	12	13	14

# Data Interpreter

Text  
Text

- If we want more specifics on what the Data Interpreter did, we can **click “Review Results”**. This will open an **Excel file** describing the changes.
- In the **Resolved Incidents** tab, we see which fields are being used as **headers**, in **red**, and which are considered **data**, in **green**

File Data Server Window Help

Connections Add

Data Prep - Flights Microsoft Excel

Sheets

Cleaned with Data Interpreter

Review the results. (To undo changes, clear the check box.)

Ideal

Irregular Delimiter

Resolved Incidents

Tiers

New Union

Resolved Incidents (Data Prep - Flights)

Resolved Incidents

Sort fields Data source order

Abs	#	#	#	#	#
Resolved Incidents	Resolved Incide...	Resolved Incide...	Resolved Incide...	Resolved Incide...	R
Employee	1/1/2014	1/2/2014	1/3/2014	1/4/2014	1
B-002	4	1	5	2	

Key for Understanding the Data Interpreter Results

Use the key to understand how your data source has been interpreted.  
To view the results, click a worksheet tab.  
Note: Tableau never makes changes to your underlying data source.

**Key:**

- Data is interpreted as column headers (field names).
- Data is interpreted as values in your data source.
- Data derived from an Excel merged cell is interpreted as value in your data source.
- Data is ignored and not included as part of your data source.
- Data has been excluded from your data source.

Note: To search for all excluded data, use CRTL +F on Windows or Command F on the Mac, and then type "\*\*\*DATA REMOVED\*\*\*".

Flights Data Summary												
This report was generated on 1-1-15												
Employee	1/1/2014	1/2/2014	1/3/2014	1/4/2014	1/5/2014	1/6/2014	1/7/2014	1/8/2014	1/9/2014	#####	#####	#####
B-002	4	1	5	2	3	0	3	1	2	0	2	5 Data
E-055	1	2	1	3	4	1	4	0	2	1	4	0 Data
E-075	14	17	16	15	18	16	14	17	12	13	14	12 Data
B-066	4	4	5	2	5	0	0	2	0	1	0	3 Data
C-025	17	13	17	18	17	17	12	15	17	17	14	15 Data
E-030	2	2	1	1	0	3	5	5	0	2	4	1 Data
C-001	14	14	14	14	13	18	17	14	13	18	15	14 Data
E-038	4	1	0	4	0	2	5	0	2	2	2	2 Data
C-054	2	5	4	4	2	3	0	5	5	5	3	5 Data
A-081	3	2	4	5	2	2	2	4	1	4	2	0 Data
B-031	14	14	14	14	15	13	15	14	12	16	12	18 Data
D-019	2	3	0	0	4	4	1	2	5	0	5	5 Data
E-096	2	0	4	4	5	3	3	0	5	4	2	0 Data

# Pivot and Split

Ideal tab

Date	Employee	Resolved Incidents
1/1/2014	B-002	4
1/1/2014	E-055	1
1/1/2014	E-075	14
1/1/2014	B-066	4
1/1/2014	C-025	17
1/1/2014	E-030	2
1/1/2014	C-001	14
1/1/2014	E-038	4
1/1/2014	C-054	2
1/1/2014	A-081	3
1/1/2014	B-031	14
1/1/2014	D-019	2
1/1/2014	E-096	2
1/1/2014	D-026	0
1/1/2014	F-022	3

Resolved Incidents tab

Employee	1/1/2014	2/1/2014	3/1/2014	4/1/2014	5/1/2014
B-002	4	1	5	2	3
E-055	1	2	1	3	4
E-075	14	17	16	15	18
B-066	4	4	5	2	5
C-025	17	13	17	18	17
E-030	2	2	1	1	0
C-001	14	14	14	14	13
E-038	4	1	0	4	0
C-054	2	5	4	4	2
A-081	3	2	4	5	2

The screenshot shows the Tableau Data Prep interface. On the left, there's a 'Connections' section with 'Data Prep - Flights' selected. Below it is a 'Sheets' section with 'Ideal' highlighted. A red arrow points from the text 'Drag ideal icon' to the 'Ideal' tab in the Sheets list. To the right, there's a large text area with the heading 'Data Prep - Flights' and a placeholder 'Drag tables here'. Another red arrow points from the text 'Drag tables here' to the placeholder area.

- The “Ideal” tab shows how we **wish** the data would be formatted – like a database table.  
“理想”选项卡显示了我们希望数据的格式——就像数据库表一样。
- However, sometimes we receive data that looks more like what we see in the “Resolved Incidents” tab. 然而，有时我们收到的数据看起来更像我们在“已解决事件”选项卡中看到的数据。
- Use **Pivot** and **Split** features to automatically reshape the data to get them ready for analysis in Tableau. 使用Pivot和Split功能自动重塑数据，为在Tableau中进行分析做好准备。

# Pivot 在枢轴上转动

Change format from **column-per-month** layout into a **single date column** and a **single column for Resolved Incidents**. 将格式从每列每月布局更改为单个日期列和已解决事件的单个列。

- **Select all the date columns.** Click on the first, scroll if necessary, then shift click on the last.
- **Right click and select “Pivot”**
- The **pivot** feature **merges the information from the original columns and rows into two new columns** – Pivot field **names**, and Pivot field **values**.
- “**Pivot field names**” is actually “**Dates**”, click to open the menu and select rename. Similarly, “**Pivot field values**” can be renamed “**Resolved Incidents**”

枢轴功能将原始列和行中的信息合并到两个新列中——枢轴字段名称和枢轴字段值。

“枢轴字段名称”实际上是“日期”，单击以打开菜单并选择重命名。同样，“数据透视场值”可以重命名为“已解决的事件”

column-per-month												
Abc	#	#	#	#	#	#	#	#	#	#	#	#
Resolved Incidents	Resolved Incide...	Resolved Incidents	Resolved Incidents	Resolved Incidents	Resolved Incidents							
Employee	1/1/2014	1/2/2014	1/3/2014	1/4/2014	1/5/2014	1/6/2014	1/7/2014	1/8/2014	1/9/2014	1/10/2014	1/11/2014	1/12/2014
B-002	4	1	5	2	3	0	3	1	2	0		
E-055	1	2	1	3	4	1	4	0	2	1		
E-075	14	17	16	15	18	16	14	17	12	13		
B-066	4	4	5	2	5	0	0	2	0	1		
G-025	17	13	17	18	17	17	12	15	17	17	14	15
E-030	2	2	1	1	0	3	5	5	0	2	4	1
G-001	14	14	14	14	13	18	17	14	13	18	15	14

Single Date Column	Single Resolved Incidents Number Column
Pivot	Pivot
Pivot Field Na...	Pivot Field Val...
1/1/2014	4 B-002
1/10/2014	0 B-002
1/11/2014	2 B-002
1/12/2014	5 B-002
1/2/2014	1 B-002
1/3/2014	5 B-002
1/4/2014	2 B-002

# Split

“Employee” field has two pieces of information – a **location** code, A, B, C, D, or E, followed by an **Employee ID** number. Split the column based on the hyphen delimiter:

- Right click “Employee” at **column header** and **select Split**
- There are now two new fields – Employee - Split 1 and Split 2 (rename to **Location** and **Employee ID** respectively)

Right click “Employee” at column header and select Split

Resolved Incidents	Pivot	Pivot
Employee	Pivot Field Names	Pivot
B-002	1/1/2014	
B-002	1/10/2014	
B-002	1/11/2014	

Location Code

Employee ID

Resolved Incidents	Calculation	Calculation
Employee	Employee - Sp...	Employee - Sp...
B-002	B	2
E-055	E	55
E-075	E	75
B-066	B	66
C-025	C	25

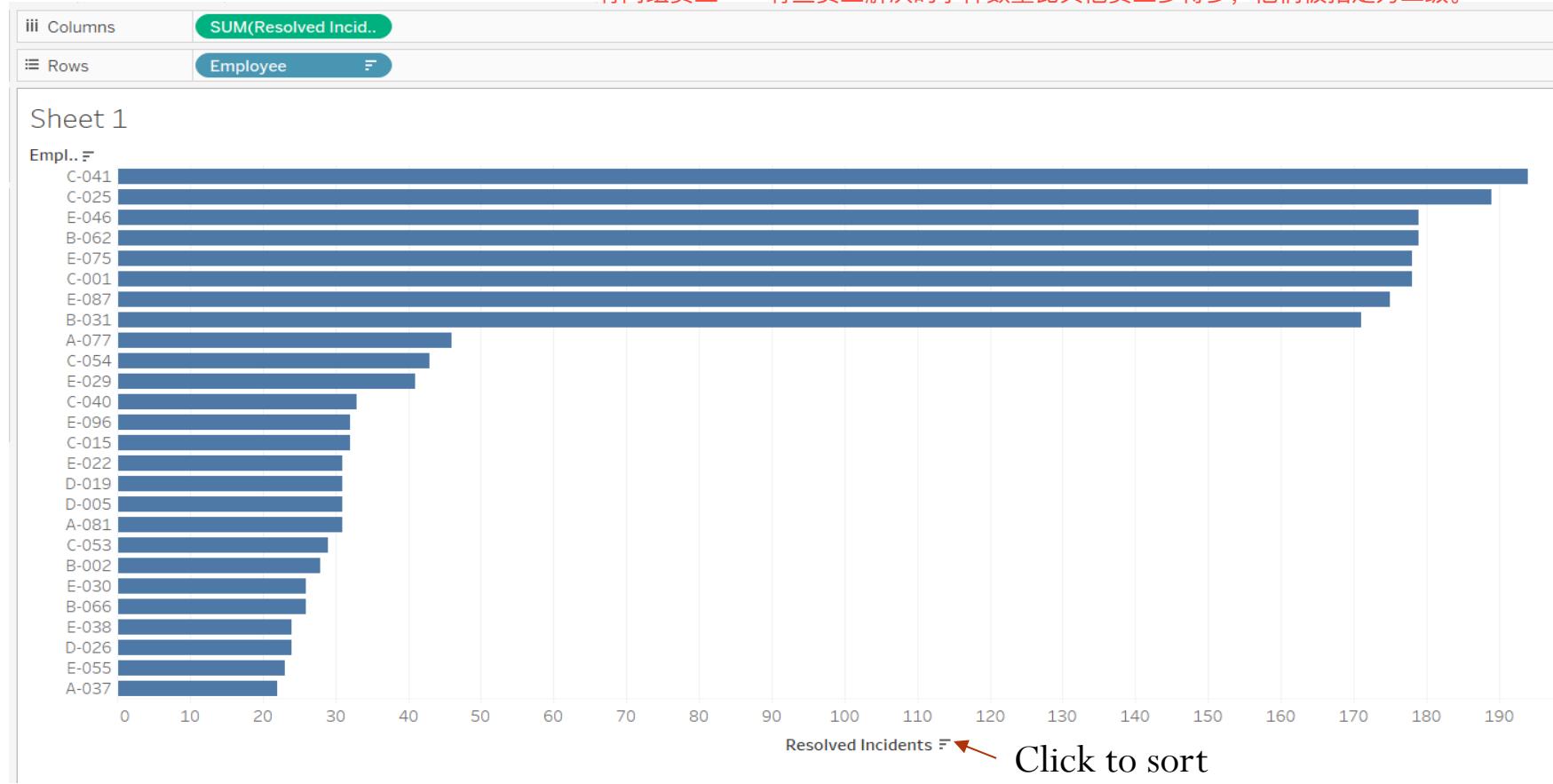
Employee – Split 1  
- Right-click rename to **Location**

Employee – Split 2  
- Right-click rename to **Employee ID**

# Custom Split

- Create a new visualization - bring “Employee” to Rows, “Resolved Incidents” to Columns, and sort based on SUM(Resolved Incidents).
- There are two groups of employees – some who resolve a much **higher number of incidents** than others, and they have a **Tier II** designation.

有两组员工——有些员工解决的事件数量比其他员工多得多，他们被指定为二级。



# Custom Split

In the **original Excel data set**, there's a tab called **Tiers**. This report adds a **-II** to the end of an employee ID if they're tier II. Because **not all rows have this -II, a standard split won't work**.

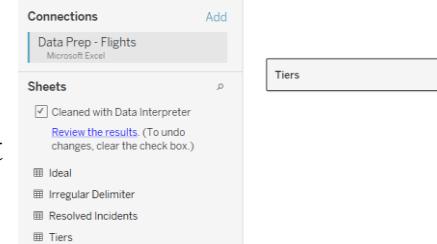
Both Split and Custom Split require **consistent delimiters**. If the data has irregular delimiters, Tableau won't be able to split out the data using these options.

拆分和自定义拆分都需要一致的分隔符。如果数据有不规则的分隔符，Tableau将无法使用这些选项拆分数据。

A	B	C	D	E	F	G	H	I	J	K	L	M	N
Flights Data Summary													
This report was generated on 1-1-15													
Employee	1/1/2014	2/1/2014	3/1/2014	4/1/2014	5/1/2014	6/1/2014	7/1/2014	8/1/2014	9/1/2014	#####	#####	#####	Header
B-002	4	1	5	2	3	0	3	1	2	0	2	5	Data
E-055	1	2	1	3	4	1	4	0	2	1	4	0	Data
E-075-II	14	17	16	15	18	16	14	17	12	13	14	12	Data
B-066	4	4	5	2	5	0	0	2	0	1	0	3	Data
C-025-II	17	13	17	18	17	17	12	15	17	17	14	15	Data
E-030	2	2	1	1	0	3	5	5	0	2	4	1	Data
C-001-II	14	14	14	14	13	18	17	14	13	18	15	14	Data
E-038	4	1	0	4	0	2	5	0	2	2	2	2	Data
C-054	2	5	4	4	2	3	0	5	5	5	3	5	Data
A-081	3	2	4	5	2	2	2	4	1	4	2	0	Data
B-031-II	14	14	14	15	13	15	14	12	16	12	18	18	Data
D-019	2	3	0	0	4	4	1	2	5	0	5	5	Data
E-096	2	0	4	4	5	3	3	0	5	4	2	0	Data
D-026	0	2	0	2	5	3	1	0	0	2	5	4	Data
E-022	3	3	4	3	4	2	0	3	2	3	3	1	Data
C-015	1	5	3	5	2	1	3	3	1	1	5	2	Data

# Custom Split

- Open a **new Tableau file** and recreate the viz, this time using the **Tiers** sheet
- Remember to use **the Data Interpreter**, and **Pivot** the dates again
- Click on the Employee column** to open the menu and select **Custom Split**
- Choose the **hyphen** delimiter
- We want to have **3 columns** => break off that 3<sup>rd</sup> column with the tier II indicator



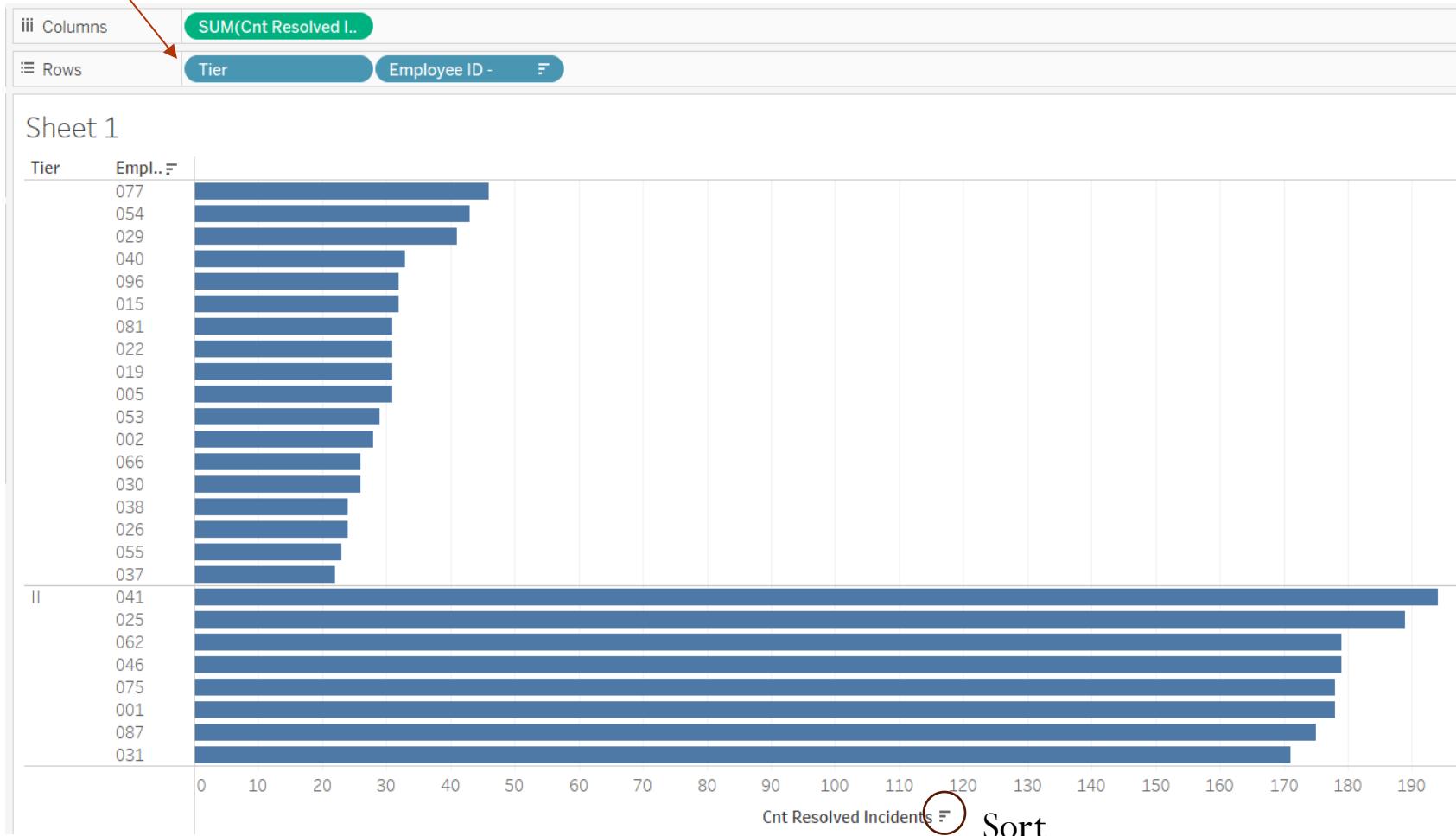
我们希望有3列=>用二级指示器断开第3列

Employee ID	Tier
E-055	055
E-055	055
E-055	055
E-075-II	075

Employee - Split 1
Employee - Split 2
Employee - Split 3

# Custom Split

Add the Tier dimension to the Rows to show the differentiated tier for the employees



# Custom Split

File Data Server Help

Connect

To a File

- Excel
- Text file
- Access
- JSON file
- PDF file
- Spatial file
- Statistical file
- More...

To a Server

- Tableau Server
- MySQL
- Oracle
- Amazon Redshift
- Microsoft SQL Server
- More...

Sample Workbooks

- Superstore
- Regional
- World Indicators

Open

- Step 1 Viz
- Global Superstore
- SUMMER TEST ...
- Enrollment Com...
- Academic Depar...
- Academic Depar...
- Course Assessm...
- Course Assessm...
- Analysis of Degr...
- ECDB Sankey
- 12 Term Major C...
- Program Array ...
- Program Array ...
- Enrollment Com...
- Enrollment Clas...

Discover

Open a Workbook

Training

- Getting Started
- Connecting to Data
- Visual Analytics
- Understanding Tableau
- More training videos...

Viz of the Week

Data Role T Models

Resources

- Blog - Get Tableau 10.4 beta to better collaborate on trusted data
- Tableau Conference 2017
- Forums

RECORDED WITH  
2017 SCREENCASTOMATIC

# Building the Charts

- Bar Charts
- Scatterplots
- Line Graphs
- Histograms
- Heat Maps

# Bar Charts

- Open the **Sample Superstore** Excel file. Drag **Orders** to data connection canvas. Analyze **Sales** (double click), **Category** (double click) and **Sub-Category** (drag to Columns, left of Category).
- Rotate it (click **Swap** in the toolbar) and sort it (click **Sort** icon).
- Analyze the chart - Furniture (especially **Chairs**) seems to be selling well (**Sales**). But what about **Profits**?

Swap columns with rows

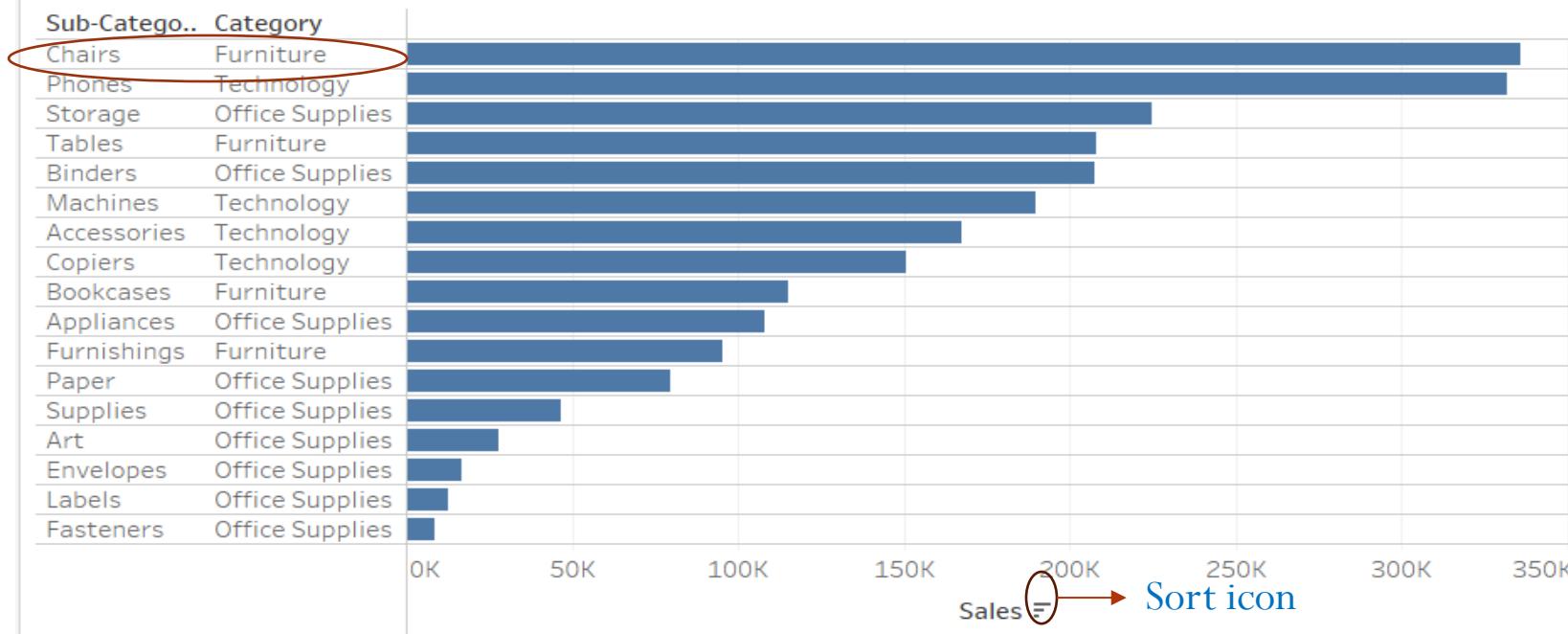


Sort ascending

Sort descending

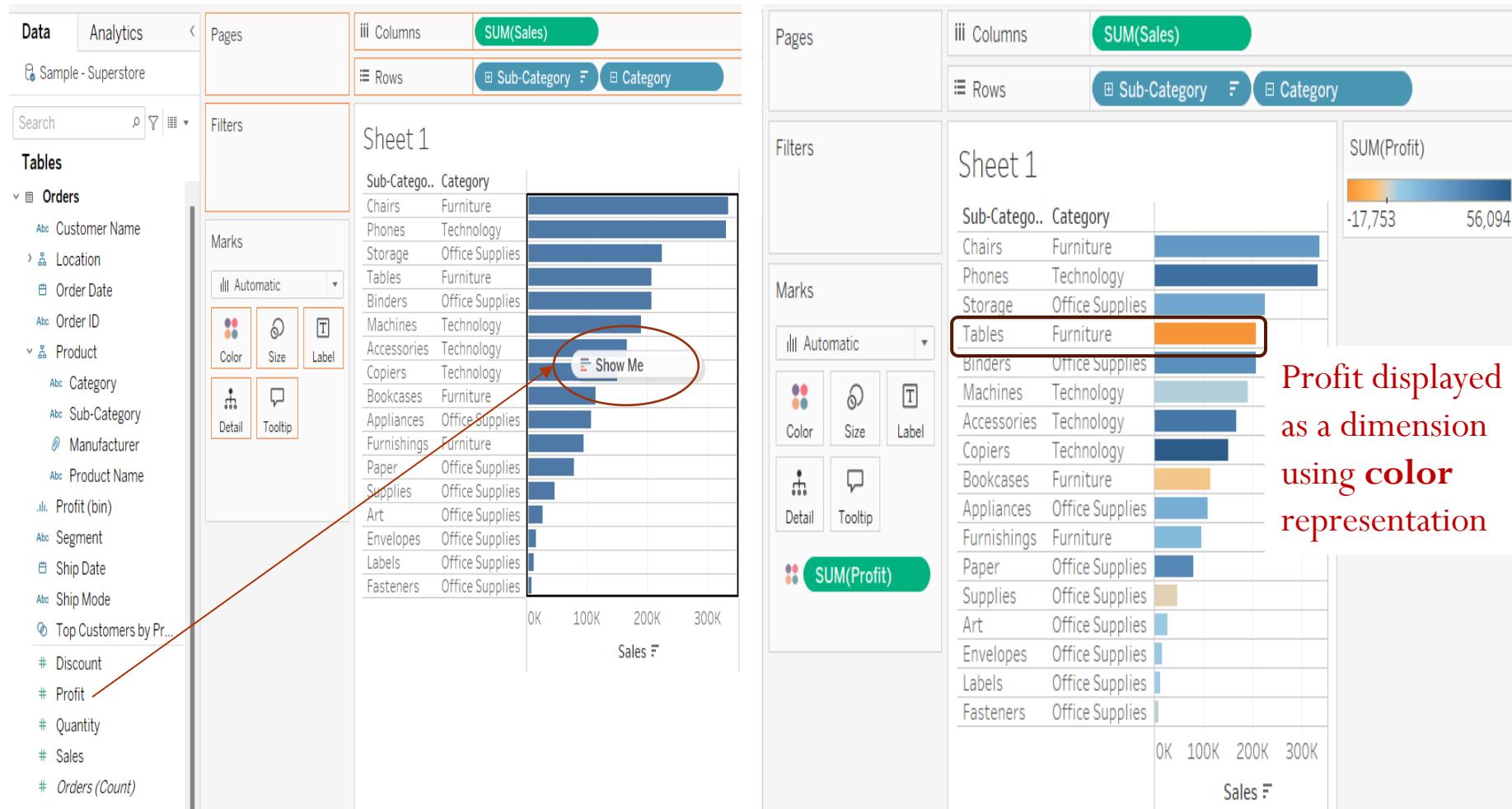
A screenshot of the Power BI ribbon. The 'Sort' icon, which is a downward-pointing arrow inside a square, is highlighted with a red oval. Other icons in the ribbon include 'File', 'Server', 'Window', 'Help', and some other sorting and filtering icons.

Sheet 1



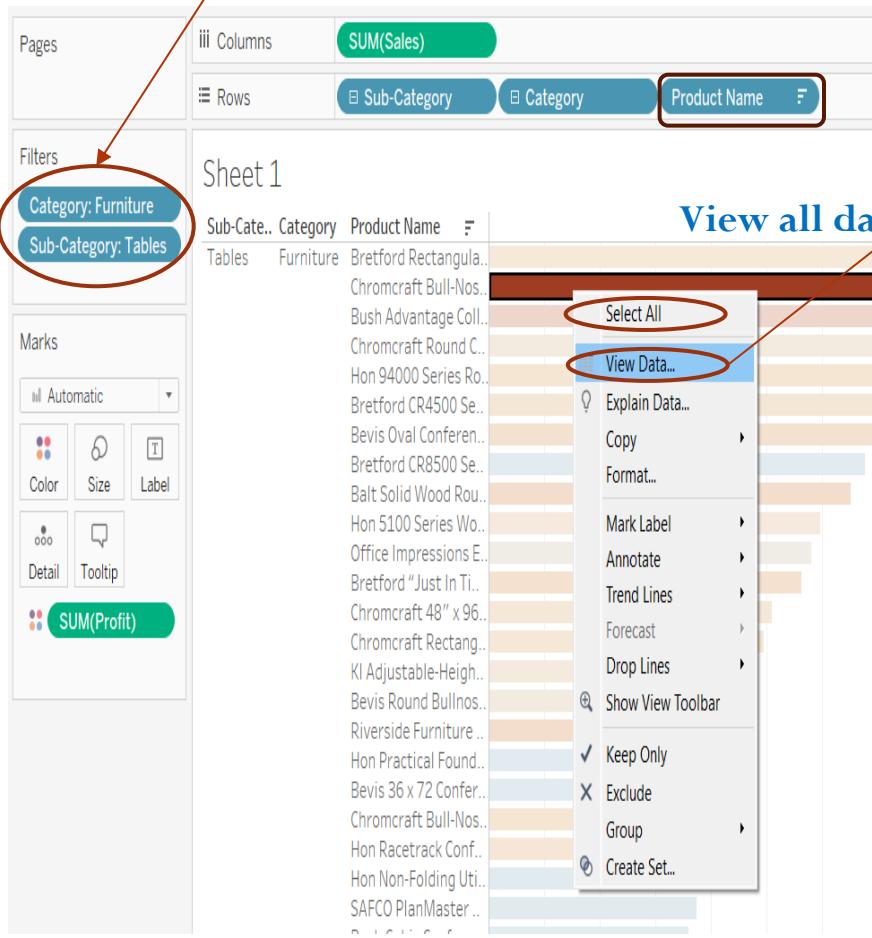
# Bar Charts

Drag **Profit** to the **Sheet Canvas** to get the “**show me**” cursor. Drop it and Tableau will show the **Color scale** for Profit. It’s clear that **Tables** are not profitable – and we are selling lots of them!



# Bar Charts

- Add filter for **Sub-Category:Tables** and further analyze at **Product level**.
- Drag out the **Product Name** to the right at the **Row shelf** and **Sort** descending.
- See actual underlying data from our top selling product here. Click the first bar...**right click** and **Select All, View Data** tab => a list of transactions for investigation.



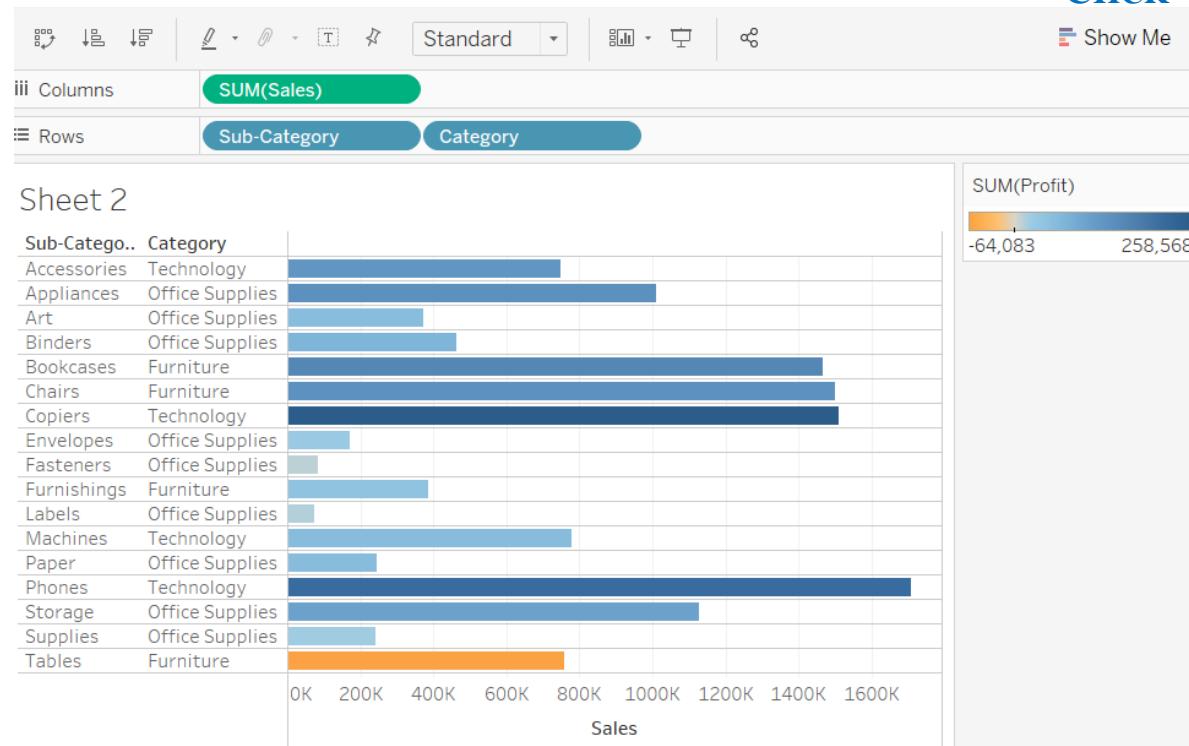
View Data: Sheet 1 (56 marks)

Summary  
56 rows 5 fields

Summary

Tables	Furniture	Anderson Hickey Conga Tabl...	#	#
Tables	Furniture	Barricks Non-Folding Utility T...	15.35	136.4
Tables	Furniture	Bevis Round Conference Roo...	-39.44	519.9
Tables	Furniture	Safco Drafting Table	34.07	638.0
Tables	Furniture	Hon 61000 Series Interactive...	-148.40	653.
Tables	Furniture	Balt Solid Wood Rectangular ...	-216.25	828.
Tables	Furniture	Riverside Furniture Stanwyck...	-415.93	917.9
Tables	Furniture	Bevis Rectangular Conferenc...	-586.84	1.007.

# Bar Charts



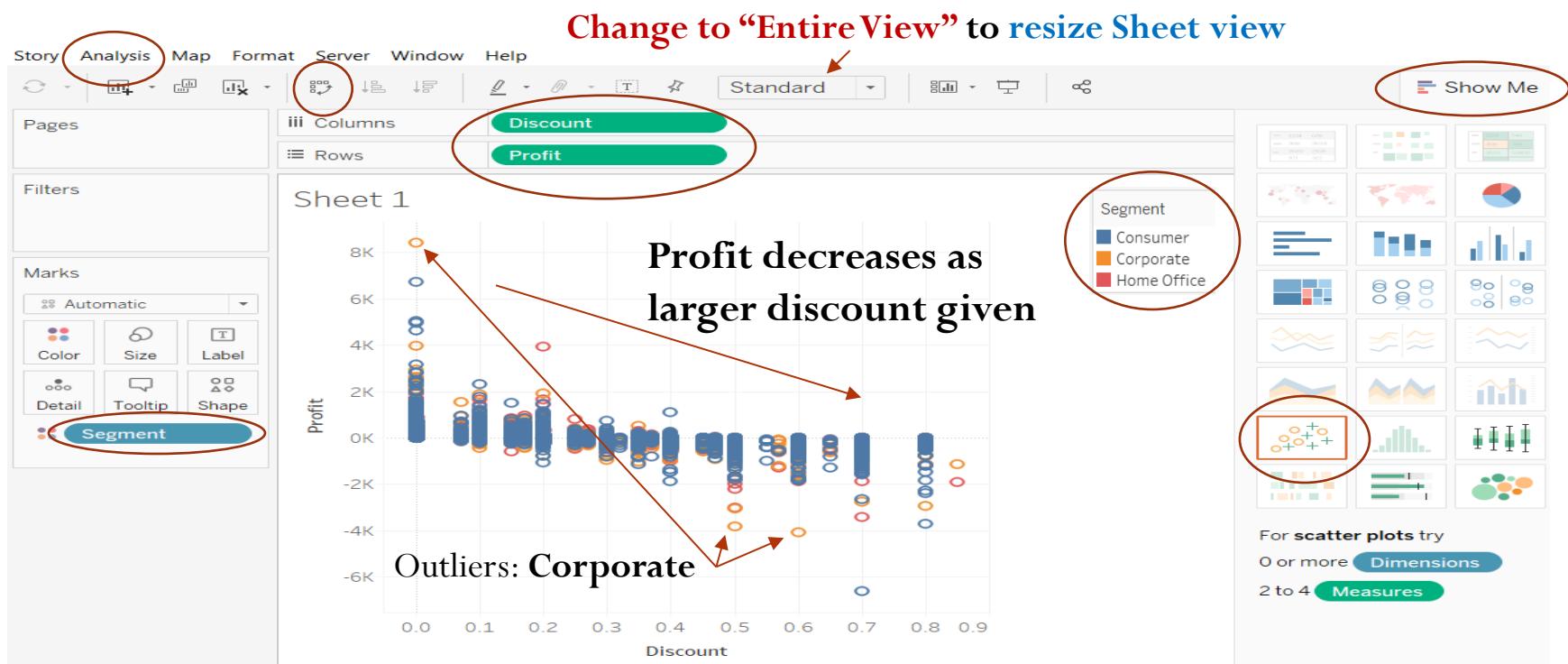
Click



- The Show Me pane provides other possibilities for visualizing the data selected – click a couple to see other options.
- Considerations for Information Visualization Design – bar charts for comparing categories
  - 信息可视化设计的注意事项-用于比较类别的条形图

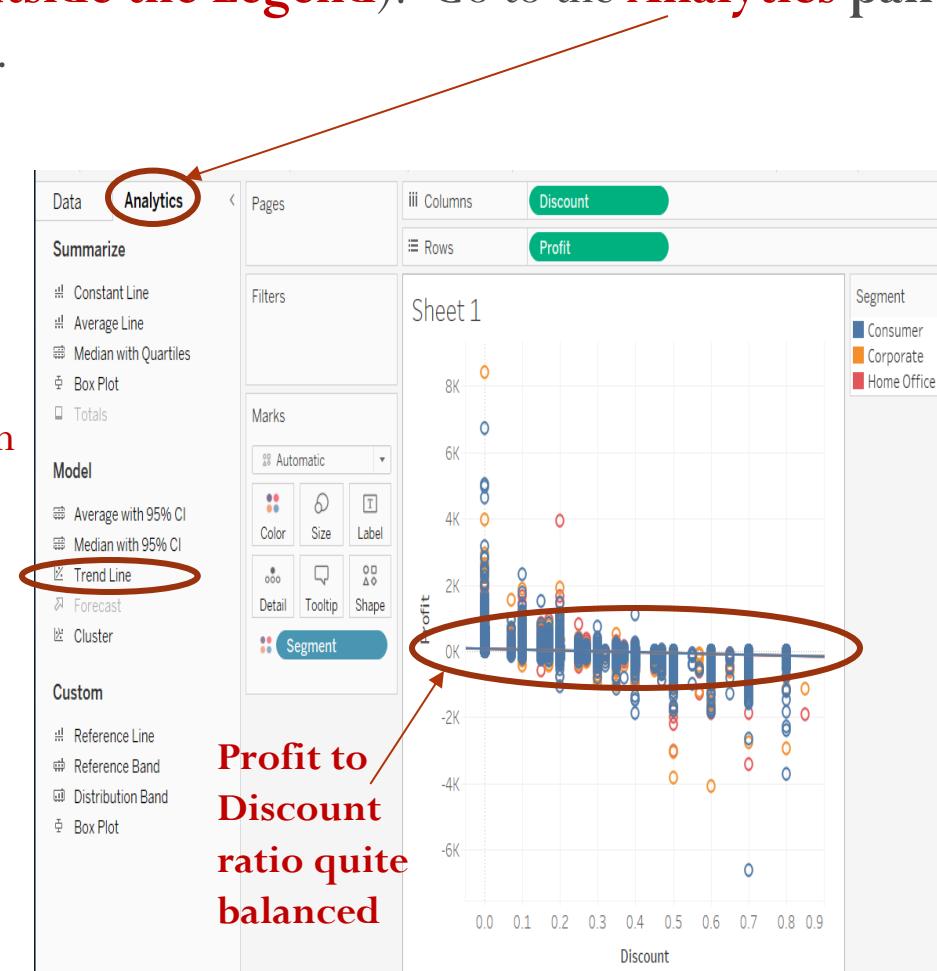
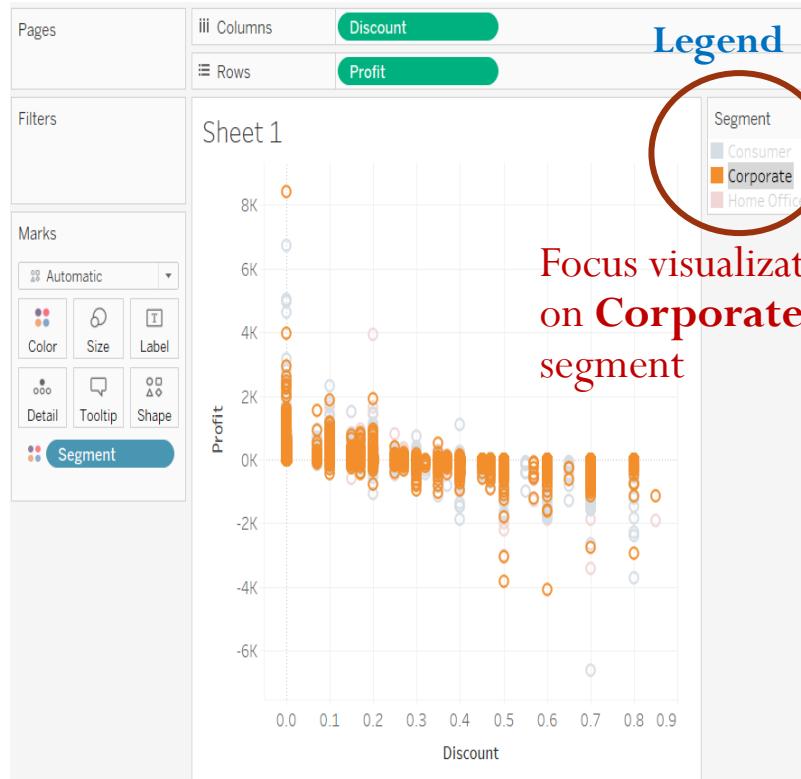
# Scatterplots

- Often used to visualize **relationships** between **numeric variables (Measures)**.
- Check if there is a relationship between **Discount** and **Profit**. Select the two fields and click **Scatterplot** on the **Show Me** menu. Go to **Analysis**, **uncheck Aggregate Measures – show all data points**.
- Swap** if needed to get Discount on the horizontal axis – looks like **Profit tends to decrease as a larger Discount is given** (confirms intuition).
- Drag **Segment** to **Color**. This makes it easy to see what segment certain outliers belong to.



# Scatterplots

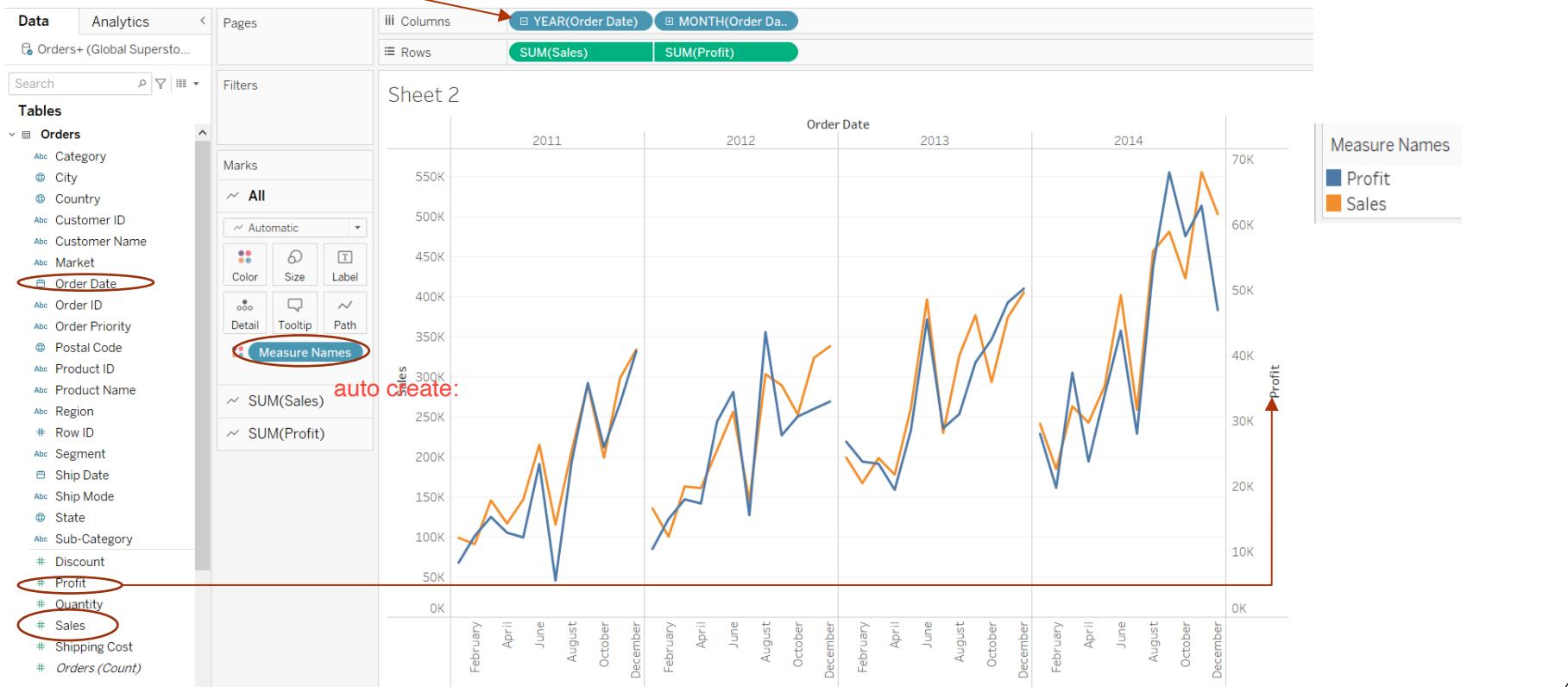
- To focus on just one segment (while leaving all the data displayed), **highlight selected values in the Legend**. E.g., focus visualization on **Corporate** segment by clicking it.
- Go back to viewing all the data (**click outside the Legend**). Go to the **Analytics** pane and drag out a **trend line** (choose **linear**).



# Line Graphs

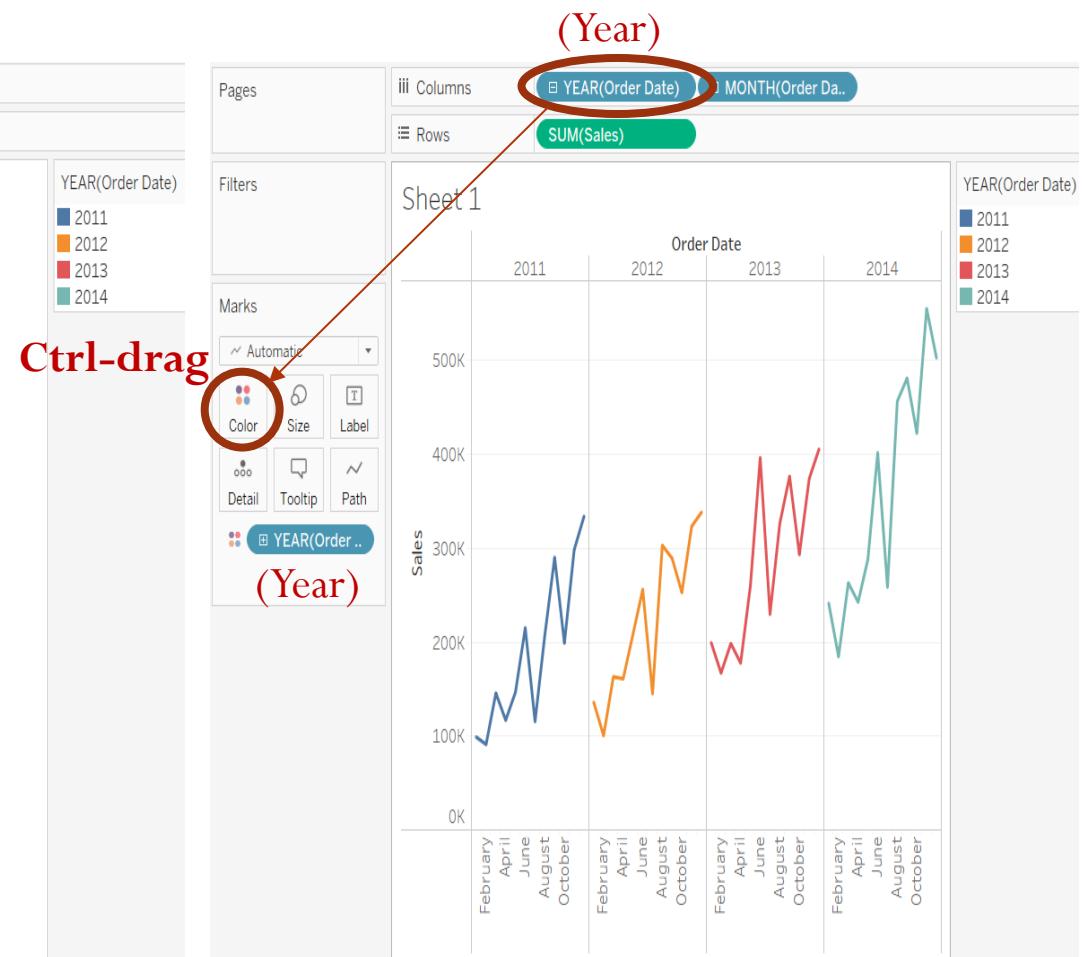
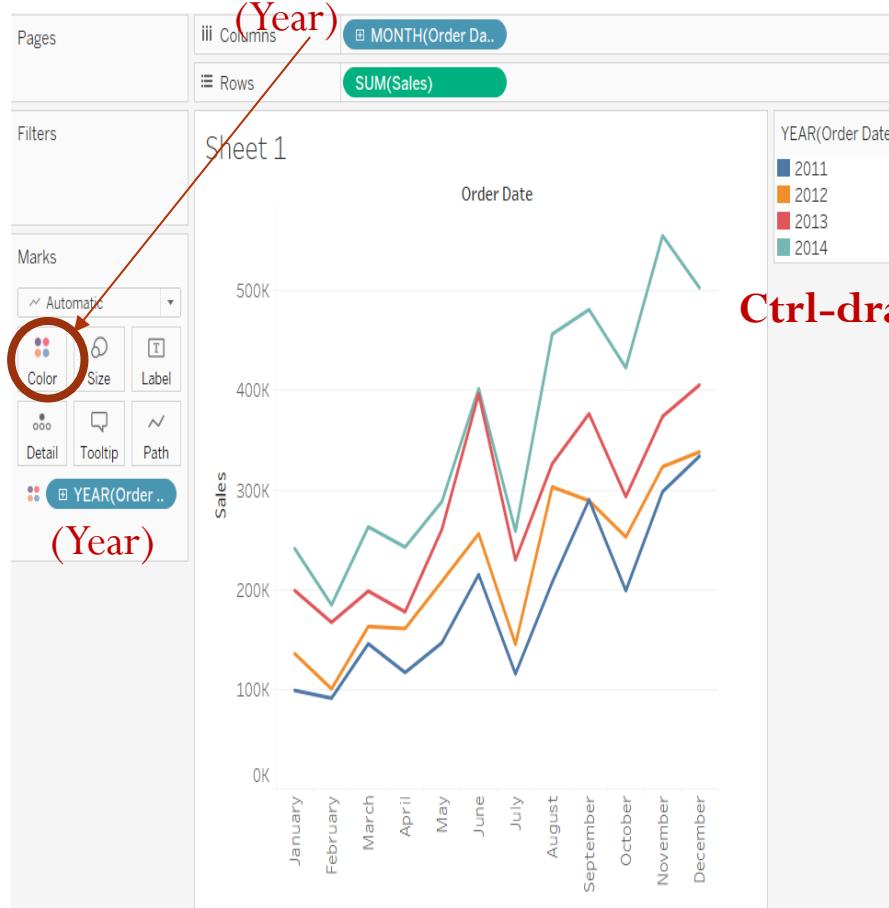
- Like Scatterplots, **line graphs** are used for analyzing **relationship between 2 variables** (scatterplot usually better for this) and often for analyzing **changes over time (time series)**.
- Let's look at **sales over time**. Double-click **Sales**, double-click **Order Date**. Expand to Quarter and then Month. **Drag Quarter out** of the viz.
- Go to **Analysis**, **check back Aggregate Measures => SUM**. Create **dual axis => drag Profit** to the **right axis**. Then **Undo** to get back to **Sales only**.

Click + to expand to get Quarter, then Month (remove Quarter subsequently)



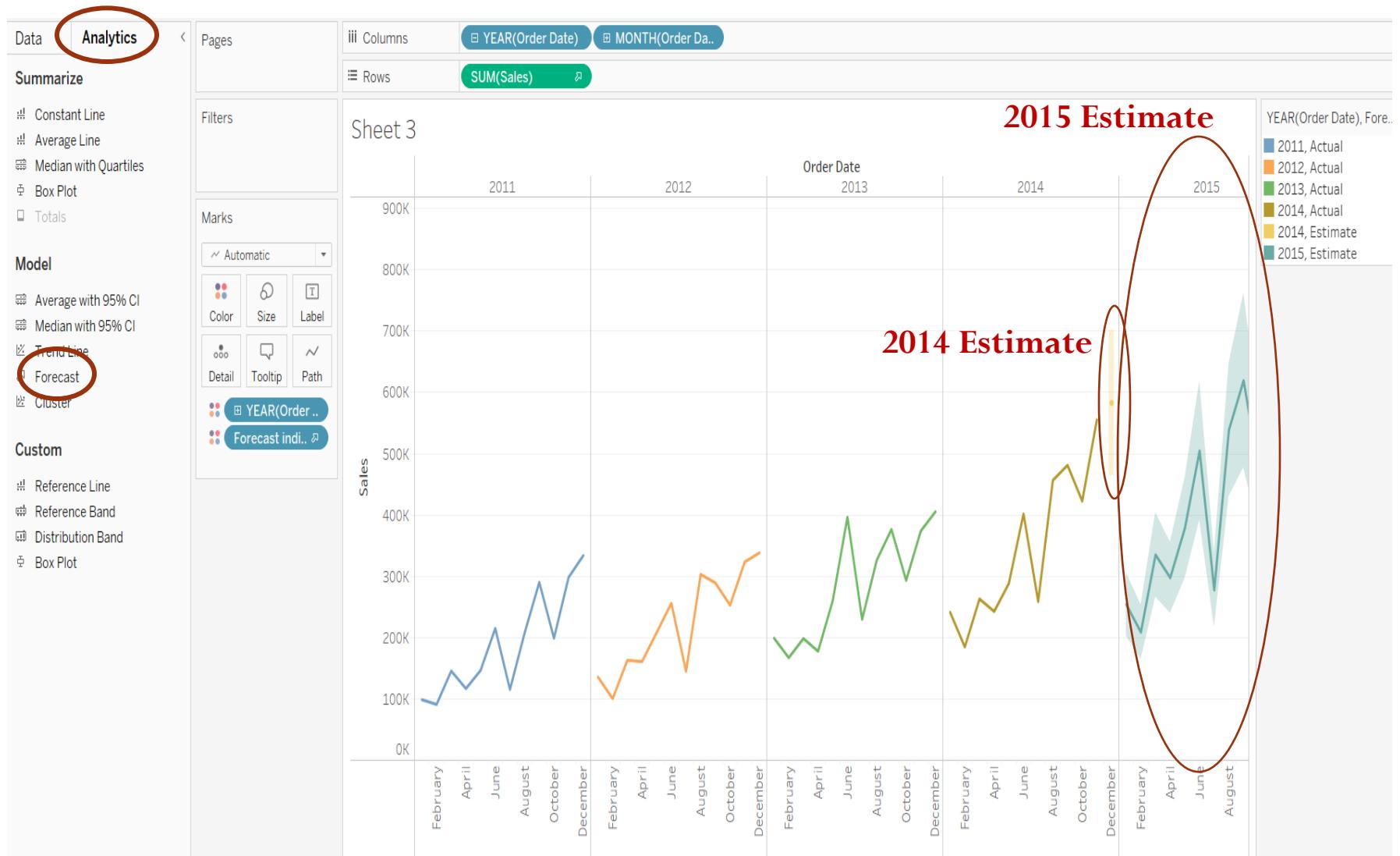
# Line Graphs

- With **discrete dates** (for this case) each part of the date is treated separately. For example, **drag Year to Color** and see direct comparisons between the years. (Note that, if you like the **current format** and just want **each year to be a different color** for clarity, you could **Ctrl-drag Year to color**).



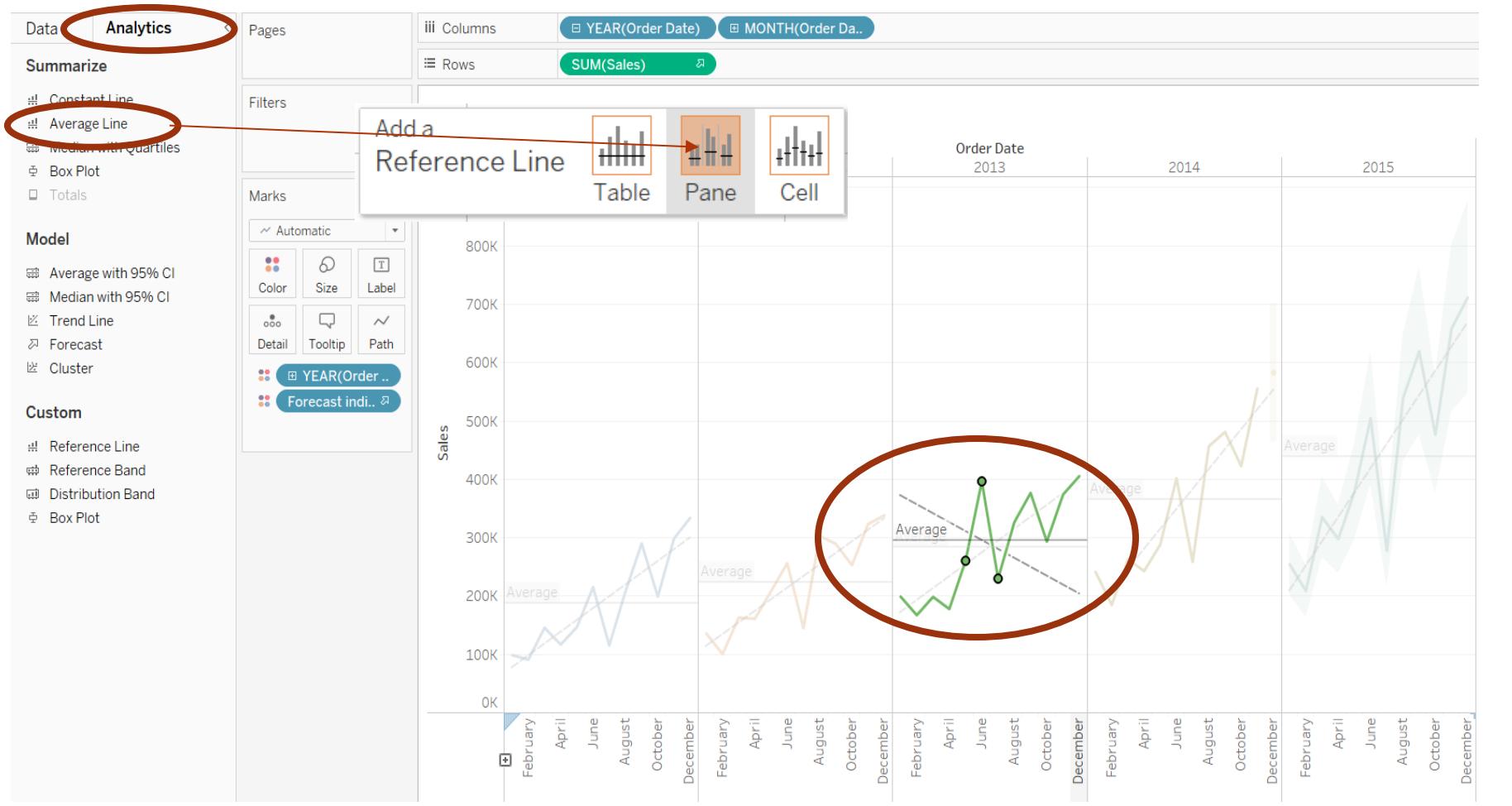
# Line Graphs

- To add a **Forecast** => go to **Analytics pane** and drag out **Forecast** to the Sheet Canvas.



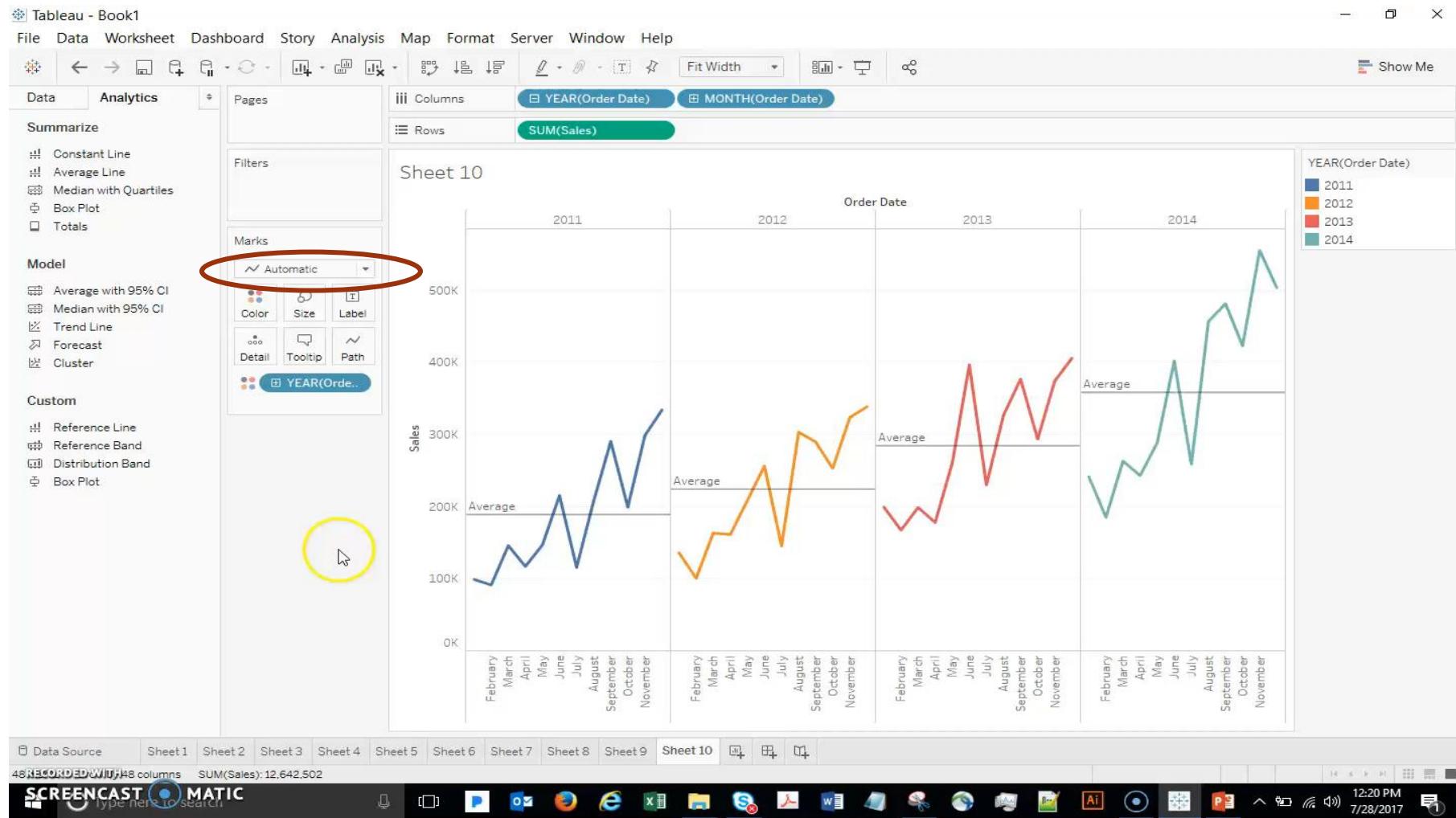
# Line Graphs

- Add reference line - drag **Average Line** (in Analytics tab) onto visualization => drop it onto the **Pane** icon. 把它放到窗格图标上
- Drill down on the data - average will dynamically adjust => **ctrl + select a few points** in the line graph and note the **new average line** (of the selected points).



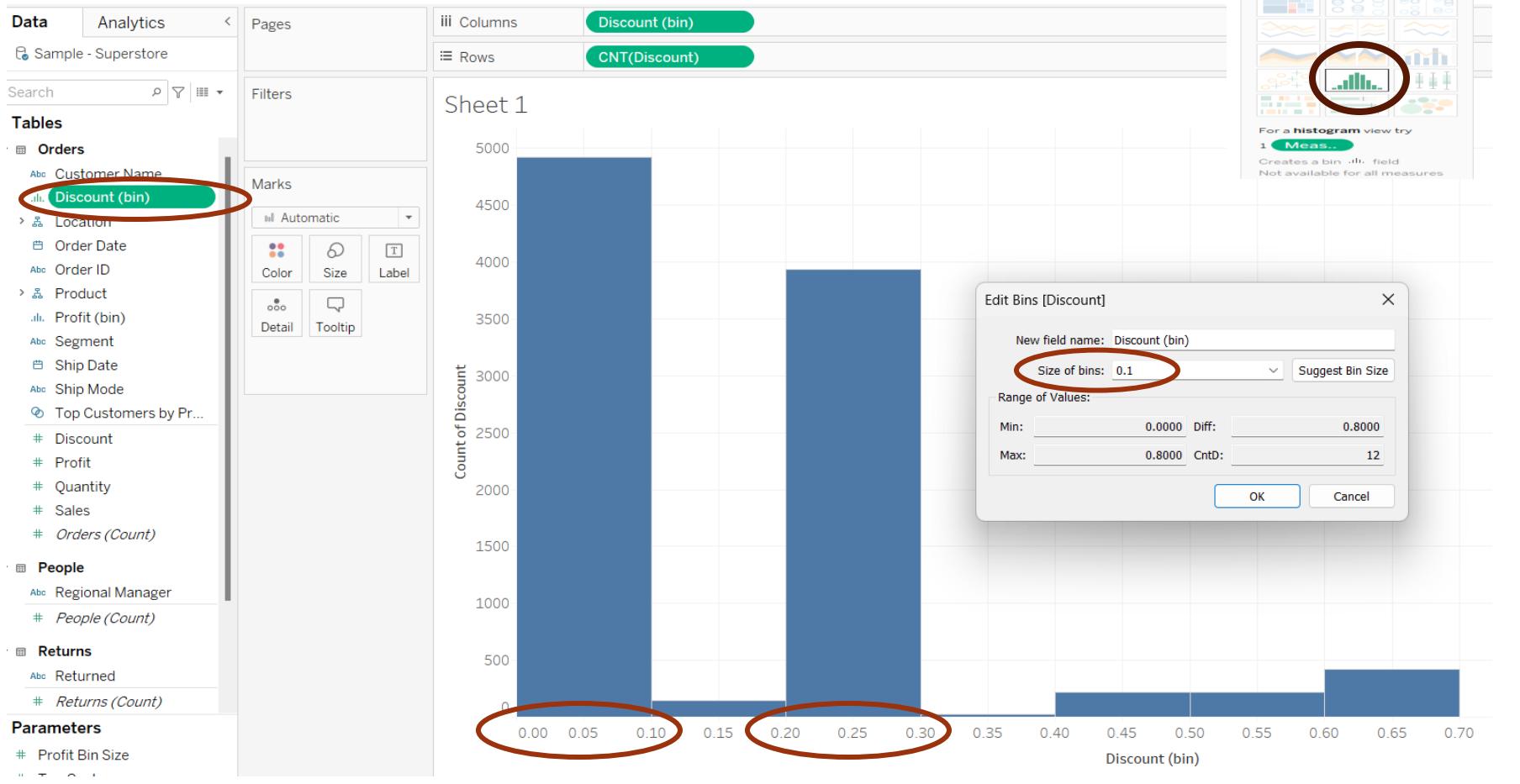
# Line Graphs – other options

- Although we are talking about line graphs => it is easy to **explore other options**: click the drop-down in the Marks card and select **Bar**, then **Circle**, and finally leave it as **Area**.



# Histograms

- Double click **Discount**. The default is a bar chart => use **Show Me** to change to **Histogram**. Right-click the new **Discount(bin)** dimension (at left panel), Edit, and change bins to **0.1** (bin range is 10% interval)
- Most discounts are between 0 and 10%, and in the 20-30% range.



# Create Calculated Field

The screenshot shows the Tableau interface with a calculated field dialog box on the left and a data view on the right.

**Calculated Field Dialog:**

- Name: Margin
- Formula:  $\text{SUM}([\text{Profit}])/\text{SUM}([\text{Sales}])$
- Message: The calculation is valid.
- Buttons: Apply (grayed out), OK (green)

**Data View (Sheet 1):**

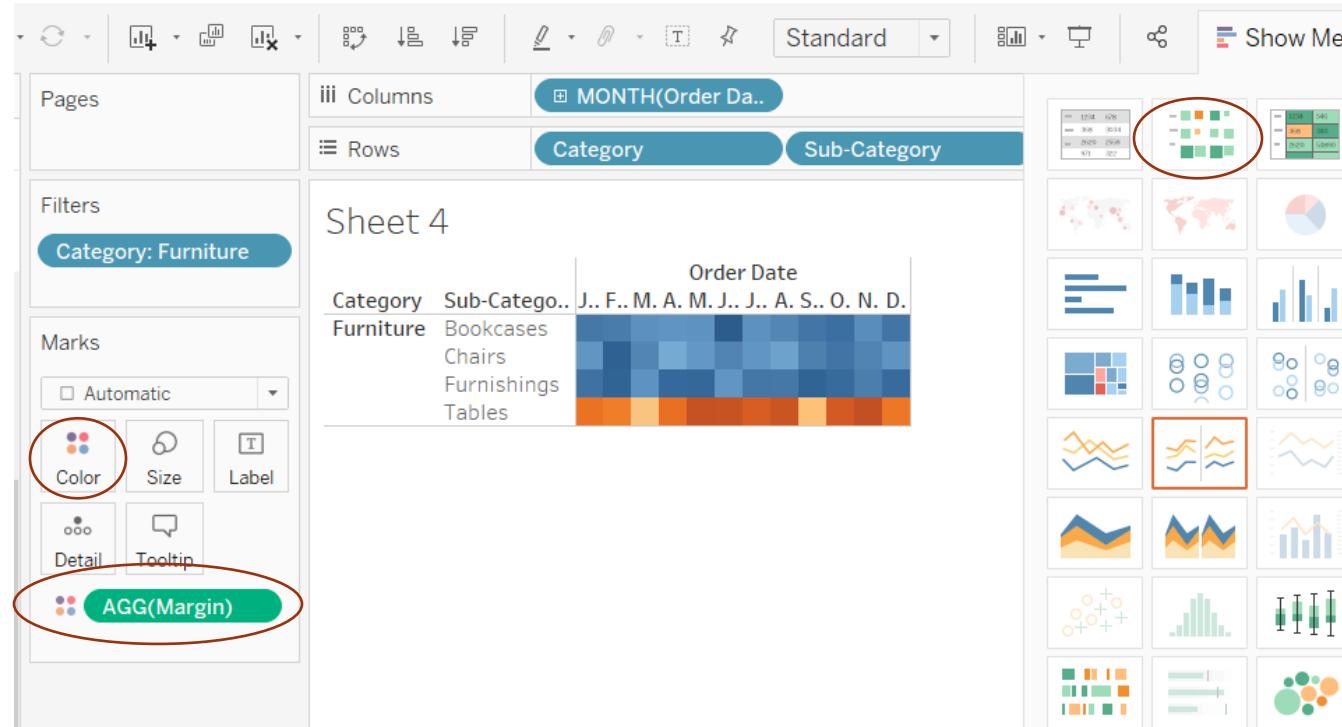
- Top navigation: Pages, Columns (MONTH(Order Da..), Rows (Category, Sub-Category).
- Filters section: Automatic Marks (Color, Size, Text, Detail, Tooltip) with AGG(Margin) selected.
- Table data:

Category	Sub-Cate..	January	February	March	April	May
Furniture	Bookcases	0.1184	0.1131	0.0866	0.0809	0.0841
	Chairs	0.0788	0.1469	0.1010	0.0550	0.0766
	Furnishi..	0.1246	0.1447	0.0833	0.1360	0.1385
	Tables	-0.0761	-0.0662	-0.0119	-0.0790	-0.1188
Office Supplies	Applianc..	0.1280	0.1663	0.1685	0.1740	0.1223
	Art	0.1352	0.1584	0.1164	0.1579	0.1533
	Binders	0.2056	0.1715	0.2833	0.1054	0.1867
	Envelopes	0.1623	0.1883	0.1684	0.1614	0.2015
	Fasteners	0.0827	0.1818	0.1415	0.1365	0.1113
	Labels	0.1994	0.1742	0.2320	0.2331	0.2088
	Paper	0.2230	0.2341	0.2639	0.2472	0.2226
	Storage	0.1201	0.0858	0.1209	0.0641	0.1098
	Supplies	0.0774	0.1207	0.0254	0.0691	0.1196
Technolo..	Accessor..	0.1806	0.1300	0.2250	0.1366	0.1965
	Copiers	0.1294	0.1518	0.1619	0.1667	0.1758
	Machines	0.0861	0.1264	0.0697	0.0796	0.1080
	Phones	0.1291	0.1839	0.1223	0.1342	0.1248

- At times it is challenging to visualize a table of numbers – or the more so for **computed (calculated) values** e.g., **margin** =>  $\text{SUM}([\text{Profit}])/\text{SUM}([\text{Sales}])$ .
- Use heat maps to visualize the **margin made on Furniture items**. Notice that margin is **not in the data** – use Calculated Fields to create the margin field.
- Analysis (Menu)…Create Calculated Field.** Name it **Margin**, formula is  $\text{SUM}([\text{Profit}])/\text{SUM}([\text{Sales}])$

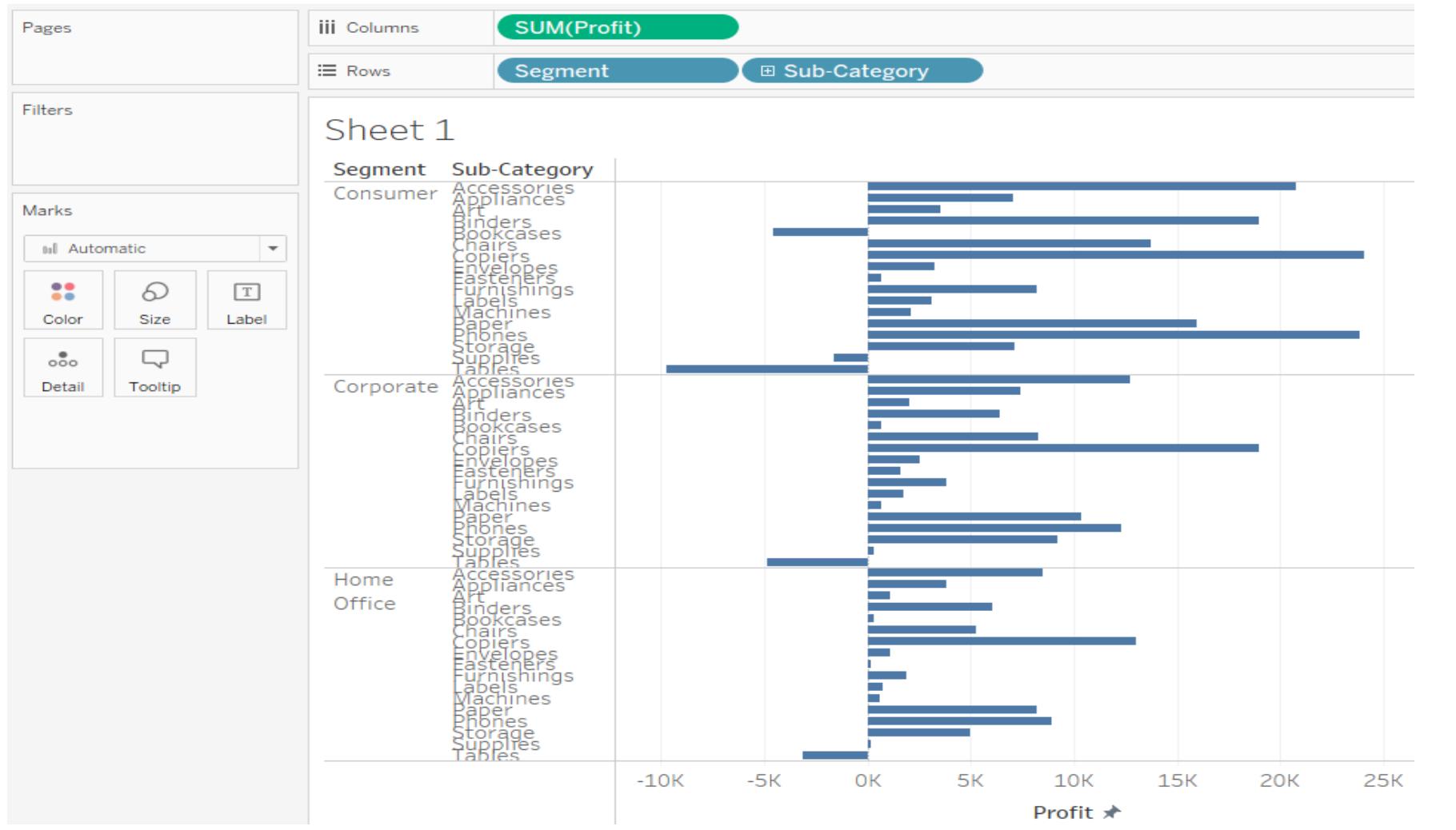
# Heat Maps

- Drag **Category** and **Subcategory** to Row, drag **Order Date** to Columns, and then change to **Month (discrete)**. Drag **Category** to Filter shelf, and select only **Furniture**
- Drag **Margin** field to the Color Marks and select **Heat Map** from Show Me. Change color to **Orange-Blue Diverging**. We see some insights – e.g. maybe we are selling tables as a loss leader and making it up on chairs?



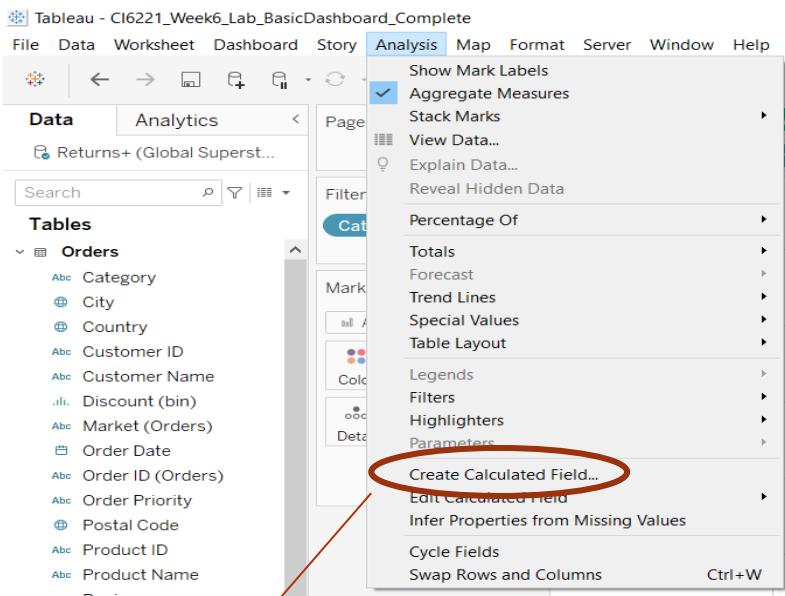
# Calculated Fields – Sign of Profit

Create a new Worksheet – **Profit** to **Columns**, **Segment** and **Subcategory** to **Rows**.



# Calculations in Tableau – Distinct Color (Pos/Neg)

- What if we want, for **emphasis**, to have one **distinct color** for **positive** and one for **negative**? We **define a formula** by creating **Calculated Fields**.
- Click **Analysis... Create Calculated Field**. Name it "**Sign of profit**". 利润的标志
- IF  $\text{SUM}([\text{Profit}]) > 0$  THEN "positive" ELSE "negative" END



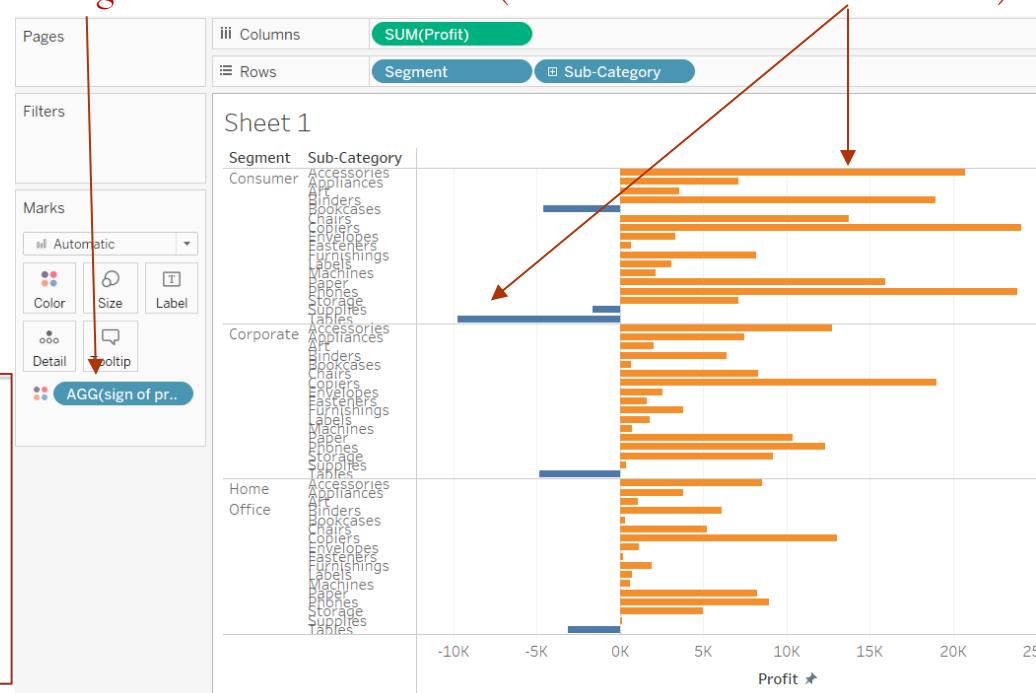
```
sign of profit

//Comments - implement sign of profit
IF SUM([Profit])>0 THEN "positive" ELSE "negative" END
```

Abc *Measure Names*  
=# Margin  
=Abc sign of profit

Little **equal sign** in front of field indicates it's a **calculated field**

Bring calculated field to color (click on color card to edit color)



# Calculations in Tableau

Tableau - Global Superstore 10-3

File Data Worksheet Dashboard Story Analysis Map Format Server Window Help

Show Me

Data Analytics Pages SUM(Profit)

Global Superstore

Dimensions

- Customers
  - Customer ID
  - Customer Name
  - Segment
- Orders
  - Order Date
  - Order ID
  - Order Priority
- Shipping
  - Ship Date
  - Ship Mode
- Location
  - Market
  - Region
  - Country
  - State
  - City
  - Postal Code
- Products
  - Category
  - Sub-Category
  - Product Name

Measures

- Discount
- Profit
- Quantity
- Sales
- Shipping Cost
- Latitude (generated)
- Longitude (generated)

Filters

Category: Furniture

Marks

Automatic

Color

Size

Label

Detail

Tooltip

Market Sub-Category

Market	Sub-Category	Profit
Africa	Bookcases	\$5,000
	Chairs	\$2,000
	Furnishings	\$1,000
	Tables	\$3,000
APAC	Bookcases	\$65,000
	Chairs	\$15,000
	Furnishings	\$12,000
	Tables	\$2,000
Canada	Bookcases	\$2,000
	Chairs	\$1,000
	Furnishings	\$500
	Tables	\$1,000
EMEA	Bookcases	\$7,000
	Chairs	\$2,000
	Furnishings	\$1,000
	Tables	\$2,000
EU	Bookcases	\$55,000
	Chairs	\$20,000
	Furnishings	\$12,000
	Tables	\$2,000
LATAM	Bookcases	\$25,000
	Chairs	\$30,000
	Furnishings	\$2,000
	Tables	\$2,000
US	Bookcases	\$10,000
	Chairs	\$25,000
	Furnishings	\$12,000
	Tables	\$2,000

(\$20,000) (\$10,000) \$0 \$10,000 \$20,000 \$30,000 \$40,000 \$50,000 \$60,000 \$70,000 Profit

28 RECORDED WITH 1 column SUM(Profit): \$285,205

SCREENCASTOMATIC Type here to search

2:42 PM 7/28/2017

# In-class Exercise

- The [Kaggle](#) dataset contains 887 records of the real Titanic passengers.
- Analyze how **Gender** and **Ticket Class** affected the **survival ratio** of passengers. Which gender and class have the highest survival?

Ticket Class      Gender and Pclass => independent variables, Survived => dependent variable



Survived	Pclass	Name	Sex	Age	Siblings/Spouses Aboard	Parents/Children Aboard	Fare
0	3	Mr. Owen	male	22		1	0 7.25
1	1	Mrs. John	female	38		1	0 71.2833
1	3	Miss. Lain	female	26		0	0 7.925
1	1	Mrs. Jacq	female	35		1	0 53.1
0	3	Mr. Willia	male	35		0	0 8.05
0	3	Mr. James	male	27		0	0 8.4583
0	1	Mr. Timot	male	54		0	0 51.8625
0	3	Master. G	male	2		3	1 21.075
1	3	Mrs. Osca	female	27		0	2 11.1333
1	2	Mrs. Nicho	female	14		1	0 30.0708
1	3	Miss. Mar	female	4		1	1 16.7
1	1	Miss. Eliza	female	58		0	0 26.55
0	3	Mr. Willia	male	20		0	0 8.05

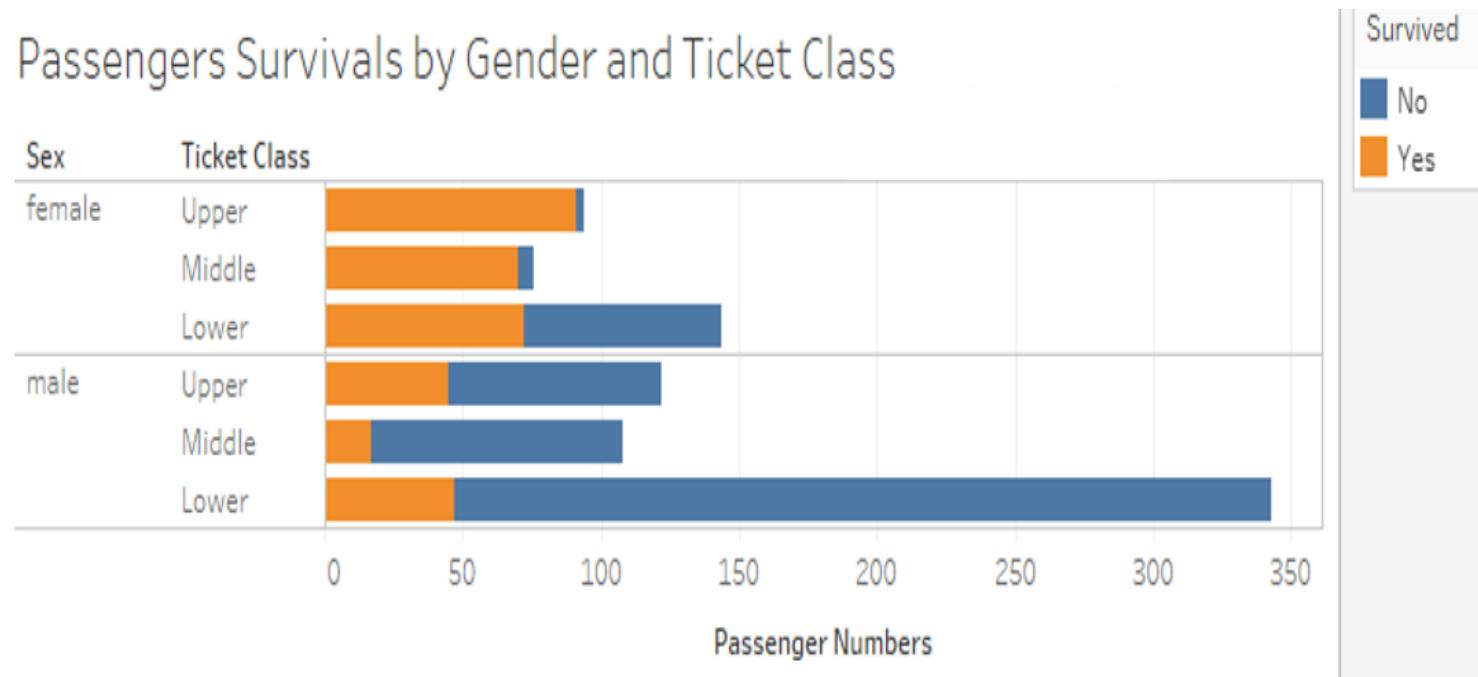
# In-class Exercise

How **Gender** and **Ticket Class** affected the **survival ratio** of passengers?

Drag “**Sex**” and “**Ticket Class**” to **Rows**

Change “**Survived**” from dimension to **count** => drag to **Columns**

Drag Dimension “**Survived**” to **Color Mark**



Which gender and class have the highest survival?

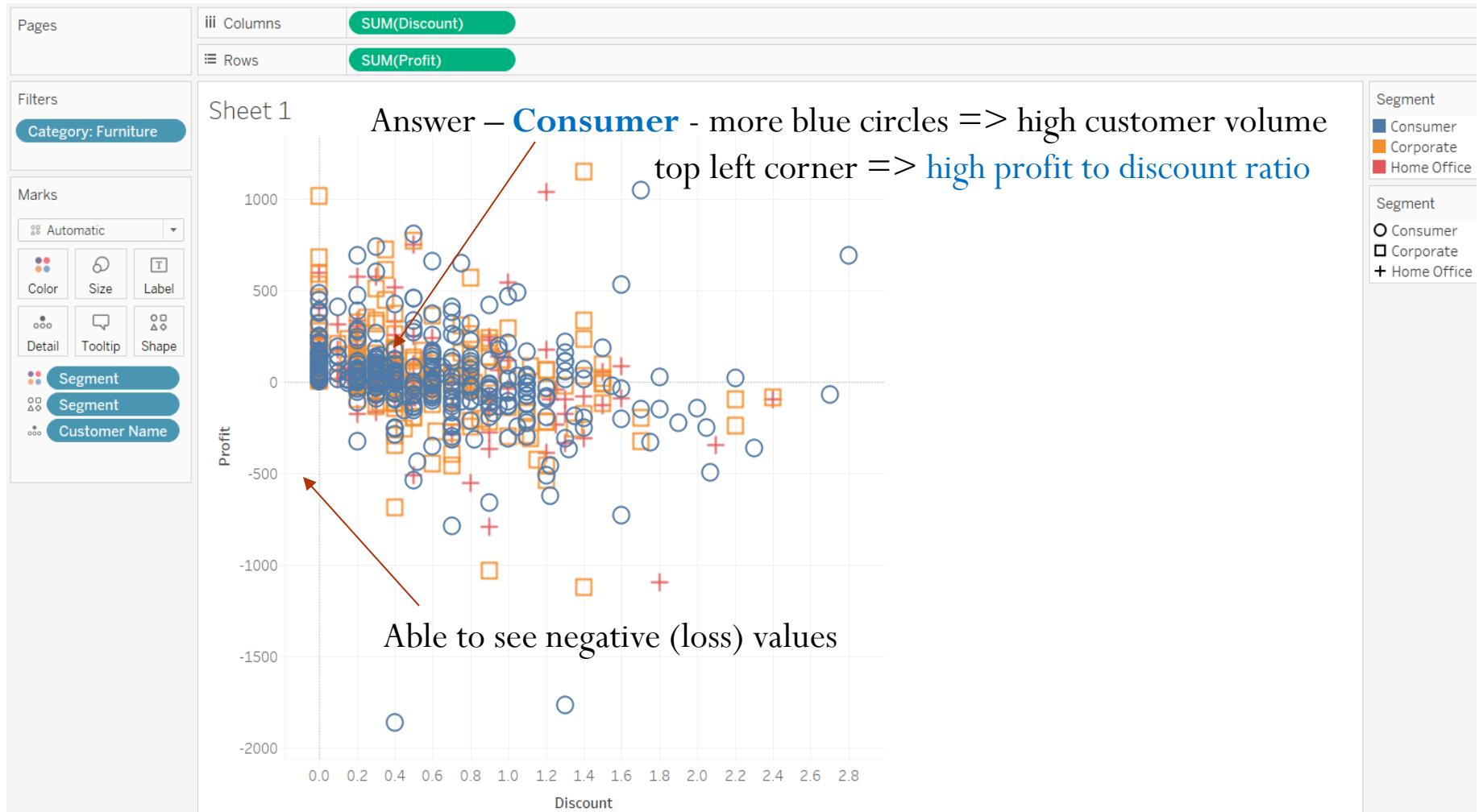
Survival information shown through **colored stacked bar** of each category – part-whole ratio details.

# In-class Exercise

- From the Sample Superstore dataset, analyze for the furniture **category** which **segment** (**consumer**, **corporate**, or **home office**) has the highest **customer volume** with highest **profit to discount** ratio (**note this is not a part-to-whole relationship**).
- Hint:
- Fields involved – **category**, **segment**, **customer volume**, **profit**, and **discount**.
- Filter – furniture category
- Main relationship  $\Rightarrow$  profit to discount
- Customer volume analysis – must drill down to customer ID level

# In-class Exercise – Solution 1

From the Sample Superstore dataset, analyze for the furniture **category** which **segment** (consumer, corporate, or home office) has the highest **customer volume** with highest positive **profit to discount ratio**.



# In-class Exercise – Solution 1

Analyze for the furniture **category** which **segment** (consumer, corporate, or home office) has the highest **customer volume** with highest **profit to discount ratio**.

- 1) Create calculated field “**Discount NULL case**” to cater for **zero discount** cases.
- 2) Create calculated field “**Profit Discount Ratio**”

Discount NULL case

```
IF [Discount] = 0 THEN 1  
ELSE [Discount]  
END
```

Profit Discount Ratio

```
[Profit]/[Discount NULL case]
```

