

Practical 1

Intro to Data Visualisation + Data Integration + Tableau





Practical 1

<u>Before Class</u>	Concepts	Intro to Data Visualisation
<u>During Class</u>	Hands-on	Intro to Data Visualisation + Integration + Tableau Cardinality - Join - Relationship
<u>After Class</u>	Hands-on	<i>Revise today's class with LMS: Apr/Oct – Week X (...)</i> <i>Do “What is IDL” on LMS</i> <i>Go through Additional Resources slides</i> - <i>Read Extra Resources on Exploratory vs Explanatory</i> - <i>Read website on cardinality Lab Q2</i> - <i>Do “Harry Potter – Which Join should I use?” quiz</i>

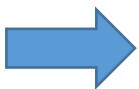
What is Data Visualisation Today?

*Data Visualisation is the act of taking **information (data)** and placing it into a **visual** context, such as a map or graph.*

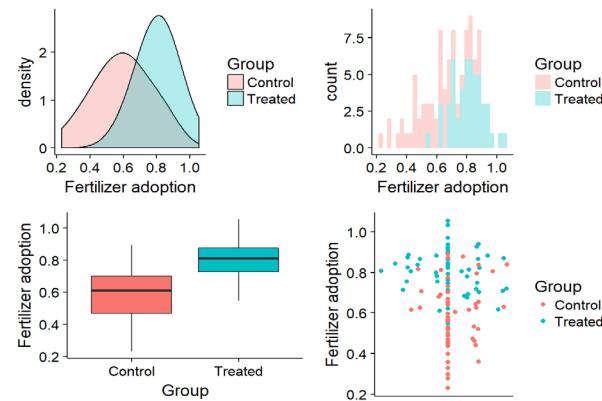
Deluge of Data

A1	B	C	D	E	F	G	H	I	J	K	L	M	
Car ID	Date Purchased	Location	Manufacturer	Date	Make	Model	Variant	Transmission	Colour	Economy	Has the car been leased?	Has the car been MPV checked?	Is the car a special edition?
1	26522	UK	VW	Golf	Petrol	Auto	Sleek	60mpg	No	Yes	No	No	
2	2633	Germany	VW	Golf	Petrol	Manual	Red	60mpg	Yes	No	No	No	
3	2624	Spain	Seat	Leon	Diesel	Auto	Green	30mpg	Yes	No	No	No	
4	2651	Italy		500	Petrol	Auto	White	30mpg	Yes	No	No	No	
5	2658	France	Renault	Clio	Petrol	Auto	Blue	40mpg	Yes	No	No	No	
6	2637	UK	Vauxhall	Corsa	Petrol	Auto	White	30mpg	Yes	No	No	No	
7	2638	UK	Vauxhall	Ampera	Petrol	Auto	Black	60mpg	Yes	No	No	No	
8	2639	UK	Vauxhall	Corsa	Diesel	Auto	Red	40mpg	Yes	Yes	No	No	
9	2640	Germany	Audi	A1	Petrol	Manual	Red	40mpg	Yes	No	No	No	
10	2641	Germany	VW	Golf	Petrol	Manual	Red	30mpg	Yes	No	No	No	
11	2651	Germany	Audi	A1	Petrol	Manual	White	30mpg	Yes	No	No	No	
12	2657	Germany	Audi	A1	Petrol	Manual	White	40mpg	Yes	Yes	No	No	
13	2623	Germany	Audi	A3	Petrol	Manual	White	40mpg	No	Yes	No	No	
14	2624	Germany	Audi	A3	Petrol	Manual	White	30mpg	No	No	No	No	
15	2625	Germany	Audi	A3	Petrol	Manual	Green	30mpg	No	No	No	No	
16	2626	Germany	VW	A7	Petrol	Manual	Blue	40mpg	No	Yes	No	No	
17	2627	Germany	VW	GTI	Diesel	Manual	Black	30mpg	No	No	No	No	
18	2628	Germany	VW	GTI	Diesel	Manual	White	30mpg	No	No	No	No	
19	2629	Germany	VW	Passat	Petrol	Manual	White	30mpg	No	No	No	No	
20	2630	Germany	Audi	A5	Petrol	Auto	White	40mpg	No	No	No	No	
21	2631	UK	Audi	A6	Diesel	Auto	White	30mpg	Yes	No	No	No	
22	2632	UK	VW	Passat	Petrol	Auto	Red	60mpg	Yes	No	No	No	
23	2633	UK	Vauxhall	Vectra	Petrol	Manual	Blue	40mpg	Yes	No	No	No	
24	2644	UK	Vauxhall	Astra	Diesel	Manual	Black	30mpg	Yes	No	No	No	
25	2645	UK	Vauxhall	Astra	Diesel	Manual	Black	30mpg	Yes	No	No	No	

Human Brain
+ Technology



Interactive +
Communicate Better



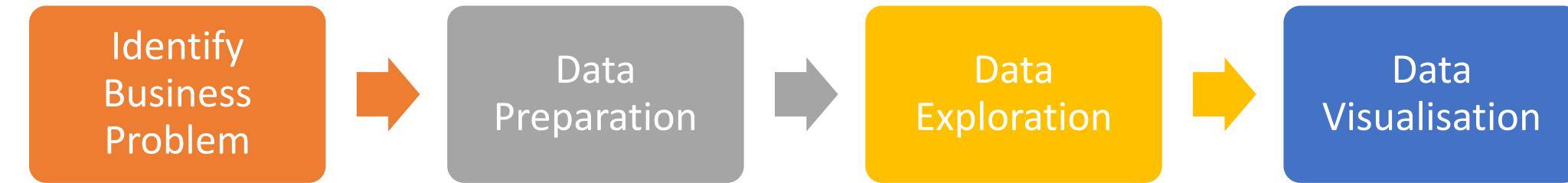
Ultimate Objective

To help managers gain improved **insight** about their business operations and make better, fact-based **decisions**.

Our Approach to Data Visualisation



How To Do Data Visualisation?



Brainstorming



Integration
Profiling
Cleaning



Making Sense
Discovering Insights

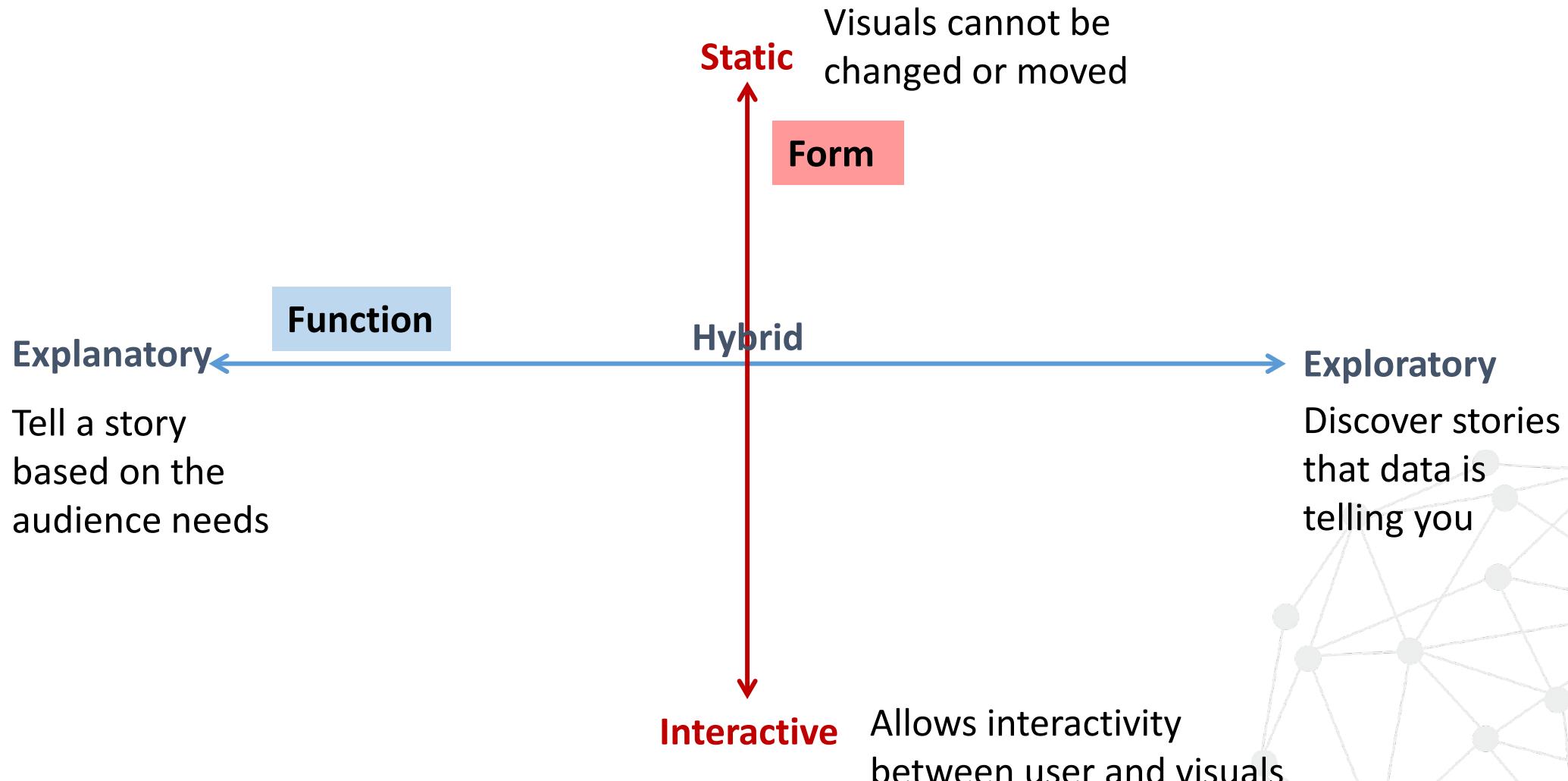


Communicating

Exploratory vs Explanatory Visualisation Static vs Interactive Visualisation



4 Broad Kinds of Data Visualisations





Types of Data Visualisation Infographics

Explanatory : Infographic is an explanatory visualization as it tells the stories that the analyst wants to explain.

Interactivity : None - **static** in nature.

Message : Audience extract what they wish get out of it.

DID YOU KNOW?

8 out of 100 teens report having serious depression.

That's 2 out of every 25 teens. 

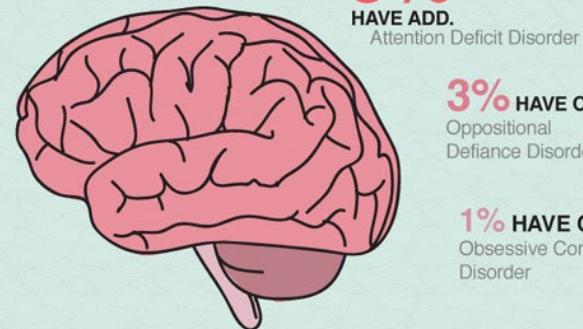


8% of all teens have an anxiety disorder.

5% HAVE ADD.
Attention Deficit Disorder

3% HAVE ODD
Oppositional Defiance Disorder

1% HAVE OCD
Obsessive Compulsive Disorder



only 38% with mood disorders receive help.

only 15% with substance abuse problems get the help they need.

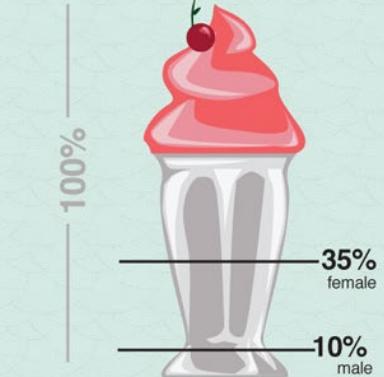
only 13% of Eating Disorder sufferers get help.

OVER **20%** OF YOUNG ADULTS HAVE A MENTAL ILLNESS

35% of teenaged girls have an eating disorder.

That's 7 out of every 25 teen girls.

10% of all teens suffering from an eating disorder are **male**.



4000

young Canadian teens commit suicide every year.

1 in 5
will get the help they need.



Types of Data Visualisation Dashboard

Dashboard is **Hybrid visualization** that falls somewhere **in-between** explanatory and exploratory visualisation.



It is an **explanatory** visual as it tells the stories that the analyst wants to explain.

Interactivity : Yet it **allows** the audience to **explore** and have some **interactivity** with it.

Message : Audience extract what they wish to get out of it.

Types of Data Visualisation

Data Story – Live vs Video Presentation

Data story is **explanatory** in nature.



Live Presentation

Interactivity : **High**

Message : Analyst has **control** over the message that they want the audience to get.



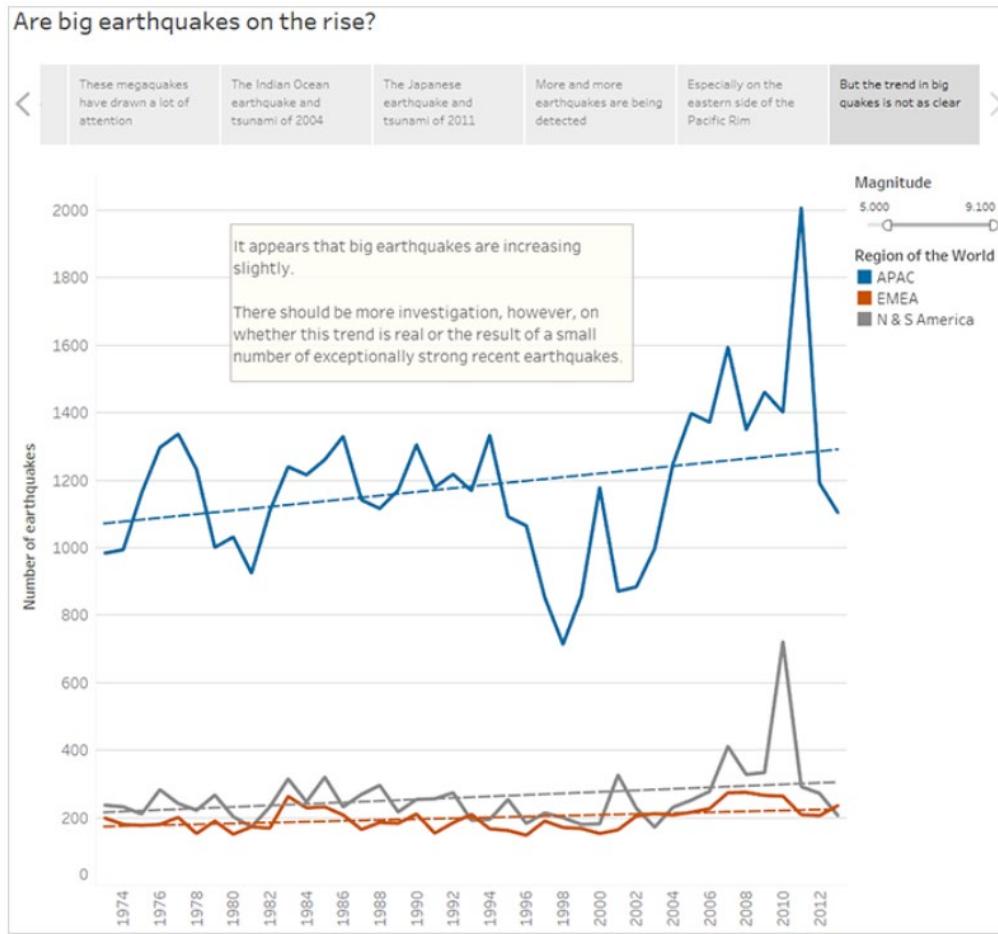
Video Presentation

Interactivity : Some
(it depends)

Message : Analyst has control over the message that they want the audience to get.

<https://youtu.be/jbkSRLYSOjo>

Data Story – Digital Data Storytelling



Digital Data Story is an **explanatory** visual as it tells the stories that its designer wants to explain.

Interactivity : Yet it **allows** the audience to have some **interactivity** with it.

Message : The designer has more **control** over the message that they want the audience to get



What Does It Mean To Us?

- New and more meaningful way to look at data;
- Identify and react to insights faster;
- Important to first ask yourself who makes up your audience and what do they need.





P01a Intro to DV + DI: Explanatory vs Exploratory / Static vs Interactive

Q1. Go to this website <https://informationisbeautiful.net/> and explore the visuals.

1. World's Biggest Data Breaches & Hacks

<https://informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/>

2. Which is The Best Performing Marvel Movie?

<https://informationisbeautiful.net/visualizations/which-is-the-best-performing-marvel-movie/>

3. Most Common Passwords

<https://informationisbeautiful.net/visualizations/top-500-passwords-visualized/>

4. Mountain Out of Molehills – A timeline of media inflamed fears

<https://informationisbeautiful.net/visualizations/mountains-out-of-molehills/>

5. Star Wars Last Jedi – Rotten Tomato – Audience vs Critique

<https://informationisbeautiful.net/visualizations/star-wars-last-jedi-one-of-the-biggest-rotten-tomatoes-audience-vs-critics-score-splits-ever/>

6. Countries with better human development show less corruption and more political freedom

<https://d33wubrfki0l68.cloudfront.net/13ecfaaf-d8c9-4284-ad13-31e5ede4153b/world-viz-raphael-halloran.png>

a) Tell us what the visualisation is about.

b) Decide if the visual is :

- Explanatory – Static
- Exploratory – Static
- Explanatory – Interactive
- Exploratory – Interactive





Intro to Data Integration and Transformation



Intro

Data often comes in fragmented pieces.

Before we begin any Data Visualisation, we have some ground work to do.

We need to bring all our data tables together and transform them into the desired shape.

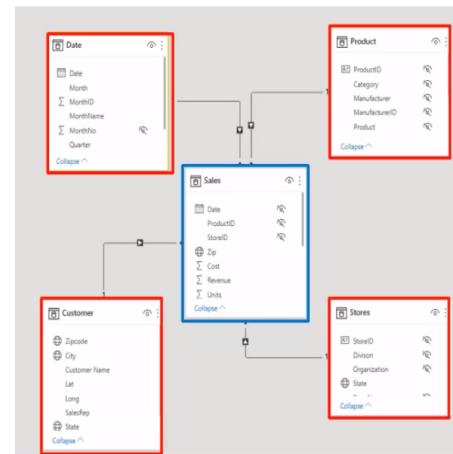


Data Integration

There are two ways of doing data integration :

(1) Creating relationships between data sources to get one unified view (a.k.a data model).

- The data tables are linked by a common field.
- The data tables are not physically combined (i.e. they remain separate).



Creating Relationship





Data Integration

There are two ways of doing data integration :

(2) Combining data by merging rows or columns of data to get one single, combined table.

- Similar terms used for combining rows : "append" or "concatenate" or "union". (Tableau uses "union".)
- Similar terms used for combining columns : "merge" or "join". (Tableau uses "join").

ID	First Name	Last Name	Food
A00001	Mickey	Mouse	Biryani
A00016	Donald	Duck	Durian

ID	First Name	Last Name	Food
B00072	Buzz	Lightyear	Satay
B00165	Sheriff	Woody	Rojak
C00092	Potato	Head	Laksa

ID	First Name	Last Name	Food
A00001	Mickey	Mouse	Biryani
A00016	Donald	Duck	Durian
B00072	Buzz	Lightyear	Satay
B00165	Sheriff	Woody	Rojak
C00092	Potato	Head	Laksa

Combining Rows

ID	First Name	Last Name
B00072	Buzz	Lightyear
B00165	Sheriff	Woody
C00092	Potato	Head

ID	Food
B00072	Satay
B00165	Rojak
C00092	Laksa

ID	First Name	Last Name	Food
B00072	Buzz	Lightyear	Satay
B00165	Sheriff	Woody	Rojak
C00092	Potato	Head	Laksa

Combining Columns



Data Integration

There are two ways of doing data integration :

- (1) Creating relationships between data sources to get one unified view (a.k.a data model).
- (2) Combining data by merging rows or columns of data to get one single, combined table.

To decide between **Join** or **Relationship** when merging columns, we need to first understand the term **Cardinality** in database language, which we will go into more details in the next section.





Data Transformation

Once the data has been integrated, there may also be a need to transform to a shape that is conducive for the software to process efficiently.

This is usually done by **Unpivoting**.

We will look into this in more details next week.



Data Integration : Cardinality



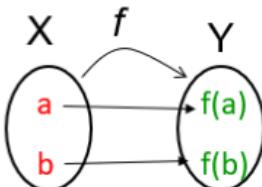
Heard of one-to-one function before?

Topic of Functions in LOMA

When is a Function Considered One-to-One?

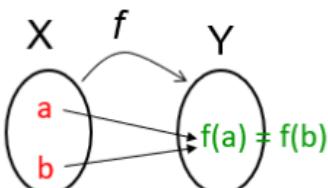
f is one-to-one

- if every input has a unique output; or
- if every element in the domain has a distinct image.



f is not one-to-one

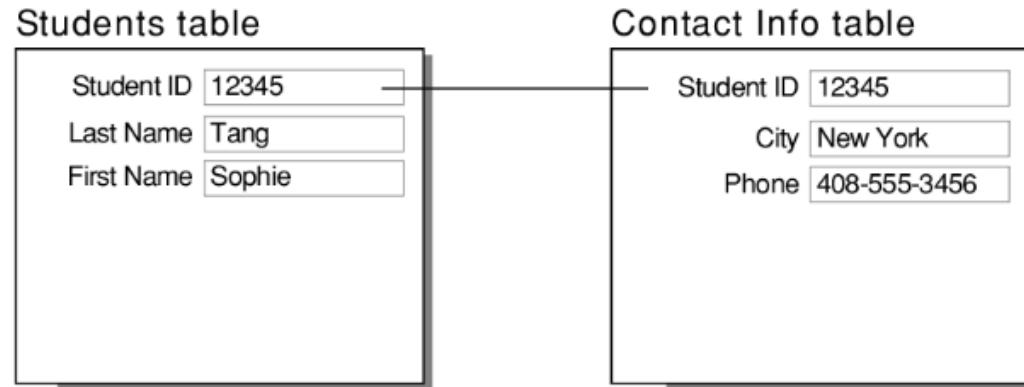
- if some elements in the domain share the same image.



- In database language, **cardinality** is the word used to describe such relationships between two data tables.
- Do the tables have one-to-one cardinality, or one-to-many cardinality, or do they have many-to-many cardinality?



What cardinality is this?



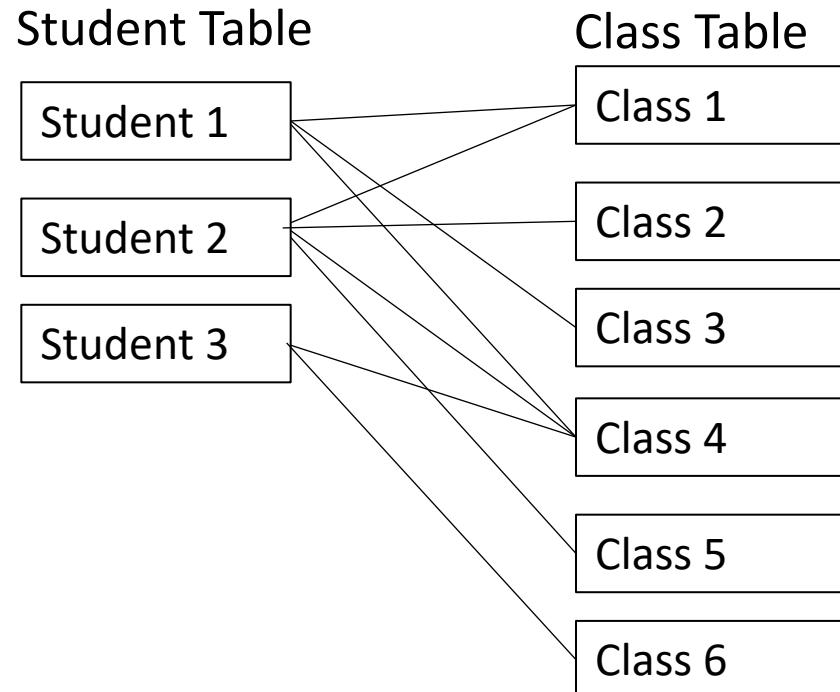
- a. one-to-one
- b. one-to-many
- c. many-to-many



Multiple Choice



What cardinality is this?



- a. one-to-one
- b. one-to-many
- c. many-to-many

 Multiple Choice

https://fmhelp.filemaker.com/help/18/fmp/en/index.html#page/FMP_Help%2Fone-to-one-relationships.html%23



What cardinality is this?

Customers table

Customer ID	12345
Name	Tang

Orders table

Order ID	B204
Customer ID	12345
Order ID	B391
Customer ID	12345
Order ID	B448
Customer ID	12345

- a. one-to-one
- b. one-to-many
- c. many-to-many



Multiple Choice



Data Integration using Join



Are the data tables related? What's the cardinality?

sales_fact_1997

product_id	time_id	customer_id	promotion_id	store_id	store_sales	store_cost	unit_sales
337	371	6280	0	2	1.5	0.51	2
1512	371	6280	0	2	1.62	0.6318	3
963	371	4018	0	2	2.4	0.72	1

sales_fact_1998

product_id	time_id	customer_id	promotion_id	store_id	store_sales	store_cost	unit_sales
173	748	2094	54	1	4.29	1.8447	3
1119	748	2094	54	1	9.51	3.5187	3
1242	748	2094	54	1	7.92	2.8512	4

store

store_id	store_type	region_id	store_name	store_street_address	store_city	store_state	store_country	store_manager	grocery_sqft
0	HeadQuarters	0	HQ	1 Alameda Way	Alameda	CA	USA	NULL	NULL
1	Supermarket	28	Store 1	2853 Bailey Rd	Acapulco	Guerrero	Mexico	Jones	17475
2	Small Grocery	78	Store 2	5203 Catanzaro Way	Bellingham	WA	USA	Smith	22271
3	Supermarket	76	Store 3	1501 Ramsey Circle	Bremerton	WA	USA	Davis	24390

time_by_day_1997_1998

time_id	the_date	the_day
367	7/1/1997	3
368	8/1/1997	4
369	9/1/1997	5



Short Answer



Combining Data

Sometimes, there is a need to **physically combine** two or more data tables by merging rows or columns of data to get **one single table**.

Similar terms used for **combining rows** : "**append**" or "**concatenate**" or "**union**". (Tableau uses "**union**").

Similar terms used for **combining columns** : "**merge**" or "**join**".
(Tableau uses "**join**").





Combining Rows

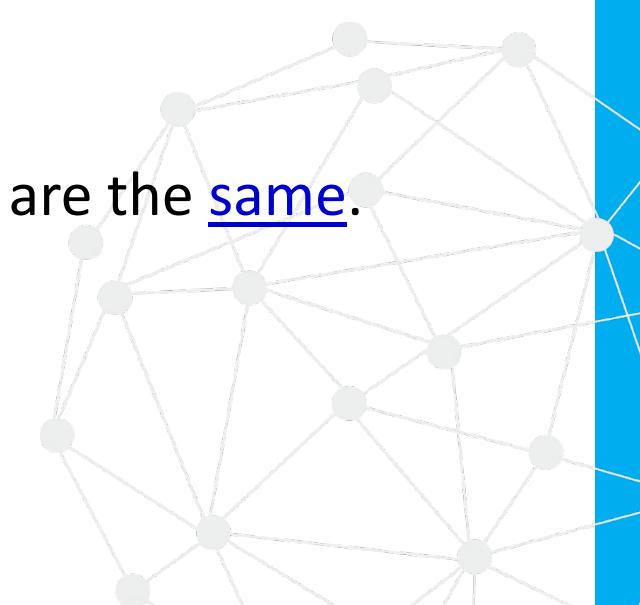
To add new rows (from a second set of data) to the first set of data.

ID	First Name	Last Name	Food
A00001	Mickey	Mouse	Biryani
A00016	Donald	Duck	Durian

ID	First Name	Last Name	Food
B00072	Buzz	Lightyear	Satay
B00165	Sheriff	Woody	Rojak
C00092	Potato	Head	Laksa



ID	First Name	Last Name	Food
A00001	Mickey	Mouse	Biryani
A00016	Donald	Duck	Durian
B00072	Buzz	Lightyear	Satay
B00165	Sheriff	Woody	Rojak
C00092	Potato	Head	Laksa



This can be done if the column names of the two data tables are the same.



Combining Columns

To add any additional field from the second table to the first table.

ID	First Name	Last Name
B00072	Buzz	Lightyear
B00165	Sheriff	Woody
C00092	Potato	Head

ID	Food
B00072	Satay
B00165	Rojak
C00092	Laksa

ID	First Name	Last Name	Food
B00072	Buzz	Lightyear	Satay
B00165	Sheriff	Woody	Rojak
C00092	Potato	Head	Laksa

The common field is “ID”.



It is thus important to identify the **common field**.

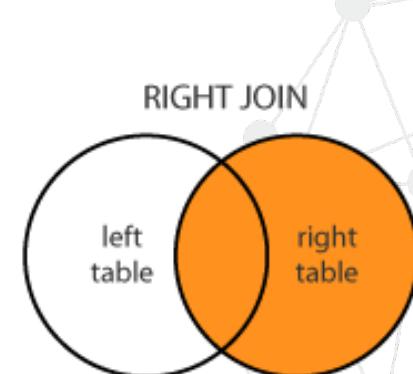
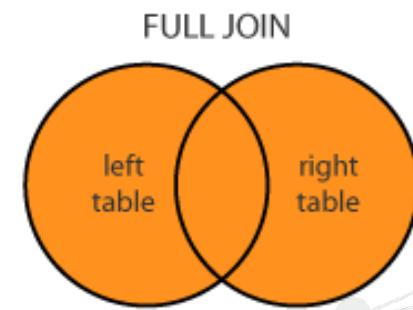
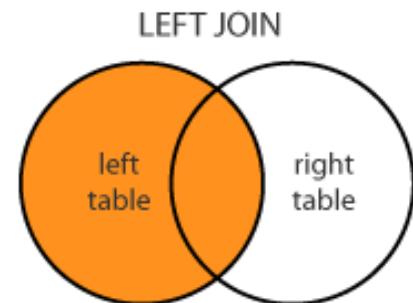
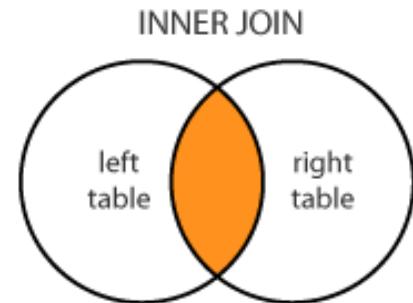


Types of Joins

When joining columns, there are a few different ways we can choose to determine which rows make it into the final table.

The four common types of joins are :

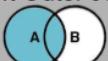
1. Inner Join
2. Right Outer Join
3. Left Outer Join
4. Full Outer Join



Types of Join : Inner Join

Join data records with identical values in the join fields.

Left Outer Join



Right Outer Join



Inner Join



Full Outer Join



Left Anti Join



Right Anti Join



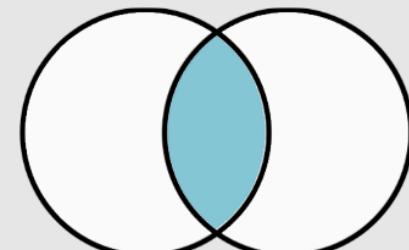
Full Anti Join



Table A

Customer ID	Customer Name	IsJoined
1	Brad Pitt	
2	Tom Hanks	✓
3	Julia Roberts	
4	Angelina Jolie	✓
4	Jen Aniston	✓

Only matching rows from both tables



Inner Join

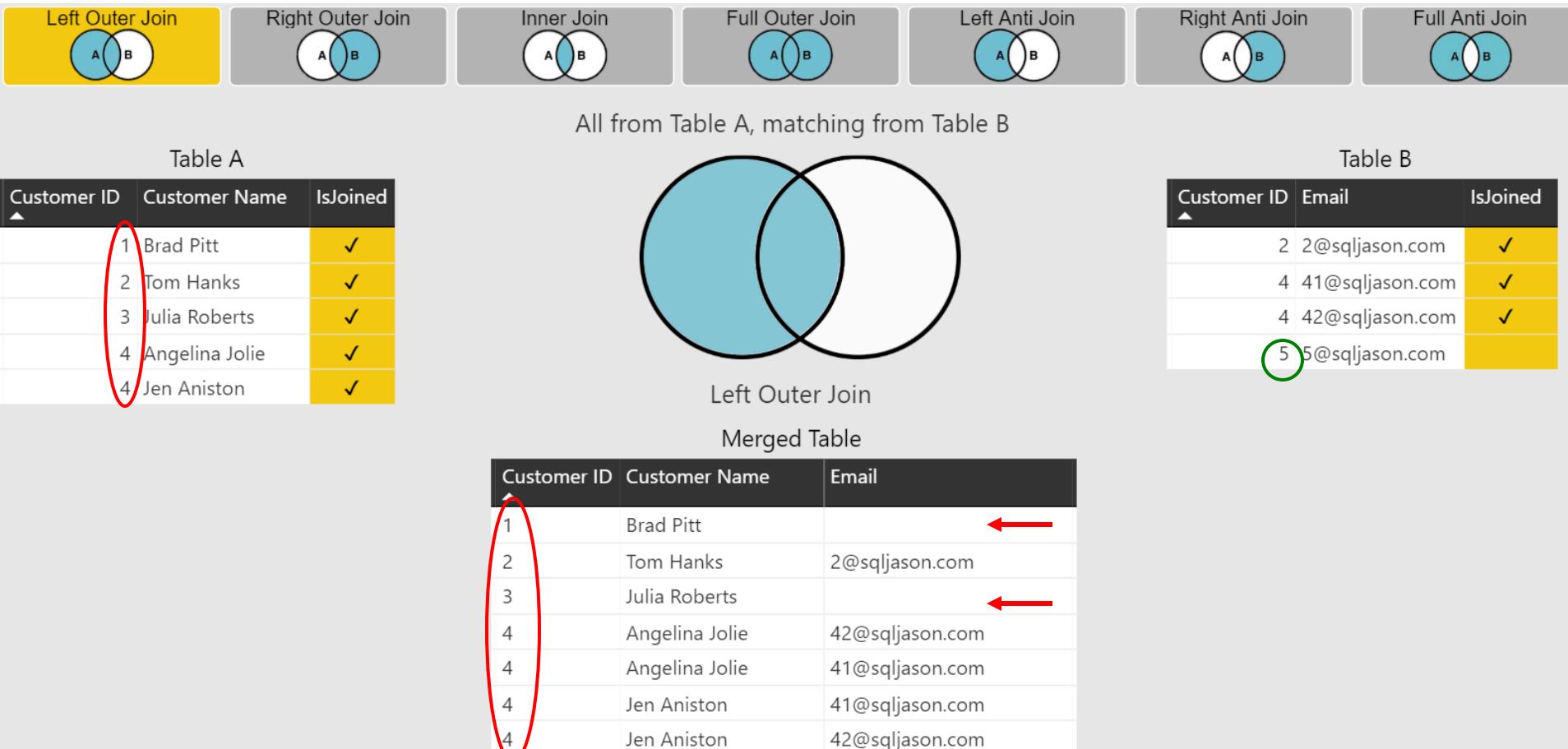
Merged Table

Customer ID	Customer Name	Email
2	Tom Hanks	2@sqljason.com
4	Angelina Jolie	42@sqljason.com
4	Angelina Jolie	41@sqljason.com
4	Jen Aniston	41@sqljason.com
4	Jen Aniston	42@sqljason.com

<https://community.powerbi.com/t5/Data-Stories-Gallery/Visualizing-Merge-Join-Types-in-Power-BI/td-p/219906>

Types of Join: Left Outer Join

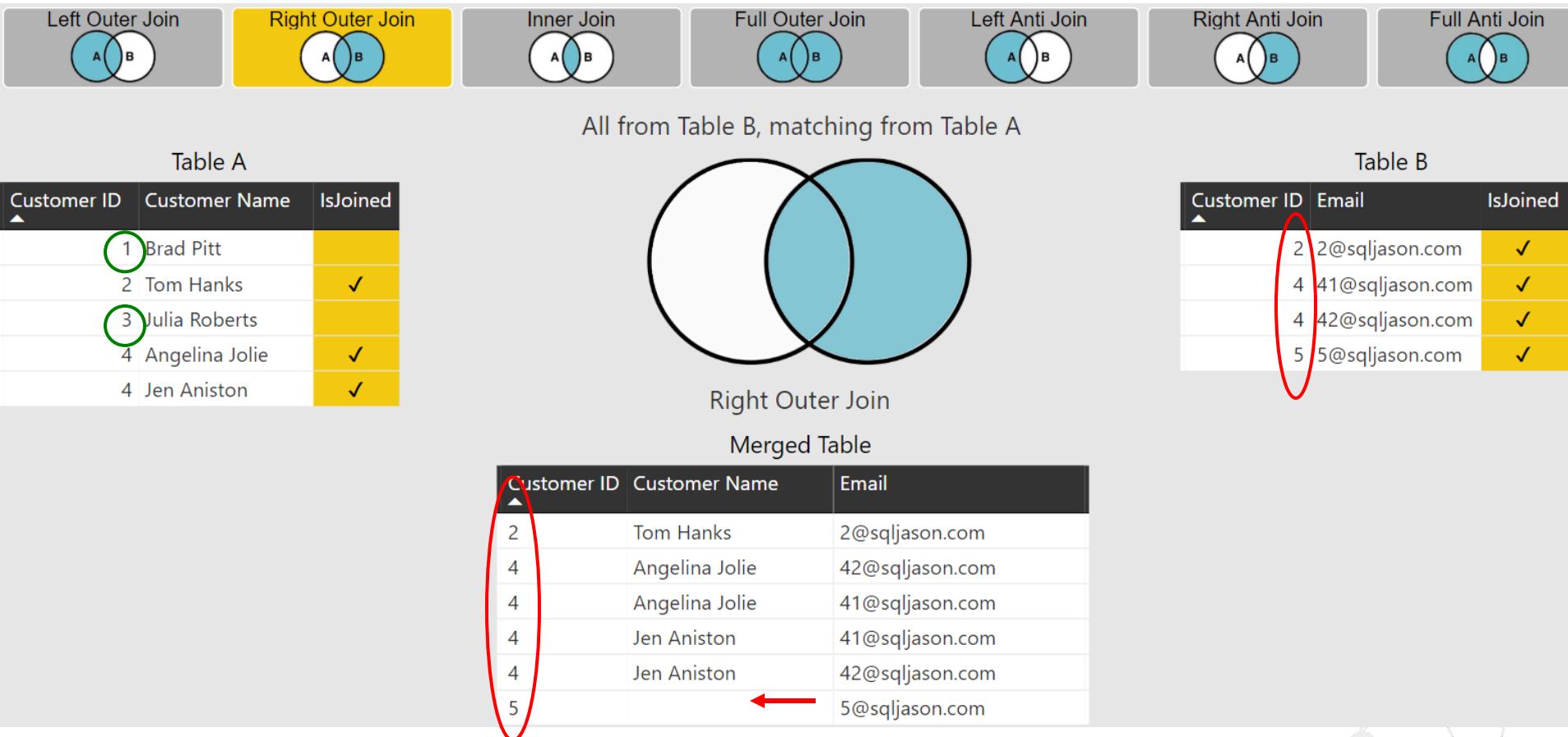
All data records in the left table occur in the result, even if no corresponding data record is found in the right table.



<https://community.powerbi.com/t5/Data-Stories-Gallery/Visualizing-Merge-Join-Types-in-Power-BI/td-p/219906>

Types of Join : Right Outer Join

All data records in the right table occur in the result, even if no corresponding data record is found in the left table.



Types of Join : Full Outer Join

This join does more than the left outer join; all data records on the second table occur in the result, even if there is no corresponding data record on the first table.

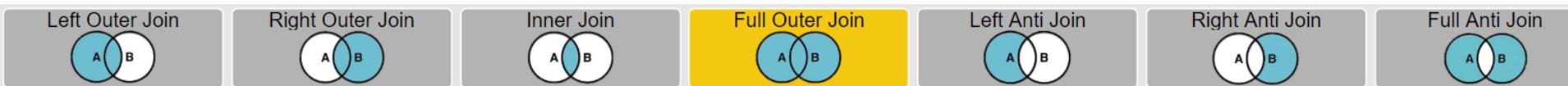
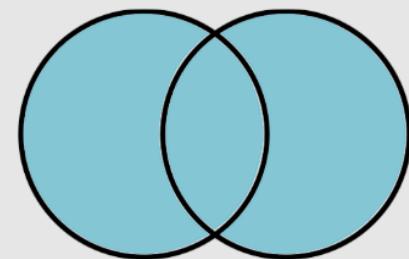


Table A

Customer ID	Customer Name	IsJoined
1	Brad Pitt	✓
2	Tom Hanks	✓
3	Julia Roberts	✓
4	Angelina Jolie	✓
4	Jen Aniston	✓

All rows from both tables



Full Outer Join

Table B

Customer ID	Email	IsJoined
2	2@sqljason.com	✓
4	41@sqljason.com	✓
4	42@sqljason.com	✓
5	5@sqljason.com	✓

Merged Table

Customer ID	Customer Name	Email
1	Brad Pitt	
2	Tom Hanks	2@sqljason.com
3	Julia Roberts	
4	Angelina Jolie	42@sqljason.com
4	Angelina Jolie	41@sqljason.com
4	Jen Aniston	41@sqljason.com
4	Jen Aniston	42@sqljason.com
5		5@sqljason.com



Summary

Data Integration using Join

- physically join the data tables together to create **one single, combined table.**

In Tableau, you can use ***Union*** to combine rows together.

In Tableau, you can use ***Join*** to combine columns together.

There are 4 common types of Join:

- (1) inner join; (2) left join; (3) right join and (4) full outer join.



Intro to Tableau Desktop

Why Tableau Desktop?

Usability - Simple and straightforward. Able to create charts quickly.

Ease of Learning - Free training videos and supported through Tableau Community Forum and user groups.

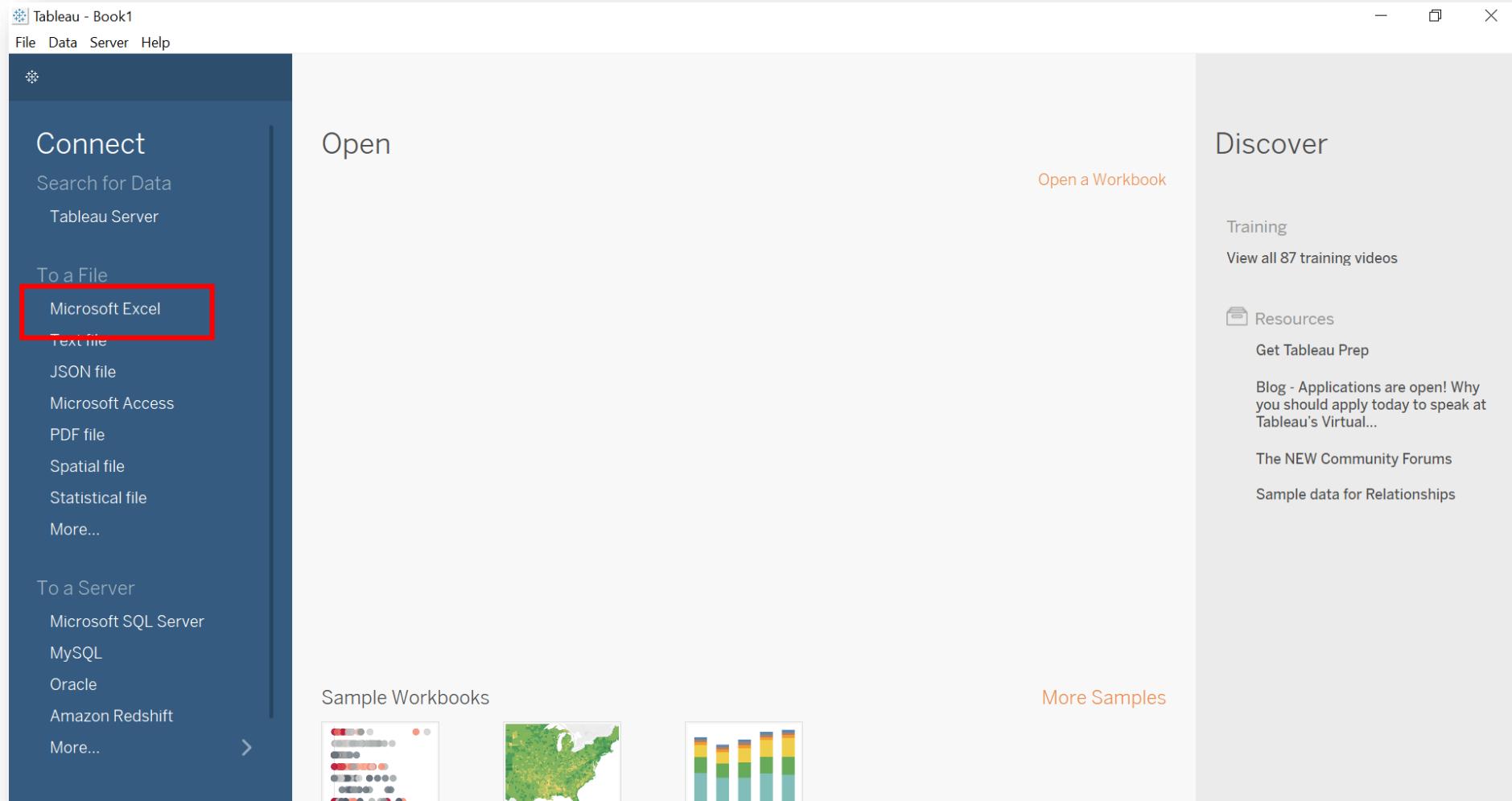
Integration - Support real time data discovery and connectivity to many different data sources.

Widely Adopted - Named as leader in Gartner Magic Quadrant for Business Intelligence And Analytics Platform



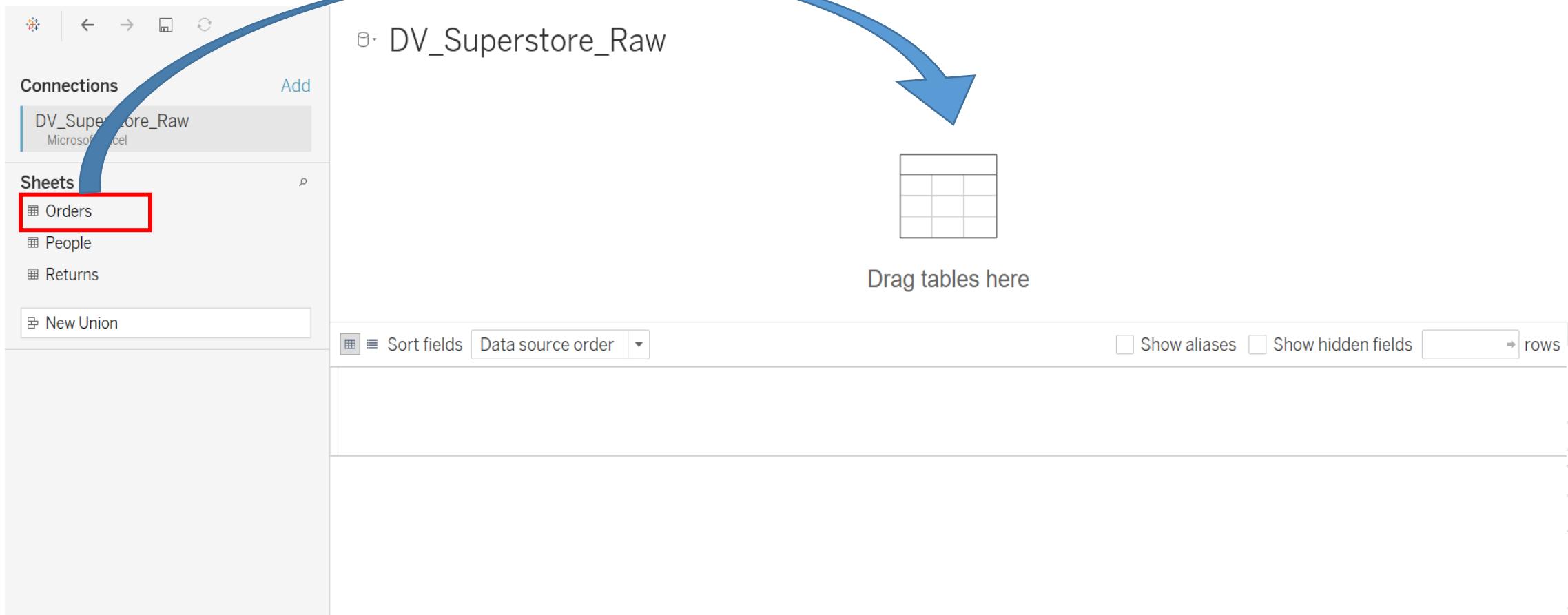
Launch Tableau Desktop

Click on Microsoft Excel



Connect to Data Source

Drag Orders table to the Drag Table Here area.



Data Source Page

Orders (DV_Superstore_Raw)

Connection: Live | Extract

Filters: 0 | Add

Orders

Need more data?

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

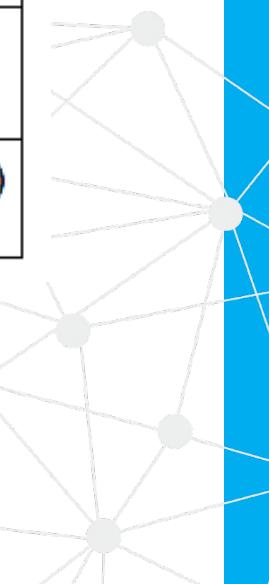
#	Abc Orders	Abc Orders	Abc Orders	Abc Orders	Abc Orders	Abc Orders	Abc Orders	Abc Orders	Abc Orders	Abc Orders	Abc Orders
Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer ...	Segment	Country	City	State	State
1	CA-2013-152...	9/11/2013	12/11/2013	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentu	
2	CA-2013-152...	9/11/2013	12/11/2013	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentu	
3	CA-2013-138...	13/6/2013	17/6/2013	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	Califo	
4	US-2012-108...	11/10/2012	18/10/2012	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale...	Florid	
5	US-2012-108...	11/10/2012	18/10/2012	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale...	Florid	



Data Type

- All fields have a data type
- The data type reflects the kind of information stored in the field
- You can change the data type

Icon	Data type
Abc	Text (string) values
📅	Date values
🕒	Date & Time values
#	Numerical values
T/F	Boolean values (relational only)
🌐	Geographic values (used with maps)



Data Source Page

Did Tableau auto-detect the data types for each column correctly?

If need be, just click on the icon (1-2-3, A-B-C etc.) and select the correct data type.

The screenshot shows the Tableau Data Source Page for the 'Orders (DV_Superstore_Raw)' connection. On the left, the 'Connections' sidebar shows 'DV_Superstore_Raw' selected. The main area displays the 'Orders' sheet with a preview of the data. A context menu is open over the 'Order ID' column, listing data types: Number (decimal), Number (whole), Date & Time, Date, String (selected), Boolean, Default, and Geographic Role. The 'String' option is highlighted with a blue selection bar and a checkmark. The preview table shows columns for Order ID, Ship Mode, Customer ID, Segment, Country, City, and State.

#	Abc	Abc	Abc	Abc	Abc	Abc	Abc	
Orders	Orders	Orders	Orders	Orders	Orders	Orders	Orders	
Row ID	Order ID	Ship Mode	Customer ID	Customer ...	Segment	Country	City	State
1	CA-2013-1	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentu
2	CA-2013-2	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentu
3	CA-2013-3	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	Califo
4	US-2012-1	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florid
5	US-2012-2	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florid

Tableau Interface

Click on Sheet 1

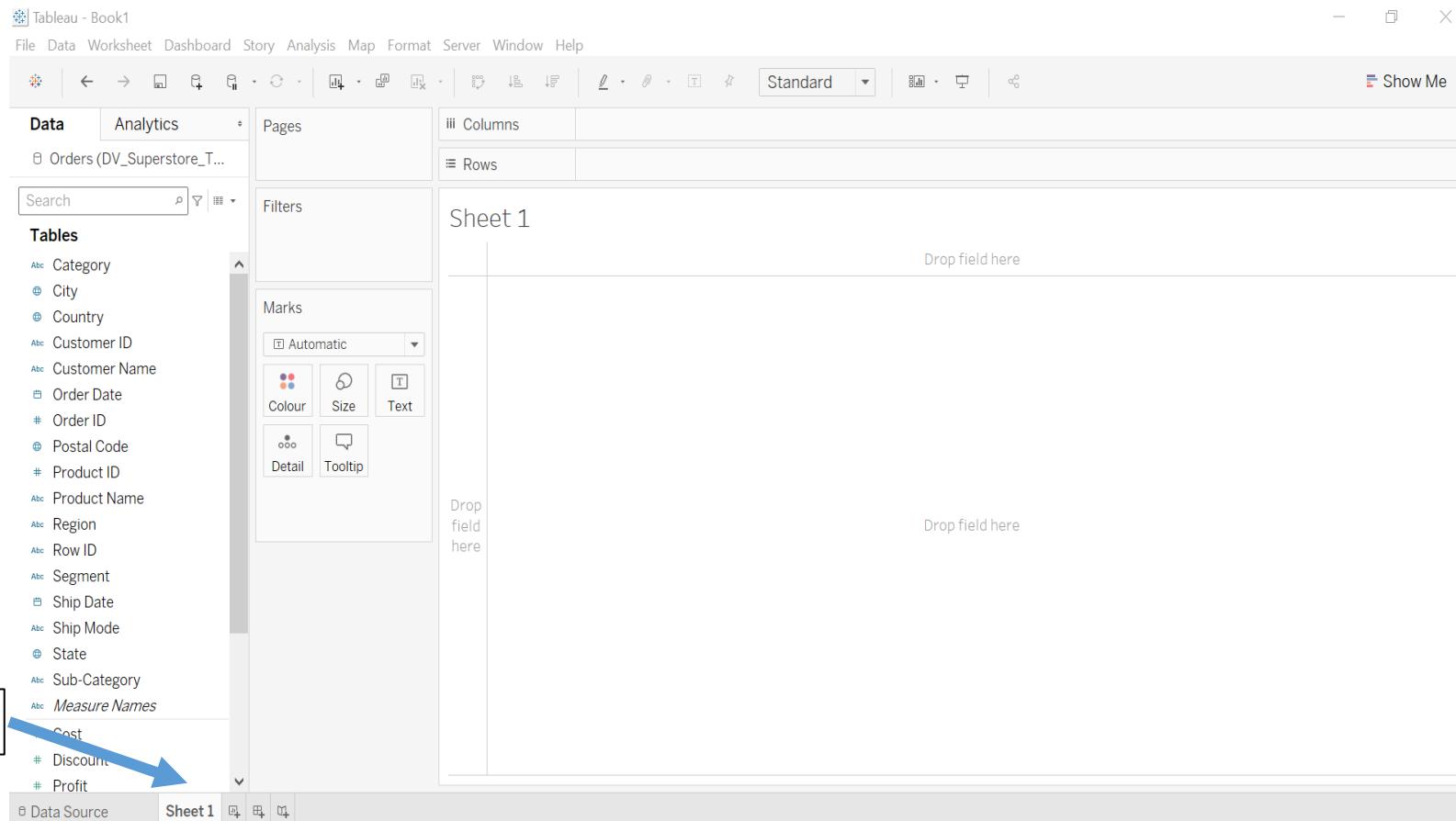


Tableau Interface

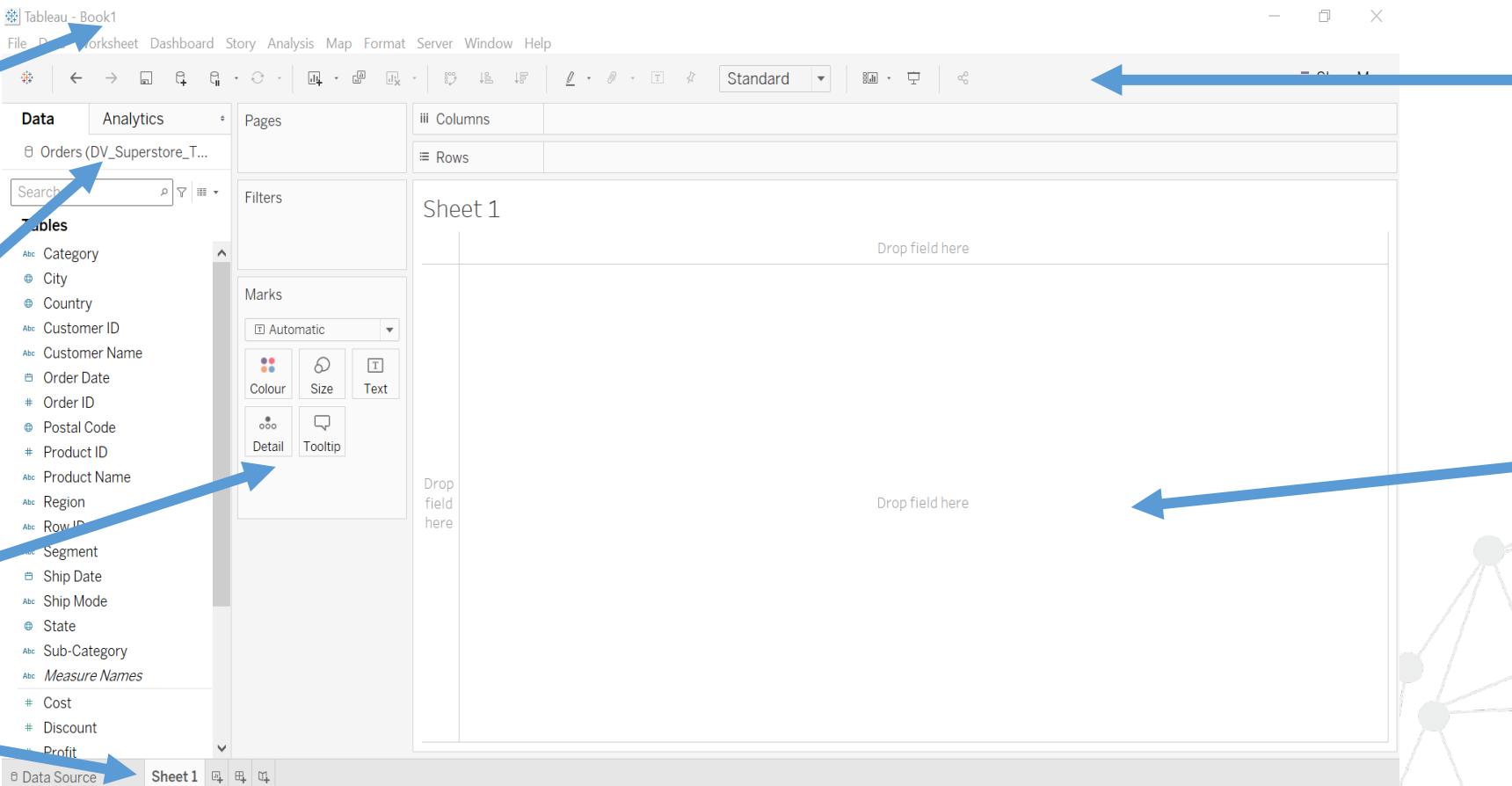
Click on Sheet 1

Workbook name

Side bar : Data Pane and Analytics Pane

Cards & Shelves used to add data into view

Sheet tabs



Toolbar to access commands, analysis and navigation tools

Workspace to create visualisations

Data Pane

Dimensions

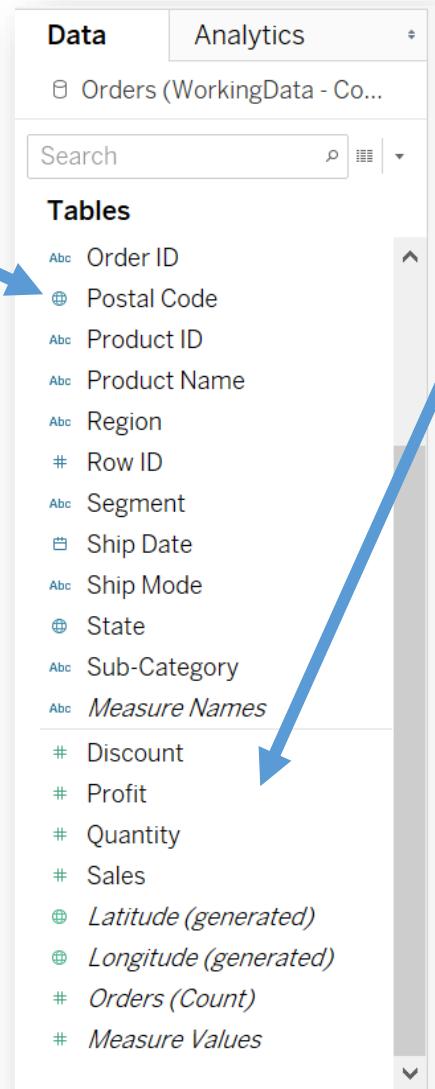
Qualitative data

E.g. Name, Address,
Postal Code, NRIC etc.

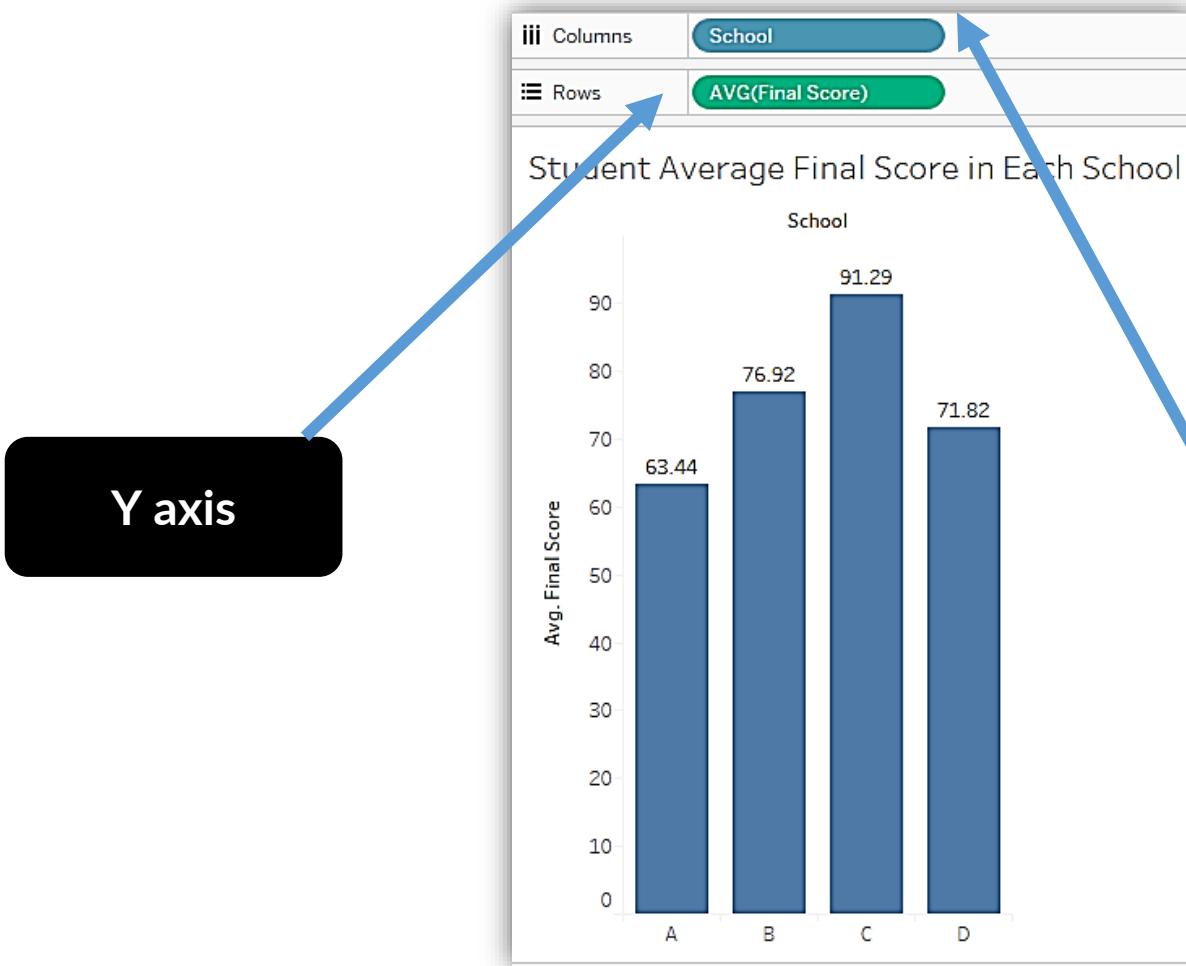
Measures

Quantitative data

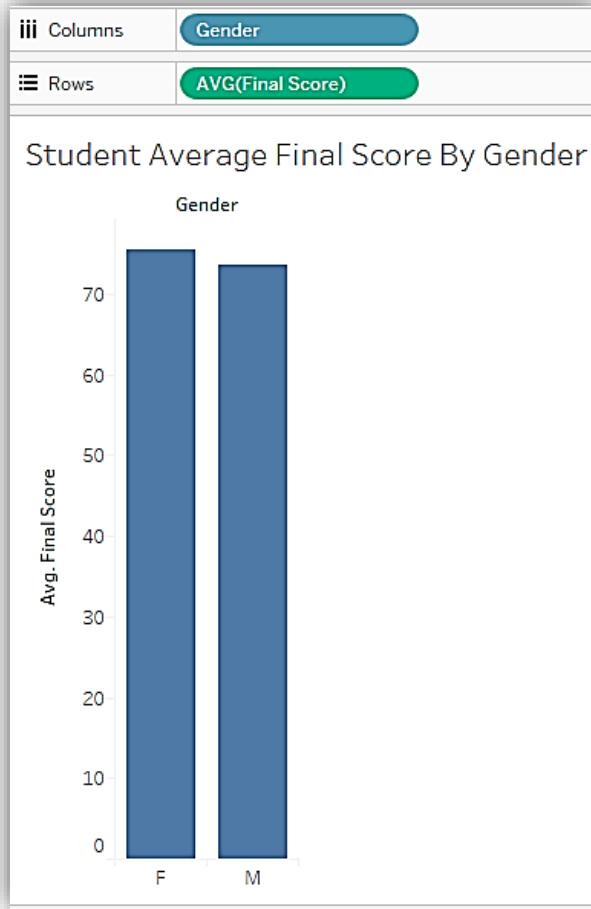
E.g. Age, Entry Score, Final
Score, Household Income etc.



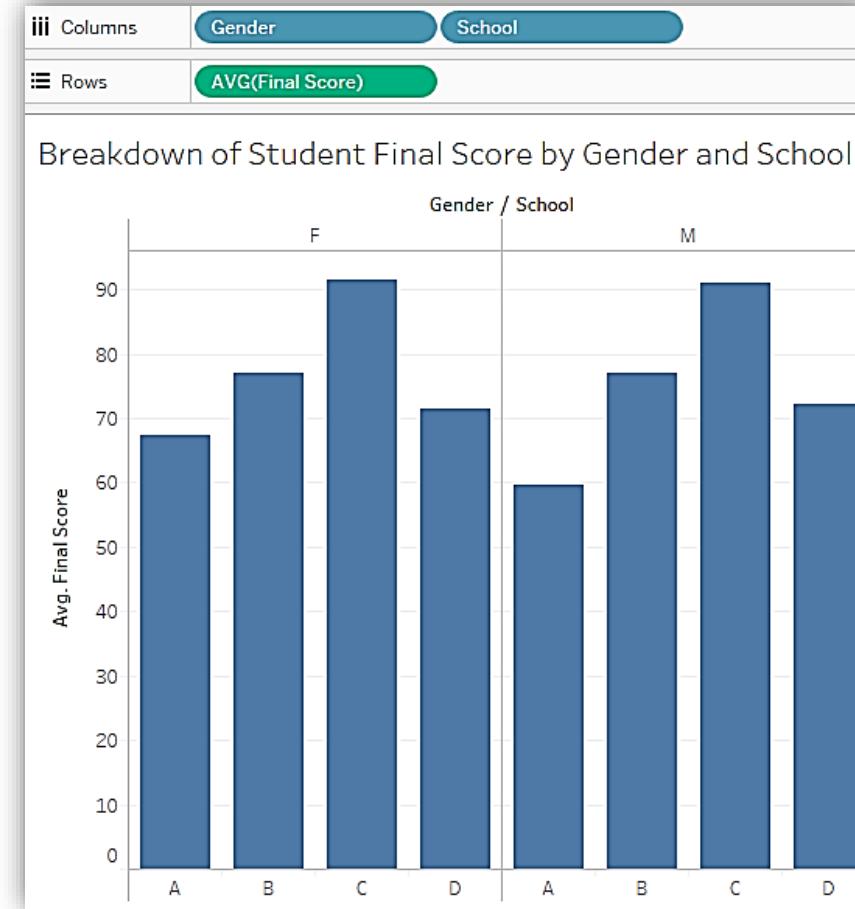
Rows and Columns



More on Rows and Columns



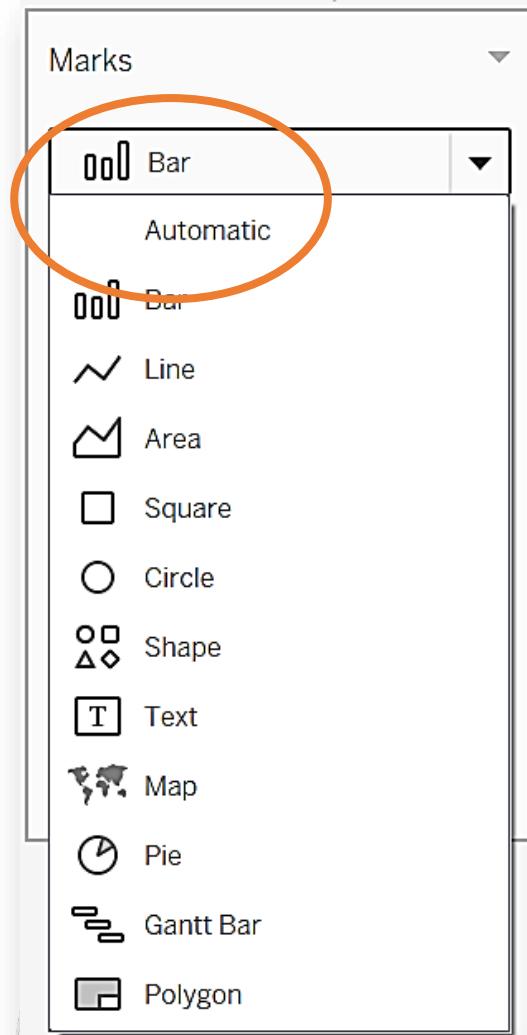
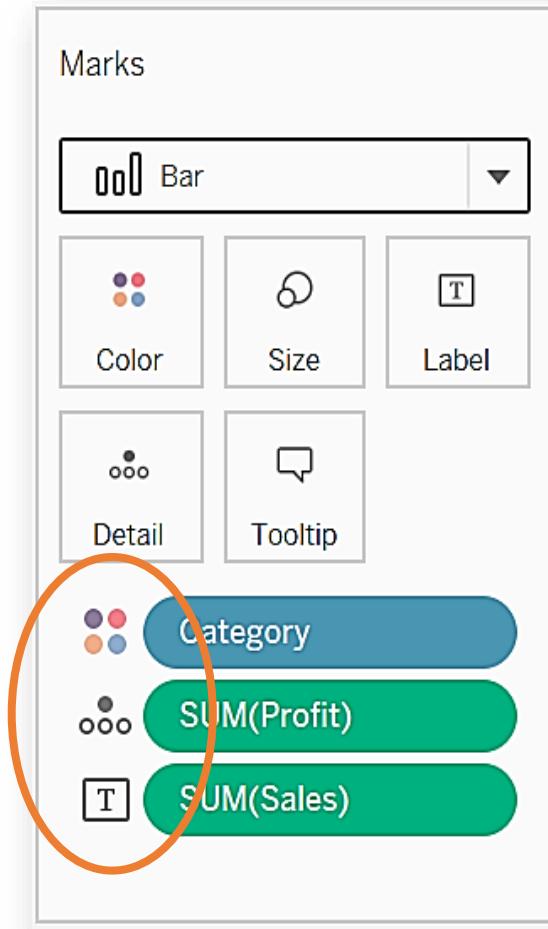
Additional Dimensions



Make Changes on the Chart

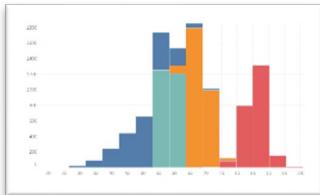
Changes to the Chart using Marks Card

- Add Color
- Make marks smaller or larger
- Add or modify Labels
- Add field to Detail
- Add field to Tooltip

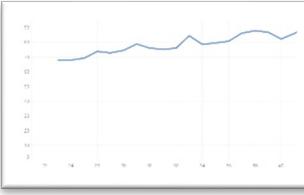


Univariate vs Bivariate/Multivariate data

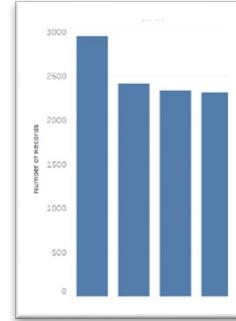
- Univariate data refers to data which has only one variable.
- Frequency distributions



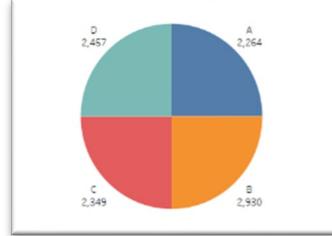
Histogram



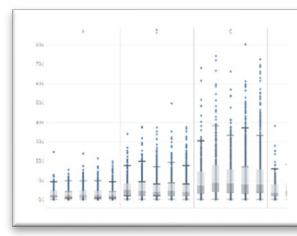
Line Chart



Bar Chart

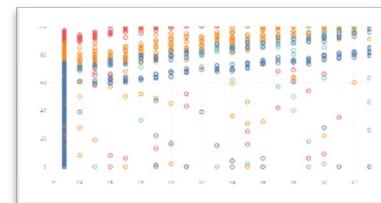


Pie Chart



Box and Whisker

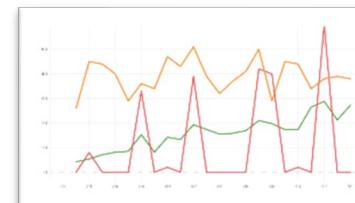
- Bivariate/Multivariate data refers to data which has two or more variables.
- Comparisons, relationships, causes, explanation



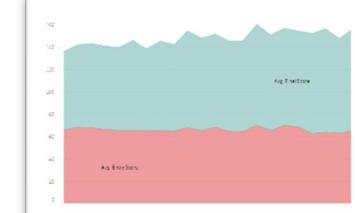
Scatter Plot



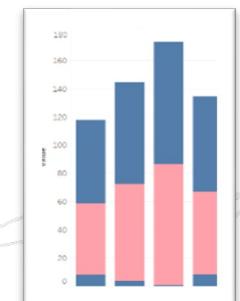
Bubble Chart



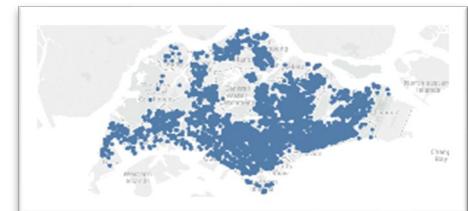
Line Chart



Area Chart



Stacked Bar



Map



'Show Me' Features

- 'Show Me' offers guidance on which chart type is best suited for the data.



Bar Chart

Line Chart

Area Chart

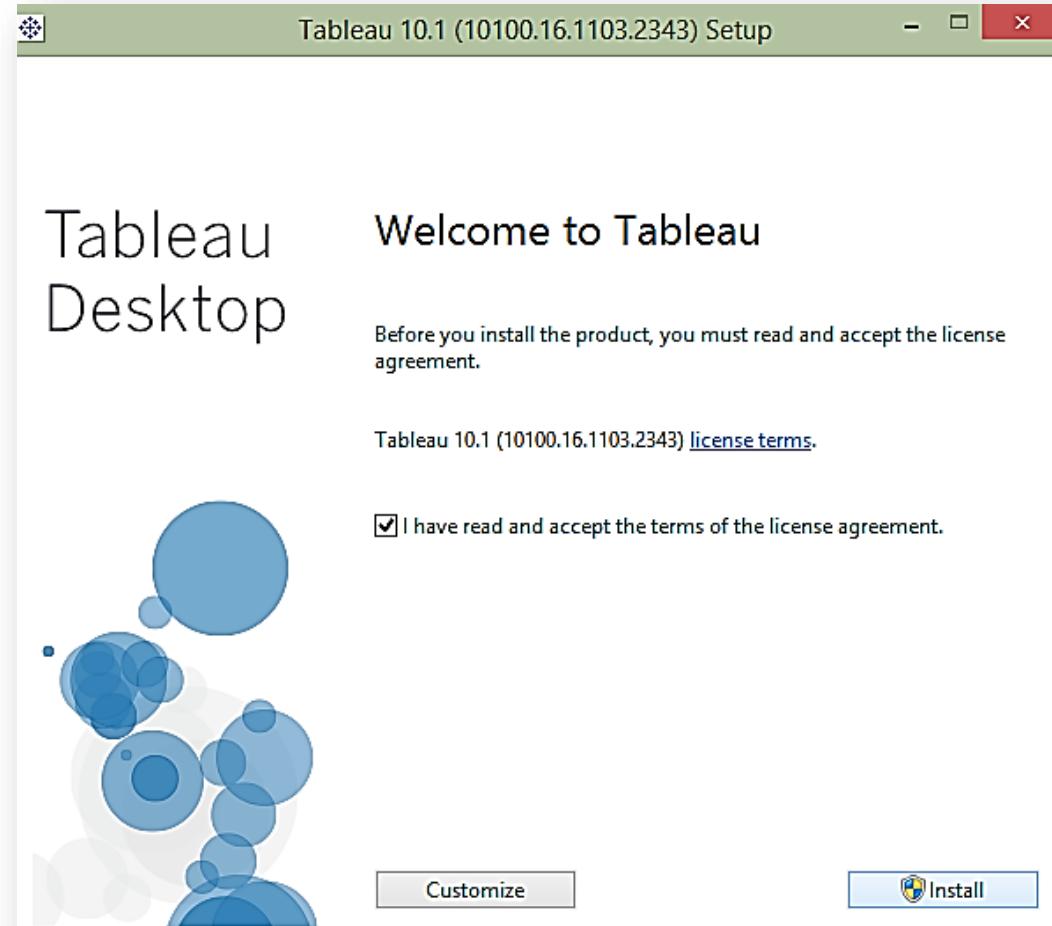
Download and Install Tableau Desktop

Download trial version of Tableau Desktop

<http://www.tableau.com/products/desktop/download>

Safekeep the email from Tableau. It will give you:

- The **Product Key** will give you :
 - 1 year free license of Tableau
- The **Access Key** will give you :
 - 1 year free access of e-learning (to prepare for certification exam)
 - 20% discount for Tableau Desktop Specialist Certification





Join and Cardinality

Using Join in Tableau

In Tableau, you can use **Join** to merge columns together.

It **physically joins** the tables together to create **one single, combined table**.

To combine the columns, you can use one of the 4 common types of join : inner join, left join, right join, full outer join.

The screenshot shows the Tableau Data Source editor. At the top, there's a tree view with a 'Join' node expanded. Below it, a message says 'Sales is made of 2 tables.' with a help icon. A dashed orange arrow points from this message to a circular join icon between two horizontal bars labeled 'Sales' and 'Qty'. Below this, a table preview shows 'Sales' with 6 fields and 16 rows. The table has columns: Name, Sales, Fields, Type, Field Name, Physical Table, and Remote Field N... . The main data table below has columns: Sales, Sales Month, Sales (Sales), Qty, Prod ID (Qty), Month (Qty), and Qty. The data in the table is as follows:

	Abc Sales	Abc Sales	# Sales	Abc Qty	Abc Qty	# Qty
Prod ID	Month	Sales (Sales)	Prod ID (Qty)	Month (Qty)	Qty	
A	Jan	100	A	Jan	12	
A	Jan	200	A	Jan	12	
A	Feb	150	A	Feb	10	
A	Feb	150	A	Feb	15	
A	Mar	100	A	Mar	12	
A	Mar	120	A	Mar	12	



Some Drawbacks

- Join can cause data loss or duplication if tables are at different levels of details.
- The joined data sources must also be fixed before analysis can begin.

When to use this?

- OK to use if you have **one-to-many** or **many-to-one** relationship.
- **Do not** use if you have **many-to-many** relationship.



Example

There are two tables : Sales and Quantity Sold.
What is the cardinality between the two tables?

Prod ID	Month	Sales
A	Jan	100
A	Jan	200
A	Feb	150
A	Mar	100
A	Mar	120
B	Jan	400
B	Feb	300
B	Feb	200
B	Mar	500
	Mar	100

Prod ID	Month	Qty
A	Jan	12
A	Feb	15
A	Feb	10
A	Mar	12
B	Jan	20
B	Jan	15
B	Feb	25
	Mar	55
	Mar	10
	Mar	12

Many-to-Many
cardinality



Example

Prod ID	Month	Sales
A	Jan	100
A	Jan	200
A	Feb	150
A	Mar	100
A	Mar	120
B	Jan	400
B	Feb	300
B	Feb	200
B	Mar	500
B	Mar	100

Prod ID	Month	Qty
A	Jan	12
A	Feb	15
A	Feb	10
A	Mar	12
B	Jan	20
B	Jan	15
B	Feb	25
B	Mar	55
B	Mar	10
B	Mar	12

What happens if you try to summarize the info from the two tables into one?

Prod ID	Month	Sales	Qty
A	Jan		
A	Feb		
A	Mar		
B	Jan		
B	Feb		
B	Mar		



Example

Prod ID	Month	Sales
A	Jan	100
A	Jan	200
A	Feb	150
A	Mar	100
A	Mar	120
B	Jan	400
B	Feb	300
B	Feb	200
B	Mar	500
B	Mar	100

Prod ID	Month	Qty
A	Jan	12
A	Feb	15
A	Feb	10
A	Mar	12
B	Jan	20
B	Jan	15
B	Feb	25
B	Mar	55
B	Mar	10
B	Mar	12

What happens if you try to summarize the info from the two tables into one?

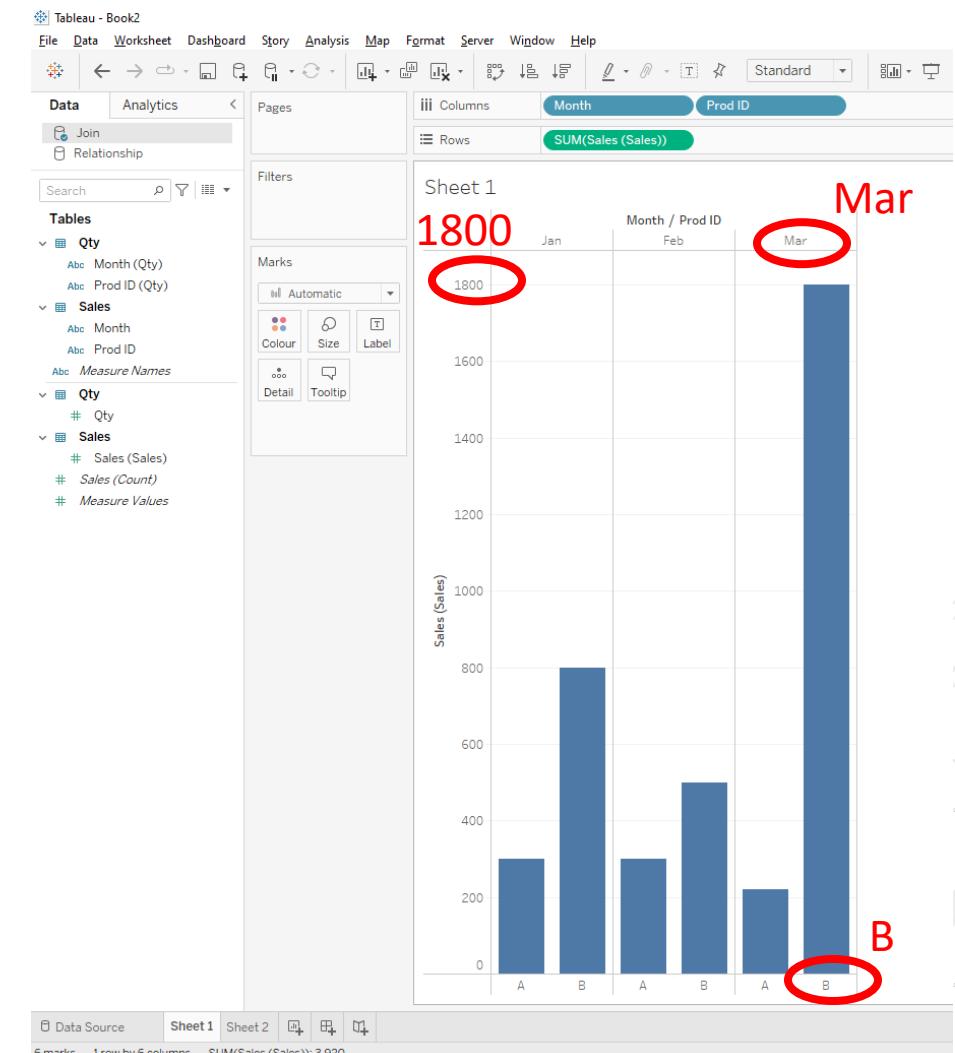
Prod ID	Month	Sales	Qty
A	Jan	300	12
A	Feb	150	25
A	Mar	220	12
B	Jan	400	35
B	Feb	500	25
B	Mar	600	77

Example - Join

What You Expect

Prod ID	Month	Sales	Qty
A	Jan	300	12
A	Feb	150	25
A	Mar	220	12
B	Jan	400	35
B	Feb	500	25
B	Mar	600	77

What You Get
using Join in Tableau



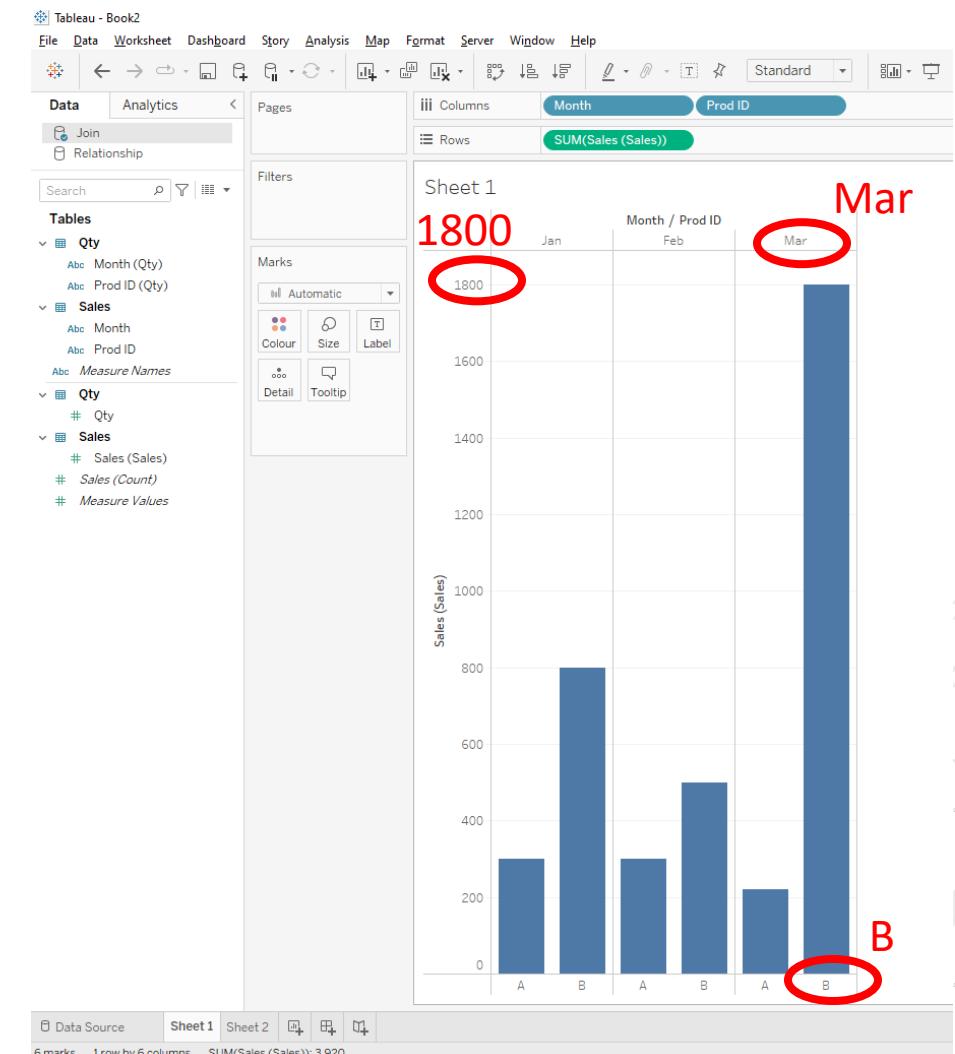
Example - Join

Where did it go wrong?

Prod ID	Month	Sales
A	Jan	100
A	Jan	200
A	Feb	150
A	Mar	100
A	Mar	120
B	Jan	400
B	Feb	300
B	Feb	200
B	Mar	500
B	Mar	100

Prod ID	Month	Qty
A	Jan	12
A	Feb	15
A	Feb	10
A	Mar	12
B	Jan	20
B	Jan	15
B	Feb	25
B	Mar	55
B	Mar	10
B	Mar	12

What You Get
using Join in Tableau





P01b - Joins - Sales: Hands-on

- On LMS, head to “Activity 2: Joins” > “Practical Questions”. Download practical sheets and datasets.
 - The database captures the sales transactions that occurs between the company and its customers. We will learn to do :
 - Data Gathering : Combine data using Joins
 - Create Calculated Field
 - Exploration using pie chart, bar chart and line chart.
- 

Data Integration Using Relationship



How are the 4 tables related? What's the Cardinality?

How do you think the main table **Articles** is related to the other 3 tables?

What's the cardinality?

Articles

	A	B	C	D	E
1	articleID	issueID	authorID	webHits	
2	1	1	8	2019	
3	2	1	5	1421	
4	3	1	4	1174	
5	4	1	4	1613	
6	5	1	3	1099	
7	6	1	8	1903	
8	7	1	7	1718	
9	8	2	1	642	
10	9	2	5	1616	
11	10	2	4	1233	
12	11	2	8	1937	
13	12	2	9	3068	
14	13	~	~	~	~

Sales

	A	B	C	D
1	saleID	issueID	lotSize	
2	1701	74	100	
3	386	18	100	
4	110	6	100	
5	1784	78	100	
6	3131	112	100	
7	1404	60	100	
8	2334	92	100	
9	2037	85	100	
10	1134	50	100	
11	586	26	100	
12	3205	114	100	
13	1763	77	100	
14	2150	88	100	

ArticleHitComments

	A	B	C	D	E
1	articleID	authorName	webHits	commentCount	
2	1	Destiny A	2019	14	
3	2	Jon Rader	1421	6	
4	3	Matt Jann	1174	8	
5	4	Matt Jann	1613	26	
6	5	Paul Seme	1099	10	
7	6	Destiny A	1903	26	
8	7	Nicole Fry	1718	21	
9	8	Jason Wig	642	8	
10	9	Jon Rader	1616	7	
11	10	Matt Jann	1233	12	

Issues

	A	B	C	D
1	issueID	PubDate		
2	1	10/24/04		
3	2	4/8/2011		
4	3	11/23/04		
5	4	4/8/2012		
6	5	12/23/04		
7	6	5/7/2001		
8	7	1/22/05		
9	8	5/6/2002		
10	9	2/21/05		
11	10	5/8/2003		
12	11	3/23/05		
13	12	5/7/2004		
14	13	4/22/05		
15	14	5/7/2005		

 Short Answer



Creating Relationship Among Data Tables

Sometimes, data come **from various sources** and you want to integrate them into **one unified view**.

In this case, there is a need to **link** the data tables together using **common field** to create a **data model**.

In the example below, the common field is customer ID.

ID	First Name	Last Name
B00072	Buzz	Lightyear
B00165	Sheriff	Woody
C00092	Potato	Head

Join by
Customer ID

ID	Food
B00072	Satay
B00165	Rojak
C00092	Laksa





Data Model

- Once the relationship among data tables has been established using the common keys, this entire **inter-related structure** is called a **data model**, which provides one unified view.



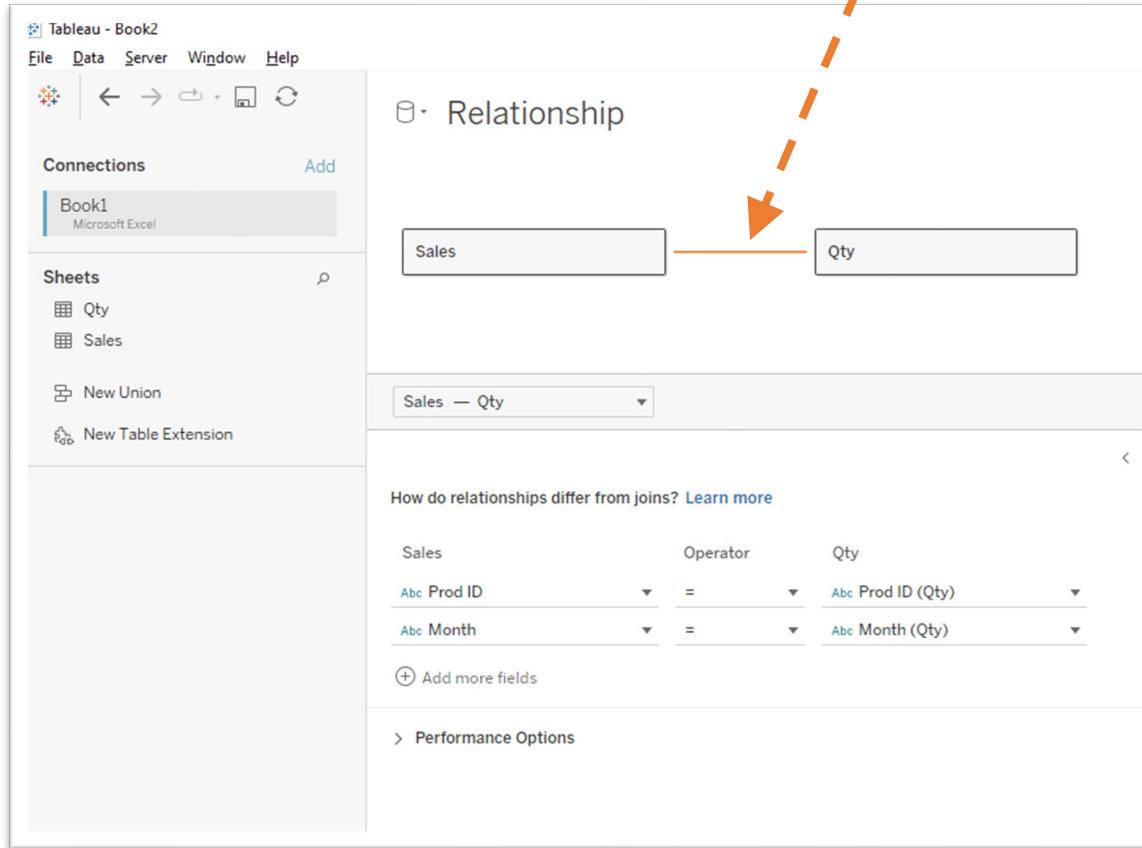
Data Model



Relationship and Cardinality

Using Relationship in Tableau

A Relationship in Tableau is a connection between two or more tables based on a common field, but does not merge the tables together.



When to use this?

This should be your default method when integrating data.

It can be used for all cardinalities, including many-to-many relationships.

Example

There are two tables : Sales and Quantity Sold.
What is the cardinality between the two tables?

Prod ID	Month	Sales
A	Jan	100
A	Jan	200
A	Feb	150
A	Mar	100
A	Mar	120
B	Jan	400
B	Feb	300
B	Feb	200
B	Mar	500
	Mar	100

Prod ID	Month	Qty
A	Jan	12
A	Feb	15
A	Feb	10
A	Mar	12
B	Jan	20
B	Jan	15
B	Feb	25
	Mar	55
	Mar	10
	Mar	12

Many-to-Many
cardinality



Example

Prod ID	Month	Sales
A	Jan	100
A	Jan	200
A	Feb	150
A	Mar	100
A	Mar	120
B	Jan	400
B	Feb	300
B	Feb	200
B	Mar	500
B	Mar	100

Prod ID	Month	Qty
A	Jan	12
A	Feb	15
A	Feb	10
A	Mar	12
B	Jan	20
B	Jan	15
B	Feb	25
B	Mar	55
B	Mar	10
B	Mar	12

What happens if you try to summarize the info from the two tables into one?

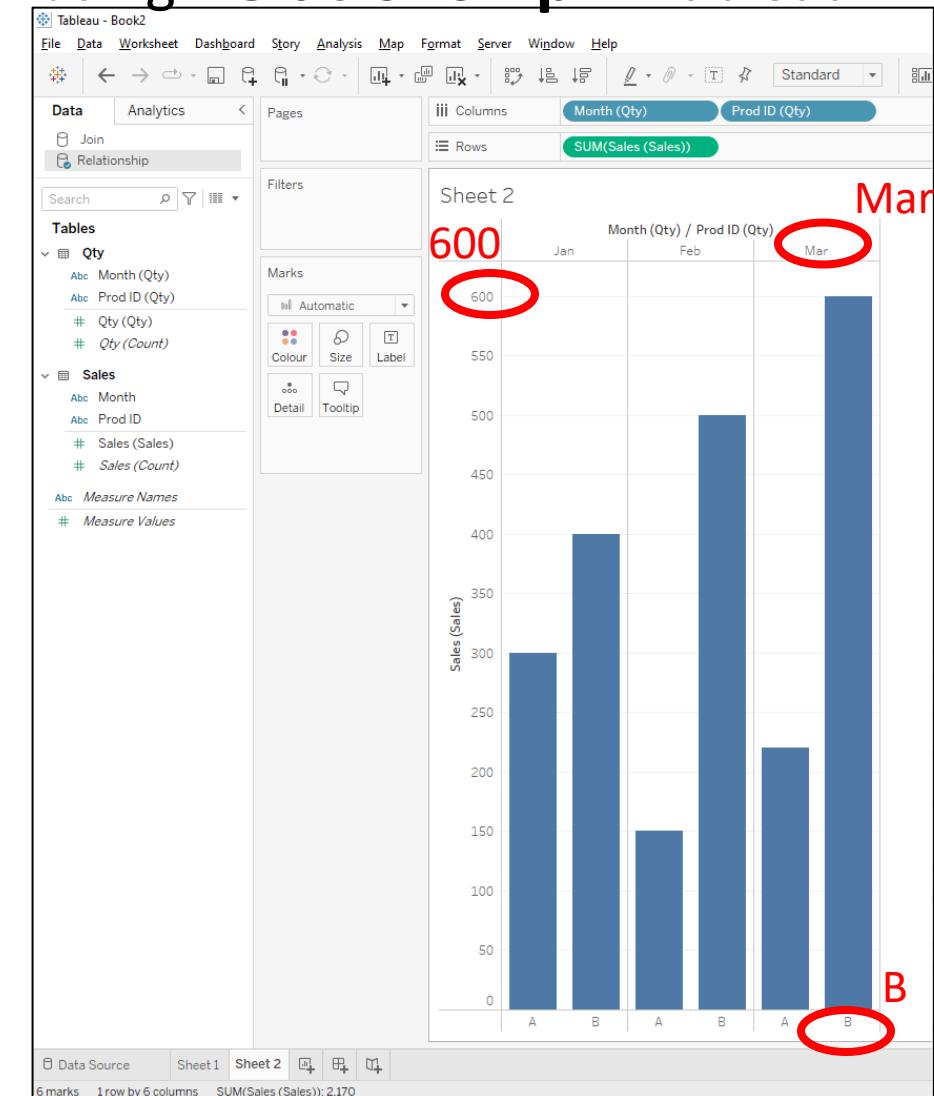
Prod ID	Month	Sales	Qty
A	Jan	300	12
A	Feb	150	25
A	Mar	220	12
B	Jan	400	35
B	Feb	500	25
B	Mar	600	77

Example - Relationship

What You Expect

Prod ID	Month	Sales	Qty
A	Jan	300	12
A	Feb	150	25
A	Mar	220	12
B	Jan	400	35
B	Feb	500	25
B	Mar	600	77

What You Get
using Relationship in Tableau





P01c – Relationships - Magazine: Hands-on

- On LMS, head to “Activity 3: Relationships > “Practical Questions”. Download practical sheets and datasets.
- The database captures info about the articles published on the web by a magazine company. We will learn to do:
 - Data Gathering : Combine data using Relationships
 - Exploration using tree map and scatter plot.





Practical 1

<u>Before Class</u>	Concepts	Intro to Data Visualisation
<u>During Class</u>	Hands-on	Intro to Data Visualisation + Integration + Tableau Cardinality - Join - Relationship
<u>After Class</u>	Hands-on	<i>Revise today's class with LMS: Apr/Oct – Week X (...)</i> <i>Do "What is IDL" on LMS</i> <i>Go through Additional Resources slides</i> - <i>Read Extra Resources on Exploratory vs Explanatory</i> - <i>Read website on cardinality Lab Q2</i> - <i>Do "Harry Potter – Which Join should I use?" quiz</i>