

Decluttering & Getting Started with Tableau

#6



MIS56 I Data Visualization

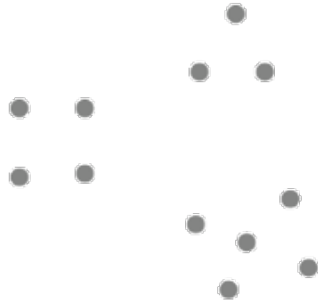


Outline

- **Decluttering: step-by-step**
 - **Tableau product portfolio**
 - **Connecting to data**
 - **Basic data prep**
 - **Dimensions and measures**
- 



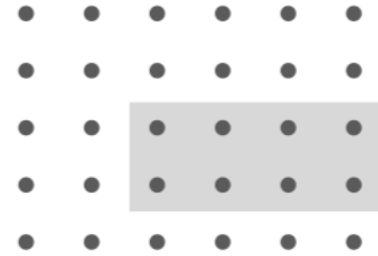
Gestalt principles of visual perception



proximity



similarity



enclosure



closure



continuity

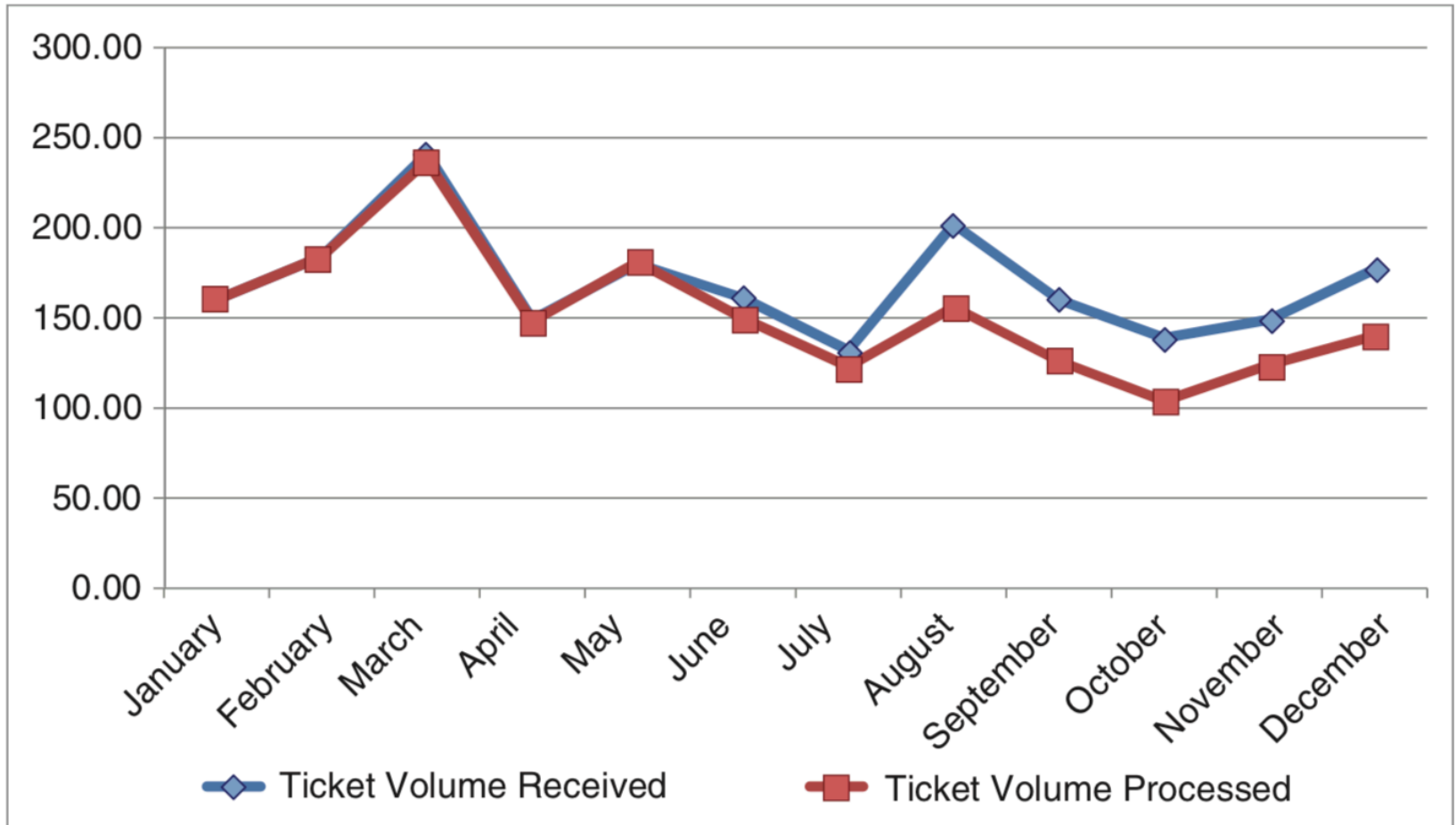


connection



Original graph

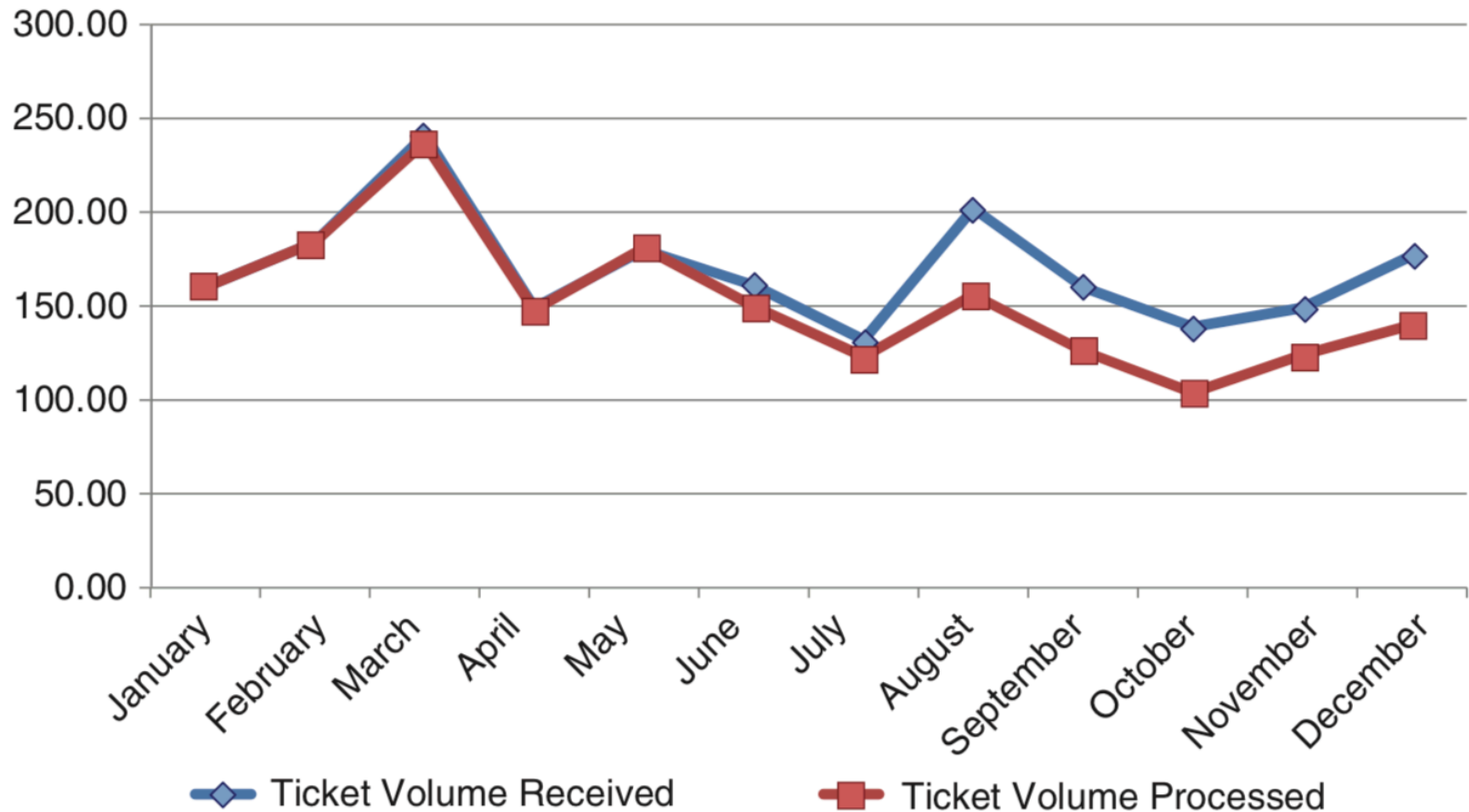
- decluttering: step by step





Remove chart border

- decluttering: step by step

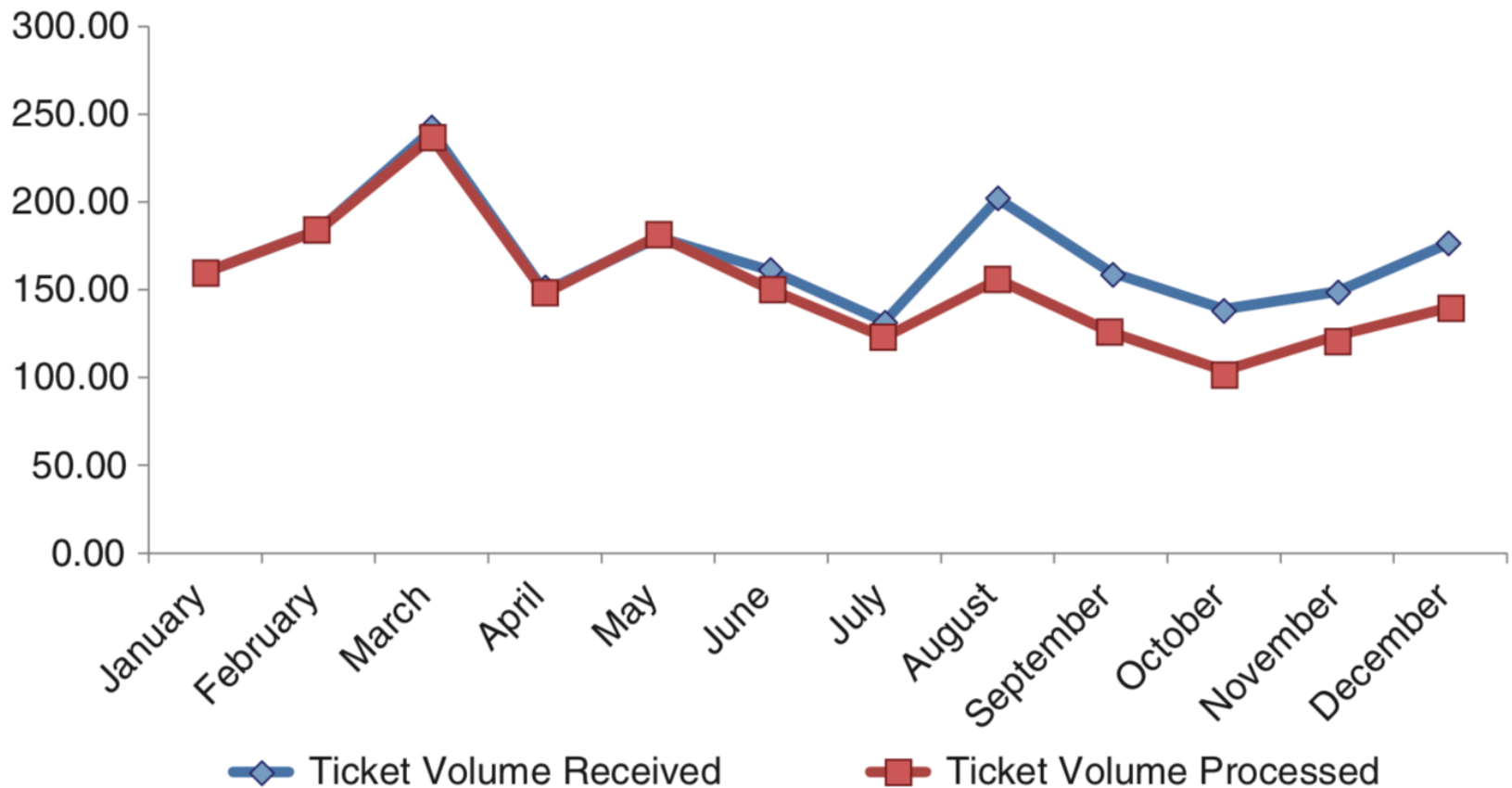


Gestalt principle of closure



Remove gridlines

- decluttering: step by step

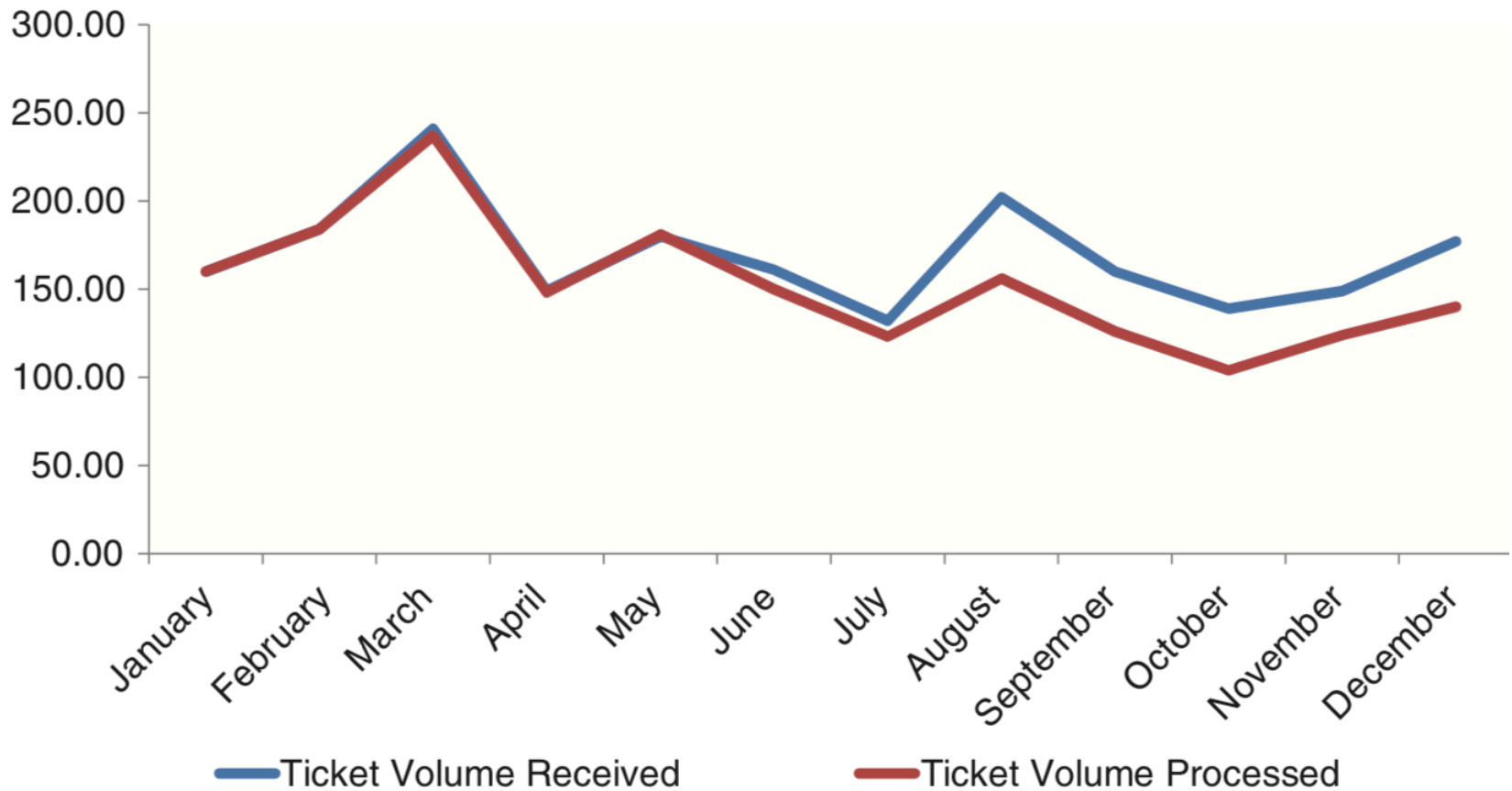


Don't let them compete visually with your data



Remove data markers

- decluttering: step by step

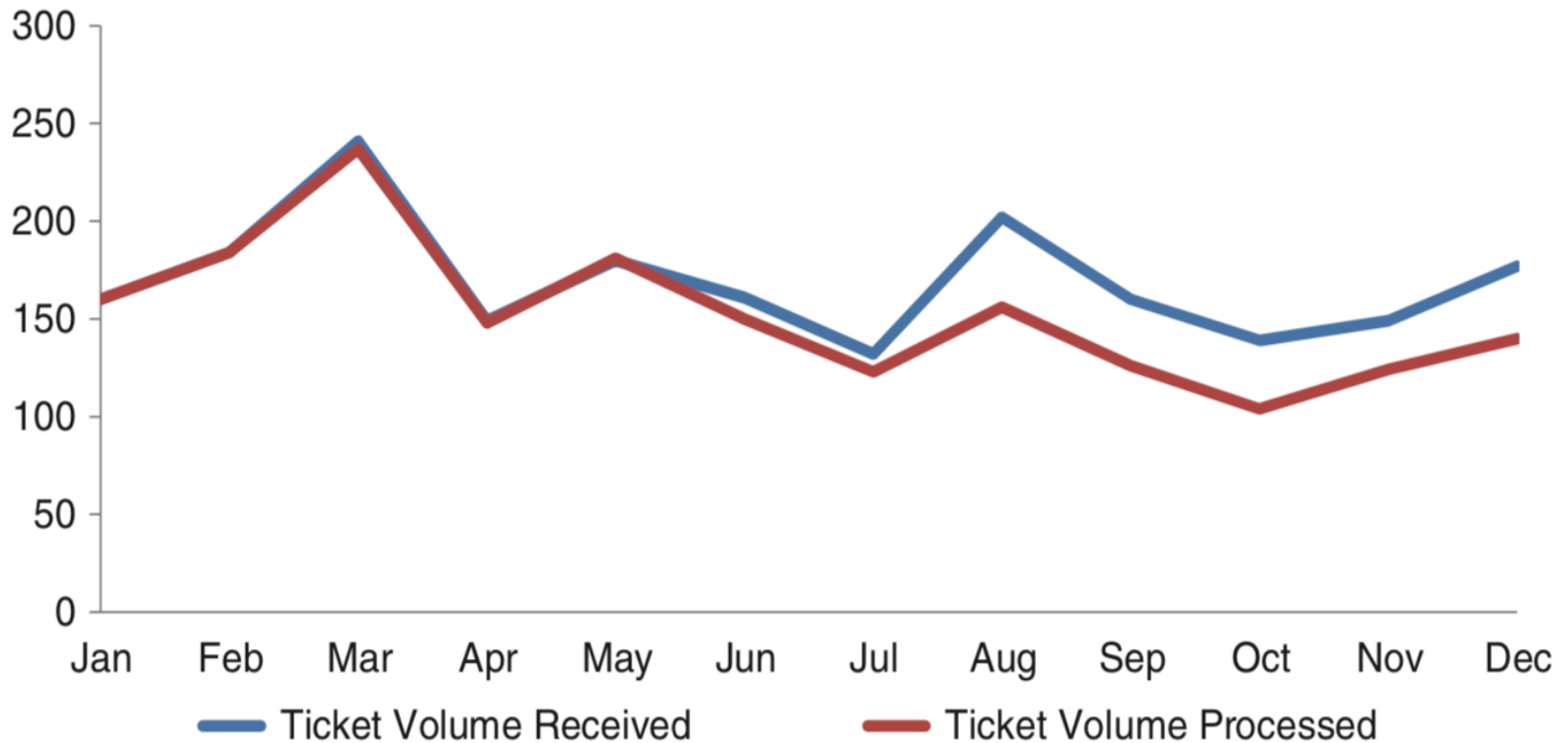


Every single element adds cognitive load on audience



Clean up axis labels

- decluttering: step by step

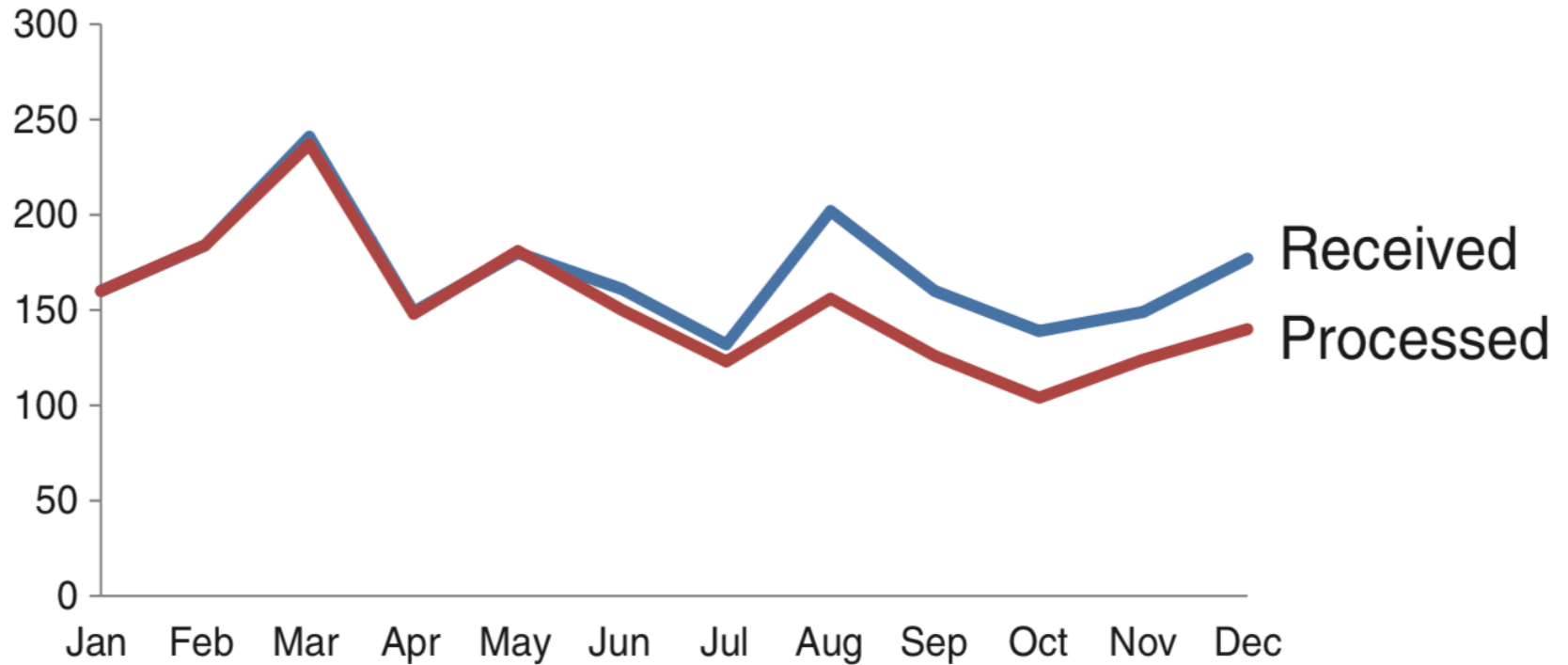


- digit 0 has no value after decimal
- diagonal elements on x-axis



Label data directly

- decluttering: step by step

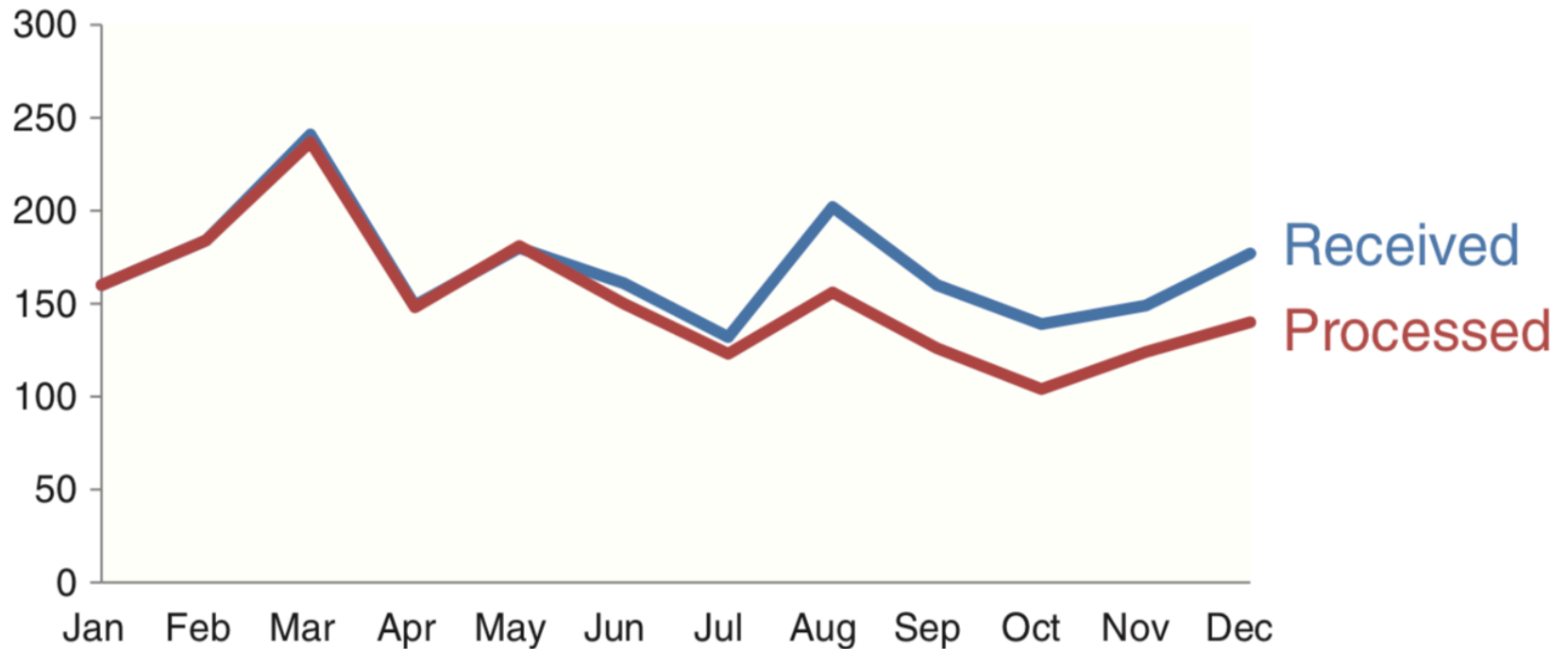


Gestalt principle of proximity



Leverage consistent color

- decluttering: step by step

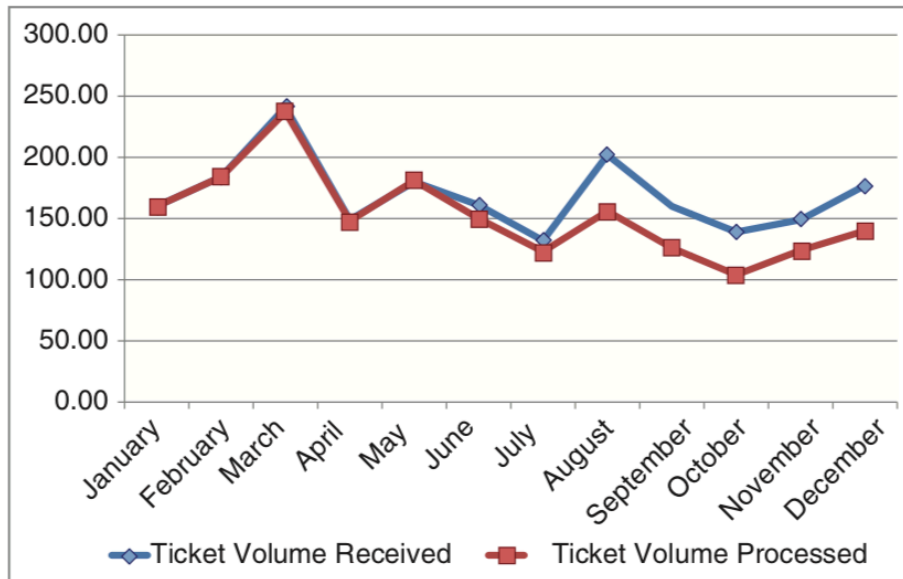


Gestalt principle of similarity

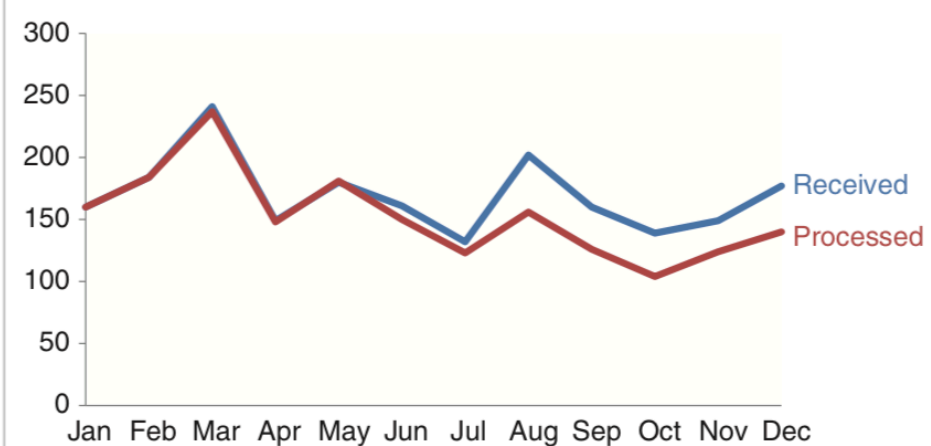


Before-and-after

- decluttering: step by step



before



after

This visual is not yet complete. But identifying and eliminating the clutter has brought us a long way in terms of reducing cognitive load and improving accessibility.

Getting started with Tableau



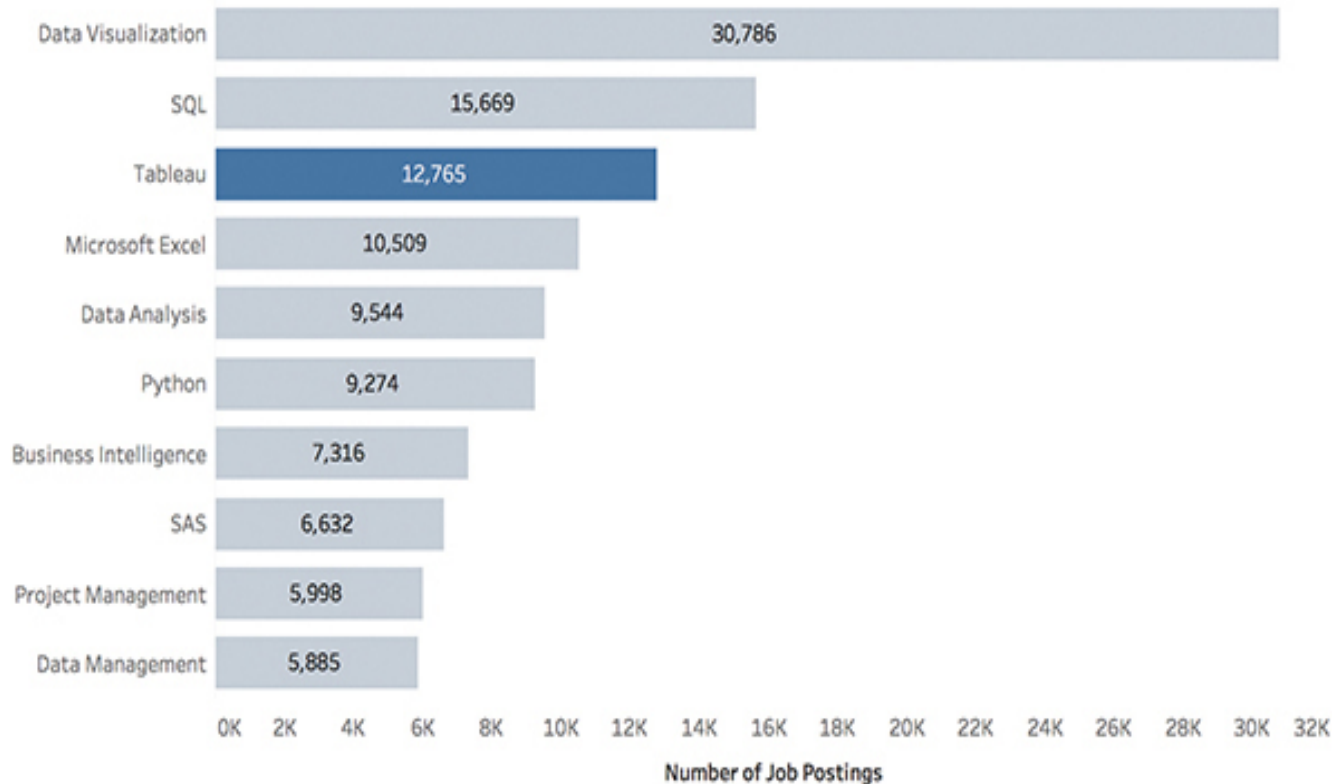


Why Tableau?

Data Visualization Top Specialized Skills

Of ~31k visualization related jobs posted between March 2017 and February 2018, ~13k listed [Tableau](#) as a desired specialized skill. Above Excel (~11k) and SAS (~7k) it is the only software listed in the top specialized skill set.

Source: Labor Insight (Burning Glass Technologies)





The Tableau product portfolio

Tableau Server

- Best suited for enterprise-wide deployments
- Intended to provide entire organizations with the ability to connect to any data source—on-premise or in the cloud—with centrally managed governance and granular security protocols to maintain balance between user flexibility and IT control



Tableau Desktop

- An application that can be used on either Windows or Mac machines
- Allows connection to data on-premise or in the cloud
- Facilitates the entire visual discovery and analytics process from connecting to data to sharing visualizations, dashboards, or interactive stories using Tableau Server or Tableau Online



The Tableau product portfolio

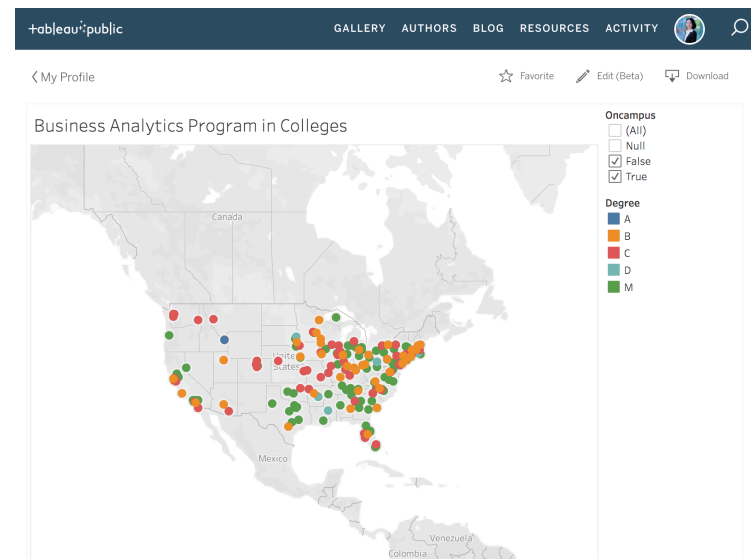
Tableau Online

- A fully cloud-hosted platform that primarily works with cloud databases



Tableau Public

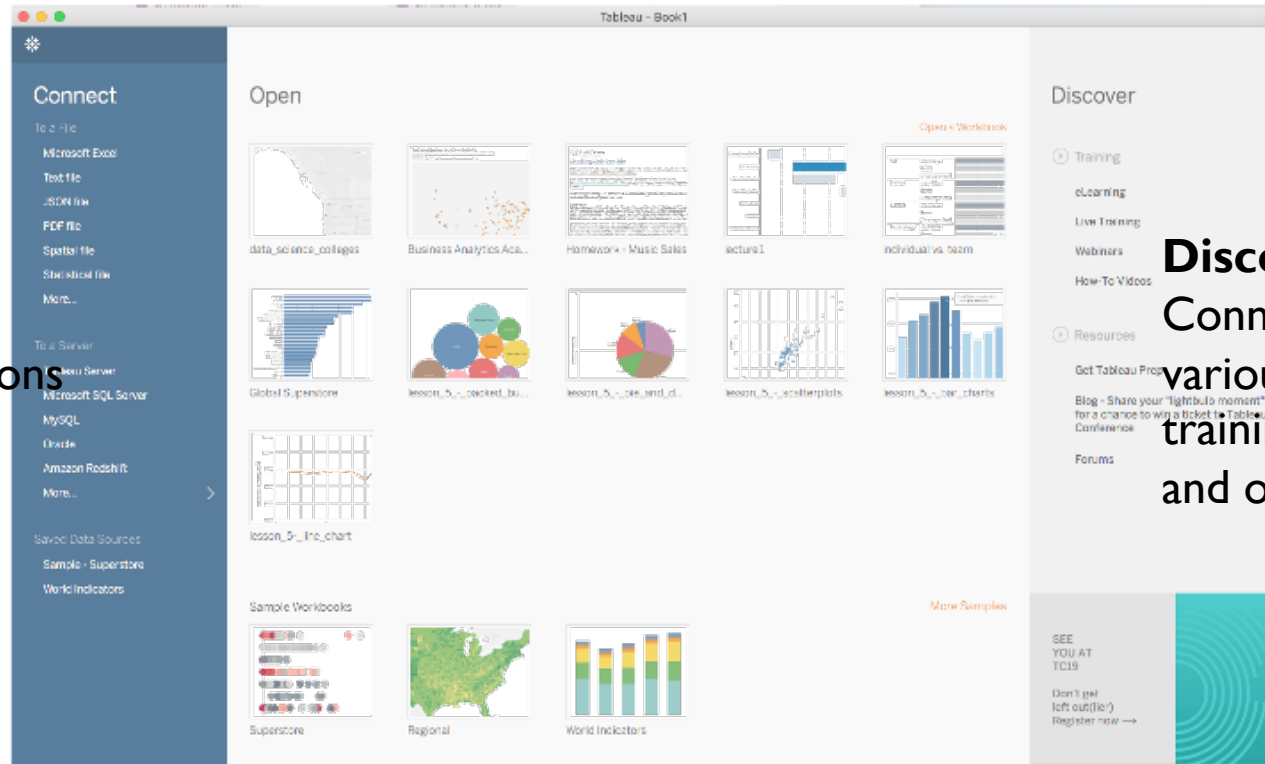
- One part data visualization hosting service, one part social networking
- A free service that allows users to publish interactive data visualizations online. These visualizations can be embedded into webpages and blogs, shared via social media or email, or made available for download to other users





Connecting to data

Open - As you create your own workbooks, recently opened workbooks appear here for quick access.



Connect –
A long list of
native connections
to various data
sources.

Discover -
Connects you to
various Tableau
training, visualization,
and other resources.

Sample workbooks - default samples provided by Tableau.

Connecting to tables

Connections:

You can add additional data sources by clicking Add. You can also edit the name of the connection or remove it as desired.

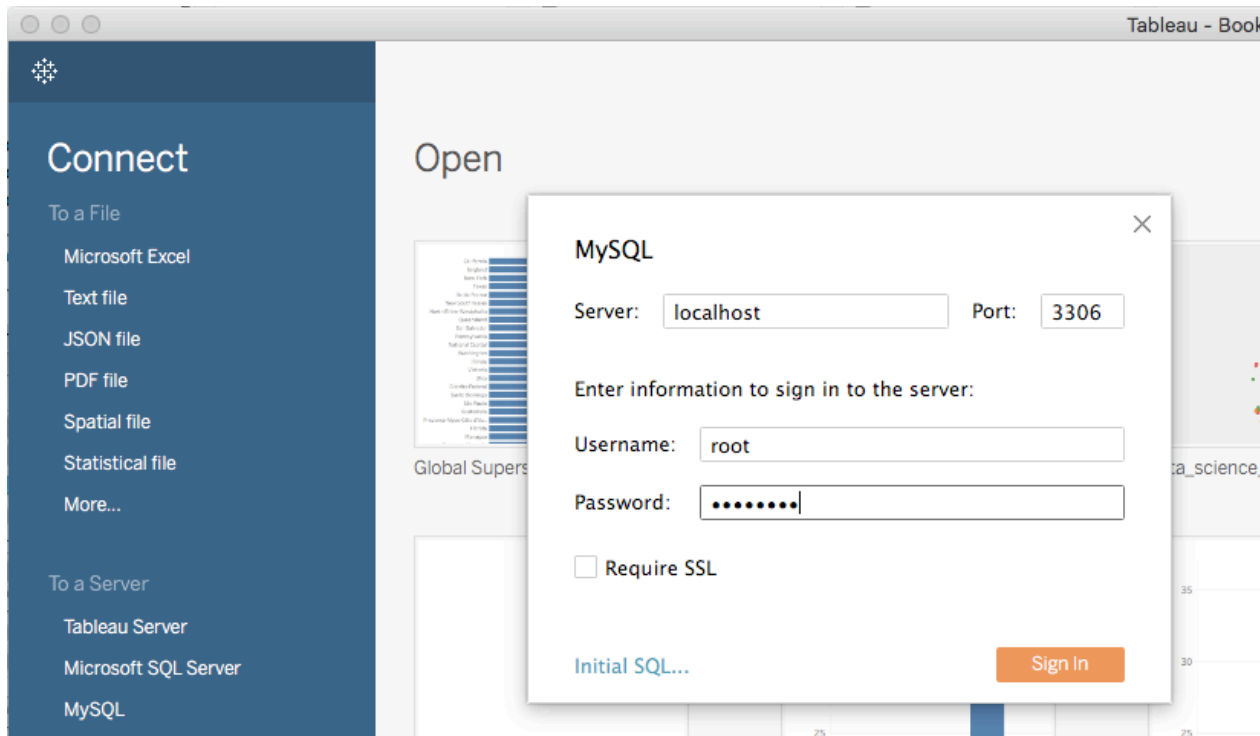
Sheets:

Display all the sheets in the Excel file, corresponding to the names of individual worksheet tabs. Sheets in Excel are treated the same as tables in a database.

The screenshot shows the Tableau 'Global Superstore' interface. On the left, the 'Connections' pane lists 'Global Superstore' (Microsoft Excel) and 'Sheets' (Orders, People, Returns, New Union). The 'Orders' sheet is selected. The main view displays a table of order data with columns: # Orders, Row ID, Order ID, Order Date, Ship Date, Ship Mode, Customer ID, Customer Name, Segment, and City. The table contains 10 rows of data. The bottom of the interface shows a navigation bar with 'Data Source', 'Bar Chart', 'Heat Maps', and a user profile 'Luis Yang'.

# Orders	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City
	32298	CA-2012-124891	7/31/2012	7/31/2012	Same Day	RH-19495	Rick Hansen	Consumer	New
	26341	IN-2013-77878	2/5/2013	2/7/2013	Second Class	JR-16210	Justin Ritter	Corporate	Woll
	25330	IN-2013-71249	10/17/2013	10/18/2013	First Class	CR-12730	Craig Reiter	Consumer	Brist
	13524	ES-2013-1579342	1/28/2013	1/30/2013	First Class	KM-16375	Katherine Murray	Home Office	Berli
	47221	SG-2013-4320	11/5/2013	11/6/2013	Same Day	RH-9495	Rick Hansen	Consumer	Daki
	22732	IN-2013-42360	6/28/2013	7/1/2013	Second Class	JM-15655	Jim Mitchum	Corporate	Sydr
	30570	IN-2011-81826	11/7/2011	11/9/2011	First Class	TS-21340	Toby Swindell	Consumer	Porii
	31192	IN-2012-86369	4/14/2012	4/18/2012	Standard Class	MB-18085	Mick Brown	Consumer	Ham
	40155	CA-2014-135909	10/14/2014	10/21/2014	Standard Class	JW-15220	Jane Waco	Corporate	Sacr
	40936	CA-2012-116638	1/28/2012	1/31/2012	Second Class	JH-15985	Joseph Holt	Consumer	Conc
	34577	CA-2011-102988	4/5/2011	4/9/2011	Second Class	GM-14695	Greg Maxwell	Corporate	Alex
	28879	ID-2012-28402	4/19/2012	4/22/2012	First Class	AI-10780	Anthony Jacobs	Corporate	Kahi

Connecting to database



- Let's connect to a MySQL database on our pc
- Remember to install the drivers (iODBC Driver Manager, MySQL ODBC connector)

Connecting to database

The screenshot shows the Tableau Desktop interface. On the left, the 'Connections' pane shows 'localhost MySQL' selected. Below it, the 'Database' dropdown is set to 'bike_station'. The 'Table' pane lists 'Stations', 'Status', 'Trips', and 'Weather'. The main view shows a data table with columns: Station Id, Name, Lat, Long, Dock Count, Landmark, Trip Id, and Duration. The table is sorted by 'Data source order'.

Tableau - Book1

Connections [Add](#)

localhost
MySQL

Database
bike_station

Table [+](#)

- Stations
- Status
- Trips
- Weather
- New Custom SQL
- New Union

Trips+ (bike_station)

Trips Stations

Sort fields Data source order

# Stations Station Id	Abc Stations Name	Stations Lat	Stations Long	# Stations Dock Count	Abc Stations Landmark	# Trips Trip Id	# Trips Duration
2	San Jose Diridon Calt...	37.3297	-121.902	27	San Jose	1049879	7,1
2	San Jose Diridon Calt...	37.3297	-121.902	27	San Jose	1049880	7,1
2	San Jose Diridon Calt...	37.3297	-121.902	27	San Jose	1050771	1
2	San Jose Diridon Calt...	37.3297	-121.902	27	San Jose	1050776	3
2	San Jose Diridon Calt...	37.3297	-121.902	27	San Jose	1051003	2
2	San Jose Diridon Calt...	37.3297	-121.902	27	San Jose	1051046	1
2	San Jose Diridon Calt...	37.3297	-121.902	27	San Jose	1051222	4

- Select database and tables in the database for visualization



Live vs. Extract

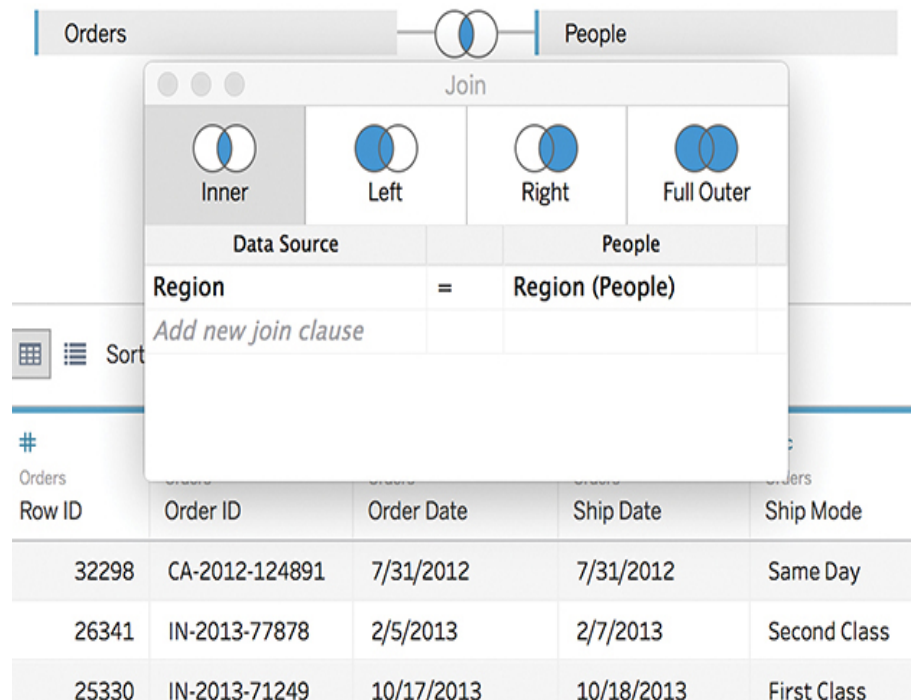
Connection

☒ Live

☐ Extract

Connection	Pros	Cons
Live	<ul style="list-style-type: none">Leverage a high-performance database's capabilitiesSee real-time changes in data	<ul style="list-style-type: none">Can result in a slower experienceSome cloud-based data sources must be extracted
Extract	<ul style="list-style-type: none">Can deter latency in a slow databaseCould reduce query load on critical systems	<ul style="list-style-type: none">Most Online Analytical Processing (OLAP) data sources cannot be extracted

Connecting to multiple tables with joins



The screenshot shows the Tableau interface with a 'Join' dialog box open. The dialog box has four tabs: 'Inner', 'Left', 'Right', and 'Full Outer'. The 'Inner' tab is selected. Below the tabs, there are two columns: 'Data Source' and 'People'. The 'Data Source' column contains the field 'Region'. The 'People' column contains the field 'Region (People)'. An equals sign '=' is placed between the two fields. Below the fields, there is a button labeled 'Add new join clause'. In the background, a table of data is visible, showing columns for 'Row ID', 'Order ID', 'Order Date', 'Ship Date', and 'Ship Mode'.

Row ID	Order ID	Order Date	Ship Date	Ship Mode
32298	CA-2012-124891	7/31/2012	7/31/2012	Same Day
26341	IN-2013-77878	2/5/2013	2/7/2013	Second Class
25330	IN-2013-71249	10/17/2013	10/18/2013	First Class

Tableau has automatically joined these tables by recognizing that Region is a common field between the two.



Overview of join types (1/2)

Inner join: Joins records where there is a matching field in both datasets. Using an inner join to combine tables produces a new virtual table that contains values that have matches in both tables.

Outer join: Joins all the records from each dataset together, even when there is no join—and rarely used. Using a full outer join to combine tables produces a table that contains all values from both tables. If a value from either table doesn't have a match with the other table, you see a null value.

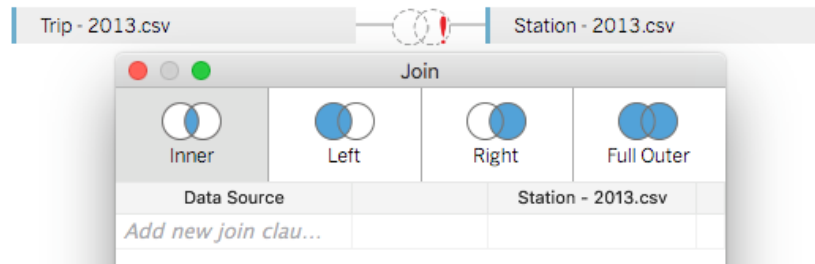


Overview of join types (2/2)


Left join: Joins records from the left and right sides of your equation when there is a match. Using a left join to combine tables produces a new virtual table that contains all values from the left table and corresponding matches from the right table. When there is no corresponding match from left to right, you will see a null value.

Right join: Joins all the records from the data on the right side of your equation and any matching records from the left side. Opposite of a left join, using a right join to combine tables produces a table that contains all values from the right and corresponding matches from the left. Likewise, when a value in the right table doesn't have a corresponding match in the left table, you see a null value.

Connecting to multiple tables with joins



- Tableau will do its best to automatically determine the best join.
- A joint error occur when Tableau can't find Tableau can't find a common column in the two tables -> **we need to specify the columns for joining the two tables**
- However, if you're unsure which type of joins your data supports, you can check the join dialog after you've connected your data. Additionally, you can adjust the join type by selecting a different join type in the Join dialog.



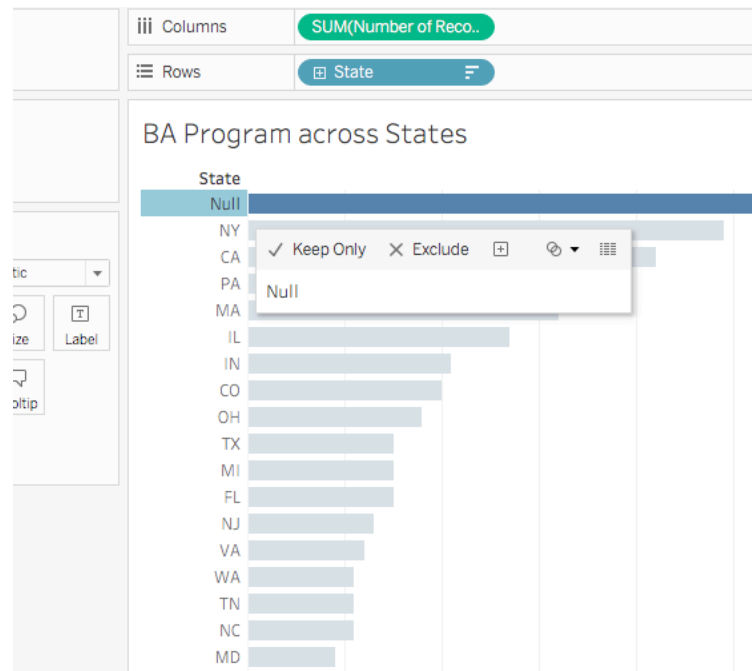
What are "NULLS"?

Country	State	City	Department	URL
DE	<i>null</i>	<i>null</i>	Statistics	http://www.statistik...
DE	<i>null</i>	<i>null</i>	<i>null</i>	http://www.informat...
CA	<i>null</i>	<i>null</i>	Business	http://business.quee...
CA	<i>null</i>	Burnaby	School of Computing ...	http://www.sfu.ca/co...
CA	<i>null</i>	<i>null</i>	<i>null</i>	https://www.cs.ualb...
CA	<i>null</i>	<i>null</i>	Mathematics and Sta...	http://www.ufv.ca/m...
CA	<i>null</i>	<i>null</i>	<i>null</i>	http://www.schulich...
BR	<i>null</i>	<i>null</i>	<i>null</i>	http://www.mackenz...
AU	<i>null</i>	<i>null</i>	<i>null</i>	http://www.deakin.e...
...

- *Null* means that some **empty cells** are in your data and Tableau is letting you know.
- **Checking fields and formatting for extraneous information is always important** when doing data analysis because you want to ensure these blank fields do not skew out results.
- A null field might indicate an **error** in the data, or some other **inaccuracy**.



What are "NULLS"?



- You'll want to exclude null fields if you don't want empty fields to show up in your data.



Basic data prep with data interpreter

Data pre-processing for visualization

The screenshot shows the Tableau Data Interpreter interface. On the left, the 'Connections' pane lists 'localhost MySQL', 'Global Superstore Microsoft Excel', and 'data_science_colleges Text file'. The 'Files' pane shows a checkbox for 'Cleaned with Data Interpreter' which is checked, with a link to 'Review the results. (To undo changes, clear the check box.)' and the file 'data_science_colleges.csv'. Below this is a 'New Union' button. On the right, a preview of the data is shown in a table format. The table has four columns: 'School', 'Program', 'Degree', and 'Online'. The data rows include Auburn University, The University of Alabama, and Arkansas Tech University.

School	Program	Degree	Online
Auburn University	Online Master of Bus...	M	False
Auburn University	Data Science	B	False
The University of Ala...	Master of Science in ...	M	True
The University of Ala...	MS in Operations Ma...	M	False
The University of Ala...	M.S. degree in Applie...	M	False
The University of Ala...	MBA with concentrat...	M	True
Arkansas Tech Unive...	Business Data Analyt...	B	False

- Tableau's built-in tool for preparing data for analysis. When you connect to an Excel sheet in Tableau, the software can recognize issues such as **missing column names**, **null values**, and so on. To remedy these and clean the file for use in analysis.



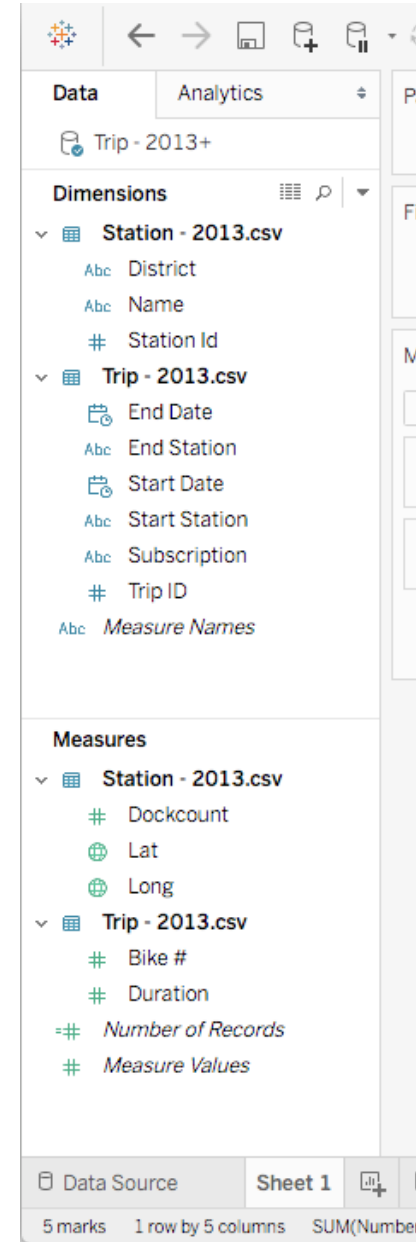
Dimensions vs measures

Dimensions

- Dimensions are things that you can group data by or drill down by.
- They are usually—but not always—categories (such as City, Product Name, or Color), and they can be grouped into strings, dates, or geographic fields.

Measures

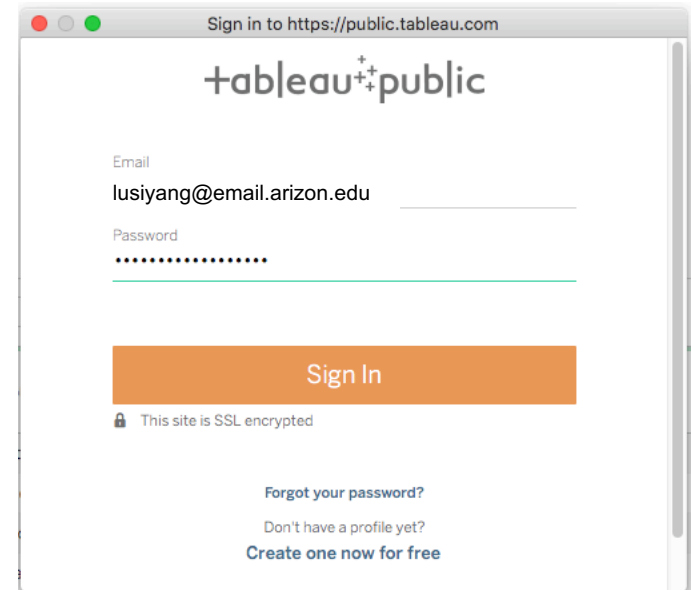
- Measures are generally numerical data on which you want to perform calculations—summing, averaging, and so on.





Task: save workbook to Tableau Public (1/2)

1. Create an account on Tableau public: <https://public.tableau.com/en-us/s/>
2. In Tableau Desktop, select **Server > Tableau Public > Save to Tableau Public**
3. Sign in using your Tableau Public account
4. Type a name for the workbook and click **Save**




Sign in to <https://public.tableau.com>

+tableau+public

Email
lusiyang@email.arizon.edu

Password
.....

Sign In

 This site is SSL encrypted

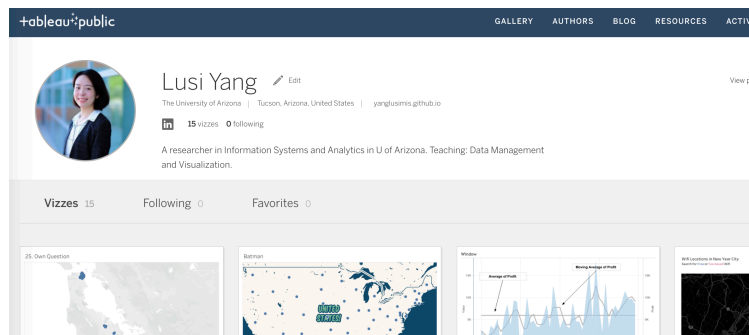
[Forgot your password?](#)

[Don't have a profile yet?](#)

[Create one now for free](#)

Task: save workbook to Tableau Public (2/2)

5. You will be able to view your visual on your profile page on Tableau Public. You can:
- download/delete/edit your visual
 - share your visual through the link provided
 - embed the visual on your own personal website



Start to build your own Tableau Public profile 😊