#6

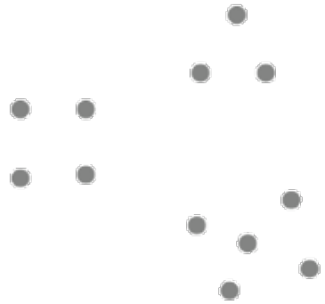
Decluttering
&
Getting Started with Tableau

MIS561 Data Visualization

Outline

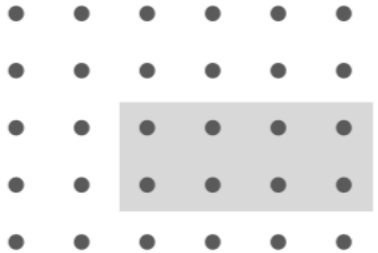
- Decluttering: step-by-step
- Tableau product portfolio
- Connecting to data
- Basic data prep
- Dimensions and measures

Gestalt principles of visual perception



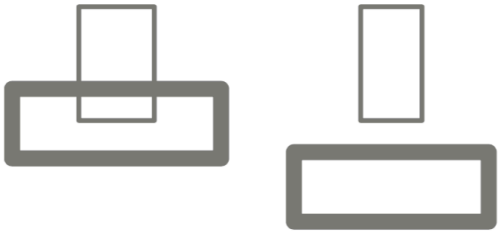
proximity

similarity



enclosure

closure

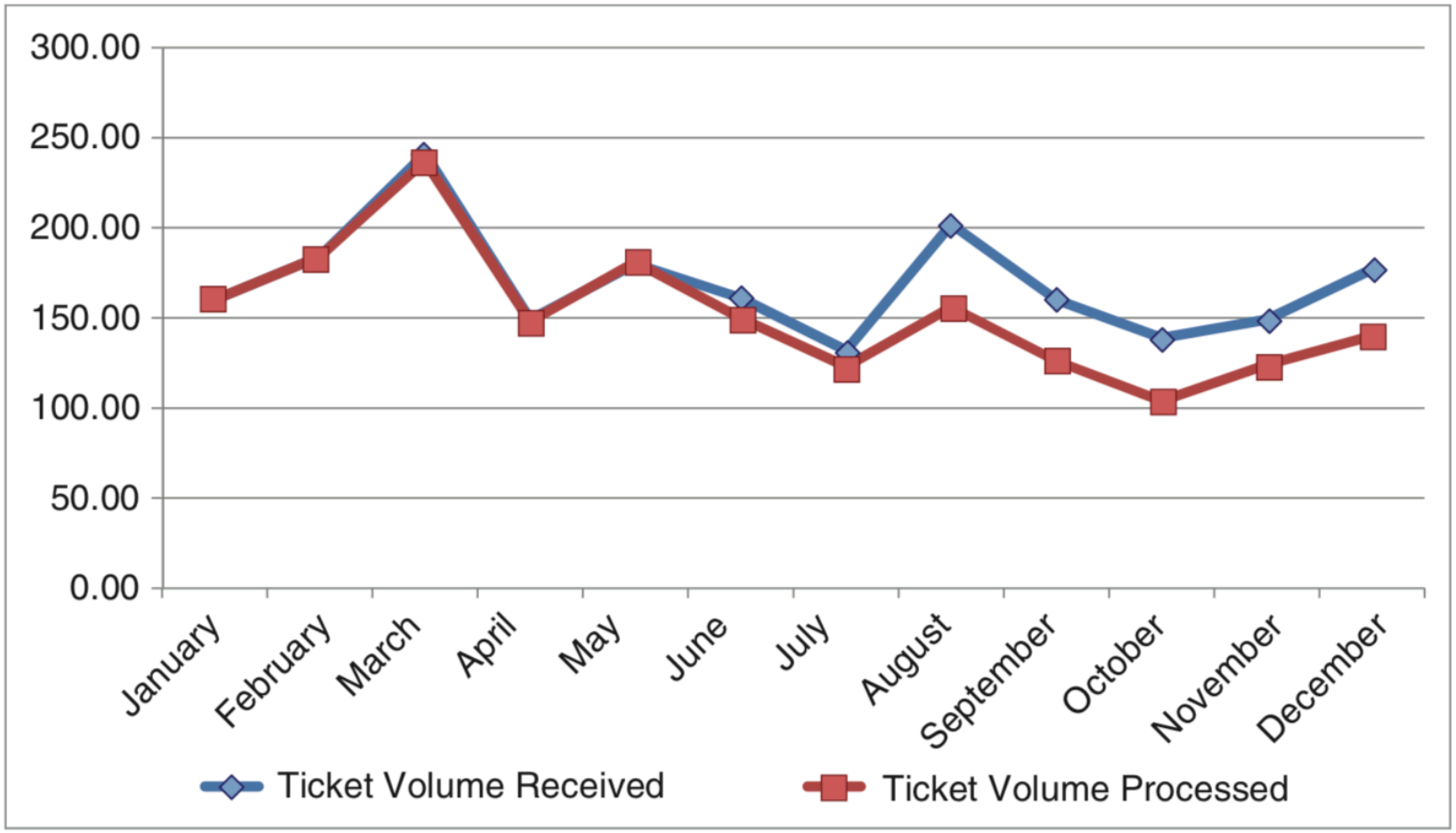


continuity

connection

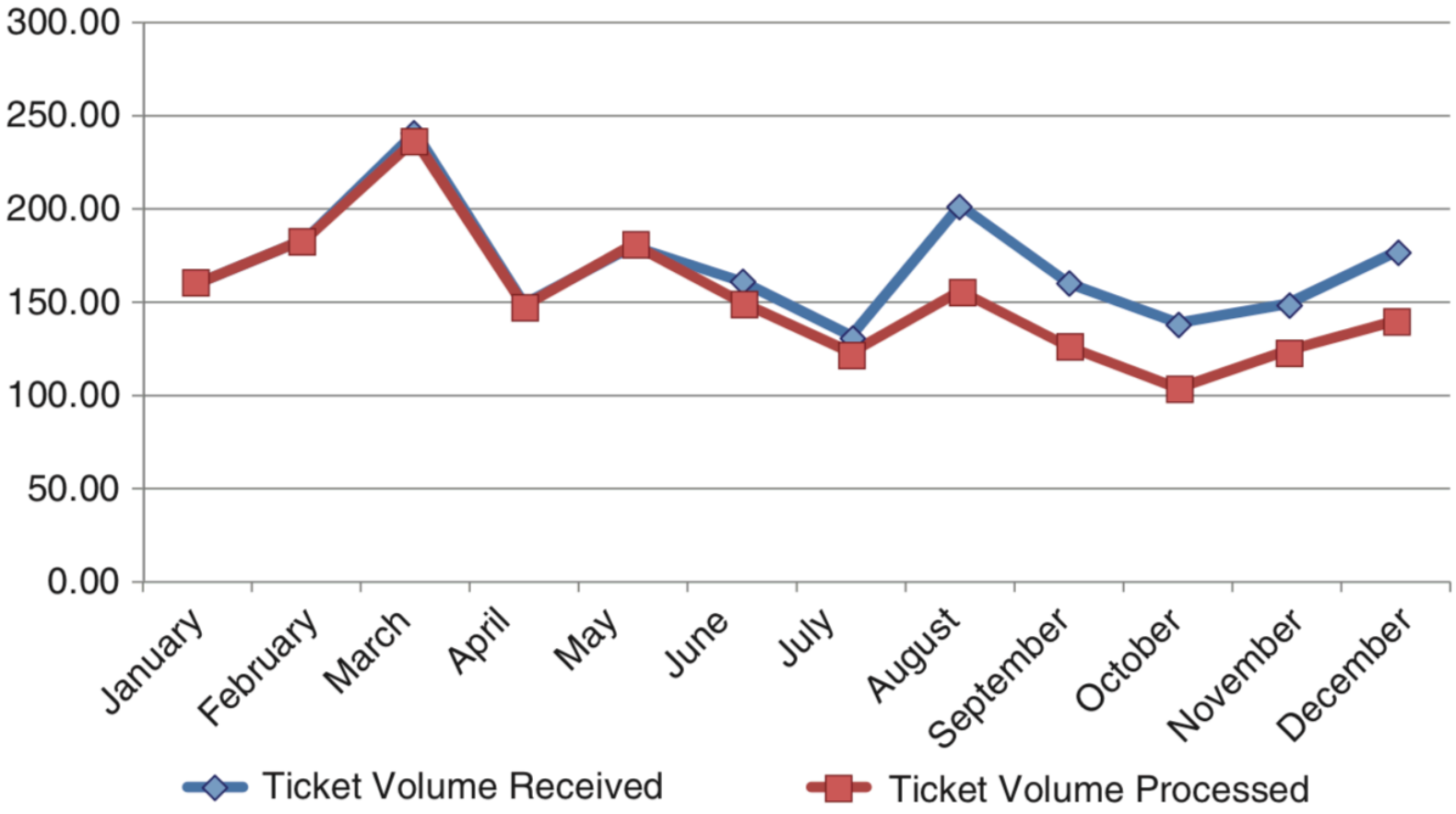
Original graph

- decluttering: step by step



Remove chart border

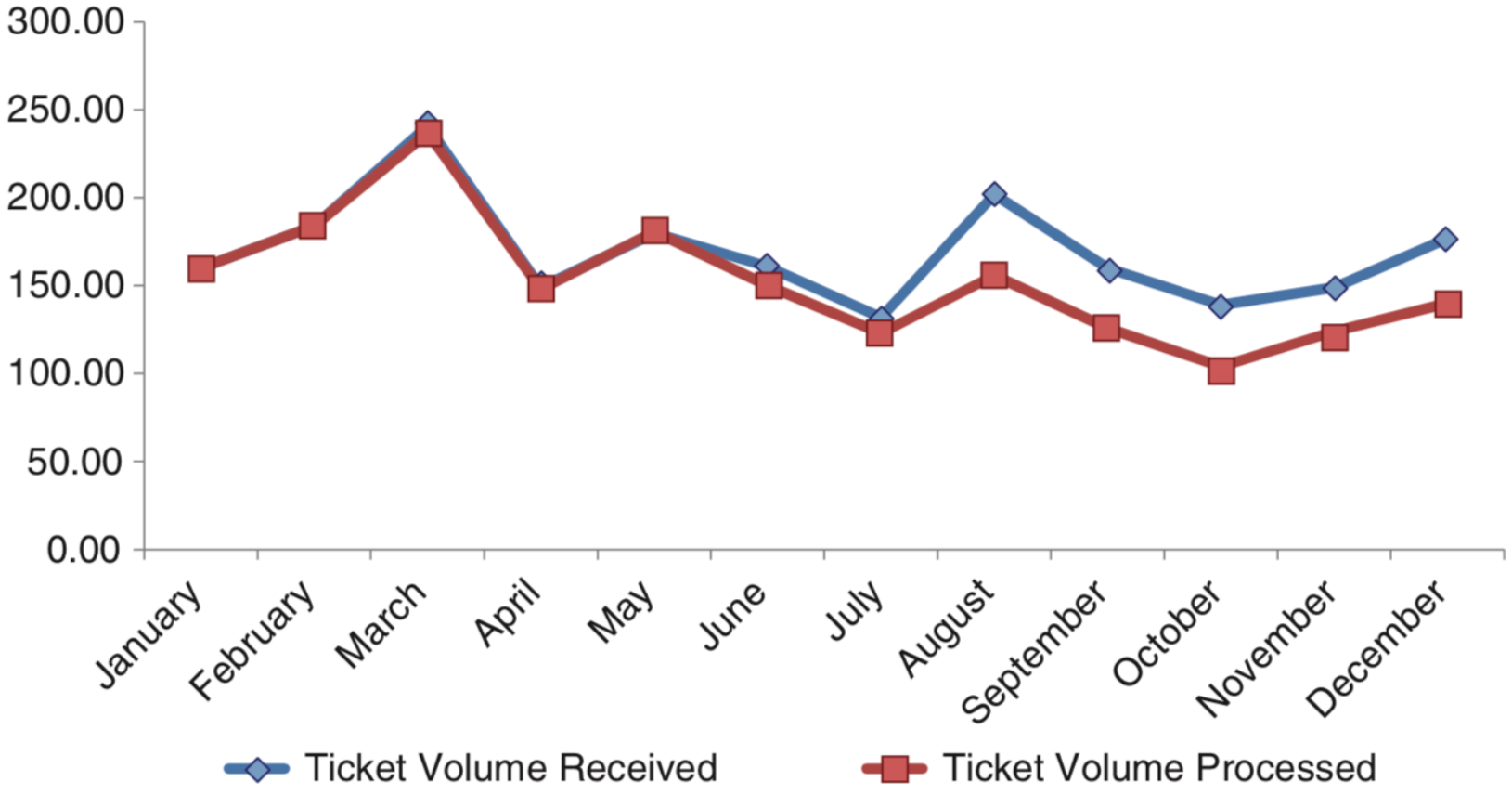
- decluttering: step by step



Gestalt principle of closure

Remove gridlines

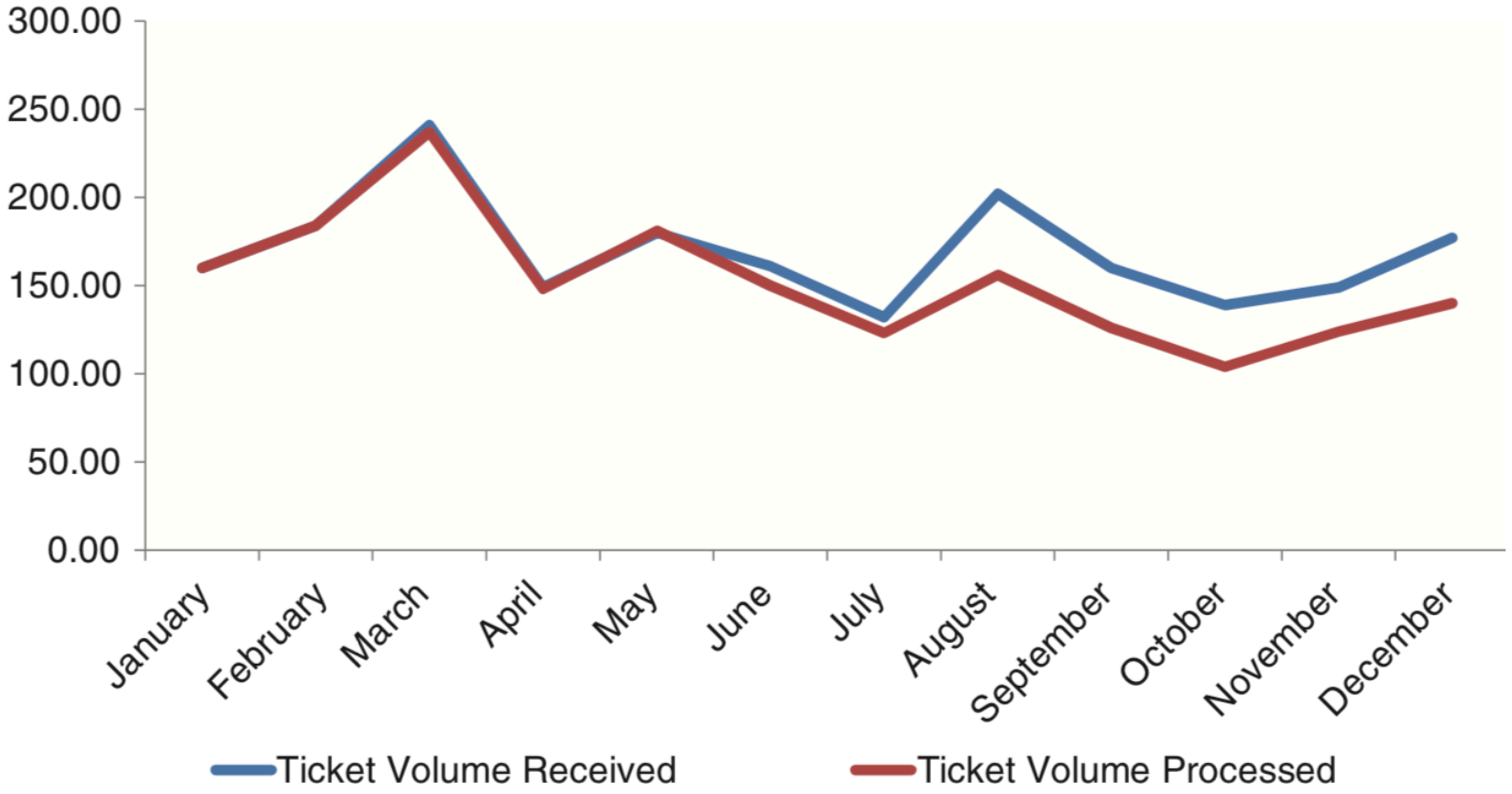
- decluttering: step by step



Don’t let them compete visually with your data

Remove data markers

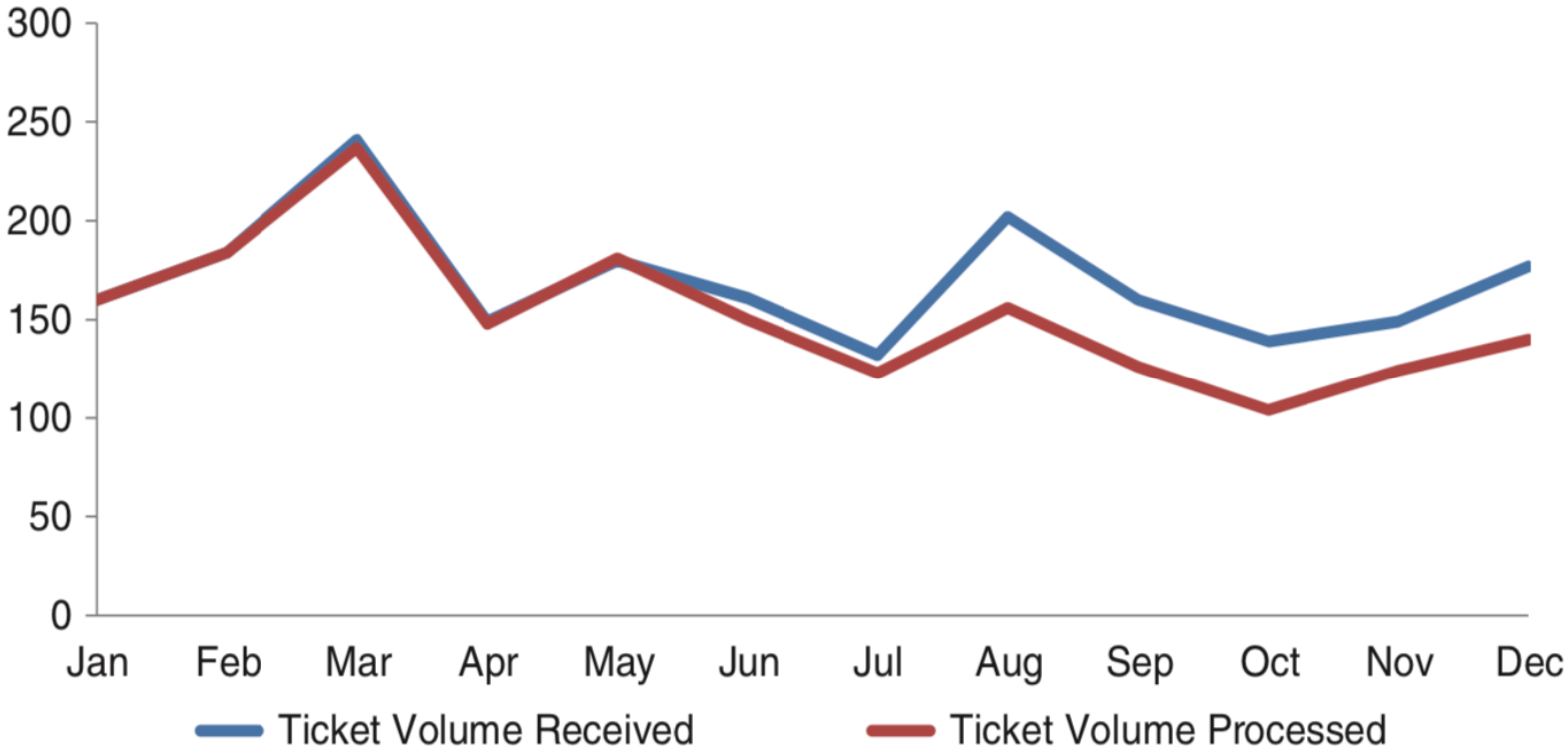
- decluttering: step by step



Every single element adds cognitive load on audience

Clean up axis labels

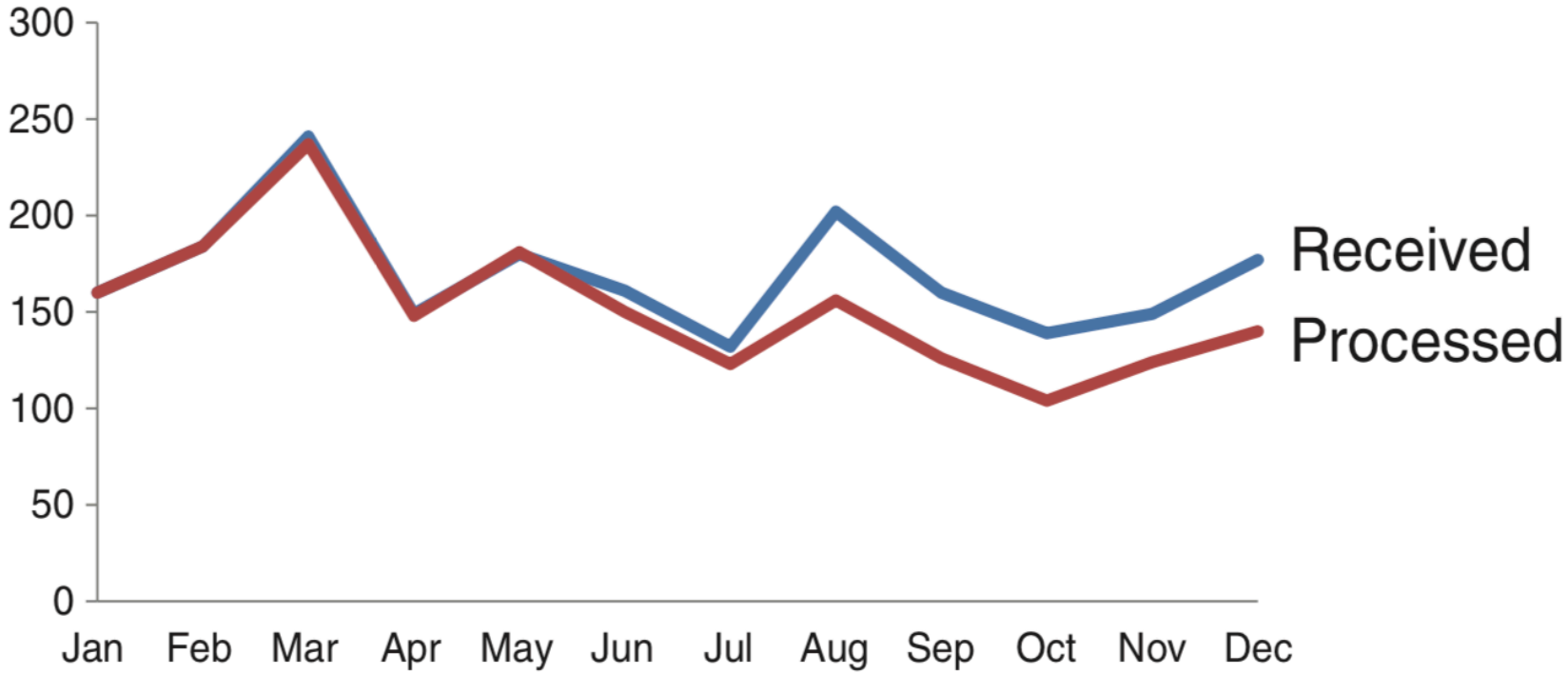
- decluttering: step by step



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

Label data directly

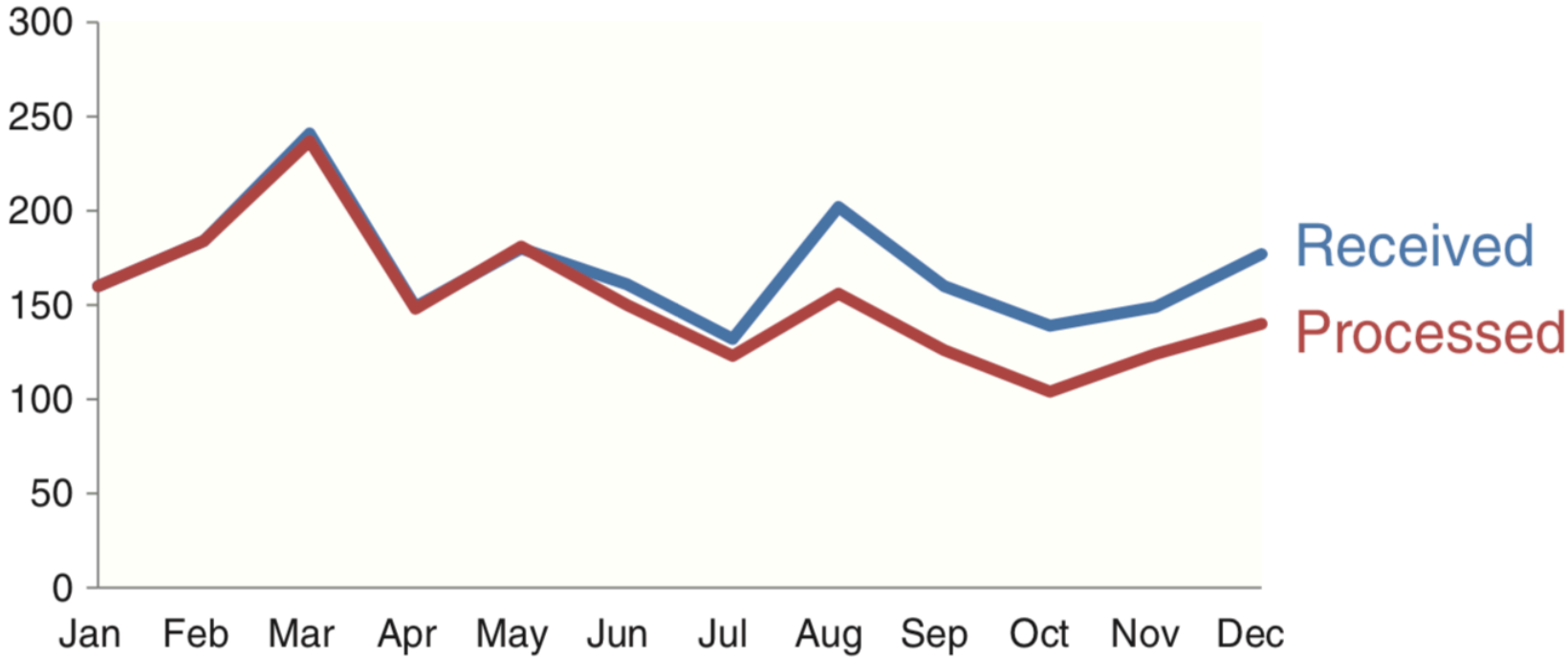
- decluttering: step by step



Gestalt principle of proximity

Leverage consistent color

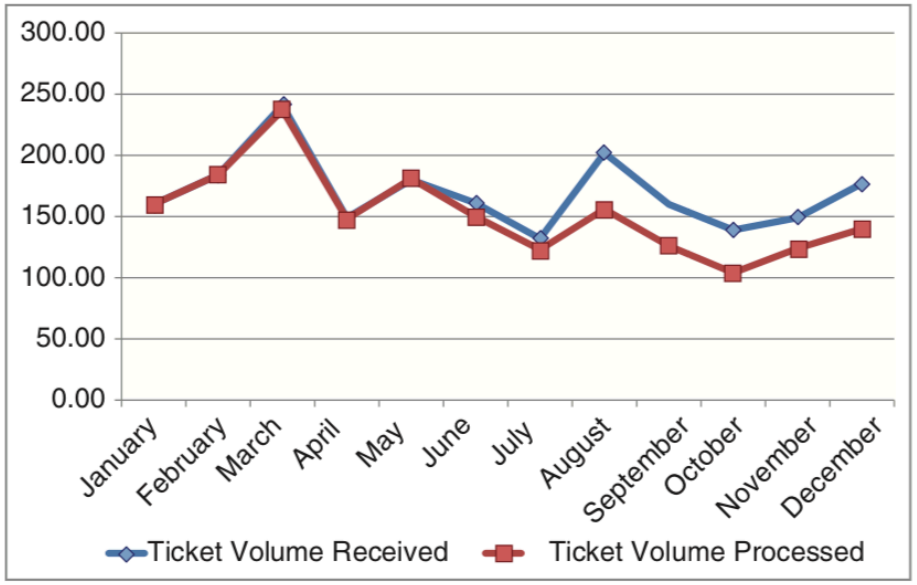
- decluttering: step by step



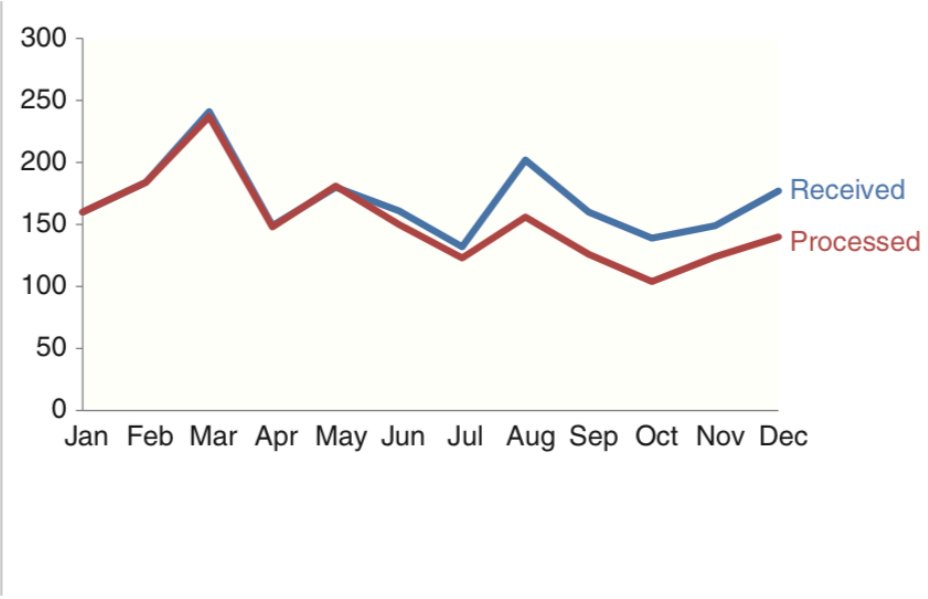
Gestalt principle of similarity

Before-and-after

- decluttering: step by step



before

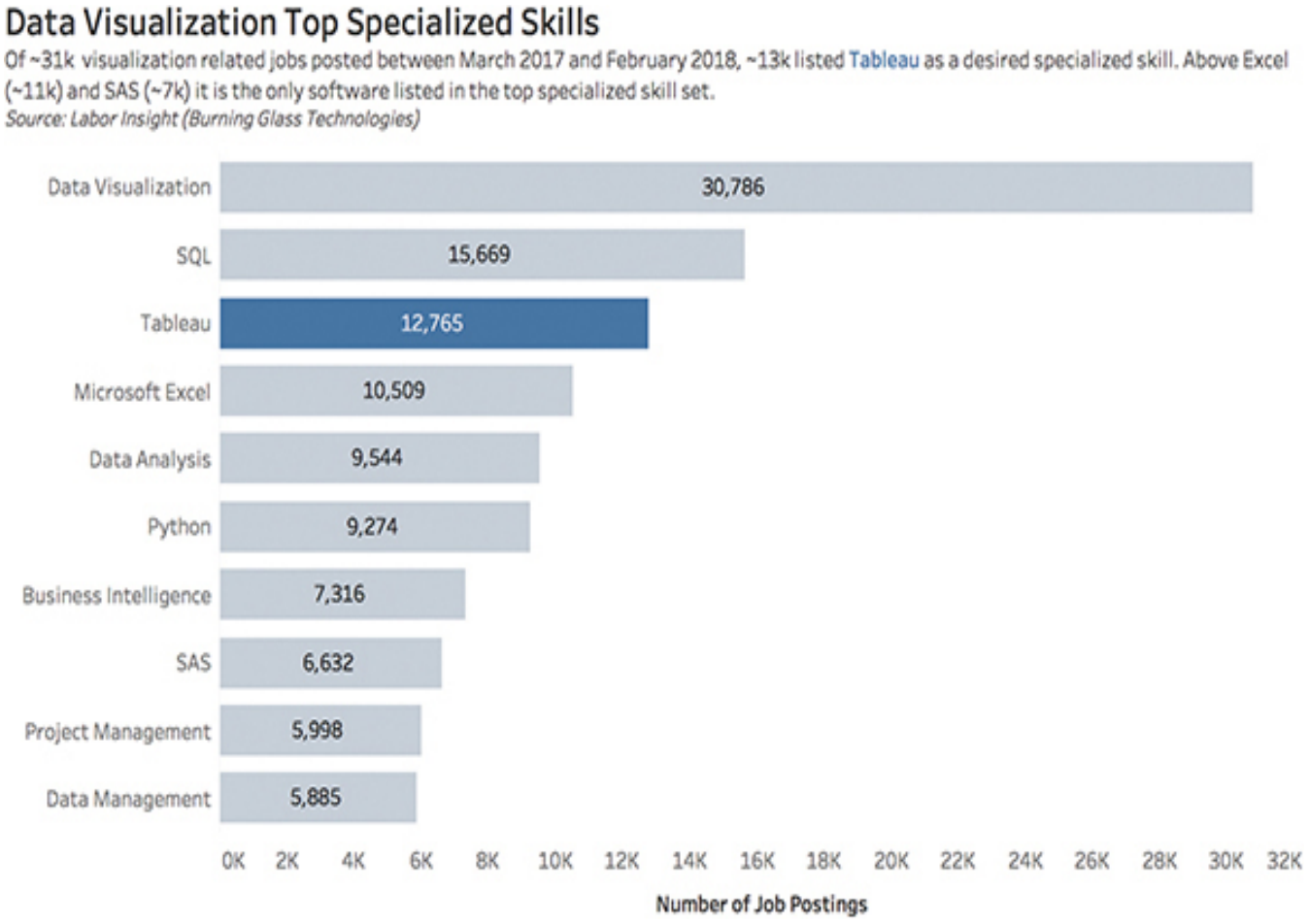


after

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)
Data Visualization
30,786
SQL
15,669
Tableau
12,765
Microsoft Excel
10,509
Data Analysis
9,544
Python
9,274
Business Intelligence
7,316
SAS
6,632
Project Management
5,998
Data Management
5,885
0K 2K 4K 6K 8K 10K 12K 14K 16K 18K 20K 22K 24K 26K 28K 30K 32K
Number of Job Postings

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

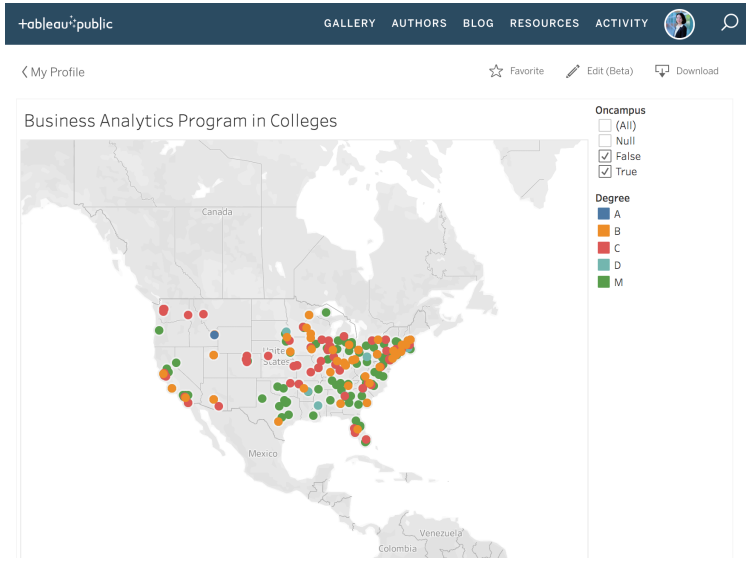
The Tableau product portfolio

Tableau Online

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

Tableau Public

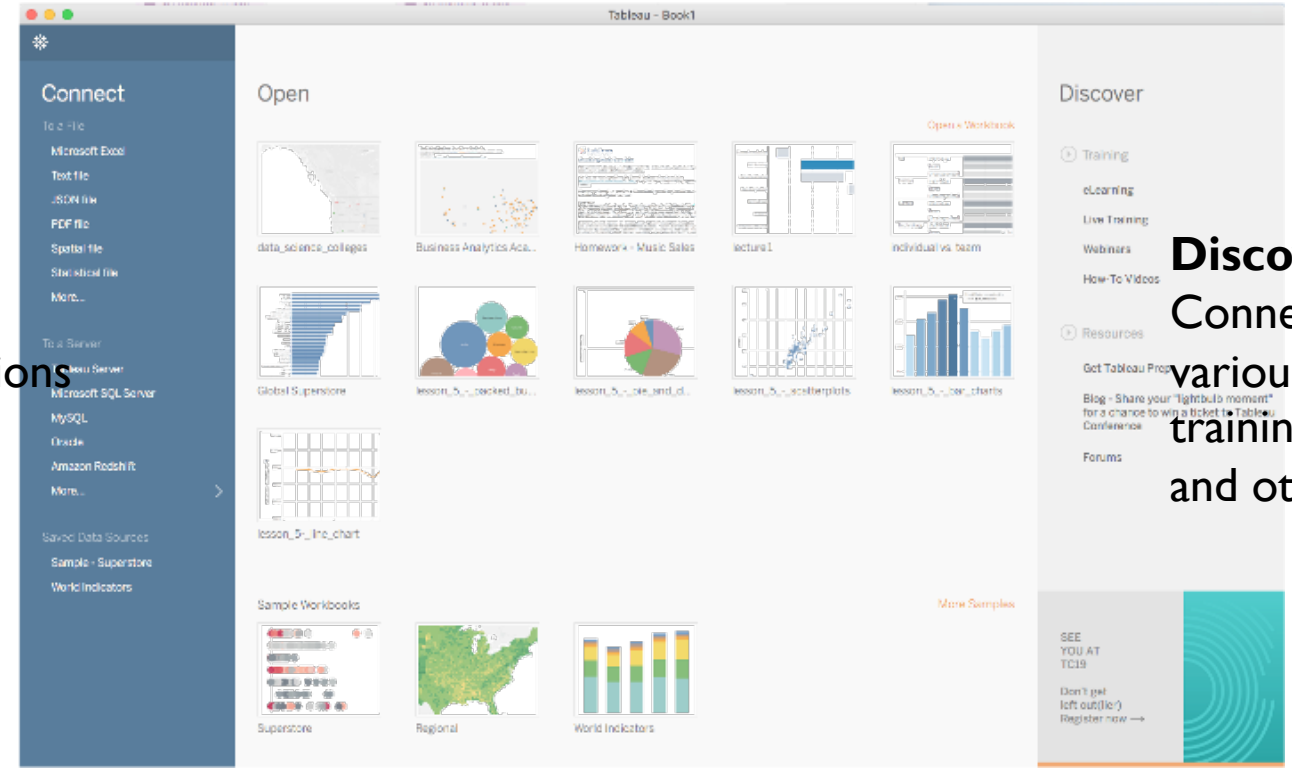
- One part data visualization hosting service, one part social networking
- A free service that allows users to publish interactive data visualizations online. These visualizations can be embedded into webpages and blogs, shared via social media or email, or made available for download to other users



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.



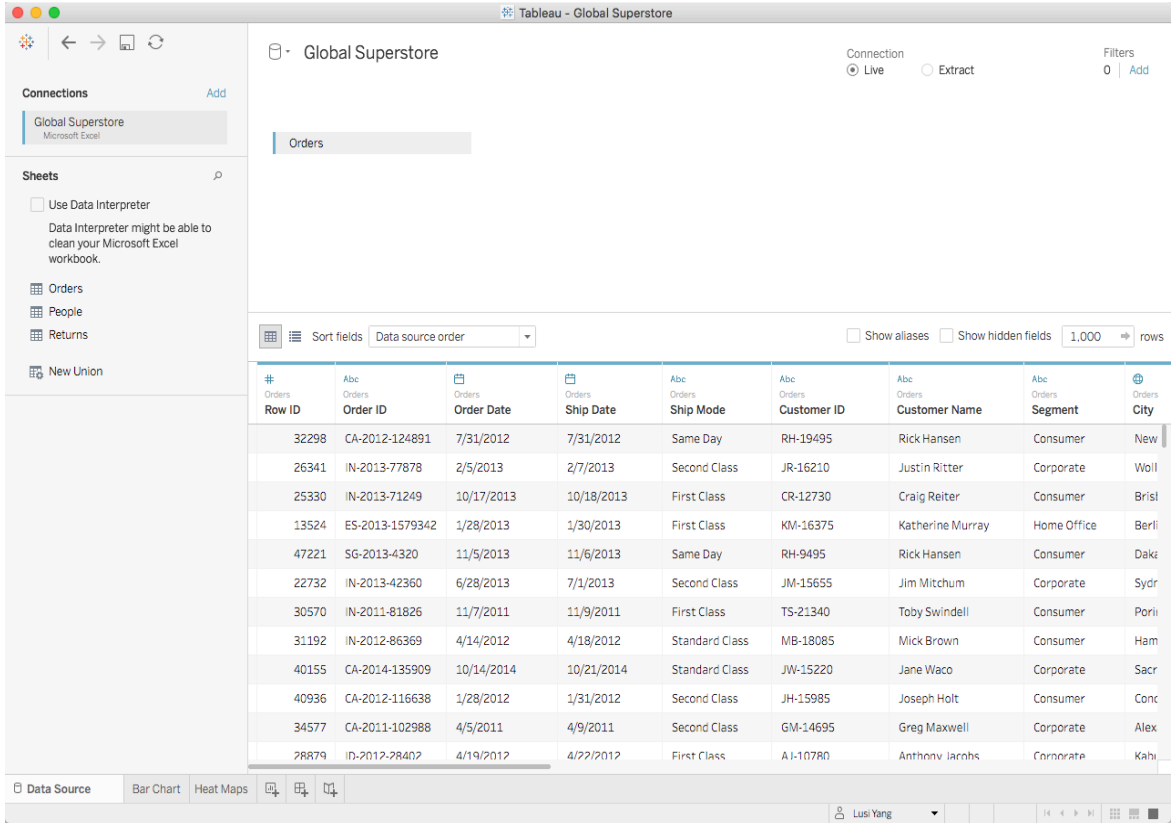
Discover - Connects you to various Tableau training, visualization, and other resources.

Sample workbooks - default samples provided by Tableau.

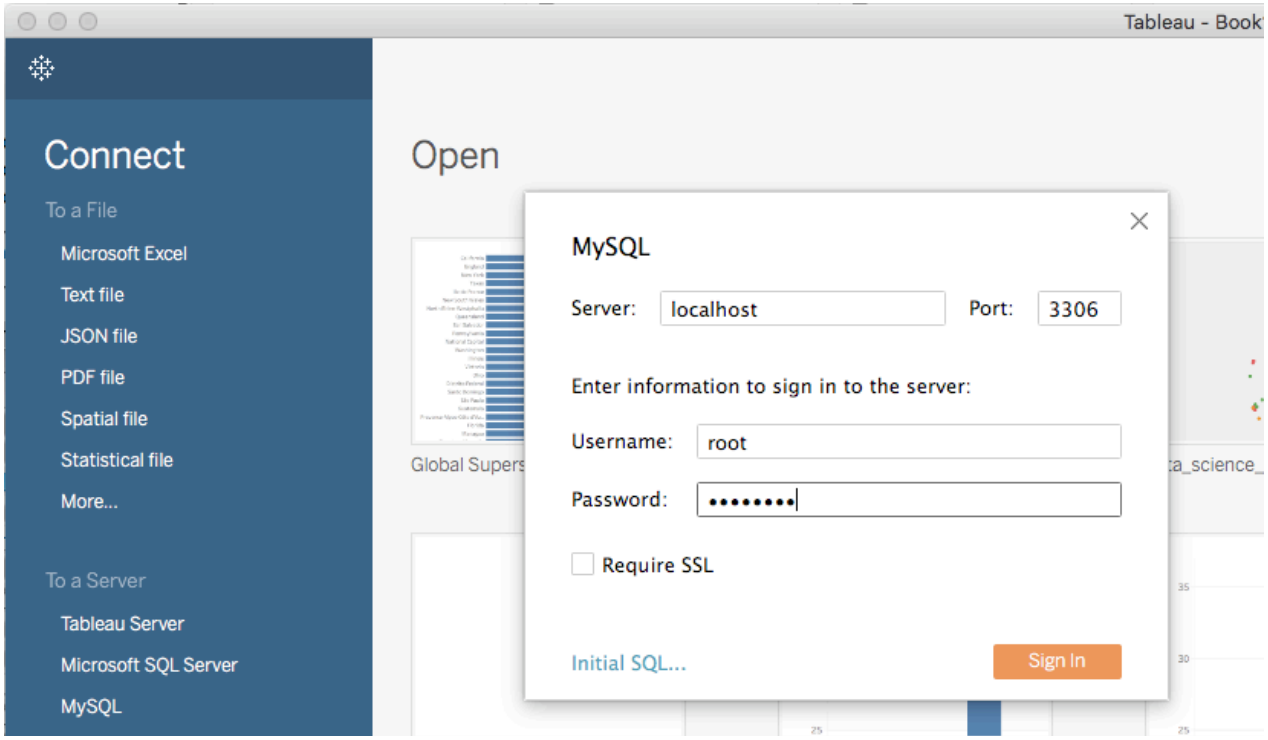
Connecting to tables

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.

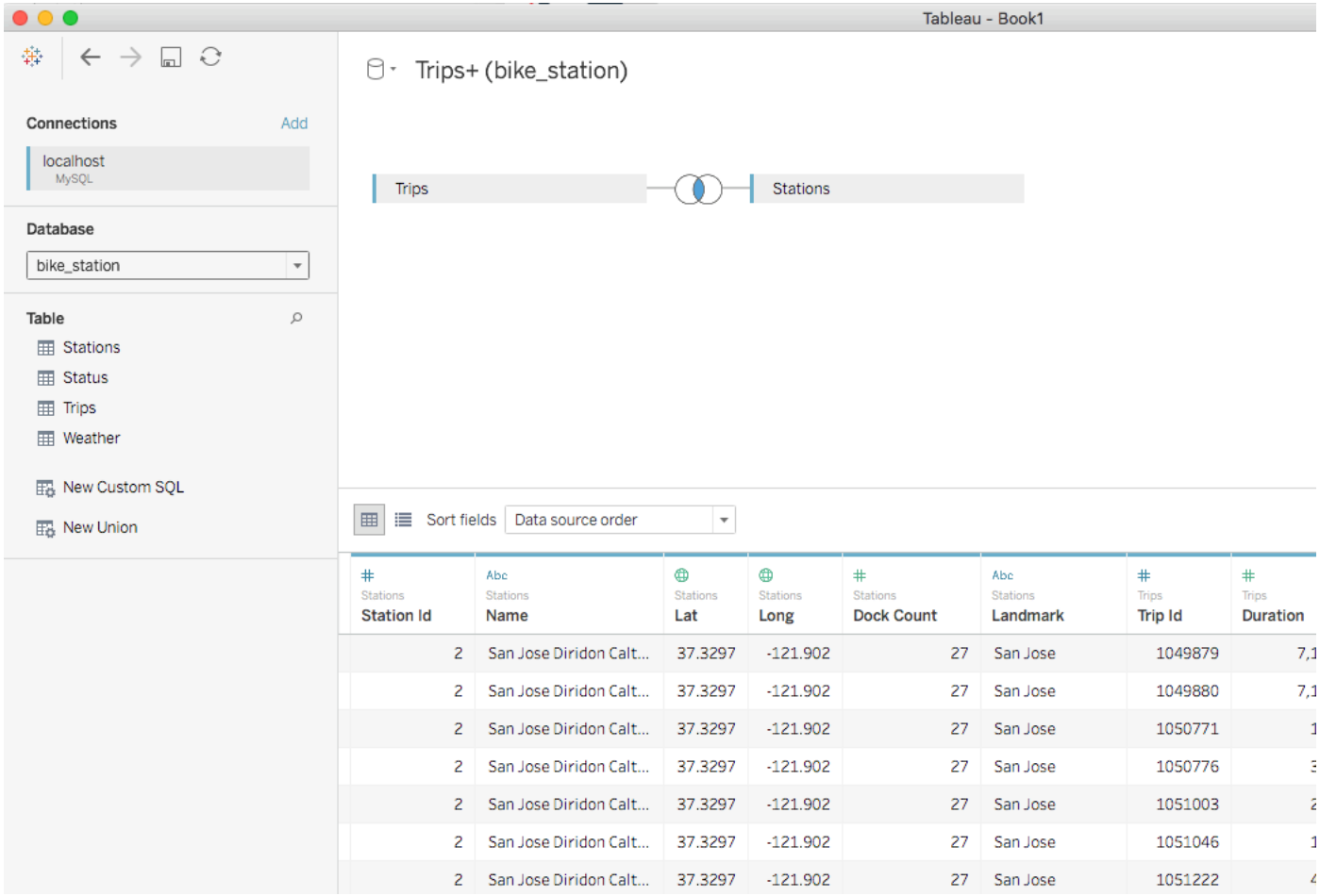


Connecting to database

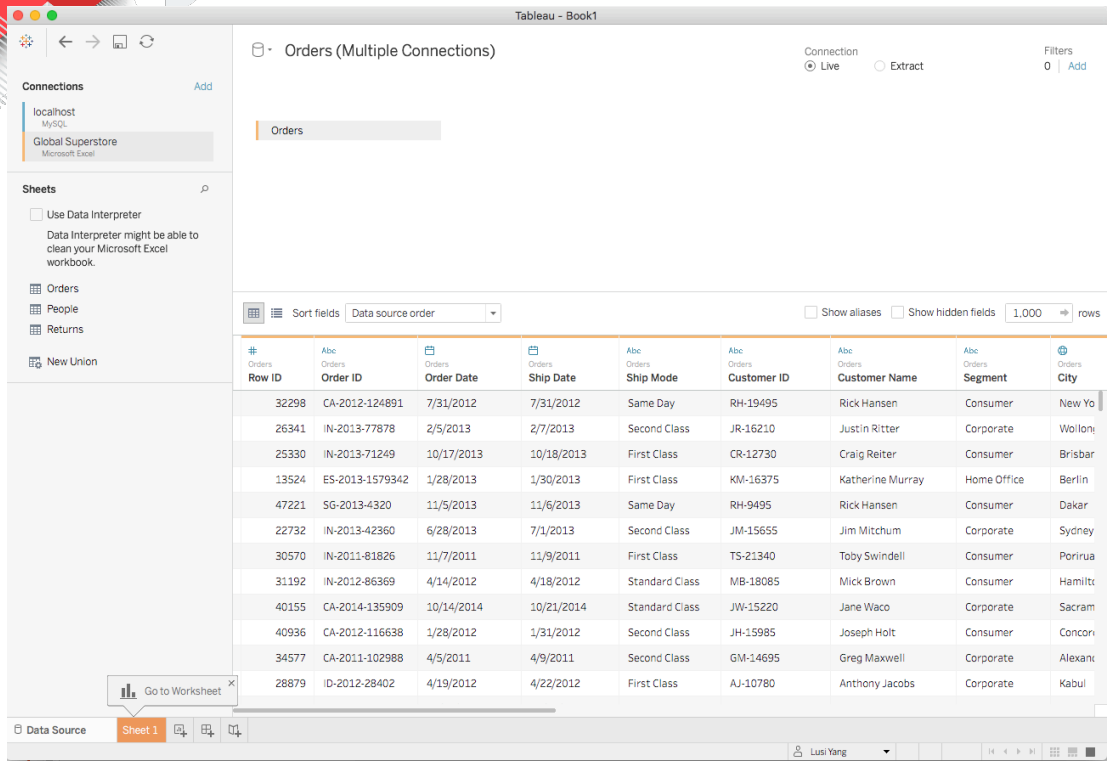


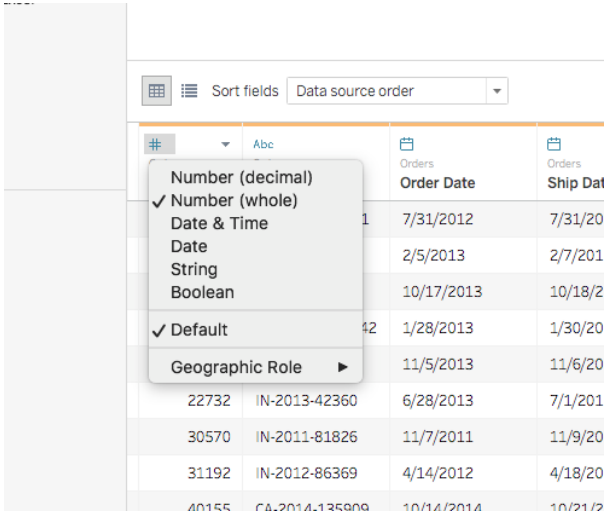
- Let's collect to a MySQL database on our pc
- Remember to install the drivers (iODBC Driver Manager, MySQL ODBC connector)

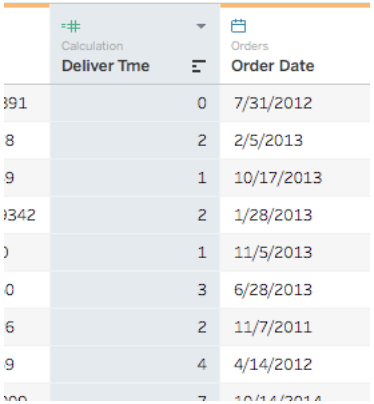
Connecting to database



- Select database and tables in the database for visualization







Populate data preview pane, and pre-process the data:

- Rename a column
- Change the default data type
- Adjust the default data sort order
- Create calculated fields
- Hide or show hidden fields

Live vs. Extract

Connection	Pros	Cons
Live	Leverage a high-performance database's capabilities See real-time changes in data	Can result in a slower experience Some cloud-based data sources must be extracted
Extract	Can deter latency in a slow database Could reduce query load on critical systems	Most Online Analytical Processing (OLAP) data sources cannot be extracted

Connecting to multiple tables with joins

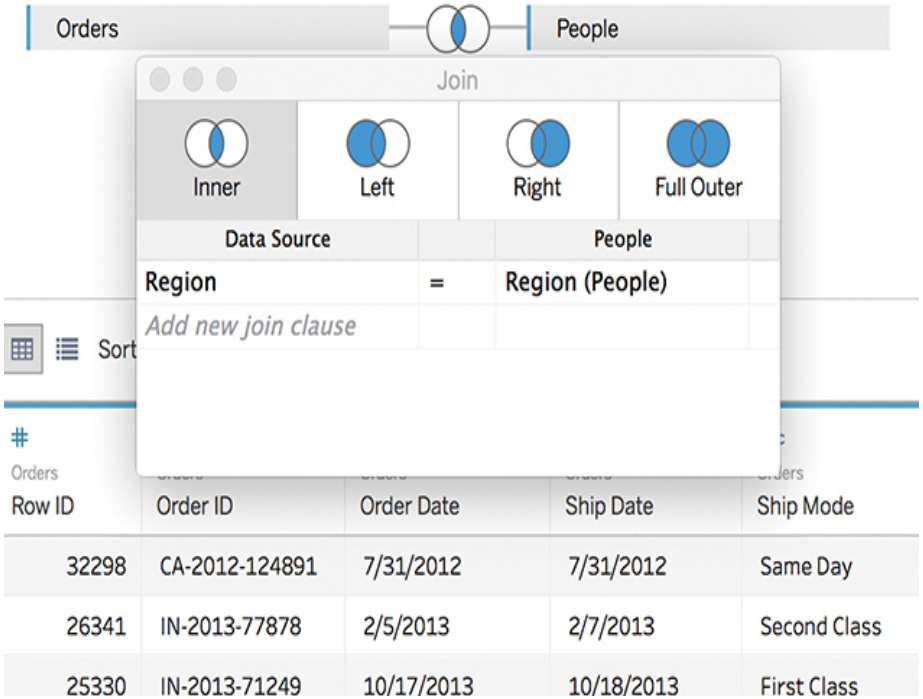


Tableau has automatically joined these tables by recognizing that Region is a common field between the two.

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.

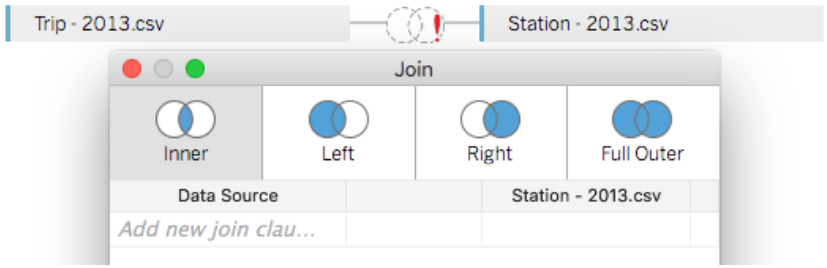
Outer join: Joins all the records from each dataset together, even when there is no join—and rarely used. Using a full outer join to combine tables produces a table that contains all values from both tables. If a value from either table doesn't have a match with the other table, you see a null value.

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



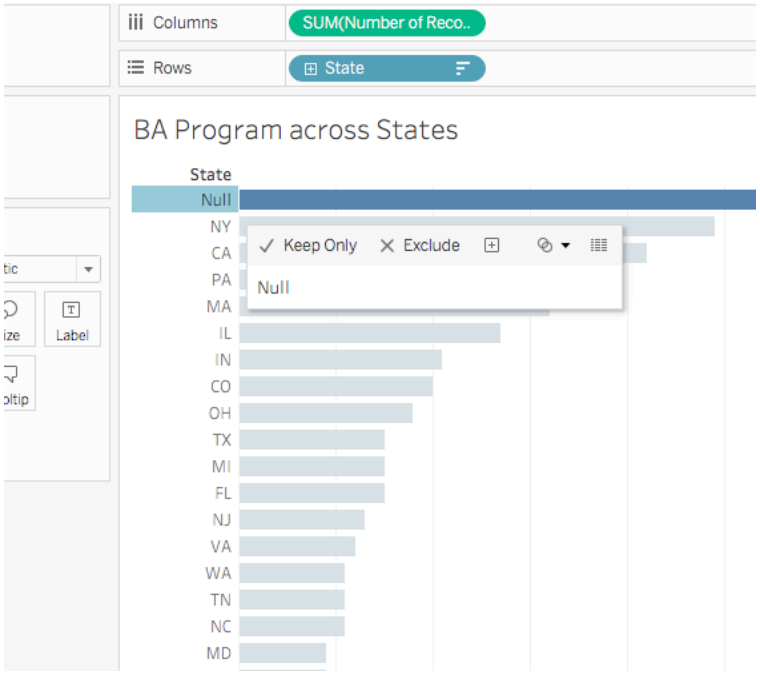
- Tableau will do its best to automatically determine the best join.
- A joint error occur when Tableau can't find Tableau can't find a common column in the two tables -> we need to specify the columns for joining the two tables
- However, if you're unsure which type of joins your data supports, you 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	*null*	*null*	Statistics	http://www.statistik...
DE	*null*	*null*	*null*	http://www.informat...
CA	*null*	*null*	Business	http://business.quee...
CA	*null*	Burnaby	School of Computing ...	http://www.sfu.ca/co...
CA	*null*	*null*	*null*	https://www.cs.ualb...
CA	*null*	*null*	Mathematics and Sta...	http://www.ufv.ca/m...
CA	*null*	*null*	*null*	http://www.schulich....
BR	*null*	*null*	*null*	http://www.mackenz...
AU	*null*	*null*	*null*	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.

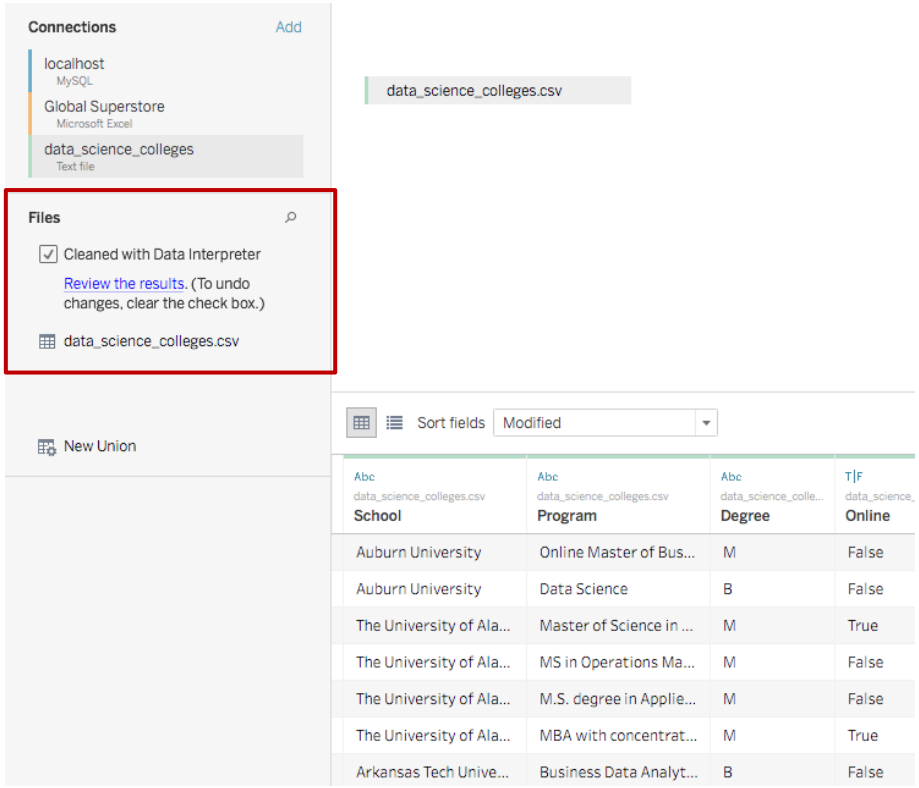
What are ”NULLS”?



- 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



- 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.

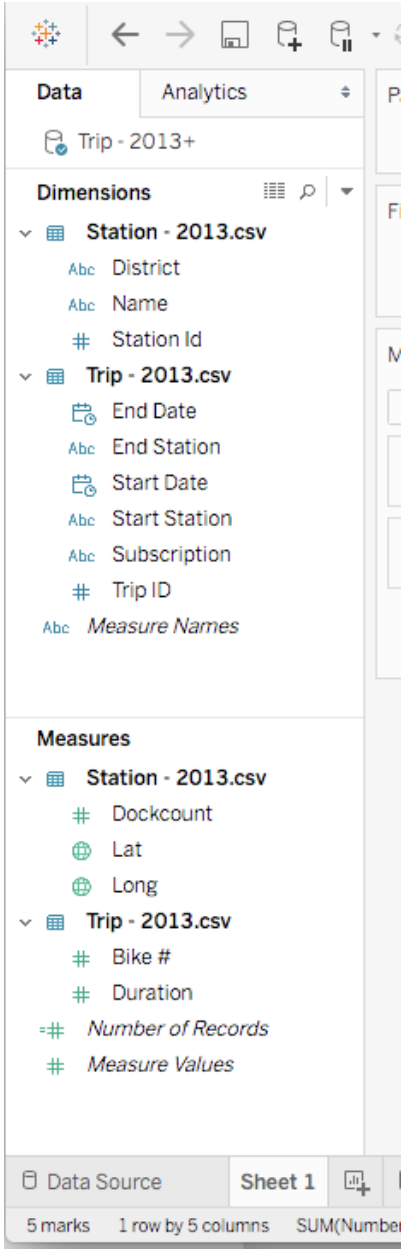
Dimensions vs measures

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.

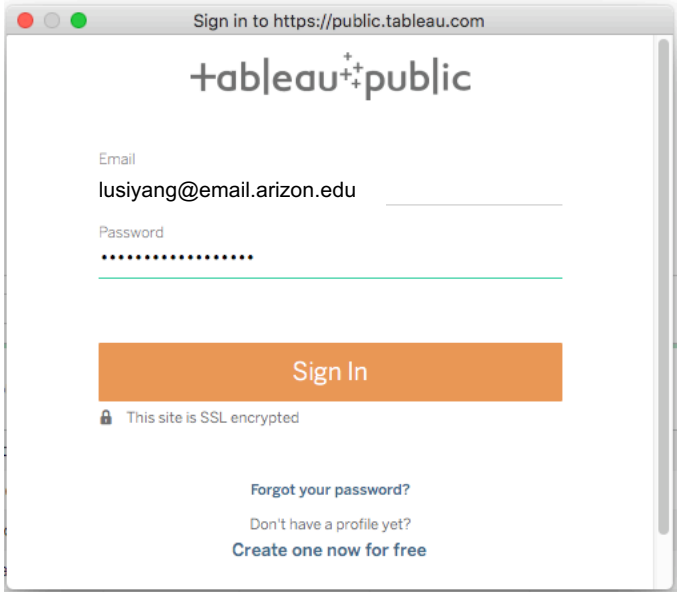
Measures

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



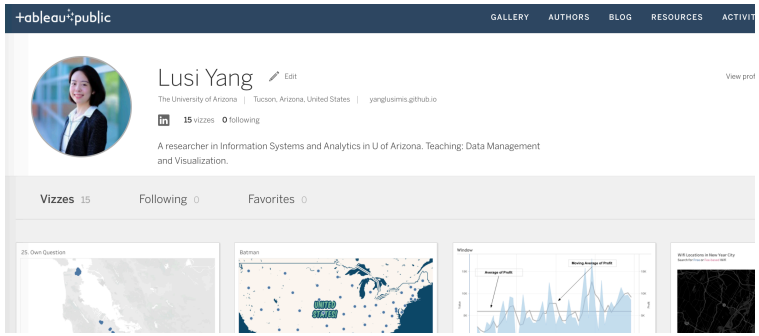
Task: save workbook to Tableau Public (1/2)

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**



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 ☺