

Data 301 Data Analytics Data Visualization

Dr. Irene Vrbik

University of British Columbia Okanagan
irene.vrbik@ubc.ca

Term 1, Summer 2019

Why learn Visualization?

- ▶ Visualization allows people to understand and extract information faster and with more accuracy than displaying text and numbers.
- ▶ A good visualization makes data more understandable and reachable to more people.
- ▶ High quality visualization encourages confidence in the data analysis and inspires people to utilize the data more effectively.

What is Data Visualization? What is Tableau?

Data visualization is the creation and presentation of visual representations of data with the goal to communicate information clearly and efficiently.

- ▶ Data visualizations include graphs, charts, images, plots, and tables.
- ▶ Data visualization is both an art and a science as it relies on both scientific data analysis and techniques as well as artistic creativity and presentation.

Tableau is a software package designed to make data visualization easy for non-expert users.

Data Visualization with Previous Tools

- ▶ We have seen data visualization in a variety of other tools including Excel, Python, and R.
- ▶ We saw how we could call to different packages to enhance visualization capabilities and improve the program's base ability to effectively communicate the information to end users and increase simplicity in creating the visualizations.
- ▶ Each program has it's advantages and disadvantages and there is no one perfect software package for data visualization as you must trade-off experience, time, and appearance.

Data Visualization with Excel

In Excel, we saw examples of charts including pivot charts, sparklines, and visual formatting of cells.



Figure: Examples of data visualizations using Excel

Data Visualization in Python

Using Python a variety of charting libraries are available including matplotlib and ggplot.

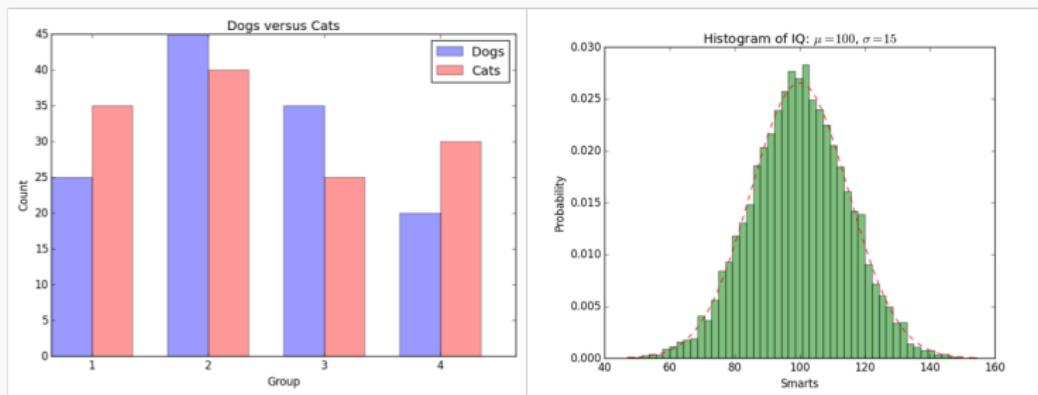


Figure: Examples of data visualizations using Python

Data Visualization in R - Qualitative Data

Using in R we can create visualization for qualitative data in form of bar charts, frequency tables, pie charts,

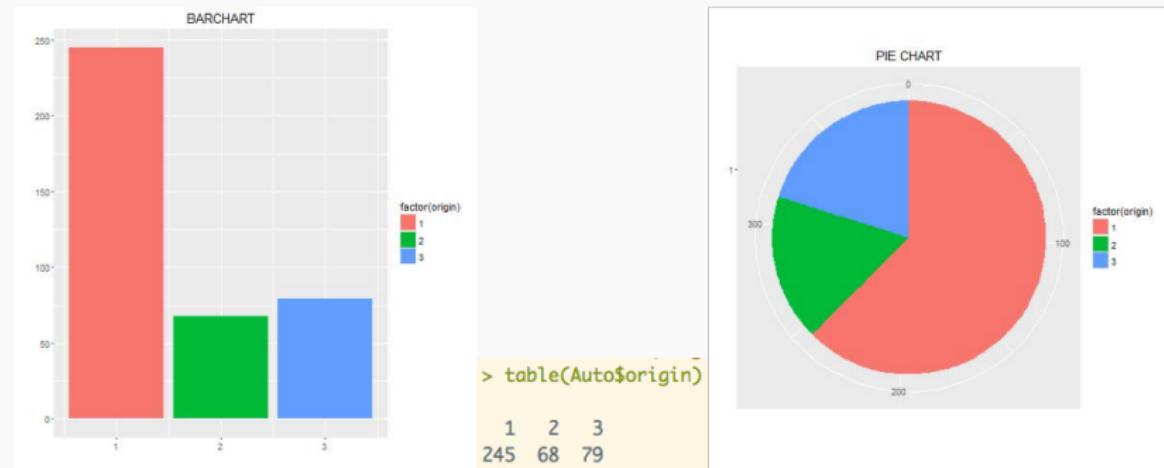


Figure: Examples of qualitative data visualizations using ggplot in R

Data Visualization in R - Quantitative Data

Using in R we saw examples of how we could represent quantitative data, eg. histograms, boxplots

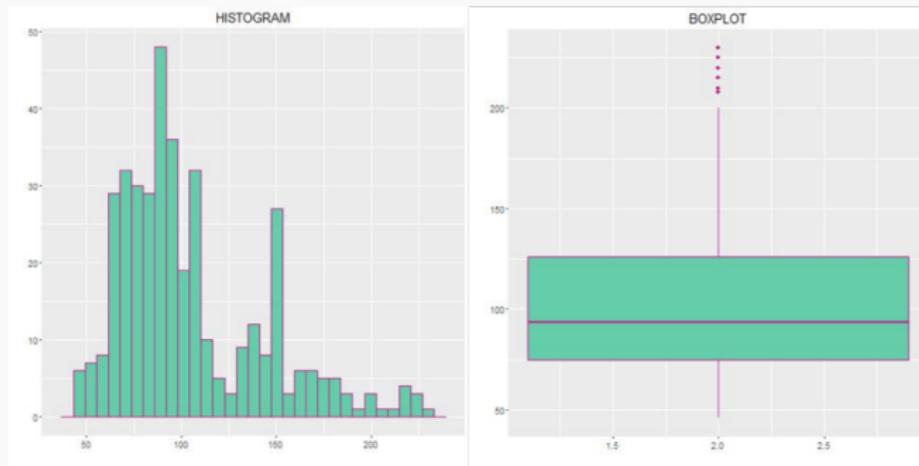


Figure: Examples of quantitative data visualizations using R

- ▶ The last two slides demonstrate how the importance of the type of data being inputted into our graphics.
- ▶ Namely the collection of graphs we could create with qualitative data is different from the collection of graphs we could make with quantitative data.
- ▶ *Tableau* is a very powerful tool for data visualization, that will do a lot of the heavy lifting and allow us to gain deeper insights into our data.
- ▶ Tableau will help us guide us along the way by segregating qualitative from quantitative fields and provide suggestions on useful types of graphics depending on the inputted fields.

Introduction to Tableau

- ▶ Tableau was founded in 2003 as a spin-off from Stanford University by Chris Stolte, Christian Chabot and Pat Hanrahan.
- ▶ The goal of Tableau is "to help people see and understand their data." - Christian Chabot, Tableau CEO
 - ▶ General goal: make less unknown and more known.
 - ▶ Graphical representation of data can allow us to see patterns that we may not necessarily see by examining the raw data.
- ▶ Tableau has desktop and server (enterprise) products as well as Tableau Public allowing sharing of data sets.

Getting Started

- ▶ We will be using the desktop version of [Tableau](#)
- ▶ Visit [this](#) site to obtain a license or to view examples of its use.
- ▶ We will be working with the Superstore data sets (Global Superstore Returns 2016.csv and Global Superstore Orders 2016.xlsx).
- ▶ The quickest way to download this data is to click the "Getting Started" video located on Tableau [home page](#) or [here](#) (you will need to register and sign in before you can see the videos).

Tableau Home page

The screenshot shows the Tableau Home page with a red circle highlighting the "Discover" section.

Connect

- To File:
 - Microsoft Excel
 - Text file
 - JSON file
 - PDF file
 - Spatial file
 - Statistical file
 - More...
- To a Server:
 - Tableau Server
 - Microsoft SQL Server
 - MySQL
 - Oracle
 - Amazon Redshift
 - More...
- Saved Data Sources:
 - Sample - Superstore
 - World Indicators

Open

Welcome
Your 14-day trial has started. Here are some great ways to learn Tableau.

Discover

- Training**
 - Getting Started
 - Connecting to Data
 - Visual Analytics
 - Understanding Tableau
 - More training videos...
- Sharing**
 - Learn more about ways to share
- Resources**
 - Get Tableau Prep
 - Blog: The Density Mark type brings a new kind of heatmap to Tableau
 - Forums

Watch these great demos if you are interested to learn more!

Sample Workbooks

- Superstore
- Regional
- World Indicators
- More Samples

VIZ OF THE WEEK
Analyzing "Um's" →

Home Page Tableau

- ▶ On the right hand panel of the Home page, you will notice a series of instructional videos.
- ▶ These are a great resource to learn more about the features available in Tableau.
- ▶ See the training videos which use the Superstore data sets
 - ▶ Global Superstore Orders 2016.xlsx
 - ▶ Global Superstore Returns 2016.csv

See finished workbook: `getting_started_finished.twbx` file

- ▶ On the left hand panel, you will notice that there are several ways that we can connect with data:
 - ▶ stored locally on our computer (eg. excel files, .csv)
 - ▶ on a server (eg. MySQL)

Home Page Tableau

- ▶ This requires that you have a license or trial key.
- ▶ Students qualify for a free license through the Tableau for Students Program if currently enrolled at an accredited academic institution (undergraduate/post-graduate level).
- ▶ The lab computers in SCI 234 should have Tableau installed (make sure you are using the latest version: Tableau 2019)
- ▶ Sidenote: the Professional Version of Tableau is priced at \$1,999 + maintenance fee per license.
- ▶ There is also a free version of their software called [Tableau Public](#).

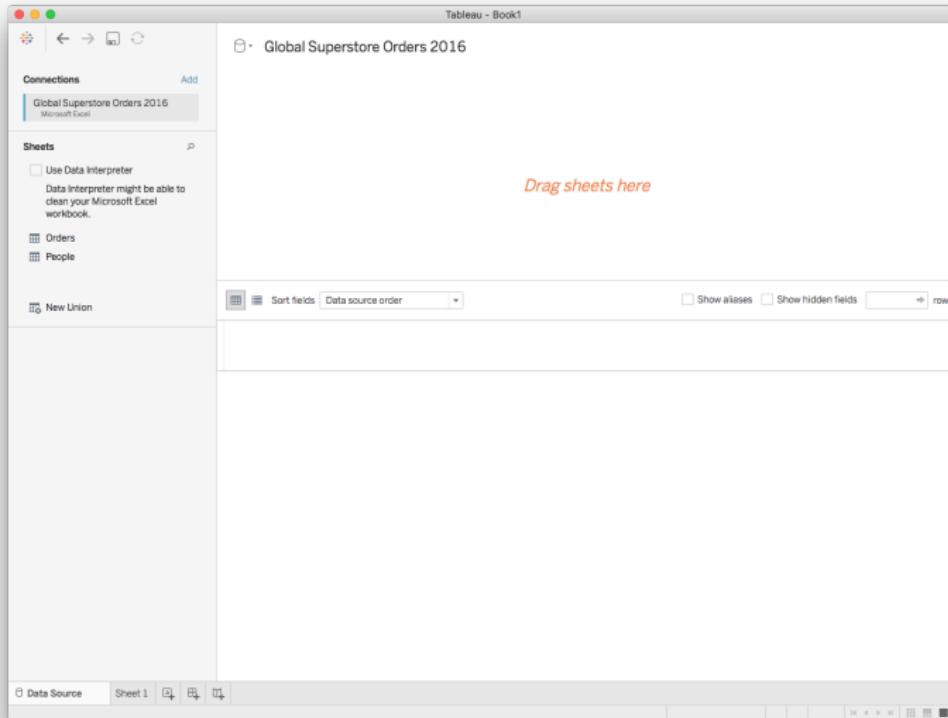
Superstore data

- ▶ This excel file contains two tabs (ie two worksheets):
 - Orders** which contains the products and transactions of customers.
 - People** Customer information

- ▶ To begin, we read this file into Tableau by navigating to the Home Page > Connect (left panel) > To a file > Microsoft Excel.

- ▶ This will open a data source page, from which we can drag and add different data sources.

Tableau Workspace



Example Connecting to Excel

The screenshot shows the Tableau software interface. At the top, it says "Tableau - Book1 - Tableau license expires in 14 days". Below that, there's a toolbar with icons for file operations. The main area is titled "Orders+ (Multiple Connections)". On the left, there's a sidebar with "Connections" (Global Superstore Orders 2016) and "Sheets" (Orders, People). A checkbox for "Use Data Interpreter" is checked, with a note below it: "Data Interpreter might be able to clean your Microsoft Excel workbook." The main workspace displays a table with columns: Row ID, Order ID, Order Date, Ship Date, Ship Mode, Customer ID, and Customer Name. The data includes rows for various orders from different dates and customers like Aaron Bergman, Justin Ritter, and Katherine Murray. At the bottom, there are buttons for "Data Source" and "Sheet 1".

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name
40098	CA-2014-AB1001514...	2014-11-11	2014-11-13	First Class	AB-100151402	Aaron Bergman
26341	IN-2014-R1621074...	2014-02-05	2014-02-07	Second Class	JR-162107	Justin Ritter
25330	IN-2014-CR1273074...	2014-10-17	2014-10-18	First Class	CR-127307	Craig Reiter
13524	ES-2014-KM1637548...	2014-01-28	2014-01-30	First Class	KM-1637548	Katherine Murray
47221	SG-2014-RH949511...	2014-11-05	2014-11-06	Same Day	RH-9495111	Rick Hansen
22732	IN-2014-JM1565574...	2014-06-28	2014-07-01	Second Class	JM-156557	Jim Mitchum
30570	IN-2012-TS2134092...	2012-11-06	2012-11-08	First Class	TS-2134092	Toby Swindell
31192	IN-2013-MB1808592...	2013-04-14	2013-04-18	Standard Class	MB-1808592	Mick Brown
40099	CA-2014-AB1001514...	2014-11-11	2014-11-13	First Class	AB-100151402	Aaron Bergman
36258	CA-2012-AB1001514...	2012-03-06	2012-03-07	First Class	AB-100151404	Aaron Bergman
36259	CA-2012-AB1001514...	2012-03-06	2012-03-07	First Class	AB-100151404	Aaron Bergman
28879	ID-2013-AJ1078014...	2013-04-19	2013-04-22	First Class	AJ-107801	Anthony Jacobs

Example Connecting to Excel

- ▶ If we have related data in another data source, we can create an integrated data source by adding a connection, by clicking the “Add” link in the top left corner.
- ▶ For example, we can create a database-like JOIN with data from another source (eg. we can connect the Orders from Global Superstore Orders 2016.xlsx to the Returns from Global Superstore Returns 2016.csv)

Tableau Workspace

The screenshot shows the Tableau desktop interface with the following details:

- Connections:** A connection named "Global Superstore Orders 2016" is selected and highlighted with an orange circle. There is also an "Add" button.
- Sheets:** The "Orders" sheet is currently active, indicated by a grey bar at the top of the main workspace.
- Data Source:** A tooltip "Go to Worksheet" points to the "Data Source" button at the bottom left.
- Worksheet Content:** The main area displays a table titled "Orders" with the following columns and data:

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment
40098	CA-2014-AB100151...	2014-11-11	2014-11-13	First Class	AB-100151402	Aaron Bergman	Consumer
26341	IN-2014-JR1621074...	2014-02-05	2014-02-07	Second Class	JR-162107	Justin Ritter	Corporate
25330	IN-2014-CR1273074...	2014-10-17	2014-10-18	First Class	CR-127307	Craig Reiter	Consumer
13524	ES-2014-9M163754B...	2014-01-28	2014-01-30	First Class	KM-1637548	Katherine Murray	Home Office
47223	SG-2014-RH949511...	2014-11-05	2014-11-06	Same Day	RH-9495111	Rick Hansen	Consumer
22732	IN-2014-JM1565574...	2014-06-28	2014-07-01	Second Class	JM-156557	Jim Mitchum	Corporate
30570	IN-2012-TS2134092...	2012-11-06	2012-11-08	First Class	TS-2134092	Toby Swindell	Consumer
31192	IN-2013-MB1808692...	2013-04-14	2013-04-18	Standard Class	MB-1B08592	Mick Brown	Consumer
40099	CA-2014-AB100151...	2014-11-11	2014-11-13	First Class	AB-100151402	Aaron Bergman	Consumer
36258	CA-2012-AB1001514...	2012-03-06	2012-03-07	First Class	AB-100151404	Aaron Bergman	Consumer

Tableau Workspace

Notice here that we have created a **left join** (on Order ID) which we can change by clicking on the Venn Diagram image.

The screenshot shows the Tableau Data Source editor for 'Orders+ (Multiple Connections)'. It displays a join operation between two datasets: 'Orders' and 'Global Superstore Returns 2011'. A tooltip indicates the current join type is 'Left'. The 'Join' button is highlighted with a blue circle. Below the join area, there are four options: Inner, Left, Right, and Full Outer. The 'Left' option is selected. The 'Order ID' field from both datasets is listed under the 'Data Source' column. The preview pane shows a sample of 1,000 rows from the joined datasets. The columns include Order ID, Order Date, Ship Date, and Ship Mode. A 'Go to Worksheet' button is visible at the bottom left.

We can also change meta-data in the excel-like spreadsheet.

Tableau Workspace

Clicking on the 'Sheet' tab (orange button in the lower left hand corner) will bring us to this workspace:

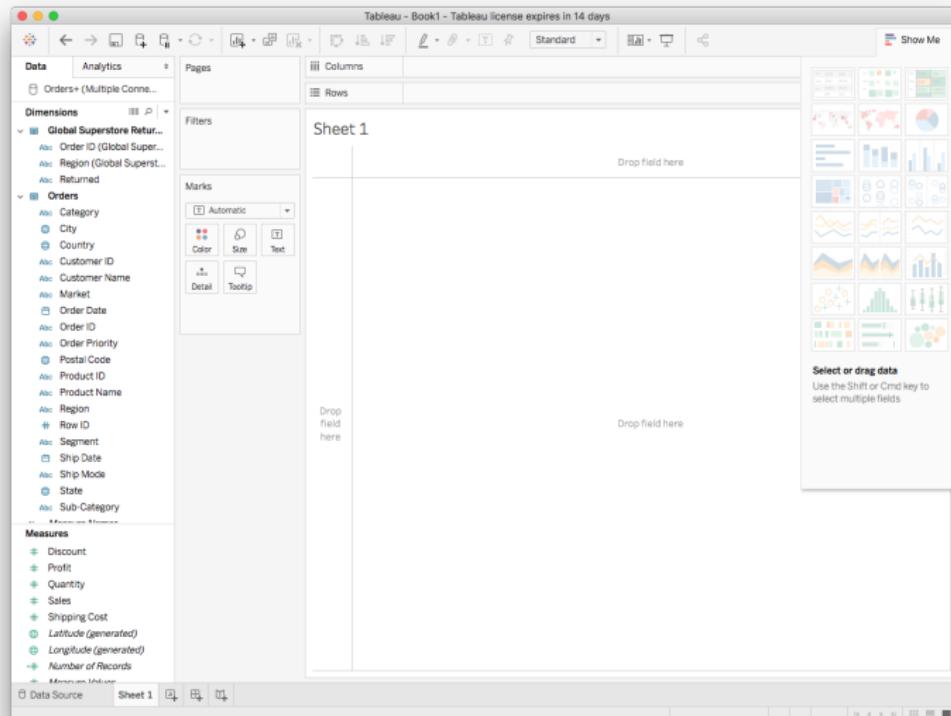


Tableau Features

- ▶ Supported data types: text  , dates  , numbers  , geographical coordinates  (latitude/longitude), Boolean 
- ▶ There are a number of aggregation functions available in Tableau: sum, average, max, count, variance, etc.
- ▶ There are many built-in functions for numeric and string manipulation.
- ▶ Calculated fields can be created and are proceeded by an equal sign: 
- ▶ Visualizations are created by dragging these fields (or “pills”) to desired shelves, . . .

Tableau Terminology

A **pill** is a field (i.e column in your data set) that can be placed in the visualization. They are stored one of two **field types**

Dimension or categorical/discrete information are colour-coded using *blue* pills

Measures, that is, metrics containing our quantitative information are colour-coded using *green* pills



Dimensions (eg. date, customer) are usually what is used for creating *labels* while the Measures (eg price) are the metrics we want to analyze (this data is often continuous).

Tableau Terminology

A **shelf** is a location to put a pill.

- ▶ Column shelf, row shelf, filter shelf
- ▶ Row and column shelves are similar to Pivot tables in Excel but with built-in visualization.

Dragging pills to different shelves will change the way our graphics appears. To remove or undo a move, simply drag the pill off the shelf to delete, or press the undo button. The Clear Sheet  button will erase the entire graphic

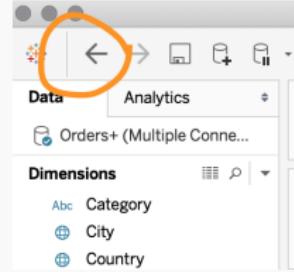
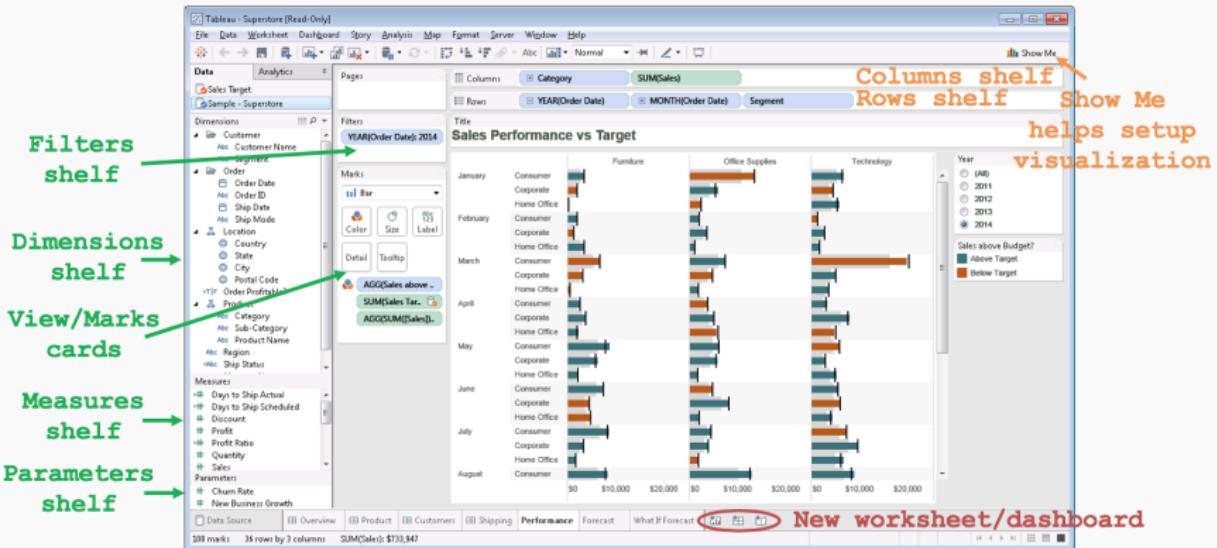


Tableau Workspace Items



View Cards

View or shape cards allows control of colour, shape, and size. They also enable filtering, labeling, and ability to add details on demand.

Color expresses discrete or continuous values

Size expresses discrete or continuous values

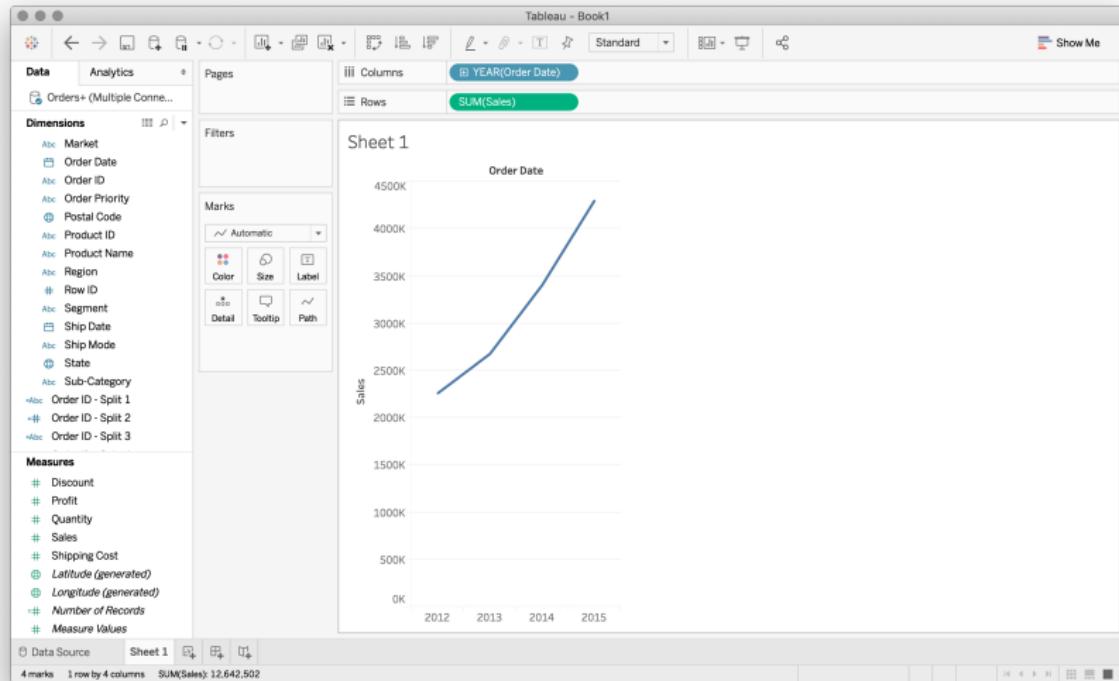
Label one or more fields can be expressed as label on marks

Detail disaggregates the marks plotted

Tooltip makes fields available to tooltips without
disaggregating data (shows info when hovering)

Shape expresses discrete or continuous fields

Multiple fields can be placed on the colour, label, detail, and tooltip buttons.

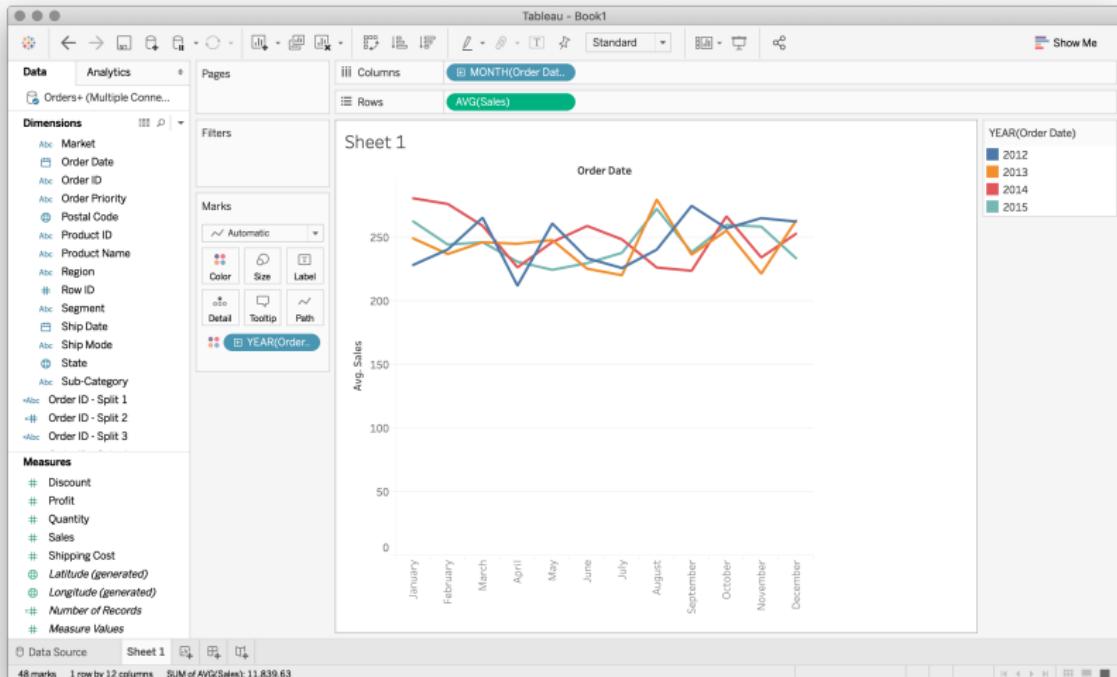


- ▶ Notice how Tableau will guess at useful aggregations of our data.

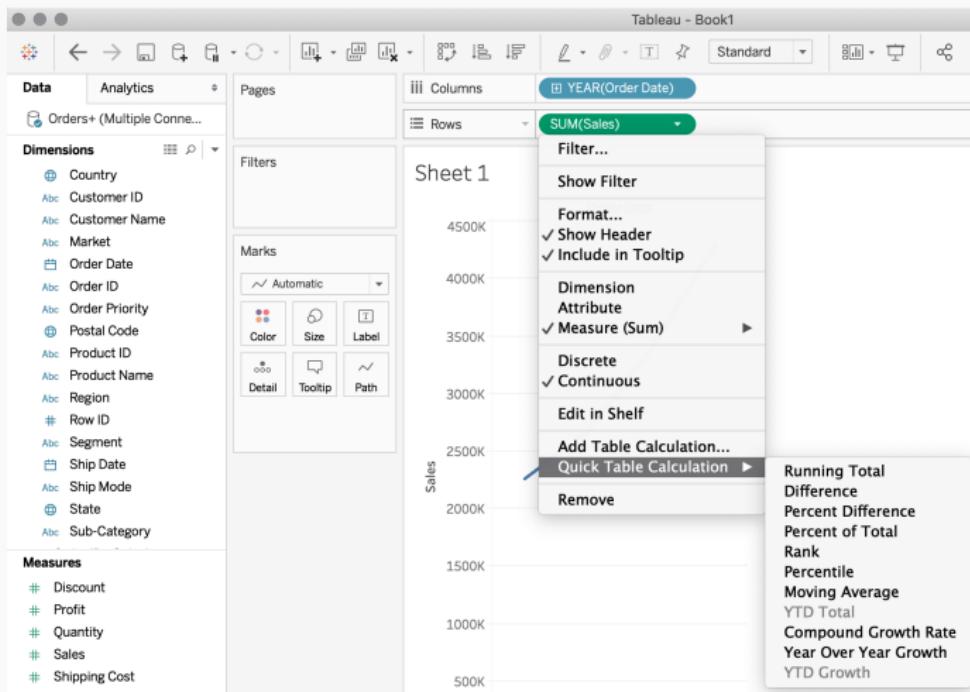
- ▶ We can change these fields by clicking the pills located on each shelf
 - ▶ eg. change SUM to MEDIAN by Measure (Sum) > Median .
 - ▶ eg. change the sales per year to sales per quarter or sales per month (see next slide for details)

For example, we may want to see *average sales month-to-month* for each year.

- ▶ Click the + sign to the Left of the blue pill, to granulate the dates into finer subcategories (ie. Quarters)
- ▶ Drag year to the right of Quarter so that we see each quarter yearly (rather than each year quarterly)
- ▶ Then drag year to 'Color' to change this side by side line graph to a single line graph with a legend.
- ▶ Change Quarter to Month to get the next level of granularity for these dates
- ▶ Click the downarrow on the green pill and change Measure (Sum) > Average.



There are a number of other 'quick calculations' we have to choose from:



Show Me Button

- ▶ The *Show Me* button suggests visualization to use based on your current dimensions and measures.
- ▶ Choose the desired dimensions and measures by selecting them with your cursor while holding down the Ctrl key (window) or Cmnd key (mac).
- ▶ As your selections change, different chart types will be suggested for you (i.e. they will become highlighted) .
- ▶ Think of Show Me as your one-click option that will automatically place pills on shelves.

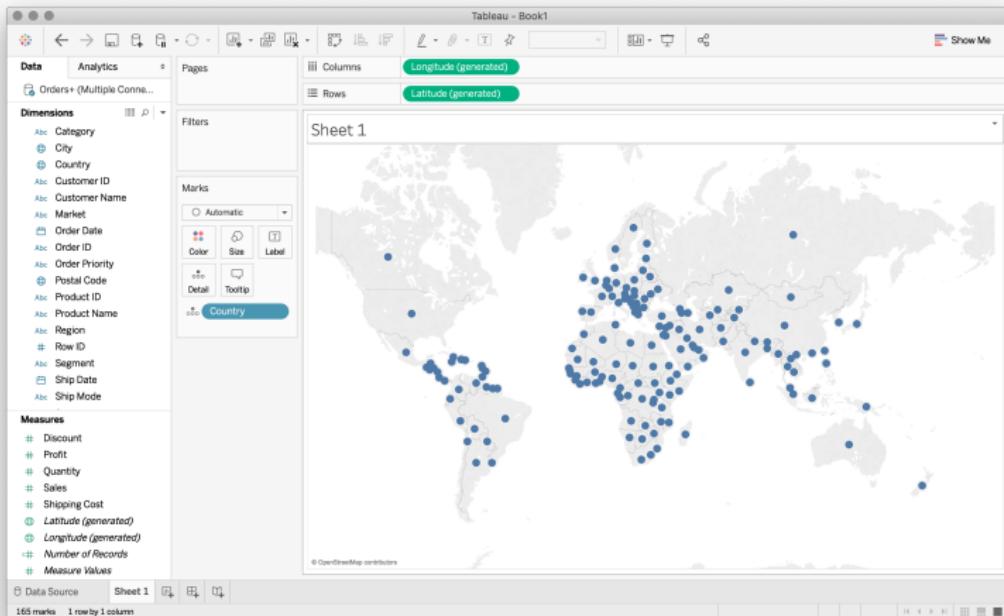
Show Me

Eg, since Country has a geographical component, a map seems like a natural choice:

The screenshot shows the Tableau interface with the title bar "Tableau - Book1 - Tableau license expires in 14 days". The left sidebar contains the "Data" pane with "Dimensions" and "Measures" sections, and the "Analytics" pane. The "Dimensions" section includes fields like Order ID, Region, Returned, Orders, Category, City, and Country. The "Measures" section includes fields like Discount, Profit, Quantity, Sales, Shipping Cost, Latitude (generated), Longitude (generated), Number of Records, and Maximum Value. The "Country" field is highlighted with a blue oval. The "Measures" section also has "Sales" highlighted with a green oval. The main workspace is titled "Sheet 1" and features a "Marks" card with "Automatic" selected. To the right of the workspace is a "Show Me" panel containing a grid of visualization icons. A tooltip for symbol maps is displayed, stating: "For symbol maps try 1 geo @ Dimension, 0 or more Dimensions, 0 to 2 Measures. May use spatial measure in place of geo dimension". The bottom navigation bar includes "Data Source", "Sheet 1", and other standard Tableau navigation icons.

Show Me

For geographic data (small globe icon), Tableau automatically generates center-point geocodes (longitude/latitude).



Show Me

We can click and drag 'Sales' onto the map to include this information in the graphic (depicted by size of the dot, also visible when hovering over the country)

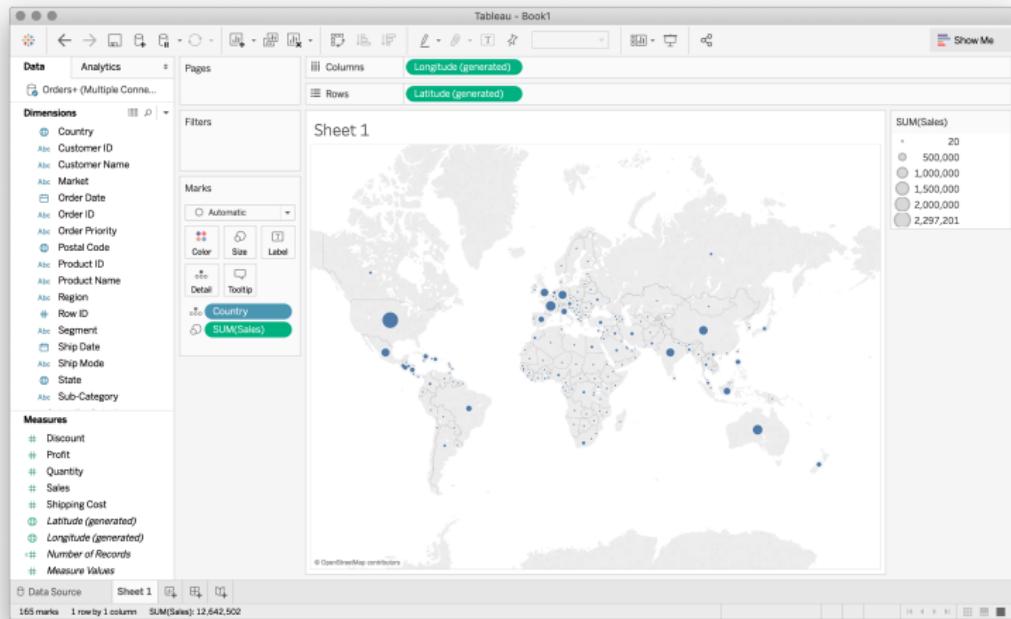


Tableau Question

Example

How many of the following statements are TRUE?

1. In Tableau blue pills are continuous.
2. A shelf is a location to place a pill.
3. The Show Me button will suggest visualizations for you.
4. A pill for a dimension may be on more than one shelf at the same time.

A) 0

B) 1

C) 2

D) 3

E) 4

Tableau Question

Answer

How many of the following statements are TRUE?

1. In Tableau blue pills are continuous. **X**
2. A shelf is a location to place a pill.
3. The Show Me button will suggest visualizations for you.
4. A pill for a dimension may be on more than one shelf at the same time.

A) 0

B) 1

C) 2

D) 3

E) 4

Tableau Question

Answer

How many of the following statements are TRUE?

1. In Tableau blue pills are continuous. X
2. A shelf is a location to place a pill. ✓
3. The Show Me button will suggest visualizations for you.
4. A pill for a dimension may be on more than one shelf at the same time.

A) 0

B) 1

C) 2

D) 3

E) 4

Tableau Question

Answer

How many of the following statements are TRUE?

1. In Tableau blue pills are continuous. X
2. A shelf is a location to place a pill. ✓
3. The Show Me button will suggest visualizations for you. ✓
4. A pill for a dimension may be on more than one shelf at the same time.

A) 0

B) 1

C) 2

D) 3

E) 4

Tableau Question

Answer

How many of the following statements are TRUE?

1. In Tableau blue pills are continuous. X
2. A shelf is a location to place a pill. ✓
3. The Show Me button will suggest visualizations for you. ✓
4. A pill for a dimension may be on more than one shelf at the same time. ✓

- A) 0 B) 1 C) 2 D) 3 E) 4

Try it: Tableau Visualizations

Example

- ▶ Install Tableau. Use trial version or [Tableau for students](#)
- ▶ Start Tableau. Use the sample.twbx file or the Superstore example and explore the visualizations.
- ▶ Try create any visualization of the data.

Tableau - Data Sources

Tableau can connect to a wide variety of data sources including:

- ▶ Microsoft Excel and Access
- ▶ Text files (txt, csv)
- ▶ Relational databases (MySQL, SQL Server, Oracle, PostgreSQL)
- ▶ NoSQL databases (MongoDB)
- ▶ Parallel and analytical databases (Greenplum, Vertica, Teradata)
- ▶ Other ODBC sources (note JDBC is not supported)

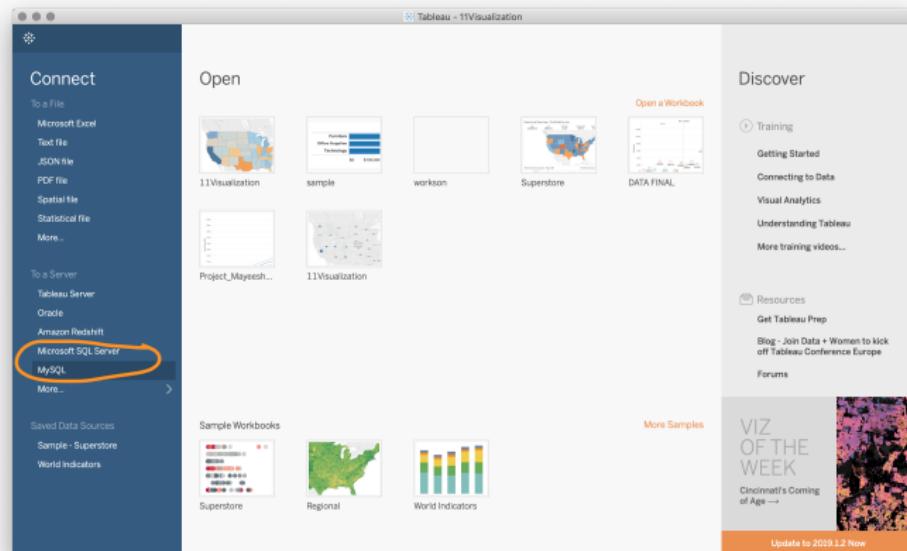
Connecting to relational databases

Connecting to a relational database like MySQL and Microsoft SQL requires:

- ▶ Driver (often need to [download](#) from database vendor)
- ▶ Database connection information

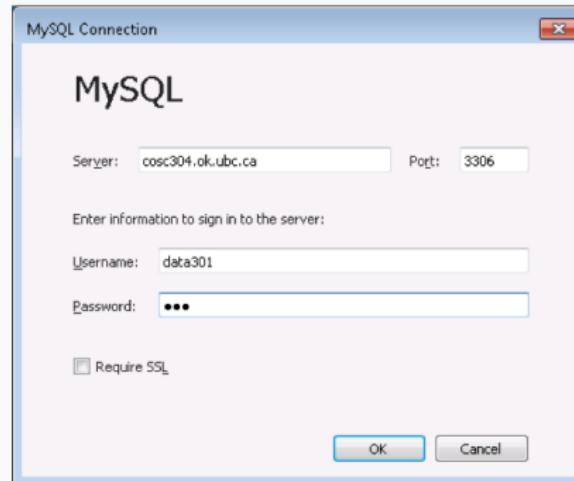
Connecting to relational databases

Once you have the necessary driver, you simple click on the appropriate option on the left hand panel of the home screen:



Connecting to relational databases

At this point you will be asked to fill in the required fields in a pop up window like this:



Using the above credentials with password `ubc` should grant you access to the `WorksOn` database that we have studied in our SQL unit.

In order to connect to this database you need to be on **ubcsecure** wifi *on campus*.

To access it from home:

- ▶ Download VPN tool at: <https://myvpn.ok.ubc.ca/> and follow instructions there.
- ▶ Launch Cisco AnyConnect and enter myvpn.ok.ubc.ca as the host and click Connect.
- ▶ Enter CWL to authenticate.

Connecting to MySQL

Tableau - Book1 - Tableau license expires in 14 days

Connections Add
cosc304.ok.ubc.ca MySQL

Database WorksOn

Table dept emp proj worksn

New Custom SQL New Union

workson+ (WorksOn)

Connection Live Extract

Filters 0 | Add

workson -> emp -> dept
workson -> proj

Sort fields Data source order Show aliases Show hidden fields 10 rows

Dno (Dept)	Dname	Mgrno	Eno (Emp)	Ename	Bdate	Title	Salary	Sup
null	null	null	E1	J. Doe	1975-01-05	EE	30,000.00	E2
null	null	null	E2	M. Smith	1966-06-04	SA	50,000.00	E5
null	null	null	E2	M. Smith	1966-06-04	SA	50,000.00	E5
null	null	null	E4	J. Miller	1950-09-01	PR	20,000.00	E6
null	null	null	E5	B. Casey	1971-12-25	SA	50,000.00	E8
null	null	null	E3	A. Lee	1966-07-05	ME	40,000.00	E7
null	null	null	E7	R. Davis	1977-09-08	ME	40,000.00	E8
null	null	null	E3	A. Lee	1966-07-05	ME	40,000.00	E7
null	null	null	E6	L. Chu	1965-11-30	EE	30,000.00	E7
null	null	null	null	null	null	null	null	

Go to Worksheet

Data Source Sheet 1

Connect or Extract Data

Tableau has its own internal data engine. There are two options when retrieving data to visualize:

1. Direct connect to source to get live data
 - ▶ Can refresh data using F5 or selecting refresh menu item
 - ▶ May be faster depending on data set/visualization
2. Extract and import data into Tableau's data engine
 - ▶ May get a performance improvement as data is local
 - ▶ May set certain scheduled times to extract and keep data up to date
 - ▶ Portability (as consumer of report does not need access to data source)

Tableau Data Source Question

Example

How many of the following statements are TRUE?

1. Tableau can connect to relational databases.
2. Tableau can process data in text and Excel files.
3. Tableau can JOIN information across tables from multiple sources.
4. Tableau will try to identify types and relationships from the data sources.

A) 0

B) 1

C) 2

D) 3

E) 4

Tableau Data Source Question

Answer

How many of the following statements are TRUE?

1. Tableau can connect to relational databases. ✓
2. Tableau can process data in text and Excel files.
3. Tableau can JOIN information across tables from multiple sources.
4. Tableau will try to identify types and relationships from the data sources.

A) 0

B) 1

C) 2

D) 3

E) 4

Tableau Data Source Question

Answer

How many of the following statements are TRUE?

1. Tableau can connect to relational databases. ✓
2. Tableau can process data in text and Excel files. ✓
3. Tableau can JOIN information across tables from multiple sources.
4. Tableau will try to identify types and relationships from the data sources.

A) 0

B) 1

C) 2

D) 3

E) 4

Tableau Data Source Question

Answer

How many of the following statements are TRUE?

1. Tableau can connect to relational databases. ✓
2. Tableau can process data in text and Excel files. ✓
3. Tableau can JOIN information across tables from multiple sources. ✓
4. Tableau will try to identify types and relationships from the data sources.

A) 0

B) 1

C) 2

D) 3

E) 4

Tableau Data Source Question

Answer

How many of the following statements are TRUE?

1. Tableau can connect to relational databases. ✓
2. Tableau can process data in text and Excel files. ✓
3. Tableau can JOIN information across tables from multiple sources. ✓
4. Tableau will try to identify types and relationships from the data sources. ✓

- A) 0 B) 1 C) 2 D) 3 E) 4

Try it: Tableau Data Sources

Example

Use Tableau to connect to Excel and MySQL data sources.

- ▶ Start Tableau. Open up Superstore Excel data source (either XLS or TDS file) in Tableau/defaults/Datasources directory.
- ▶ **install** the MySQL ODBC connector
- ▶ Server: cosc304.ok.ubc.ca Database: data301 User: data301 Password: ubc

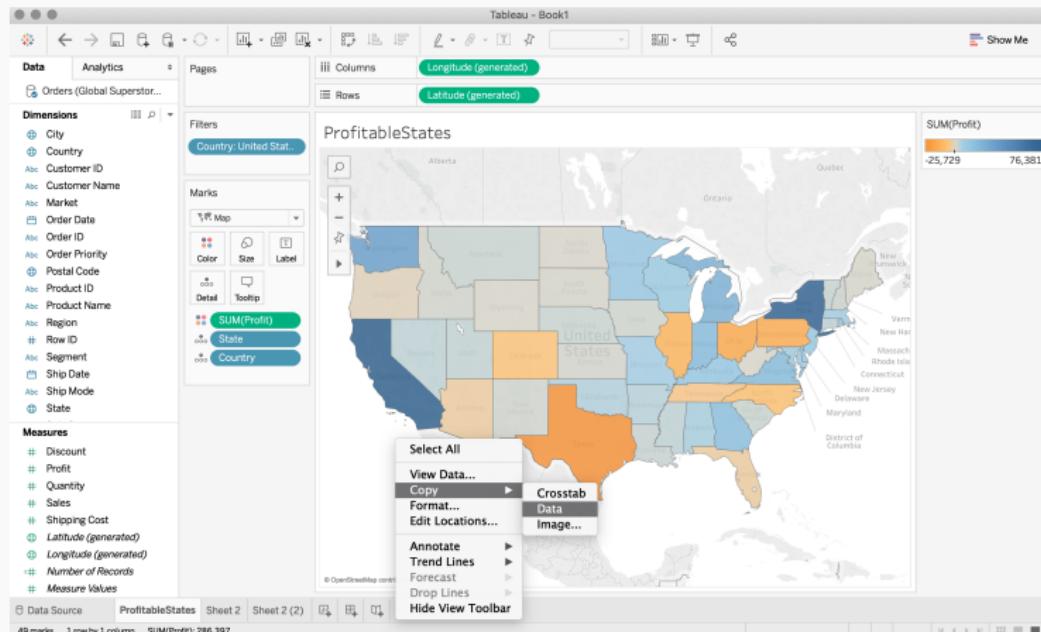
Superstore visualizations:

- ▶ Map showing profit by state. Save this sheet as State Profit
- ▶ Visualization to indicate what is the best selling product category per market. Save this sheet as Best Product
- ▶ Annotate the best selling product category for each market.

Some notes

- ▶ You can rename the tabs in Tableau in the same way we would in Excel (double click the tab and rename).
- ▶ If we want the raw data associated with a visualization, it is as simple as right clicking on the image, then selecting **Copy > Data** and pasting this information into Excel.
- ▶ Alternatively, you could go to toolbar and select **Worksheet > Duplicate as Crosstab**.

Copy data in Tableau

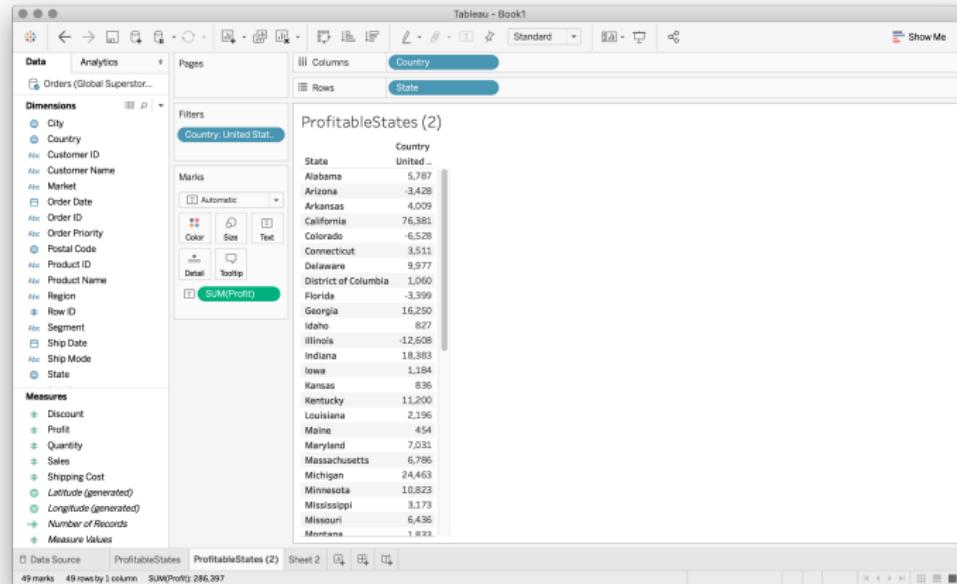


Paste into Excel

The screenshot shows a Microsoft Excel spreadsheet titled "Book1". The data is organized into columns A through S, with rows 1 through 33. Column A contains the country name ("Country") and state abbreviations ("State"). Columns B through E contain geographic coordinates: Latitude (deg), Longitude (deg), and Profit. The data includes all 50 US states. The table is styled with a light gray background and white text. The "Home" tab is selected in the ribbon.

	Country	State	Latitude (deg)	Longitude (deg)	Profit
1	United States	Alabama	32.75	-86.75	5,786.83
2	United States	Alaska	64.5	-141.5	1,627.92
3	United States	Arizona	34.9	-111.9	1,627.92
4	United States	Arkansas	34.7	-93.5	4,098.69
5	United States	California	37.25	-119.75	76,381.39
6	United States	Colorado	39	-105.5	-6,527.86
7	United States	Connecticut	41.66	-72.66	3,511.49
8	United States	Delaware	38.6716	-75.4165	9,977.37
9	United States	District Of Co	38.91	-77	1,059.59
10	United States	Florida	28.42	-81.55	3,399.30
11	United States	Georgia	32.3	-84	7,631.64
12	United States	Idaho	44.0759	-114.543	816.7231
13	United States	Illinois	40	-89.25	-17,667.89
14	United States	Indiana	40	-86.25	18,382.94
15	United States	Iowa	42	-93.5	1,183.81
16	United States	Kansas	38.5	-96.5	816.4435
17	United States	Kentucky	37.79	-84.87	11,199.70
18	United States	Louisiana	30.4	-92.3	1,014.10
19	United States	Maine	46.5	-69.24	434.4862
20	United States	Maryland	39.4968	-77.0814	7,051.18
21	United States	Massachusett	42.33	-71.82	6,785.50
22	United States	Michigan	43.1064	-84.6165	24,403.19
23	United States	Minnesota	46.25	-94.25	10,823.19
24	United States	Mississippi	32.75	-89.75	3,172.98
25	United States	Missouri	38.25	-92.5	6,486.21
26	United States	Montana	49	-104.5	1,014.13
27	United States	Nebraska	41.5	-99.75	2,057.09
28	United States	Nevada	39.25	-116.75	3,316.77
29	United States	New Hampsh	43.66	-71.49	1,706.50
30	United States	New Jersey	40.16	-74.49	9,772.91
31	United States	New Mexico	34.5	-106	1,157.12
32	United States	New York	43	-75.49	78.55
33	United States	North Carolin	35.5	-80	1.91

Duplicate as crosstab



Interactive plots

We can also make the plot interactive by right clicking a desired field and selecting "Show Filter".

For instance, if we select "Show Filter" for Country, the list of countries will appear in the right hand panel of our graphic.

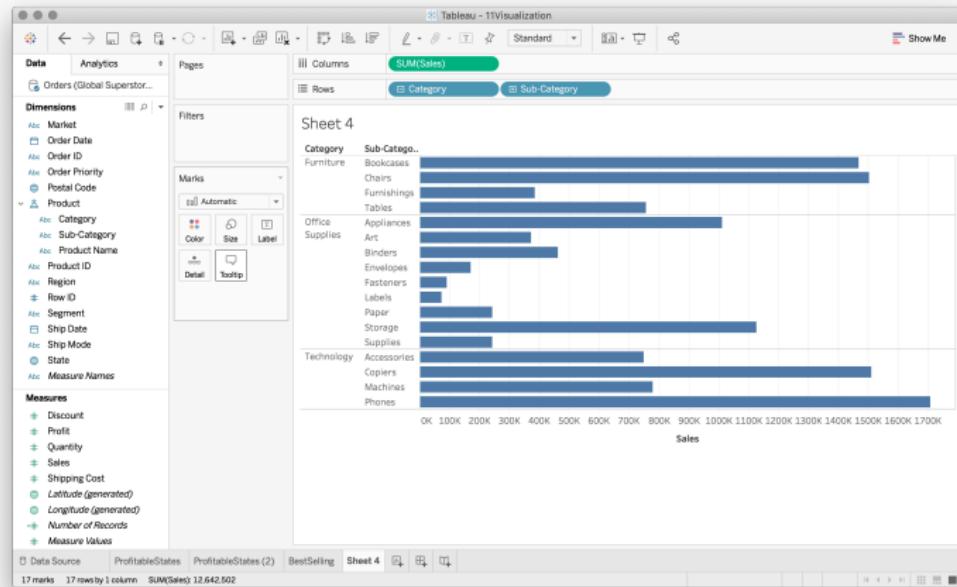
The user can now select or deselect other countries for our visualization or "viz"

Subcategories

- ▶ **Hierarchies** are groupings of data that make it easier to roll-up and drill-down into data. Examples:
 - ▶ year, quarter, month
 - ▶ country, state, city
- ▶ We can create own hierarchies by dragging dimensions on top of each other.
- ▶ For example, in the Superstore data, 'Sub-category' is a, for lack of better term, sub-category of 'Category'.
- ▶ To have Tableau recognize it as such, we simply drag the sub-category pill over-top of category (in addition, we could do that with product name).
- ▶ We might name this new pill 'Product'.

User made dimension

Notice that the 'Product' pill now has a clickable + sign to expand and collapse the levels of the hierarchy.



User made dimension

We can click the sort buttons to arrange bars by descending/ascending order.

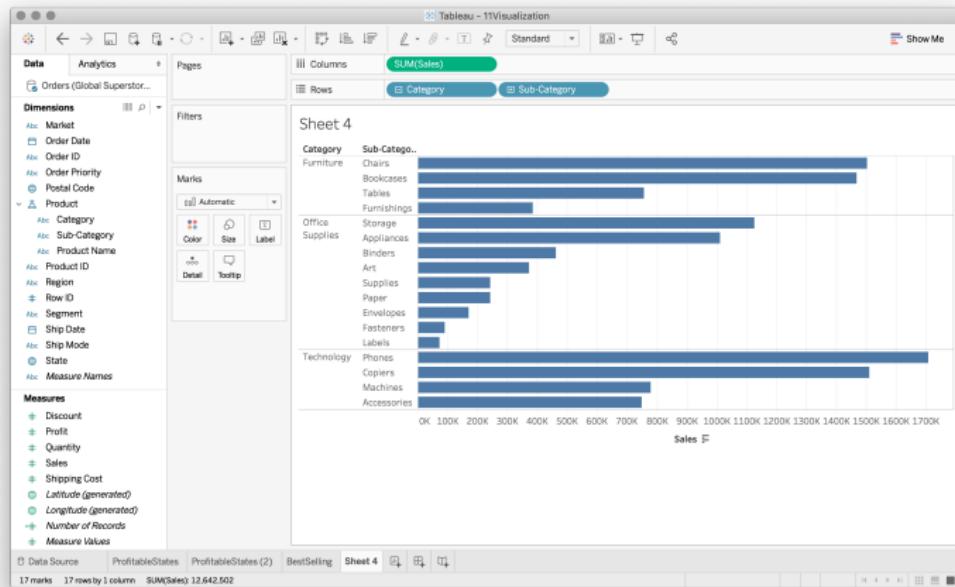


Tableau Files

Tableau Workbook (.twb) (default) saves workbook but no data

Tableau Packaged Workbook (.twbx) contains data and visualization for easier sharing

Tableau Datasource (.tds) metadata on a data source

Tableau Bookmark (.twb) one worksheet within workbook

Tableau Data Extract (.tde) compressed snapshot of data stored in column format

Joining Tables

When connecting tables **R** and **S**, there are four types of joins:

INNER JOIN row in result for each row of R that matches a row of S

LEFT OUTER JOIN row in result for each row of R that matches a row of S OR a row of R that does not match anything in S

RIGHT OUTER JOIN row in result for each row of R that matches a row of S OR a row of S that does not match anything in R

FULL OUTER JOIN row in result for each row of R that matches a row of S OR a row of R that does not match anything in S OR a row of S that does not match anything in R

See [here](#) for a visual representation of joins (for SQL but same concept [here](#)).

Join Example

Boys	
Bid	BoyName
1	Joe
2	Steve
3	Fred
5	James

Boys INNER JOIN Girls			
Bid	BoyName	Gid	GirlName
2	Steve	2	Jane
5	James	5	Fran

Boys LEFT OUTER JOIN Girls			
Bid	BoyName	Gid	GirlName
1	Joe		
2	Steve	2	Jane
3	Fred		
5	James	5	Fran

Girls	
Gid	GirlName
2	Jane
4	Sarah
5	Fran
6	Julie

Boys FULL OUTER JOIN Girls			
Bid	BoyName	Gid	GirlName
1	Joe		
2	Steve	2	Jane
3	Fred		
		4	Sarah
5	James	5	Fran
		6	Julie

Boys RIGHT OUTER JOIN Girls			
Bid	BoyName	Gid	GirlName
2	Steve	2	Jane
		4	Sarah
5	James	5	Fran
		6	Julie

Data Blending

Data blending allows "joining" data that does not reside in a single source. There are automatic and manual methods.

Automatic field names must match across sources. Will link secondary data source with primary data source.

Manual methods include ability to specify SQL statement to perform with join.

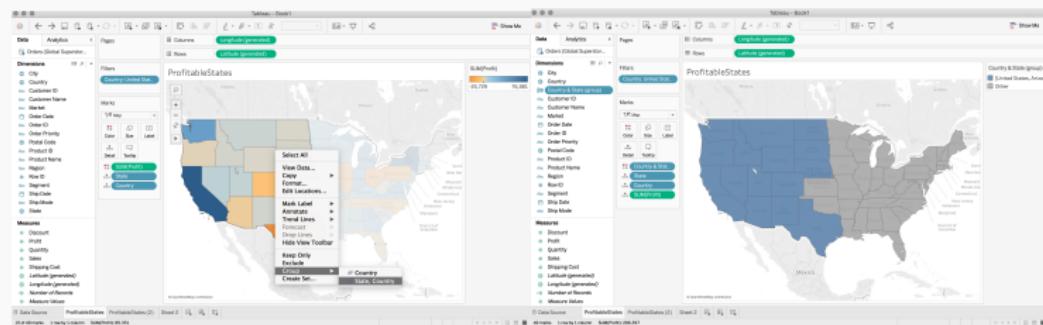
Try it: Tableau Data Sources - Joins

Using the MySQL tables in the data301 database, create some joins to connect them so it looks like this:



Dynamic Grouping/Renaming

- ▶ Dynamic grouping (also called ad hoc groups) can be created by using **Ctrl+Select** (windows) or **Cmnd+select** (mac) to select elements in visualization and select **Group** from menu.



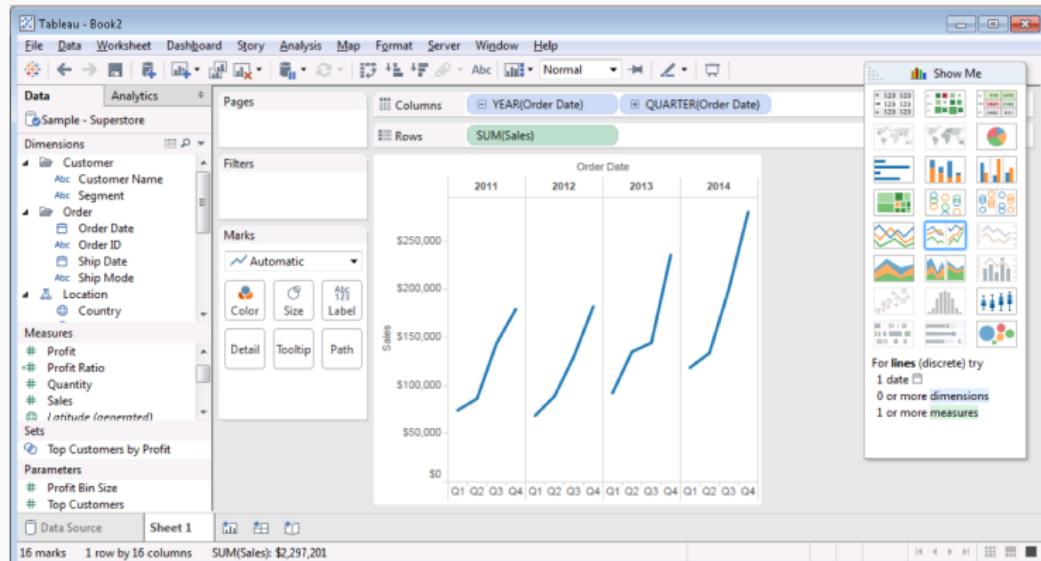
- ▶ It is also possible to rename values/labels and correct value errors.
 - ▶ Eg. change California to "CA" by right clicking on the state and selecting **Annotate > Mark** and edit the defaults

Tableau Chart Types

Chart types:

- ▶ text tables/crosstabs
- ▶ maps
- ▶ heat maps, highlight tables, tree maps
- ▶ line charts
- ▶ area fill charts and pie charts
- ▶ scatter plot, circle view, side-by-side plots (identify outliers)
- ▶ bullet graph, packed bubble, histogram, Gantt charts

Line Chart (discrete time)



Text Table (Crosstab)

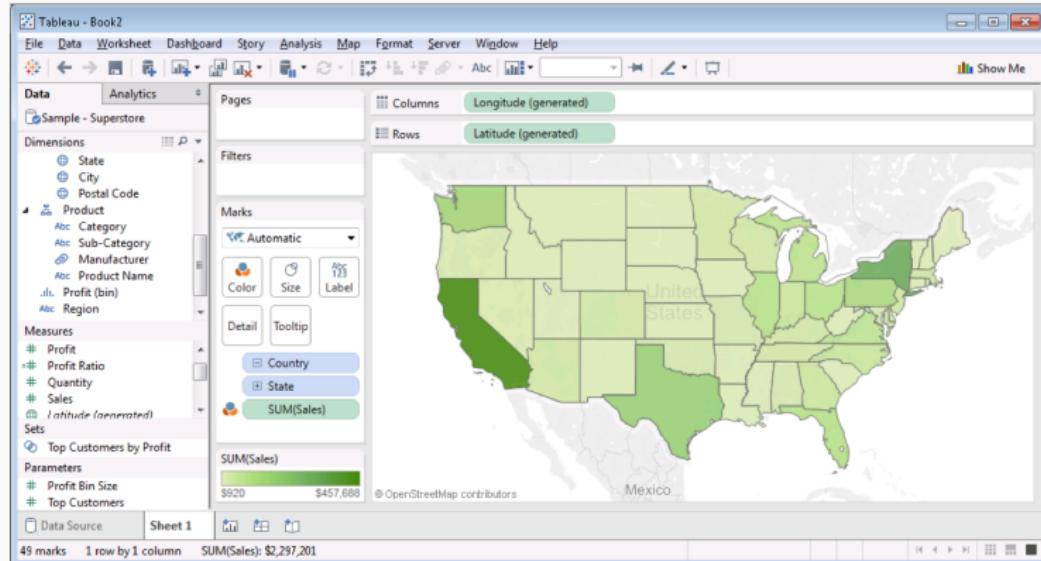
These are very similar to Pivot Tables: after selecting the Dimensions you want, drag the desired Measure directly overtop the table. N.B. you can change the aggregate function (SUM is the default in most cases)

The screenshot shows the Tableau interface with a crosstab visualization. The visualization has 'Category' in the Rows shelf and 'Sub-Category' in the Columns shelf. The data is grouped by 'Order Date' (2011, 2012, 2013, 2014). The data source is 'Sample - Superstore'. The 'Measures' shelf contains 'Sales' with the aggregate function 'SUM(Sales)' highlighted in green. The 'Marks' shelf shows options for Color, Size, and Text.

Category	Sub-Category	Order Date			
		2011	2012	2013	2014
Furniture	Bookcases	\$20,037	\$38,544	\$26,275	\$30,024
	Chairs	\$77,242	\$71,735	\$83,919	\$95,554
	Furnishings	\$13,826	\$21,096	\$27,874	\$28,915
	Tables	\$46,088	\$39,150	\$60,833	\$60,894
Office Supplies	Appliances	\$15,314	\$23,241	\$26,059	\$42,927
	Art	\$6,058	\$6,237	\$5,910	\$8,914
	Binders	\$43,488	\$37,453	\$49,485	\$72,986
	Envelopes	\$3,856	\$4,512	\$4,730	\$3,379
	Fasteners	\$661	\$545	\$960	\$858
	Labels	\$2,841	\$2,956	\$2,827	\$3,861
	Paper	\$14,835	\$15,286	\$20,638	\$27,778
	Storage	\$50,329	\$45,046	\$58,632	\$66,834
	Supplies	\$14,394	\$1,952	\$14,278	\$16,049
Technology	Accessories	\$25,014	\$40,524	\$41,896	\$59,946
	Copiers	\$10,850	\$26,179	\$49,599	\$62,899
	Machines	\$62,023	\$27,764	\$55,907	\$43,545

Bottom status bar: 68 marks, 17 rows by 4 columns, SUM(Sales): \$2,257,201

Maps



Heat Map

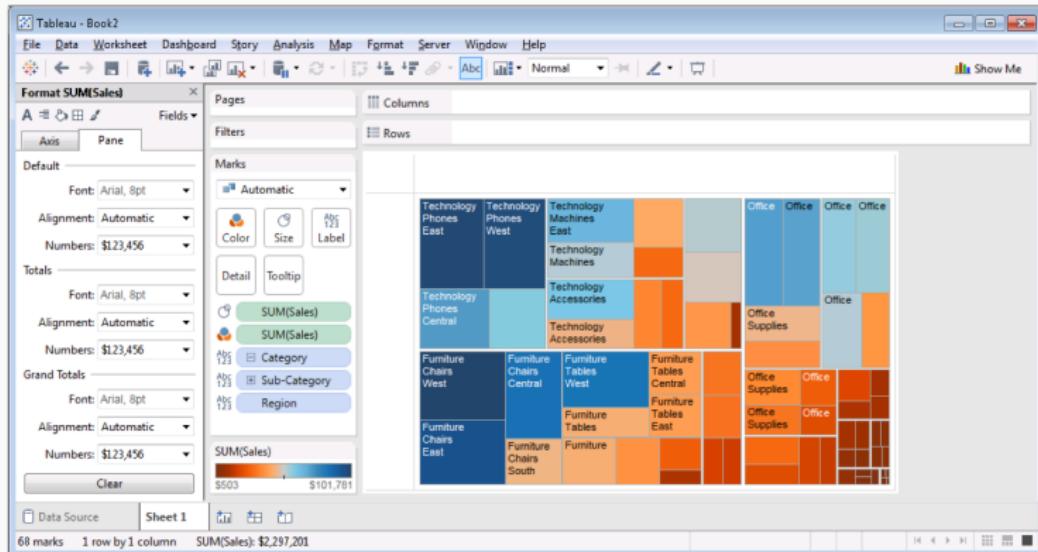
Turn the cross tab viz from slide 65 to a heat map by going to the Show Me tab and selecting the option in the top right corner.
Think of this as a complex conditional formatting of Excel cells.

The screenshot shows a Tableau desktop interface with the following details:

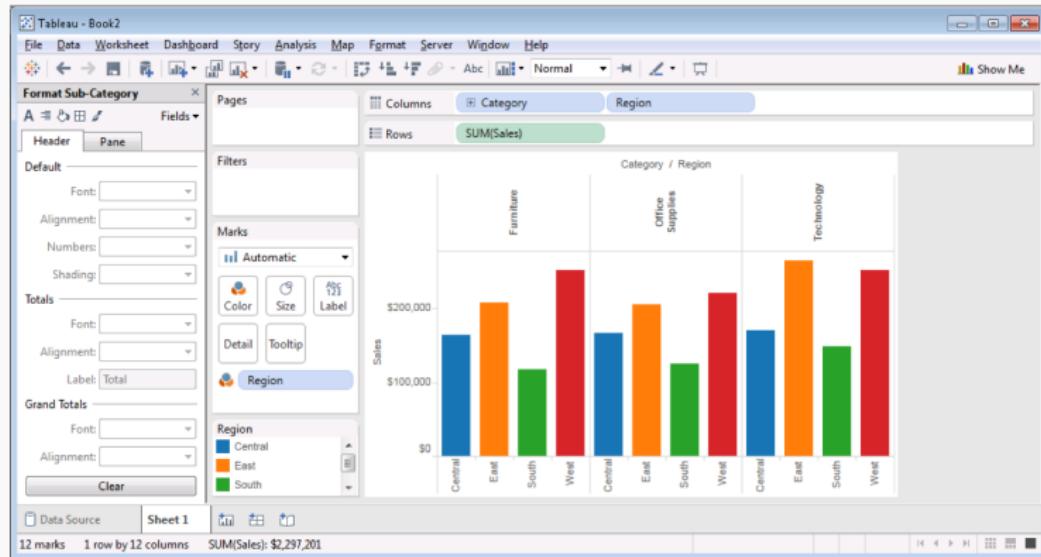
- Top Bar:** Shows the title "Tableau - Book1" and various application icons.
- Left Panel (Data Shelf):**
 - Dimensions:** Order Date, Order ID, Order Priority, Postal Code, Product ID, Region, Row ID, Segment, Ship Date, Ship Mode, State.
 - Product:** Category, Sub-Category, Product Name, Measure Names.
 - Measures:** Discount, Profit, Quantity, Sales, Shipping Cost, Latitude (generated), Longitude (generated), Number of Records, Measure Values.
- Middle Panel (Sheet 4):**
 - Rows:** Category, Sub-Category.
 - Columns:** Order Date (2012, 2013, 2014, 2015).
 - Data:** A crosstab table showing sales data for Furniture, Office Supplies, and Technology categories across years.
- Right Panel (Show Me):** A large grid of visualization options. One specific option, a heatmap icon, is circled in red.
- Bottom Panel:** Shows the data source (68 marks, 17 rows by 4 columns, SUM(Sales): 12,642,902) and navigation buttons.

Tree Map

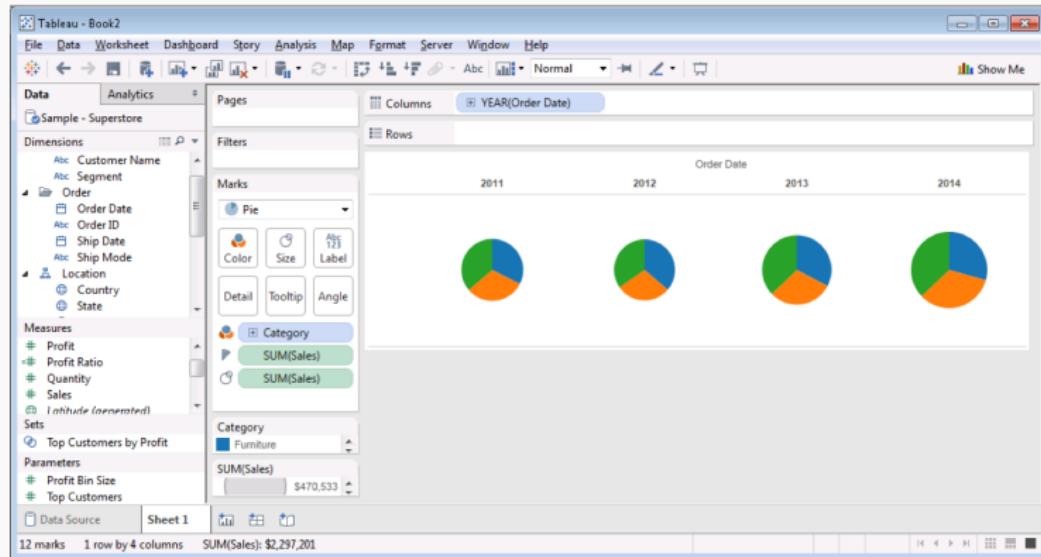
Tree maps are a way of displaying hierarchical data using nested figures, usually rectangles. To change the colours used, click 'Color' in the Marks panel and go to Edit Colors.



Bar Charts

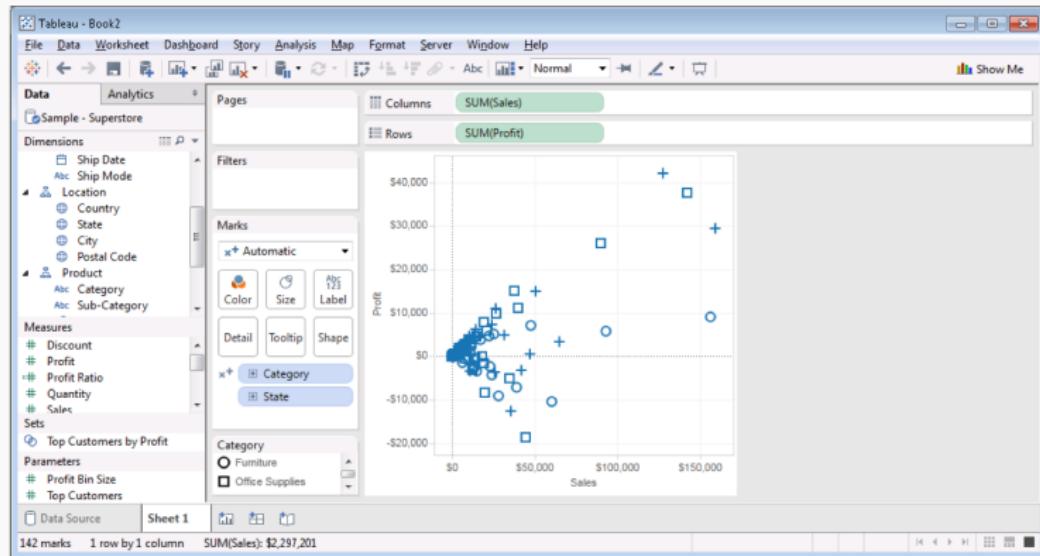


Pie Charts



Scatter Plots

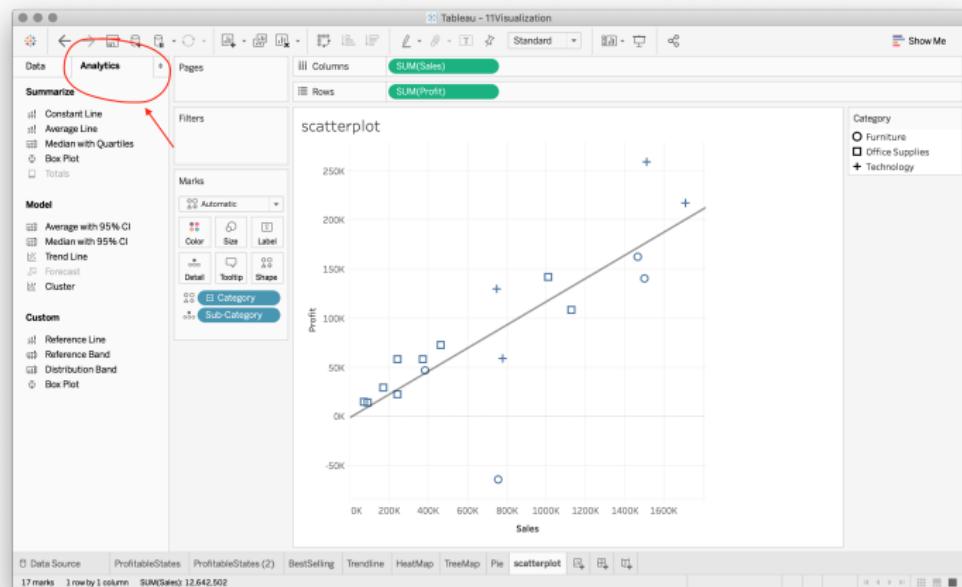
You can change the points that appear in your scatterplot by dragging pills into the **Details** button in the marks field.



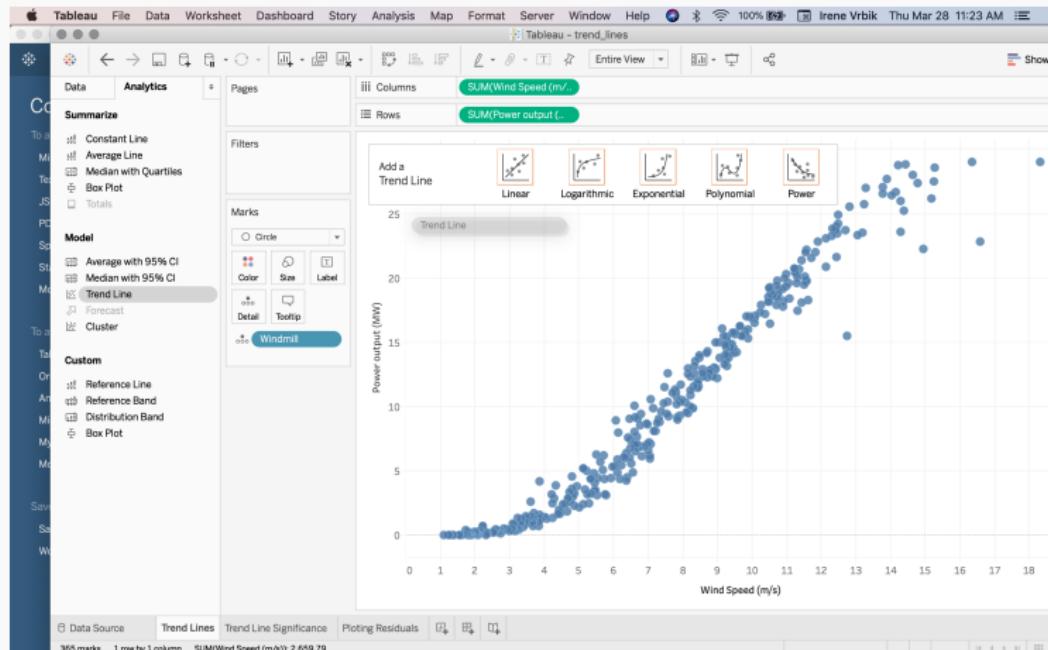
Trend Lines and Reference Lines

- ▶ Trend lines show patterns in data.
- ▶ Trend lines can be easily created by:
 - ▶ navigating the **Analytics** pane → **Model** → Trend Line (double click or drag).
 - ▶ or by right clicking on appropriate visualizations.
- ▶ These trend lines do not have to be linear.
- ▶ To remove the trend line, simple drag it off the viz (or press undo).
- ▶ If you hover over the trend line on your Viz, you can see some useful information (eg. p-value, R-squared) in the so-called "tooltip".

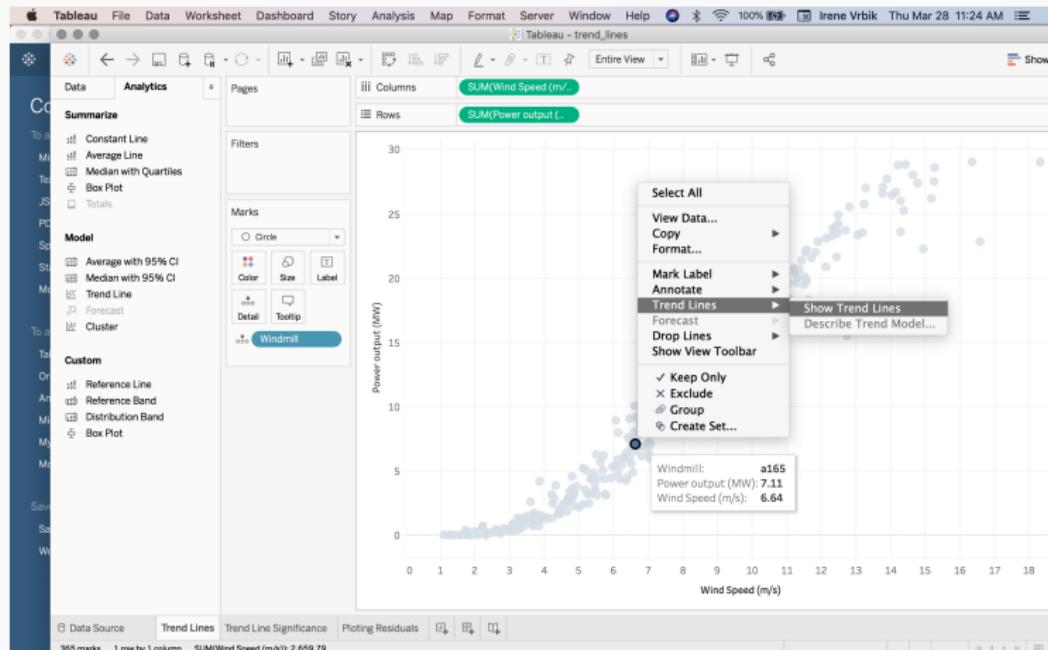
Adding Trend Lines



Adding Trend Lines



Adding Trend Lines



Trend Lines and Reference Lines

- ▶ We can get 'side-by-side' or colour-coded trendlines for multiple fields by dragging an additional dimension into the view.
- ▶ For example:
 - ▶ Navigate back to the Data pane
 - ▶ Drag **Location** onto the y-axis of the Viz (or the Row shelf) see [Figure 1](#)
 - ▶ Alternatively we could drag **Location** onto 'Color' on the **Marks** pane: see [Figure 2](#)

Adding Trend Lines

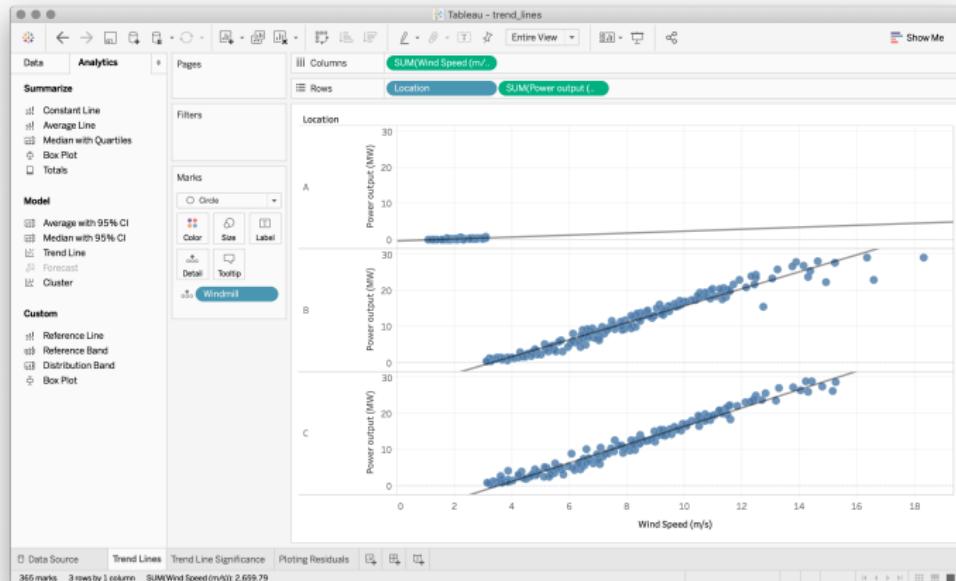


Figure: default

Adding Trend Lines

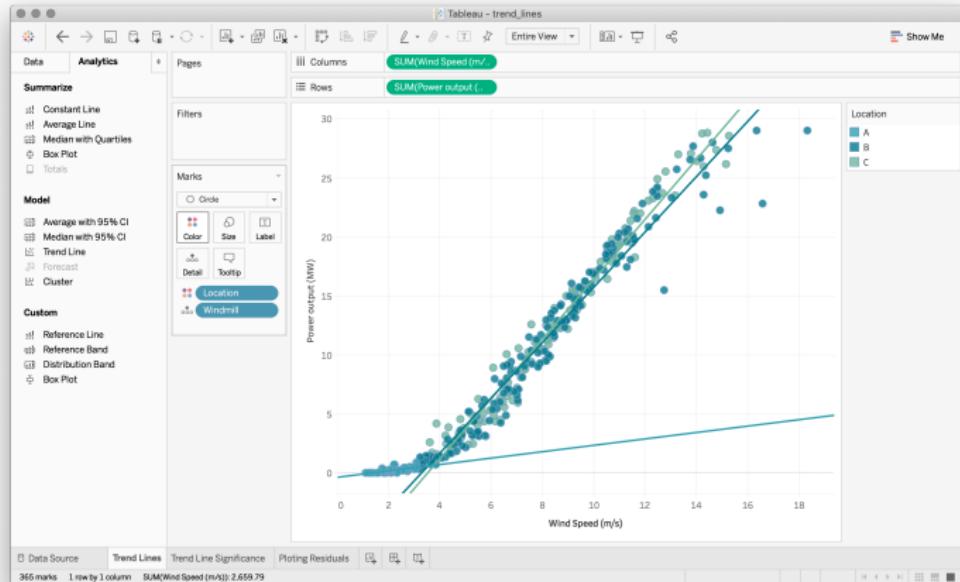
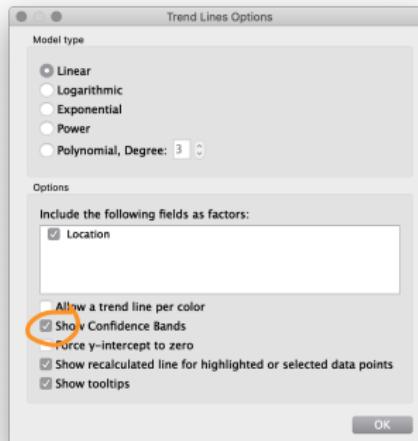


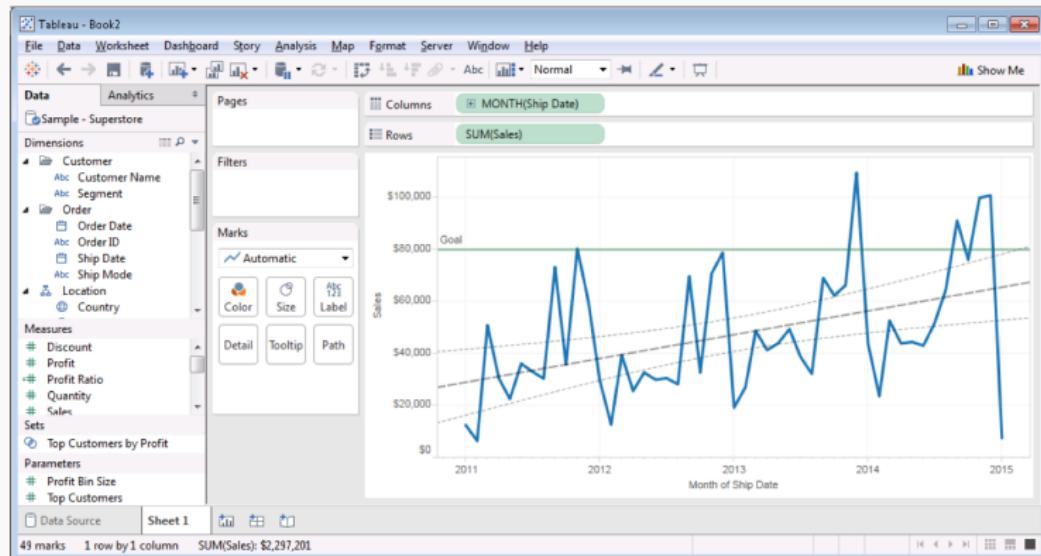
Figure: default

Reference Lines

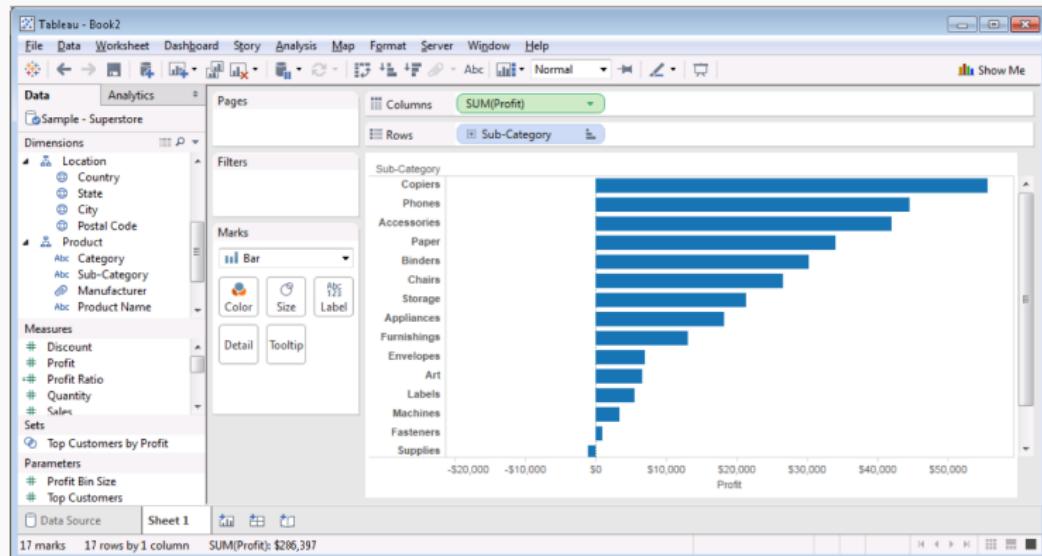
- ▶ Reference lines allow comparison with a reference (detect trends and outliers). To add a reference line:
 - ▶ go to the **Analytics** tab → **Custom** pane "Reference Line"
 - ▶ or right click on the y (or x) axis on the Viz and select **Add Reference Line**.
- ▶ To show the confidence bands, right click your trend line, **Trend Lines** → **Edit Trend Lines...** and select "Show Confidence Bands"



Adding Trend Lines and Reference Lines

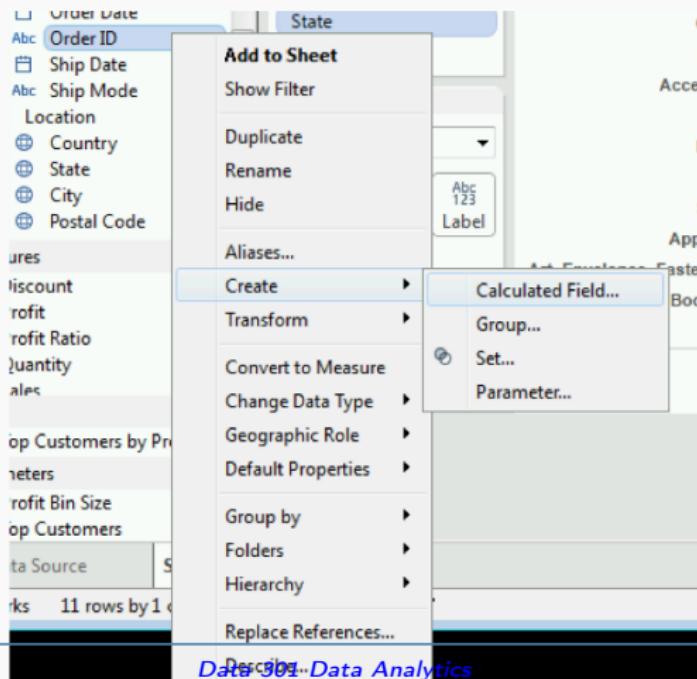


Sorting



Calculations

- ▶ Functions in tableau allow us to manipulate our data.
- ▶ We can save those calculations in a **calculated field**.
- ▶ To create a calculated field, simply right click in the data pane and select **Create → Calculated Field**

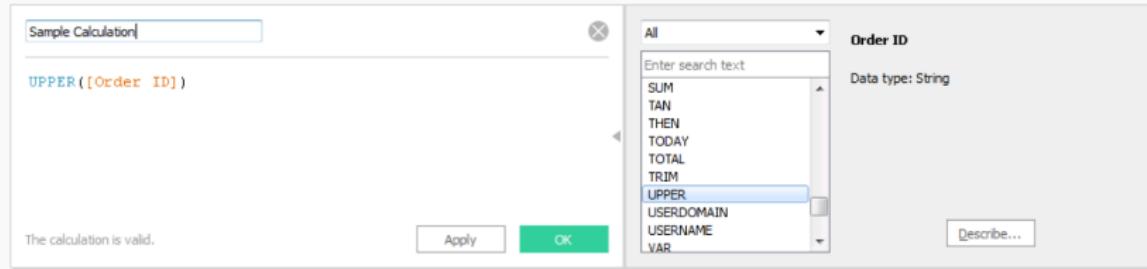


Calculations

- ▶ Available functions are displayed on the right hand panel, and our formulas (similar to what you would write in an Excel cell) is written on the left.
- ▶ You may search for functions within the right-hand panel.
- ▶ Functions work very much the same as in Excel, only now instead of referencing numbers be cell we call on them by name using [PillName].
- ▶ Notice that the name of our new calculated field will be prefaced by an equal sign 
- ▶ Now, this calculated field may be treated as any other pill.

Creating a Calculated Field

An example calling a function on a String



Creating a Calculated Field

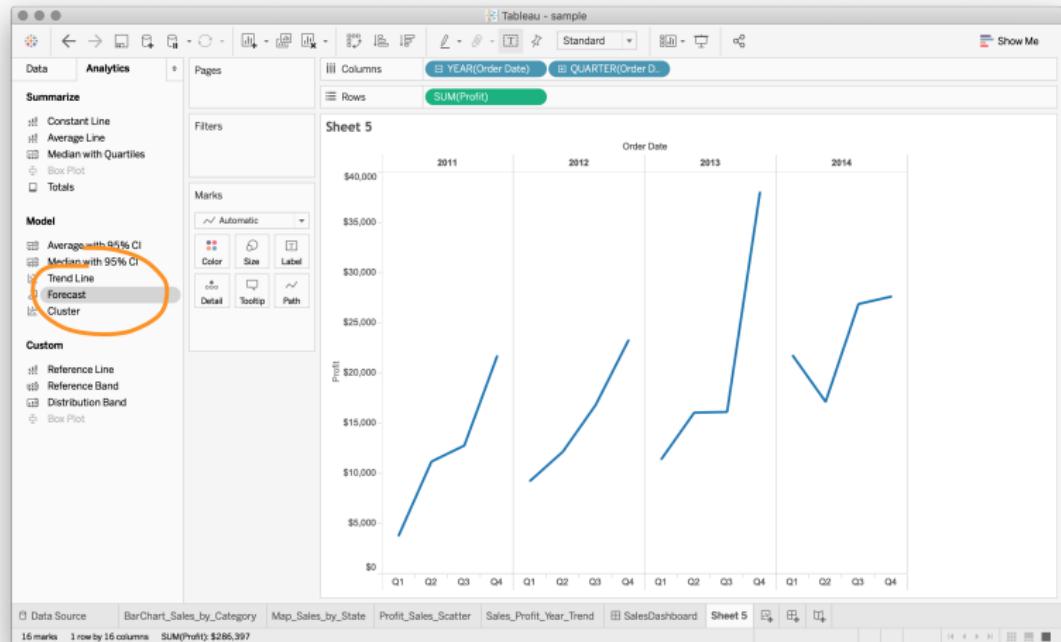
An example calling a function on a Number

The screenshot shows the Tableau Data Editor interface. In the top navigation bar, 'Tableau - 11 Visualization' is selected. The left sidebar contains 'Data' and 'Analytics' sections, with 'dept (WorksOn)' and 'dept+ (WorksOn)' under 'Data'. The 'Dimensions' section includes 'Customer Name', 'Market', 'Order Date', 'Order ID', 'Order Priority', 'Postal Code', 'Product ID', 'Product Name', 'Region', 'Row ID', 'Segment', 'Ship Date', 'Ship Mode', 'State', 'Sub-Category', and 'Measure Names'. The 'Measures' section includes '# Discount', '# Profit', '# Quantity', '# Sales', '# Shipping Cost', '(Latitude generated)', '(Longitude generated)', and '# Number of Records'. The main workspace shows a chart titled 'Ship Date' with a single data point at 'January' for 'Sales' (400K). A calculated field dialog is open, showing the formula `[Sales]+[Profit]`. The right side of the dialog lists various functions: ABS, ACOS, AND, ASCII, ASIN, ATAN, ATAN2, ATTR, AVG, CASE, CEILING, etc. The status bar at the bottom indicates '12 marks 1 row by 12 columns SUM(Sales): 12,642,502'.

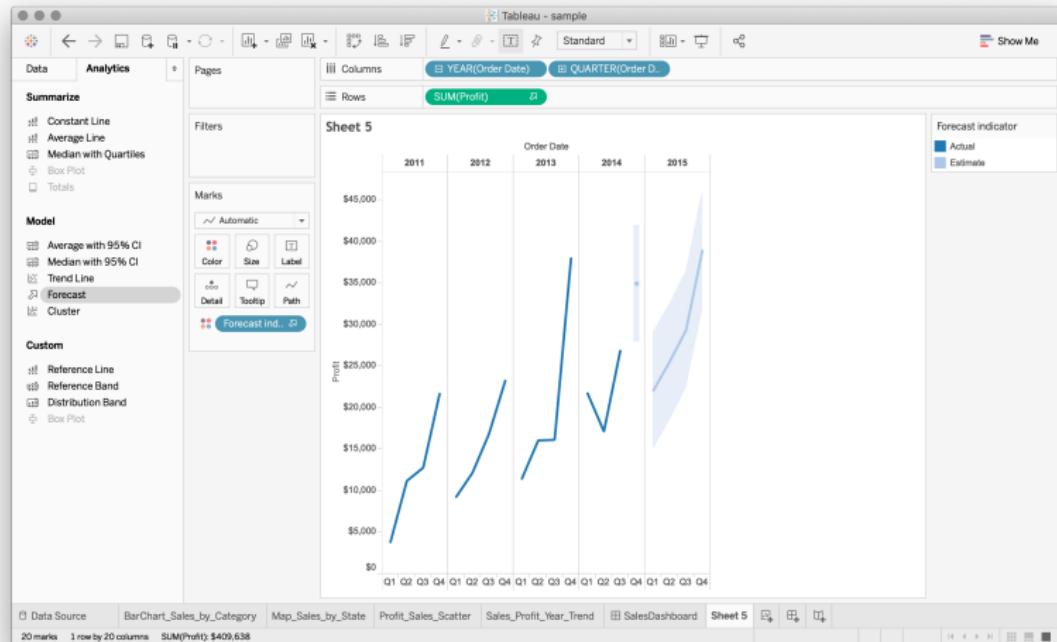
Forecasting

- ▶ Forecasts use statistical models to generate predictions for future data based on historical information.
- ▶ To build a forecast in Tableau, we need a *Date Dimension* pill and a Measure (i.e some green pill).
- ▶ Once the pills are dragged to the appropriate shelf, navigate to the Analytics tab, and select the Forecast option.
- ▶ By default, this will create a forecast prediction curve, alongside prediction bands.

Forecasting



Forecasting



Forecasting

Alternatively, we can right click on an appropriate Viz and select **Forecast → Show Forecast**

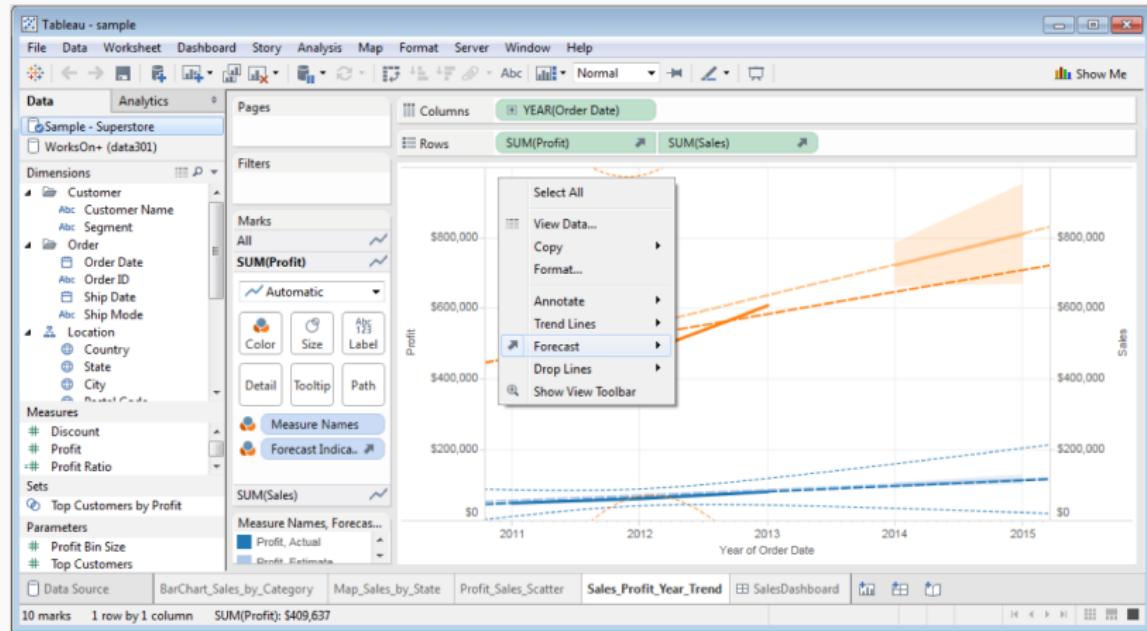


Tableau Question

Example

How many of the following statements are TRUE?

1. There can only be one pill on the row shelf.
2. A trend line can only be linear.
3. A user can group multiple items into a group in the visualization.

A) 0

B) 1

C) 2

D) 3

Tableau Question

Answer

How many of the following statements are TRUE?

1. There can only be one pill on the row shelf. **X**
2. A trend line can only be linear.
3. A user can group multiple items into a group in the visualization.

A) 0

B) 1

C) 2

D) 3

Tableau Question

Answer

How many of the following statements are TRUE?

1. There can only be one pill on the row shelf. **X**
2. A trend line can only be linear. **X**
3. A user can group multiple items into a group in the visualization.

A) 0

B) 1

C) 2

D) 3

Tableau Question

Answer

How many of the following statements are TRUE?

1. There can only be one pill on the row shelf. **X**
2. A trend line can only be linear. **X**
3. A user can group multiple items into a group in the visualization. **✓**

- A) 0 B) 1 C) 2 D) 3**

Try it: Tableau Charts

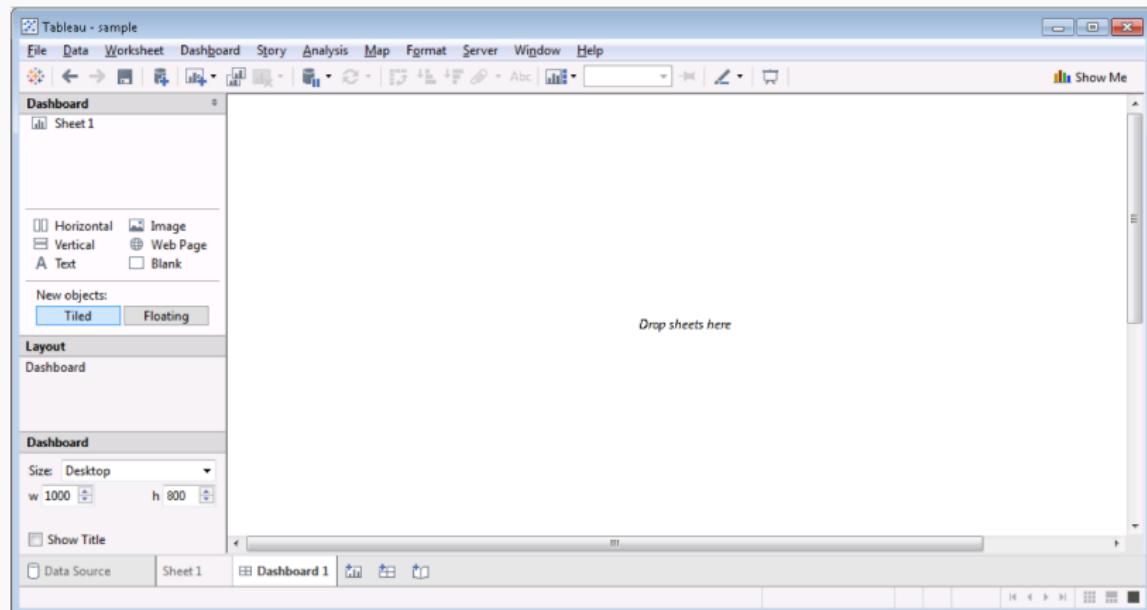
Using the Superstore data set, create a visualization for each of these chart types:

- ▶ line chart (with forecast and trend line)
- ▶ bar chart (with filters and sorting)
- ▶ pie chart (with a parameter)
- ▶ heat map (with grouping)
- ▶ scatter plot (with a calculated field)
- ▶ histogram
- ▶ circle view

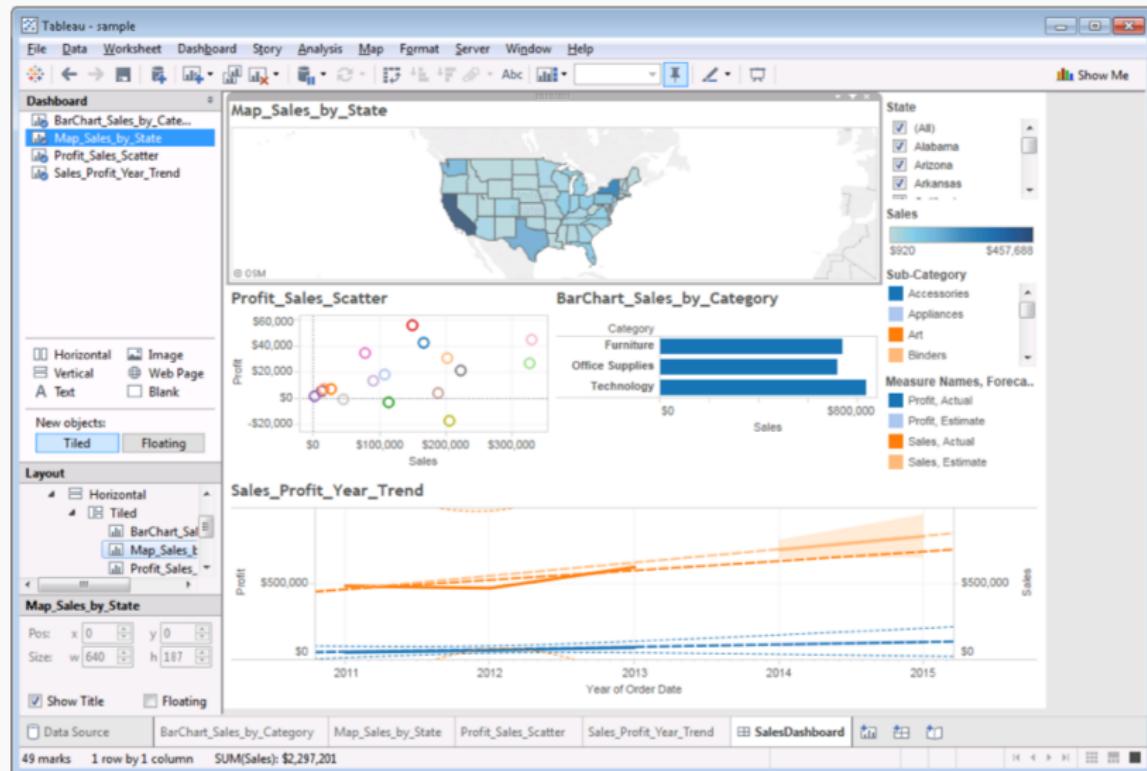
Dashboards

- ▶ A **dashboard** consists of multiple sheets organized to make information and its relationships more understandable.
- ▶ Tableau recommendation: 4-pane dashboard designs

Dashboard Starter View



Dashboard Populated with Worksheets



Try it: Tableau Dashboard

Example

Using the Superstore data set, create your own dashboard with multiple visualizations.

Conclusion

- ▶ Tableau is a software system for visualizing data sets from multiple sources using a wide-range of visualization techniques.
 - ▶ line charts, bar charts, scatter plots, heat maps, pie charts, histograms
- ▶ Visualization of data sets is critical for communicating meaning and understanding, especially for people with less understanding of the data set.

Objectives

- ▶ Explain the purpose of visualization
- ▶ List different types of visualizations available in Excel, Python, R, GIS
- ▶ List the three "types of data"
- ▶ Define: pill, shelf, view card (as used in Tableau)
- ▶ Explain the purpose of the Show Me button
- ▶ Be able to connect to Excel and relational databases using Tableau
- ▶ Compare/contrast connecting to versus extracting data with Tableau
- ▶ List and explain the different Tableau file types
- ▶ Define and compute: inner join, left outer join, right outer join, full outer join
- ▶ Use dynamic grouping and renaming to clean and correct data values in a visualization

Objectives

- ▶ List and use the different Tableau chart types: text tables, maps, heat maps, tree maps, line charts, pie charts, area charts, scatter plot, circle view, histogram, Gantt charts
- ▶ Add trend lines, references lines, quantiles to a visualization
- ▶ Create and use hierarchies
- ▶ Create and use filters
- ▶ Create calculated fields
- ▶ Use parameters to allow user-controlled visualizations
- ▶ Add forecasts to a visualization
- ▶ Organize visualizations into a dashboard