

# DATA 301 Introduction to Data Analytics Visualization

Dr. Ramon Lawrence  
University of British Columbia Okanagan  
ramon.lawrence@ubc.ca

## 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 charts, and R.

A data visualization package is selected based on its ability to effectively communicate the information to end users and the simplicity in creating the visualizations.

There is no one perfect software package for data visualization as you must trade-off experience, time, and appearance.

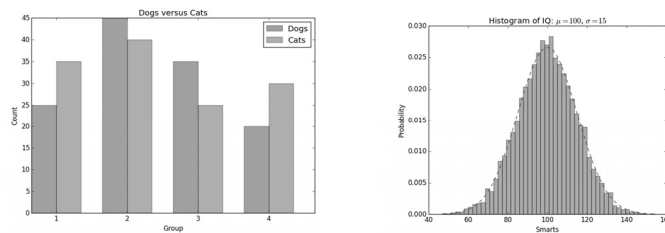
## Data Visualization in Excel

Charts including pivot charts, spark lines, and visual formatting of cells



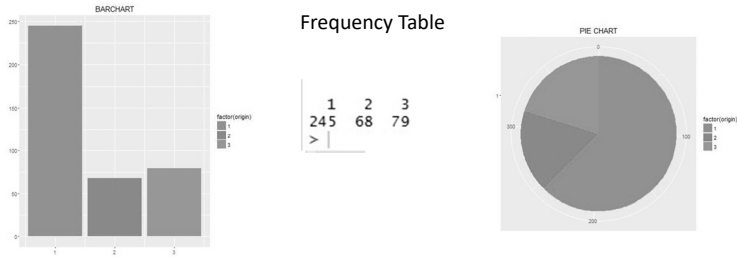
## Data Visualization in Python

Variety of charting libraries including matplotlib and ggplot



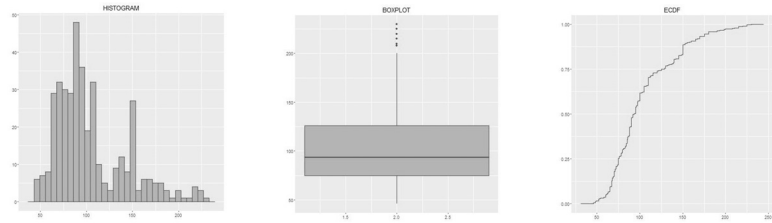
## Data Visualization in R - Qualitative Data

Qualitative data: bar chart, frequency table, pie chart



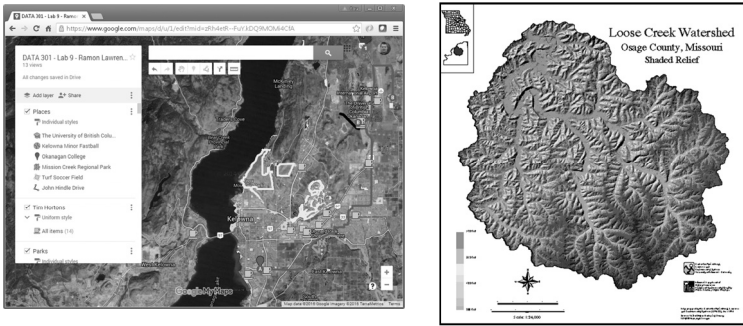
## Data Visualization in R - Quantitative Data

Quantitative data: histogram, box plot, ECDF



## Data Visualization in GIS

Maps of coordinates with overlays (markers, points, lines, regions)



## Types of Data

Data can be considered of three types:

- 1) **Known data** - monitoring and regular reporting data for visibility
- 2) **Data You Know You Need** - for understanding outliers or trends in the known data and deciding on how to act on them
- 3) **Data You Need but Do Not Know It** - information that you would have not thought about but knowing it would be very valuable (data to discover!)

Visual analytics helps with all three types of data, but especially the last two to understand trends and discover important information.

## Introduction to Tableau

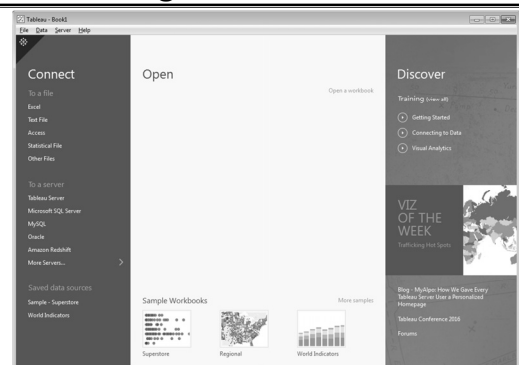
**Tableau** (<http://www.tableau.com/>) was founded in 2003 as a spin-off from Stanford University by Chris Stolte, Christian Chabot and Pat Hanrahan.

- 2015 revenue was over \$650 million with over 3000 employees

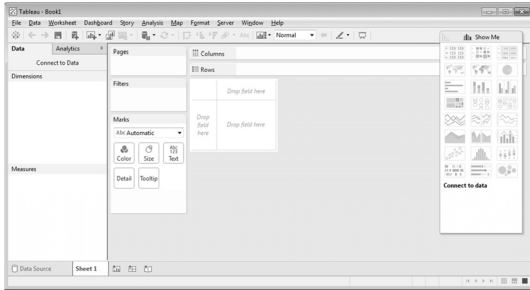
The goal of Tableau is "to help people see and understand their data."  
- Christian Chabot, Tableau CEO

Tableau has desktop and server (enterprise) products as well as Tableau Public allowing sharing of data sets.

## Tableau - Home Page



## Tableau Workspace



## Tableau Features

Supported data types: text, dates, numbers, geographical coordinates (latitude/longitude), Boolean

Aggregation functions: sum, average, max, count, variance, etc.

Many built-in functions for numeric and string manipulation.

Calculated fields can be created and are preceded by an =.

## Tableau Terminology

A **pill** is a dimension, attribute, or measure that can be placed in the visualization.

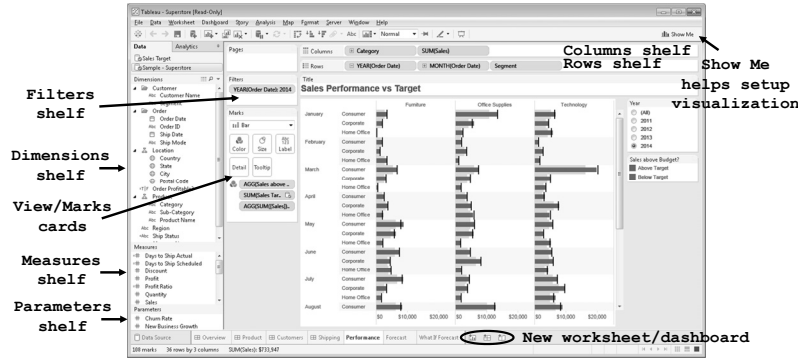
- Blue pills are discrete. Green pills are continuous.



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.

## Tableau Workspace Items



## View Cards

View or shape cards allows control of color, 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/tooltips—makes fields available to tooltips without disaggregating data
- Shape—expresses discrete or continuous fields

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

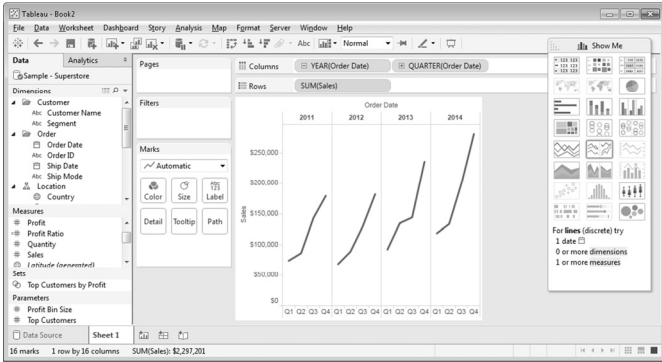
## Show Me Button

The Show Me button suggests visualization to use based on your current dimensions and measures.

It will also place pills on shelves automatically.



## Tableau Visualization: Show Me



## Tableau Question

**Question:** How many of the following statements are TRUE?

- 1) In Tableau blue pills are continuous.
- 2) The View Cards interface allows for changing color and size of features in the visualization.
- 3) A shelf is a location to place a pill.
- 4) The Show Me button will suggest visualizations for you.
- 5) 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

- 1) Install Tableau. Use trial version or student license provided in Connect.
- 2) Start Tableau. Use the sample.twbx file or the Superstore example and explore the visualizations.
- 3) Try create any visualization of the data.

## Tableau - Data Sources

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

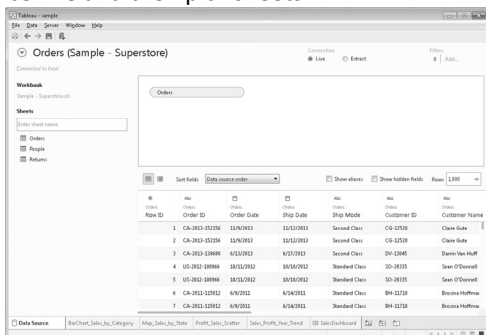
- Microsoft Excel and Access
- Text files (txt, csv)
- Relational databases (MySQL, SQL Server, Oracle, PostgreSQL)
- NoSQL databases (MongoDB)
- Parallel and analytical databases (Greenplum, Vertica, Teradata)
- Other ODBC sources (note JDBC is not supported)

A sample data source called Superstore is available in the Tableau/defaults/Datasources directory.

- File: Sample - Superstore.tds (Tableau Data Source) or
- File: Sample - Superstore.xls (Excel file)

## Example Connecting to Excel

Select Excel file and then pick sheets.



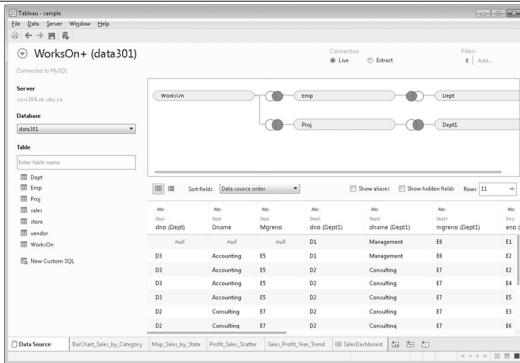
## Example Connecting to MySQL

Connecting to a relational database like MySQL requires:

- 1) Driver (often need to download from database vendor)
  - <https://www.tableau.com/support/drivers>
- 2) Database connection information



## Example Connecting to MySQL (2)



## 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 times to extract
  - Portability (as consumer of report does not need access to data source)
  - Support for functions not supported by source (e.g. Excel)

## Tableau Data Sources Question

**Question:** How many of the following statements are **TRUE**?

- 1) Tableau can connect to many relational databases.
- 2) Tableau can process data in text and Excel files.
- 3) Tableau can either leave data in data source or extract it locally.
- 4) Tableau can connect to data sources using JDBC.
- 5) 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

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 from: <https://dev.mysql.com/downloads/connector/odbc/>
- Server: cosc304.ok.ubc.ca Database: data301 User: data301 Password: ubc

Superstore visualizations:

- Map showing profit by state.
- Visualization to indicate what is the best selling product category per store.

WorksOn visualizations:

- Visualize the number of projects, employees, and hours worked per department.
- Visualize employee ages to see if age impacts if they are supervisors.

## 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

Note similarities with Excel/spreadsheet terminology.

## 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

### Join Example

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

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

### Join Question

**Question:** Given these tables, how many rows are in the result of Boys LEFT OUTER JOIN Girls ON Bid=Gid?

Boys		Girls	
Bid	BoyName	Gid	GirlName
1	Joe	1	Jane
1	Steve	1	Sarah
3	Fred	5	Fran
5	James	6	Julie
7	Ben		
7	Bishop		

- A) 9      B) 8      C) 7      D) 6      E) 0

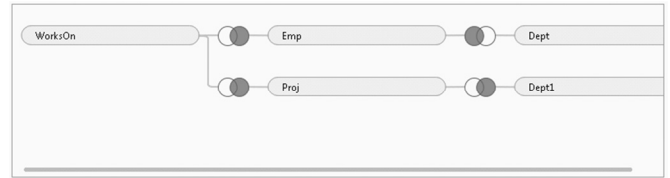
### Data Blending

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:



Create a visualization with this data set.

### Dynamic Grouping/Renaming

Dynamic grouping (also called ad hoc groups) can be created by using Ctrl+Select to select elements in visualization and select Group from menu.

It is also possible to rename values/labels and correct value errors.

### Geographic Data

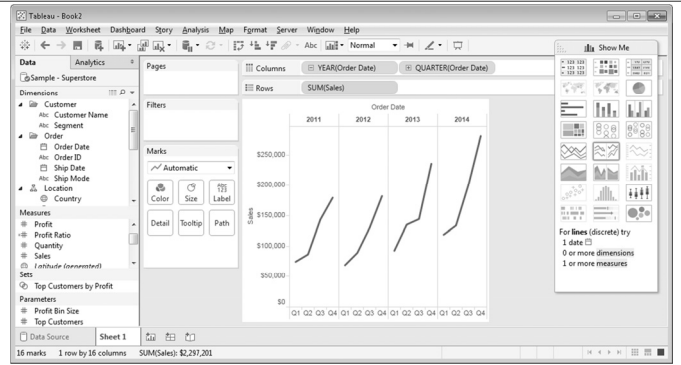
For geographic data (small globe icon), Tableau automatically generates center-point geocodes (longitude/latitude).

## Tableau Chart Types

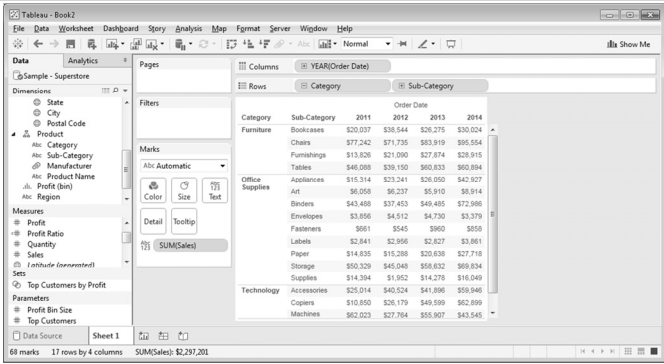
Chart types:

- 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

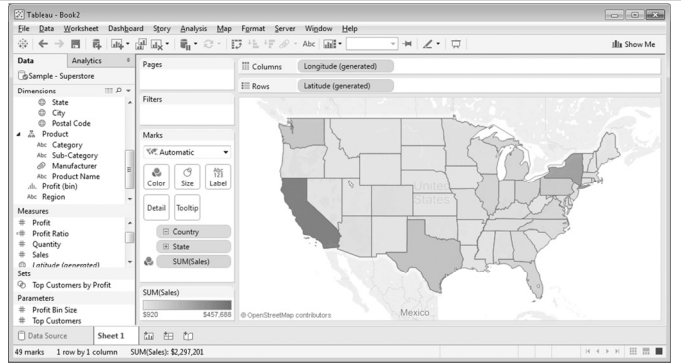
## Line Chart (discrete time)



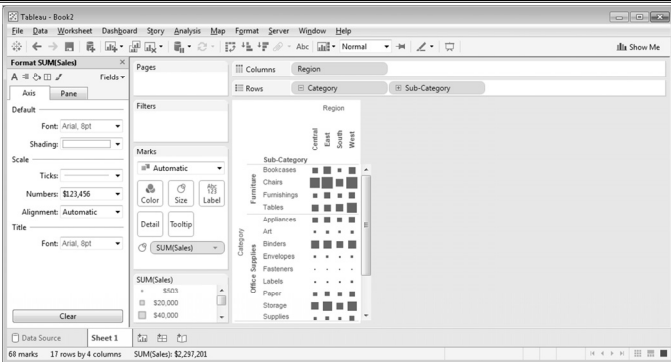
## Text Table (Crosstab)



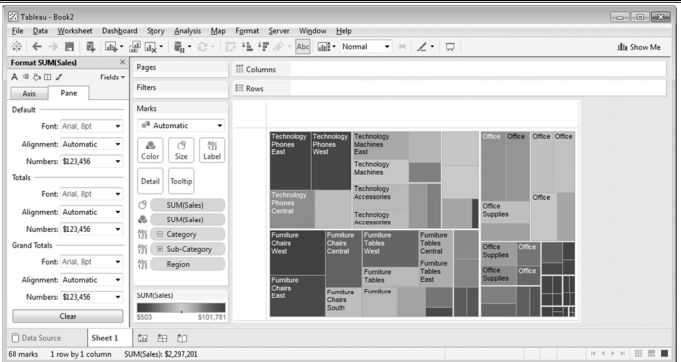
## Maps



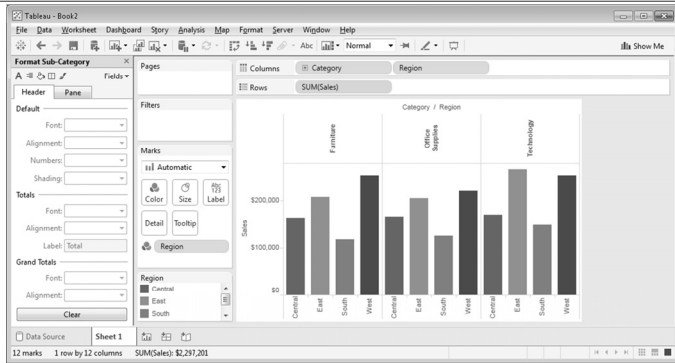
## Heat Map



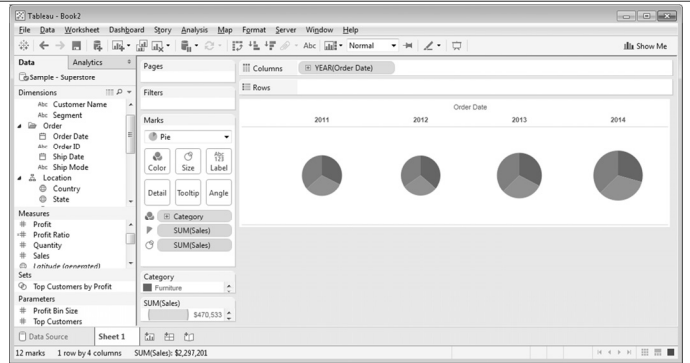
## Tree map



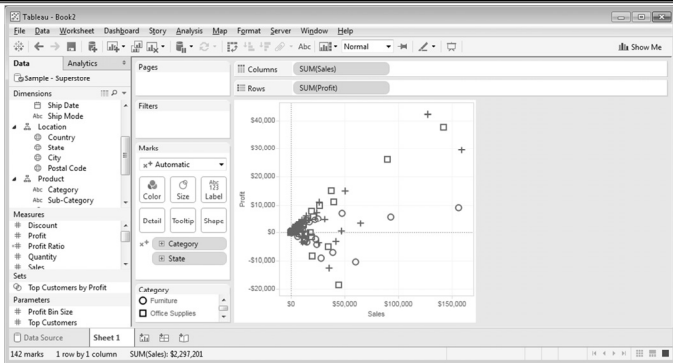
## Bar Charts



## Pie Charts



## Scatter Plots

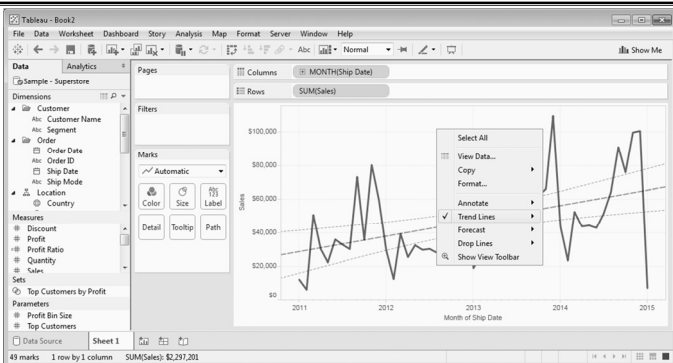


## Trend Lines and Reference Lines

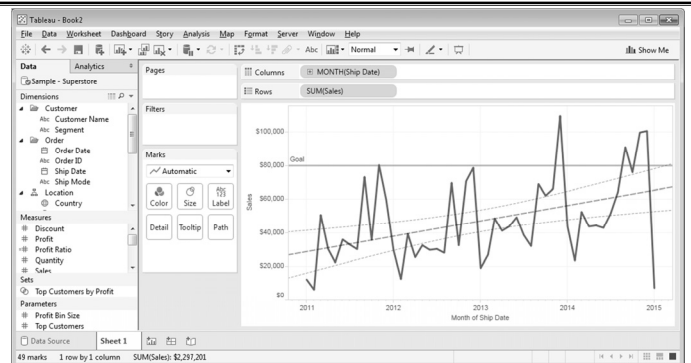
Trend lines show patterns in data using a line of best fit.

Reference lines allow comparison with a reference (detect trends and outliers).

## Adding Trend Lines

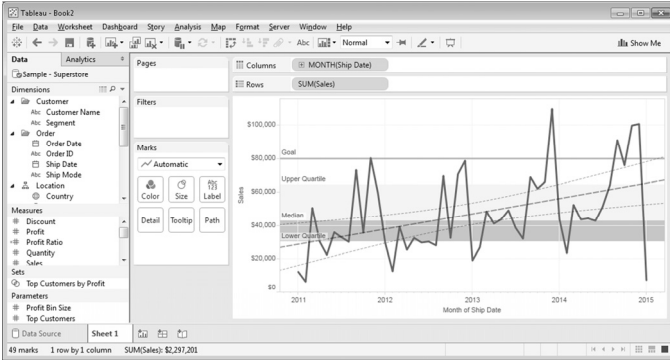


## Trend Lines and Reference Lines

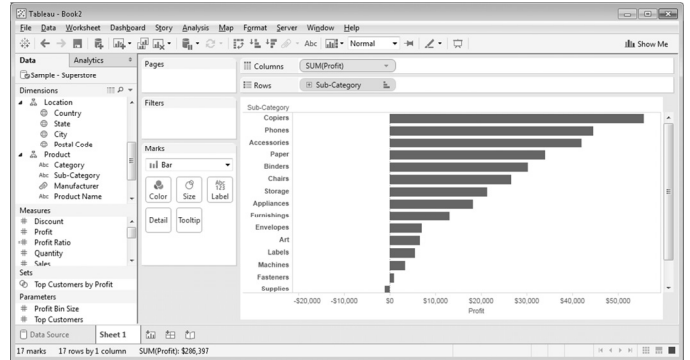




## Adding Quantiles



## Sorting



## Hierarchies

**Hierarchies** are groupings of data that make it easier to roll-up and drill-down into data.

Examples:

- category and subcategory
- year, quarter, month
- country, state, city

Can create own hierarchies by dragging dimensions on top of each other.

## Filters

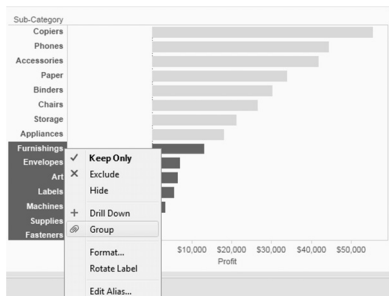
There are multiple ways to define filters:

- 1) Drag dimension into filter shelf
- 2) Quick filters allow people using report to change filters dynamically. (click on item in filter shelf and select Show Filter option)

## Grouping

Grouping allows summarizing data without using a hierarchy.

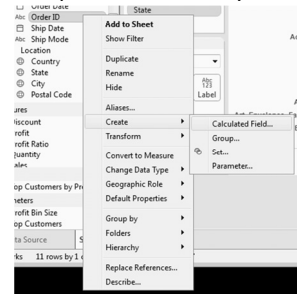
- Multi-select elements then in pop-up menu select Group



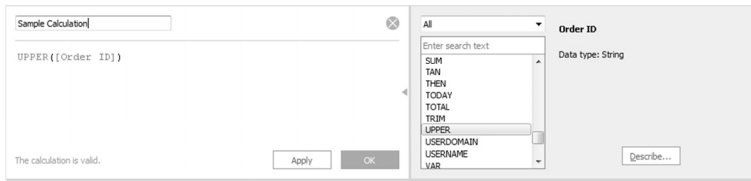
## Calculations

**Calculated fields** are performed on data source when possible.

**Table calculations** are performed locally in Tableau.



## Creating a Calculated Field



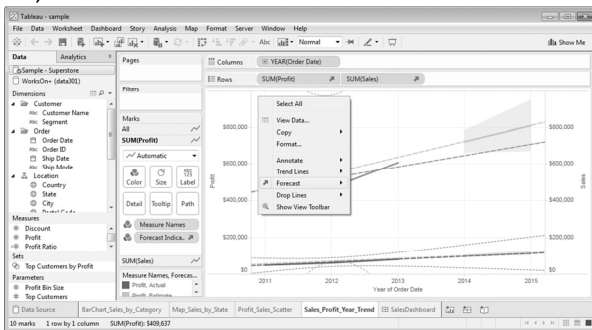
## Parameters

Calculations may have parameters.

Parameters may be exposed in the visualization so the user can control them.

## Forecasting

Right click, select Forecast then Show Forecast.



## Tableau Charts Question

**Question:** 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.
- 4) Calculated fields are calculated on the data source if possible.
- 5) Filters may be exposed to the user of the visualization just like parameters.

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

## Try it: Tableau Charts

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

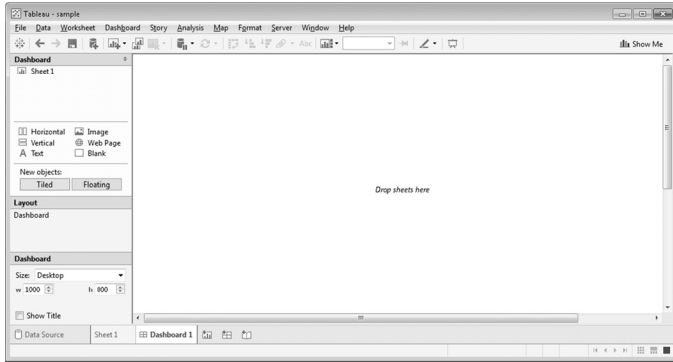
- line chart (with forecast and trend line)
- bar chart (with filters and sorting)
- pie chart (with a parameter)
- heat map (with grouping)
- scatter plot (with a calculated field)
- histogram
- circle view

## Dashboards

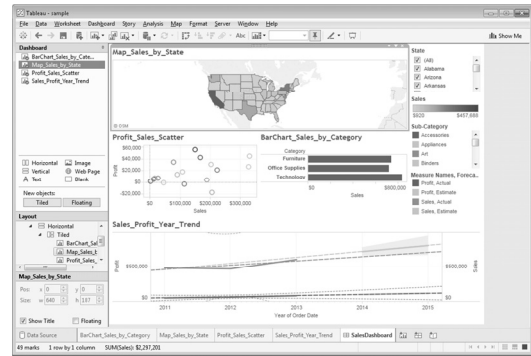
A **dashboard** consists of multiple sheets organized to make information and its relationships more understandable.

Tableau recommendation: 4-pane dashboard designs

## Dashboard Starter View



## Dashboard Populated with Worksheets



## Try it: Tableau Dashboard

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 (2)

- 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