

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.

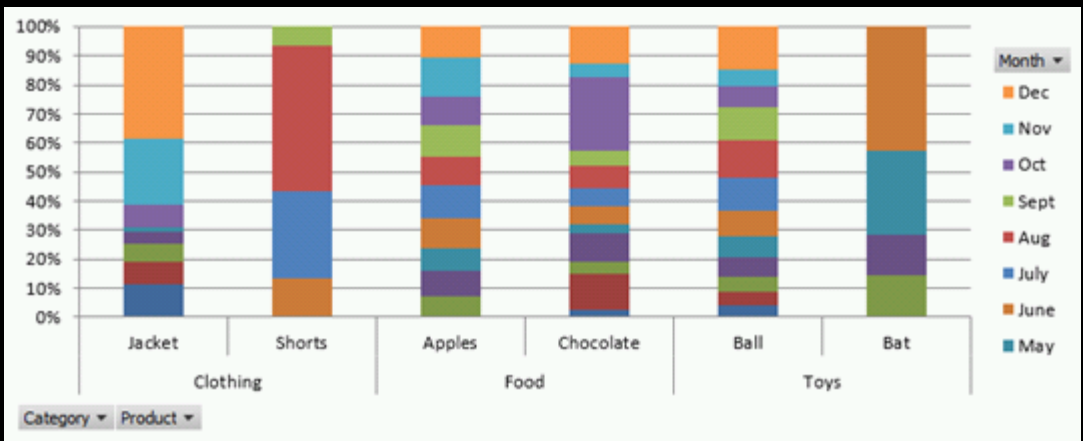
There is no one perfect software package for data visualization as you must trade-off experience, time, and appearance.

Data Visualization in Excel

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

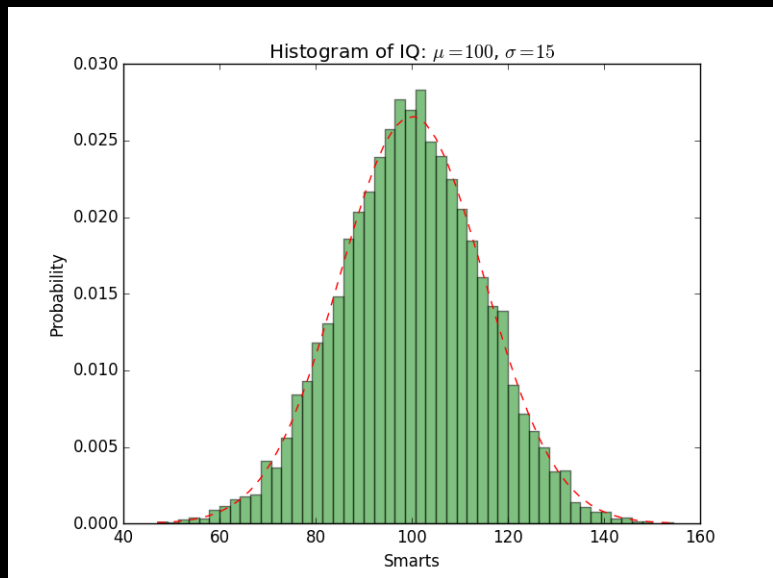
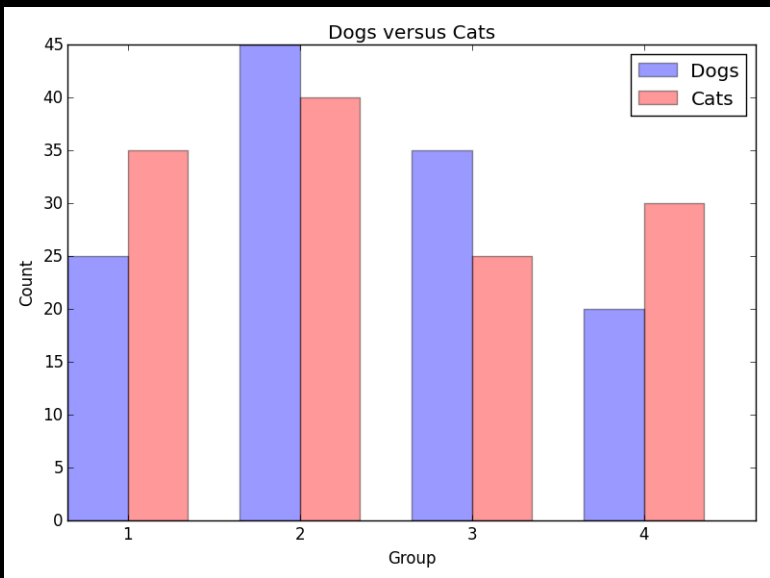
	A	B	C	D	E	F	G
1	Category	Product	Month	Volume	Price	Cost	Revenue
2	Food	Chocolate	Jan	20	\$ 2.00	\$ 1.00	\$ 40.00
3	Clothing	Jacket	Jan	15	\$50.00	\$35.00	\$ 750.00
4	Toys	Ball	Jan	55	\$ 1.00	\$ 0.50	\$ 55.00
5	Food	Chocolate	Feb	80	\$ 2.50	\$ 1.00	\$ 200.00
6	Clothing	Jacket	Feb	10	\$50.00	\$35.00	\$ 500.00
7	Toys	Ball	Feb	65	\$ 1.00	\$ 0.60	\$ 65.00
8	Food	Chocolate	Mar	30	\$ 2.00	\$ 1.00	\$ 60.00
9	Toys	Ball	Mar	70	\$ 1.00	\$ 0.40	\$ 70.00
10	Toys	Bat	Mar	10	\$75.00	\$50.00	\$ 750.00
11	Clothing	Jacket	Mar	8	\$50.00	\$30.00	\$ 400.00
12	Food	Apples	Mar	100	\$ 3.00	\$ 2.00	\$ 300.00
13					Total:		\$3,190.00

Monthly Sales



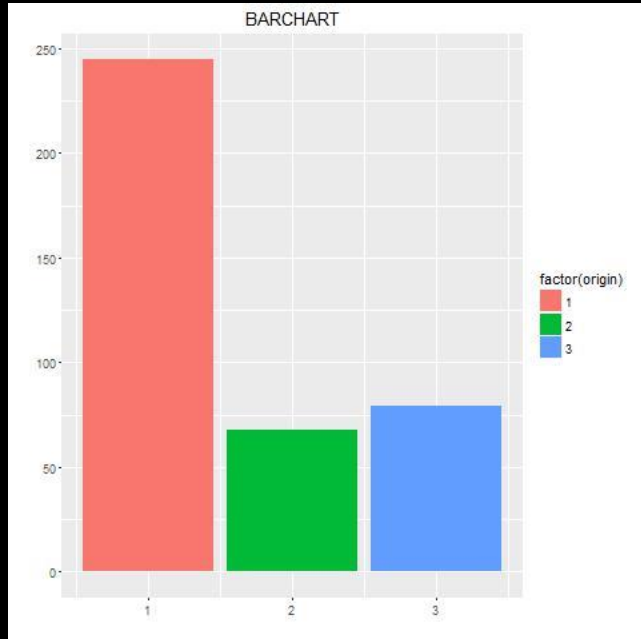
Data Visualization in Python

Variety of charting libraries including matplotlib and ggplot



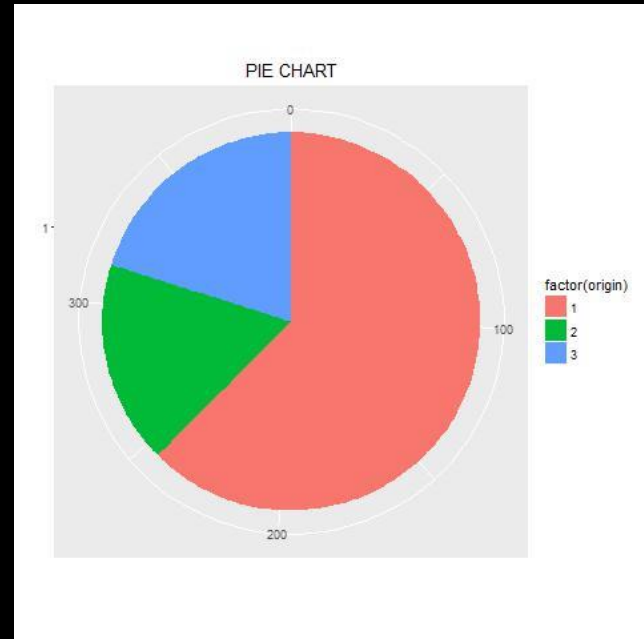
Data Visualization in R - Qualitative Data

Qualitative data: bar chart, frequency table, pie chart



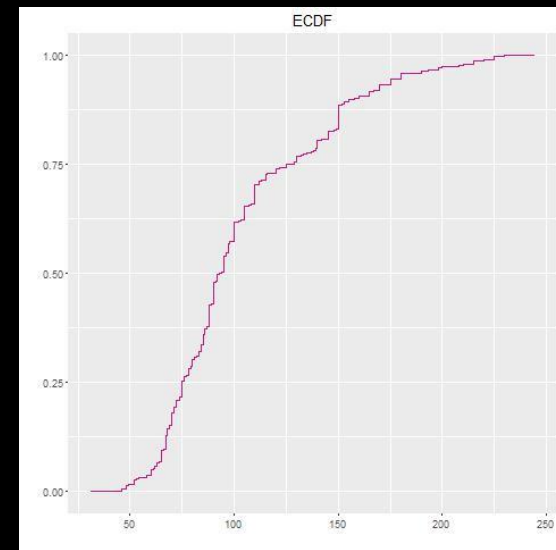
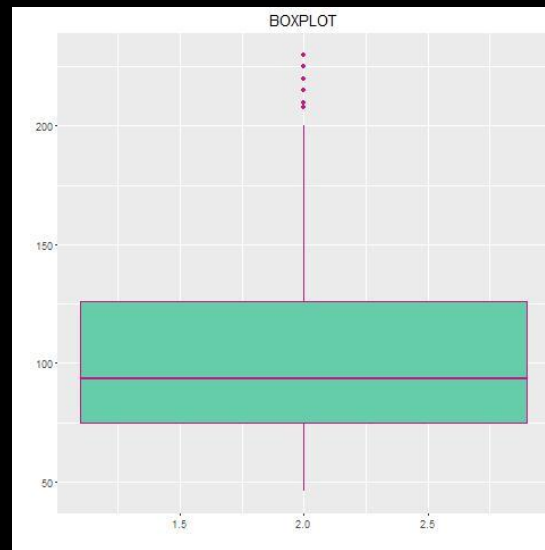
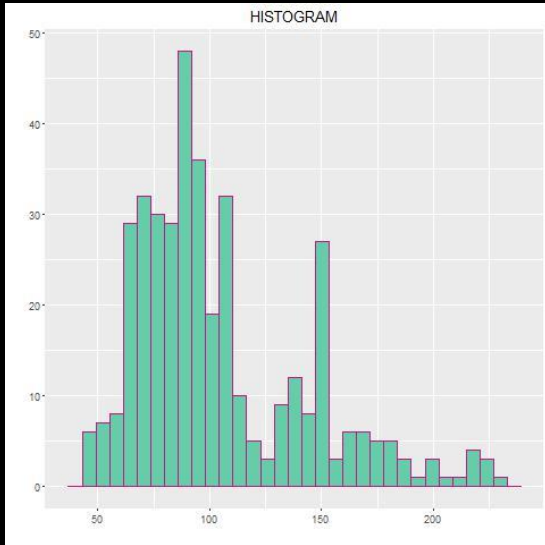
Frequency Table

```
  1   2   3
245  68  79
> |
```



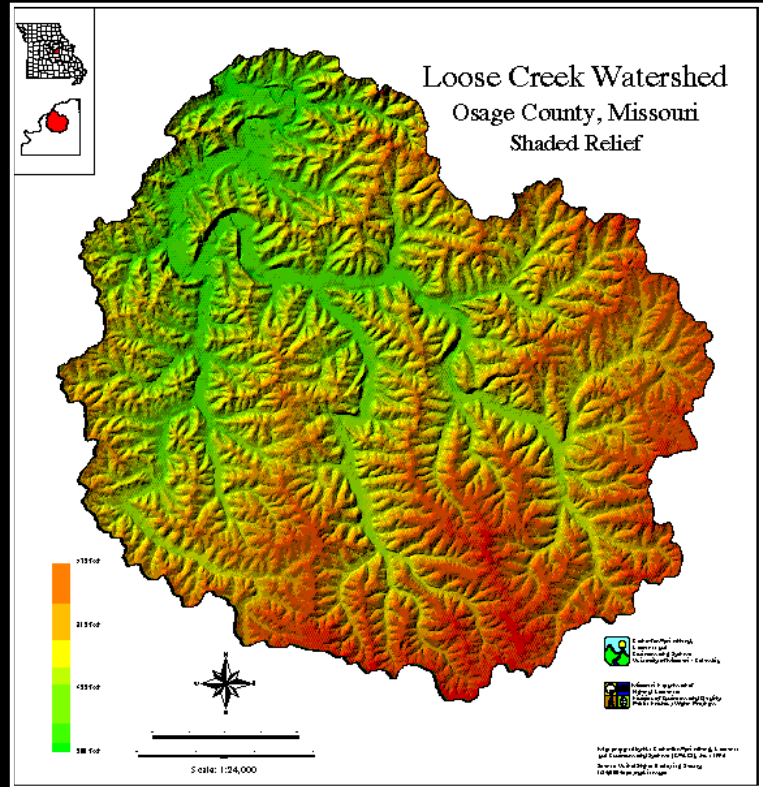
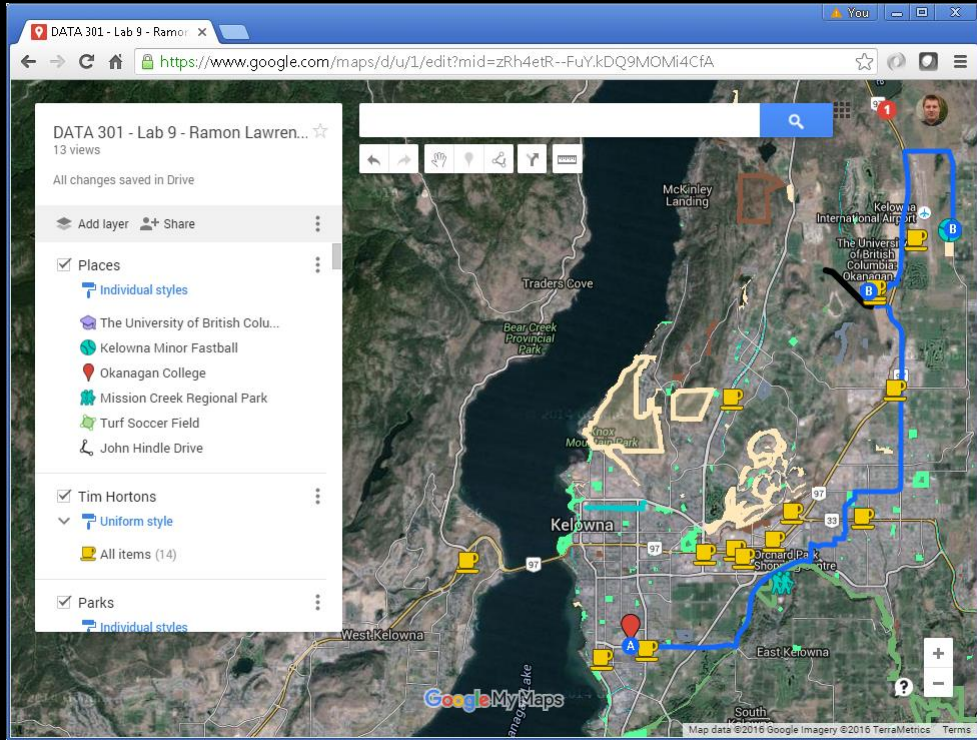
Data Visualization in R - Quantitative Data

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

The screenshot shows the Tableau Desktop home page with the following layout:

- Connect Panel (Left):**
 - To a file:** Excel, Text File, Access, Statistical File, Other Files.
 - To a server:** Tableau Server, Microsoft SQL Server, MySQL, Oracle, Amazon Redshift, More Servers... >
 - Saved data sources:** Sample - Superstore, World Indicators.
- Open Panel (Center):**
 - Open a workbook
 - Sample Workbooks:** Superstore (scatter plot), Regional (map of US), World Indicators (bar chart). Includes a [More samples](#) link.
- Discover Panel (Right):**
 - Training (view all):** Getting Started, Connecting to Data, Visual Analytics.
 - VIZ OF THE WEEK:** Trafficking Hot Spots (with world map).
 - Blog - MyAlpo:** How We Gave Every Tableau Server User a Personalized Homepage.
 - Tableau Conference 2016**
 - Forums**

Tableau Workspace

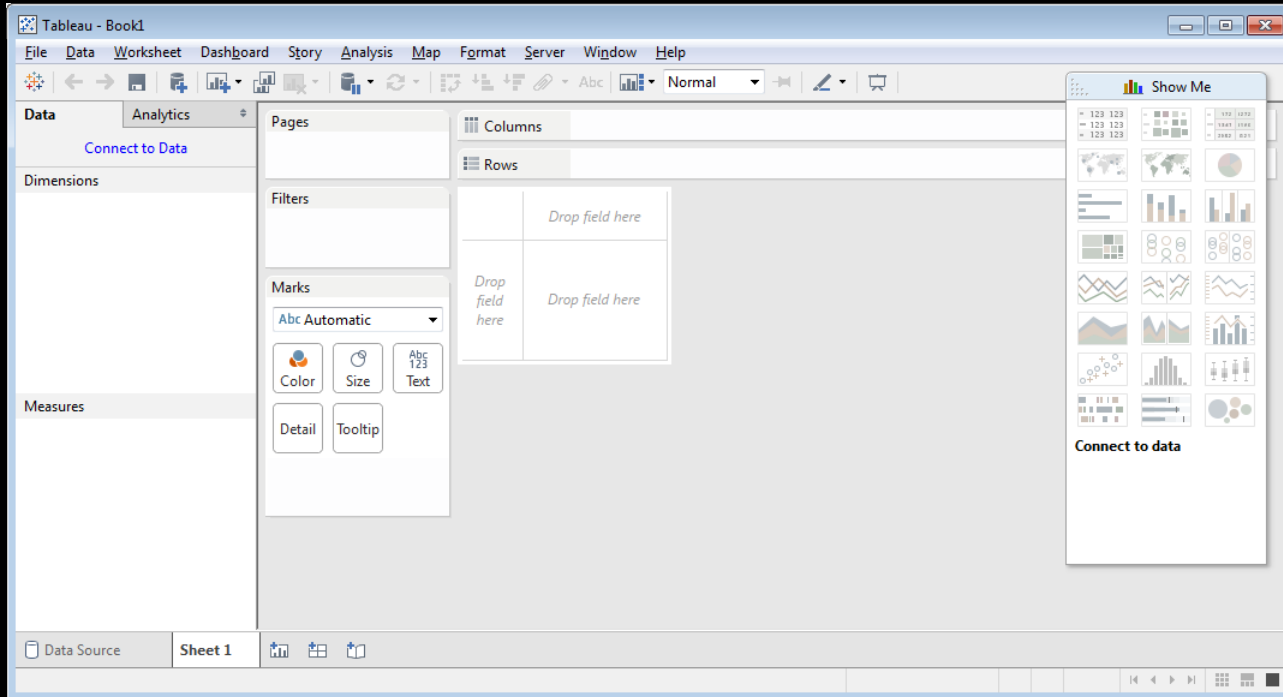


Tableau Features

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

Filters shelf →

Dimensions shelf →

View/Marks cards →

Measures shelf →

Parameters shelf →

Columns shelf

Rows shelf

Show Me helps setup visualization

New worksheet/dashboard

Sales Performance vs Target

Year	Month	Customer	Category	Actual Sales	Target
2014	January	Consumer	Furniture	\$5,000	\$10,000
		Corporate	Office Supplies	\$15,000	\$10,000
		Home Office	Technology	\$5,000	\$10,000
	February	Consumer	Furniture	\$5,000	\$10,000
		Corporate	Office Supplies	\$5,000	\$10,000
		Home Office	Technology	\$5,000	\$10,000
	March	Consumer	Furniture	\$10,000	\$10,000
		Corporate	Office Supplies	\$20,000	\$10,000
		Home Office	Technology	\$5,000	\$10,000
	April	Consumer	Furniture	\$5,000	\$10,000
		Corporate	Office Supplies	\$5,000	\$10,000
		Home Office	Technology	\$5,000	\$10,000
	May	Consumer	Furniture	\$10,000	\$10,000
		Corporate	Office Supplies	\$5,000	\$10,000
		Home Office	Technology	\$5,000	\$10,000
	June	Consumer	Furniture	\$5,000	\$10,000
		Corporate	Office Supplies	\$5,000	\$10,000
		Home Office	Technology	\$5,000	\$10,000
	July	Consumer	Furniture	\$10,000	\$10,000
		Corporate	Office Supplies	\$5,000	\$10,000
		Home Office	Technology	\$5,000	\$10,000
	August	Consumer	Furniture	\$5,000	\$10,000
		Corporate	Office Supplies	\$10,000	\$10,000
		Home Office	Technology	\$10,000	\$10,000

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.

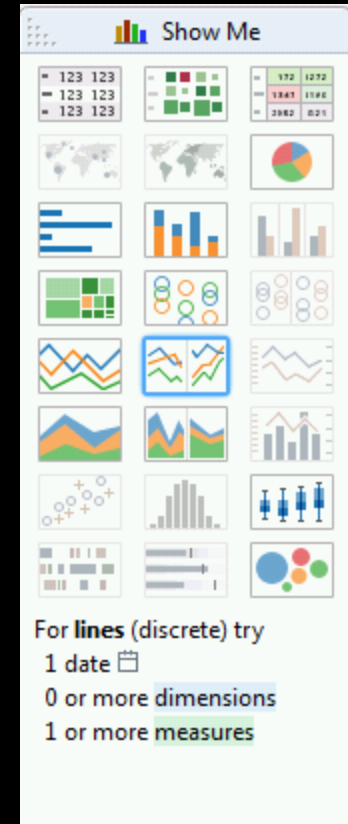
- Color—expresses discrete or continuous values
- Size—expresses discrete or continuous values
- Label—one or more fields can be expressed as label on marks
- Detail—disaggregates the marks plotted
- Tooltip/tooltips—makes fields available to tooltips without disaggregating data
- Shape—expresses discrete 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.



The screenshot shows a 'Show Me' interface with a grid of 24 visualization icons. The icons include: a table with numerical data, a heatmap, a table with numerical data, a world map, another world map, a pie chart, a horizontal bar chart, a grouped bar chart, a bar chart with error bars, a heatmap, a treemap, a treemap, a line chart with multiple series, a line chart with multiple series, a line chart with multiple series, a stacked area chart, a stacked area chart, a bar chart with error bars, a scatter plot with plus signs, a histogram, a box plot, a treemap, a treemap, and a bubble chart.

For **lines** (discrete) try
1 date 📅
0 or more **dimensions**
1 or more **measures**

Tableau Visualization: Show Me

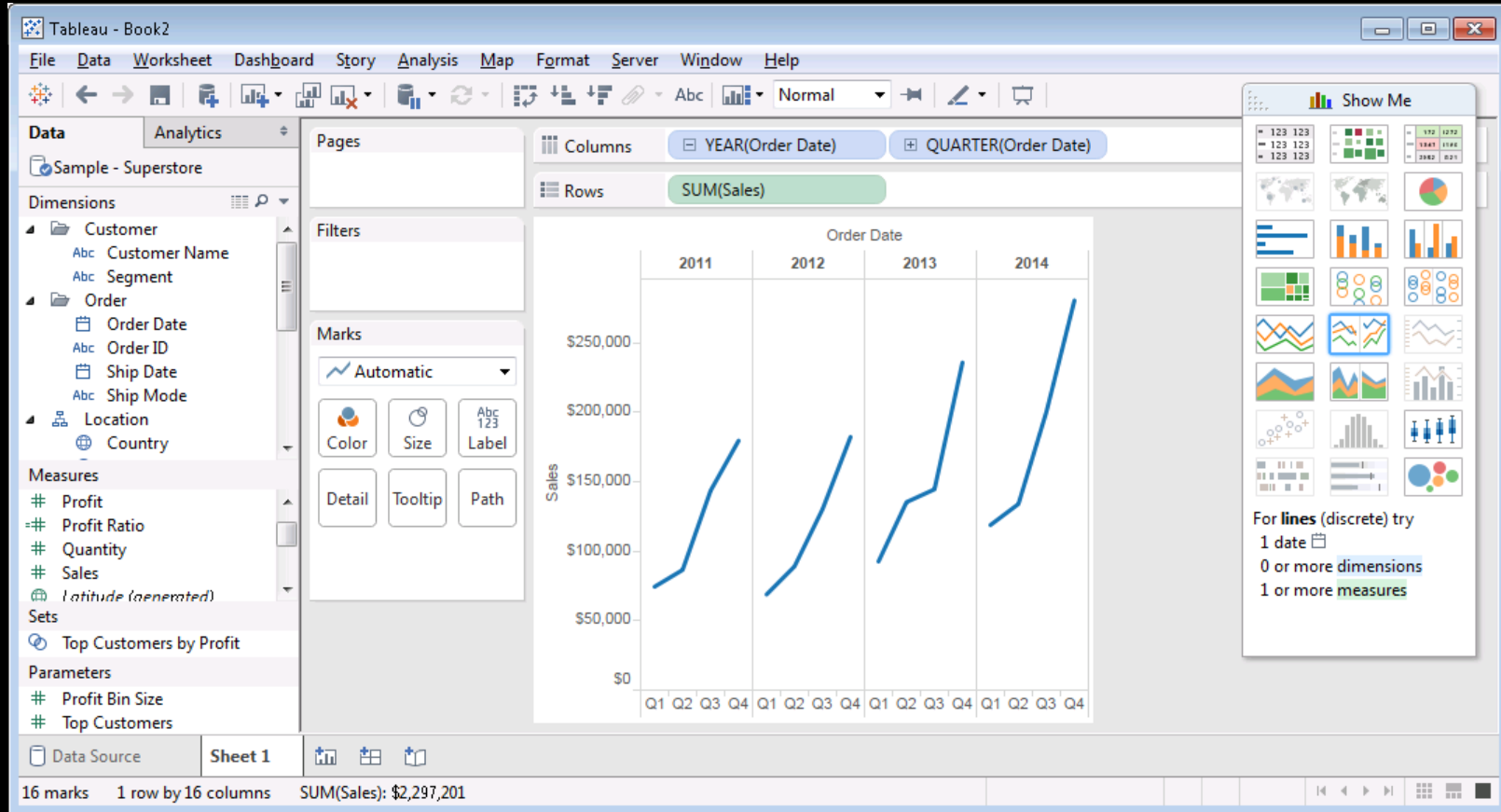


Tableau Question

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.

Tableau - sample

File Data Server Window Help

Orders (Sample - Superstore) Connection: Live Extract Filters: 0 Add...

Connected to Excel

Workbook: Sample - Superstore.xls

Sheets: Enter sheet name, Orders, People, Returns

Sort fields: Data source order Show aliases Show hidden fields Rows: 1,000

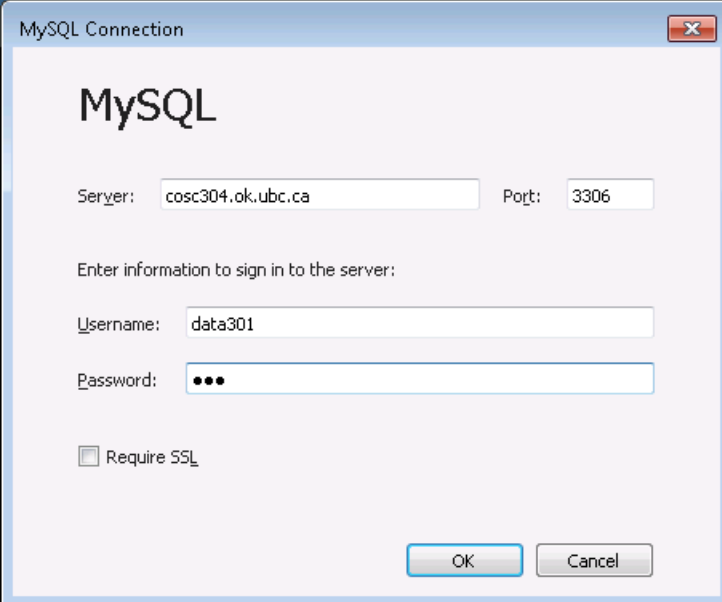
#	Orders	Orders	Orders	Orders	Orders	Orders
Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name
1	CA-2013-152156	11/9/2013	11/12/2013	Second Class	CG-12520	Claire Gute
2	CA-2013-152156	11/9/2013	11/12/2013	Second Class	CG-12520	Claire Gute
3	CA-2013-138688	6/13/2013	6/17/2013	Second Class	DV-13045	Darrin Van Huff
4	US-2012-108966	10/11/2012	10/18/2012	Standard Class	SO-20335	Sean O'Donnell
5	US-2012-108966	10/11/2012	10/18/2012	Standard Class	SO-20335	Sean O'Donnell
6	CA-2011-115812	6/9/2011	6/14/2011	Standard Class	BH-11710	Brosina Hoffma
7	CA-2011-115812	6/9/2011	6/14/2011	Standard Class	BH-11710	Brosina Hoffma

Data Source: BarChart_Sales_by_Category, Map_Sales_by_State, Profit_Sales_Scatter, Sales_Profit_Year_Trend, SalesDashboard

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



The image shows a screenshot of the 'MySQL Connection' dialog box in Tableau. The dialog has a title bar with 'MySQL Connection' and a close button. The main content area is titled 'MySQL'. It contains the following fields and options:

- Server:** cosc304.ok.ubc.ca
- Port:** 3306
- Enter information to sign in to the server:**
 - Username:** data301
 - Password:** (masked with three dots)
- Require SSL

At the bottom of the dialog are 'OK' and 'Cancel' buttons.

Example Connecting to MySQL (2)

Tableau - sample

File Data Server Window Help

WorksOn+ (data301) Connection: Live Extract Filters: 0 | Add...

Connected to MySQL

Server
cosc304.ok.ubc.ca

Database
data301

Table
Enter table name

- Dept
- Emp
- Proj
- sales
- store
- vendor
- WorksOn
- New Custom SQL

WorksOn — Emp — Dept

Proj — Dept1

Sort fields: Data source order Show aliases Show hidden fields Rows: 11

Dept dno (Dept)	Dept Dname	Dept Mgreno	Dept1 dno (Dept1)	Dept1 dname (Dept1)	Dept1 mgreno (Dept1)	Emp eno (Emp)
null	null	null	D1	Management	E8	E1
D3	Accounting	E5	D1	Management	E8	E2
D3	Accounting	E5	D2	Consulting	E7	E2
D3	Accounting	E5	D2	Consulting	E7	E4
D3	Accounting	E5	D2	Consulting	E7	E5
D2	Consulting	E7	D2	Consulting	E7	E3
D2	Consulting	E7	D2	Consulting	E7	E6

Data Source | BarChart_Sales_by_Category | Map_Sales_by_State | Profit_Sales_Scatter | Sales_Profit_Year_Trend | SalesDashboard

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

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

Join Question

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

Boys

Bid	BoyName
1	Joe
1	Steve
3	Fred
5	James
7	Ben
7	Bishop

Girls

Gid	GirlName
1	Jane
1	Sarah
5	Fran
6	Julie

A) 9

B) 8

C) 7

D) 6

E) 0

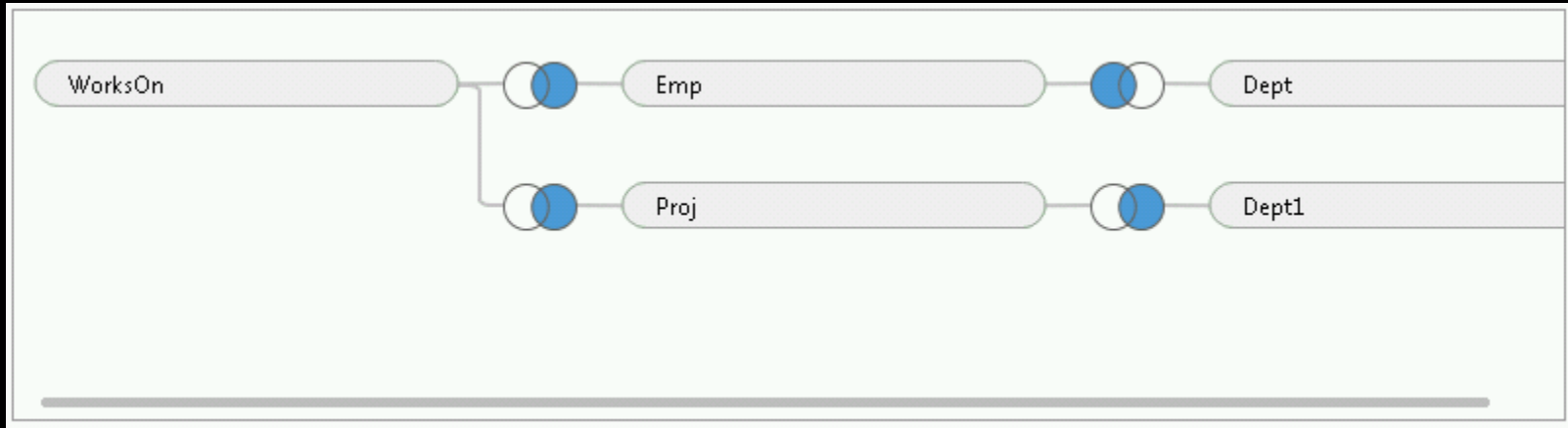
Data Blending

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

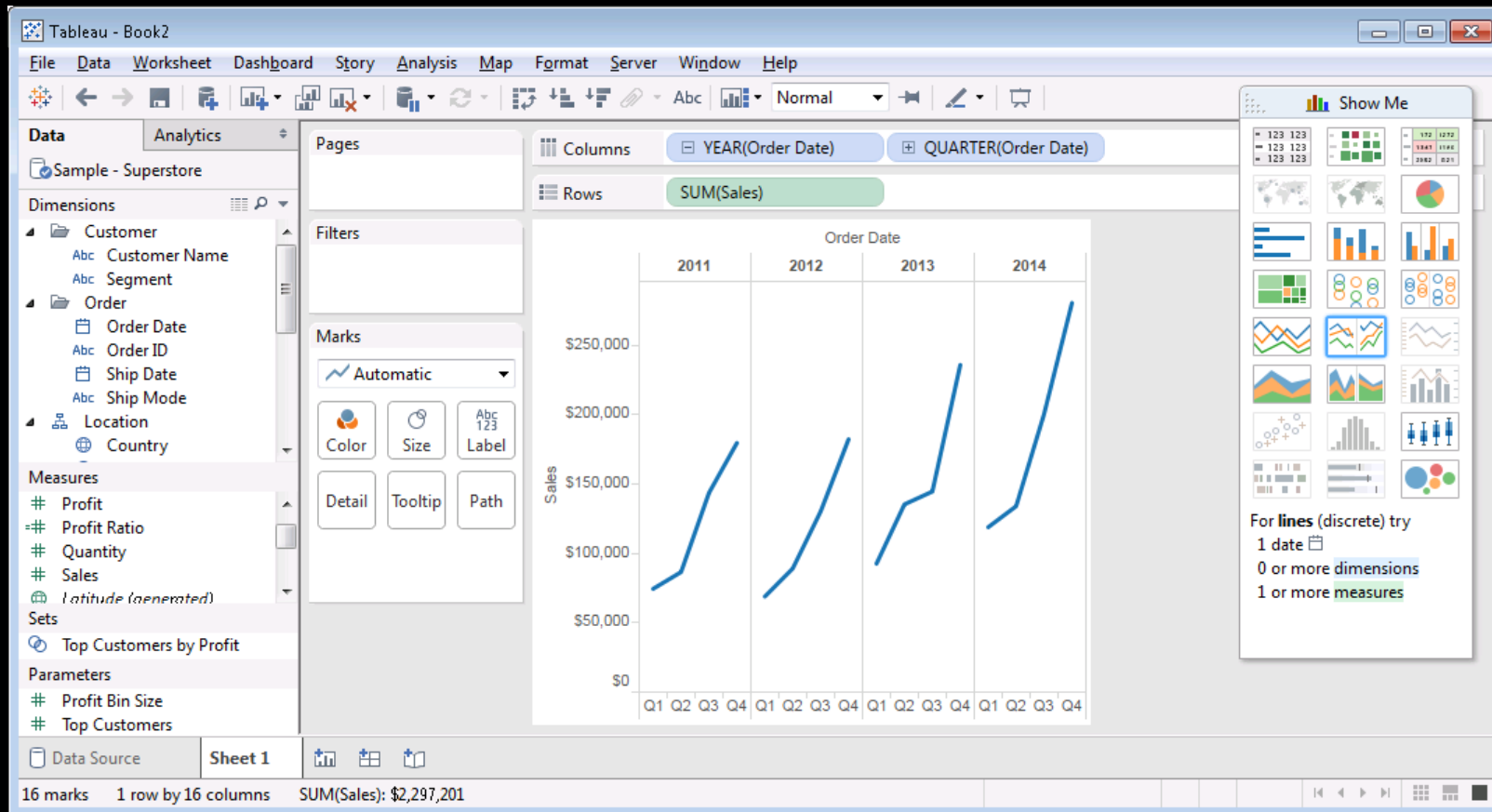
For geographic data (small globe icon), Tableau automatically generates center-point geocodes (longitude/latitude).

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)



Show Me

For lines (discrete) try

- 1 date
- 0 or more dimensions
- 1 or more measures

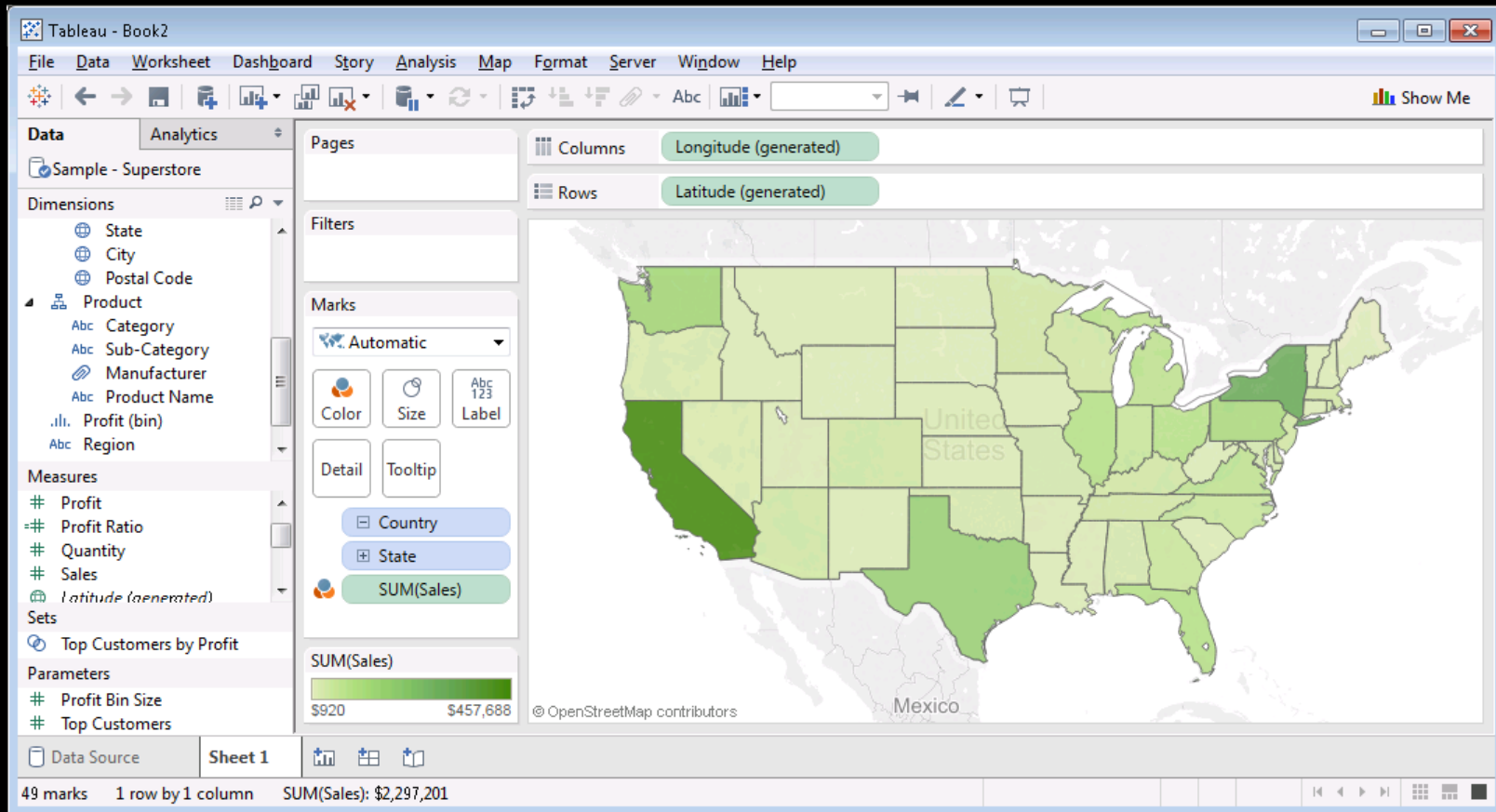
Text Table (Crosstab)

The screenshot shows the Tableau interface with a text table visualization. The data is summarized in the table below:

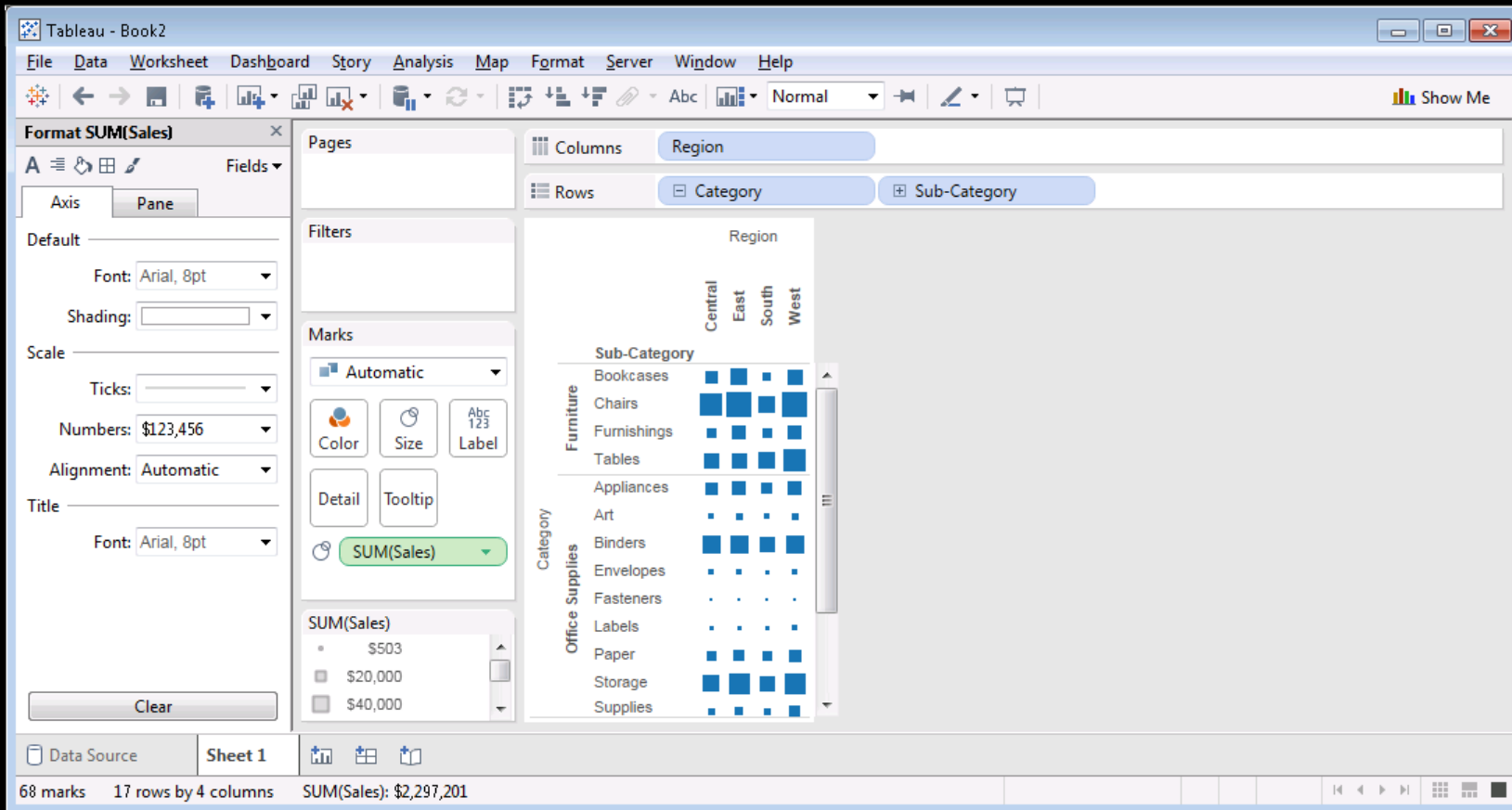
Category	Sub-Category	Order Date			
		2011	2012	2013	2014
Furniture	Bookcases	\$20,037	\$38,544	\$26,275	\$30,024
	Chairs	\$77,242	\$71,735	\$83,919	\$95,554
	Furnishings	\$13,826	\$21,090	\$27,874	\$28,915
	Tables	\$46,088	\$39,150	\$60,833	\$60,894
Office Supplies	Appliances	\$15,314	\$23,241	\$26,050	\$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
Technology	Paper	\$14,835	\$15,288	\$20,638	\$27,718
	Storage	\$50,329	\$45,048	\$58,632	\$69,834
	Supplies	\$14,394	\$1,952	\$14,278	\$16,049
	Accessories	\$25,014	\$40,524	\$41,896	\$59,946
	Machines	\$10,850	\$26,179	\$49,599	\$62,899
		\$62,023	\$27,764	\$55,907	\$43,545

The Tableau interface includes a menu bar (File, Data, Worksheet, Dashboard, Story, Analysis, Map, Format, Server, Window, Help), a toolbar with various analysis tools, and a 'Show Me' button. The left pane shows the 'Data' source as 'Sample - Superstore' and a list of dimensions (State, City, Postal Code, Product, Category, Sub-Category, Manufacturer, Product Name, Profit (bin), Region) and measures (Profit, Profit Ratio, Quantity, Sales, Latitude (generated)). The bottom status bar indicates 68 marks, 17 rows by 4 columns, and a total SUM(Sales) of \$2,297,201.

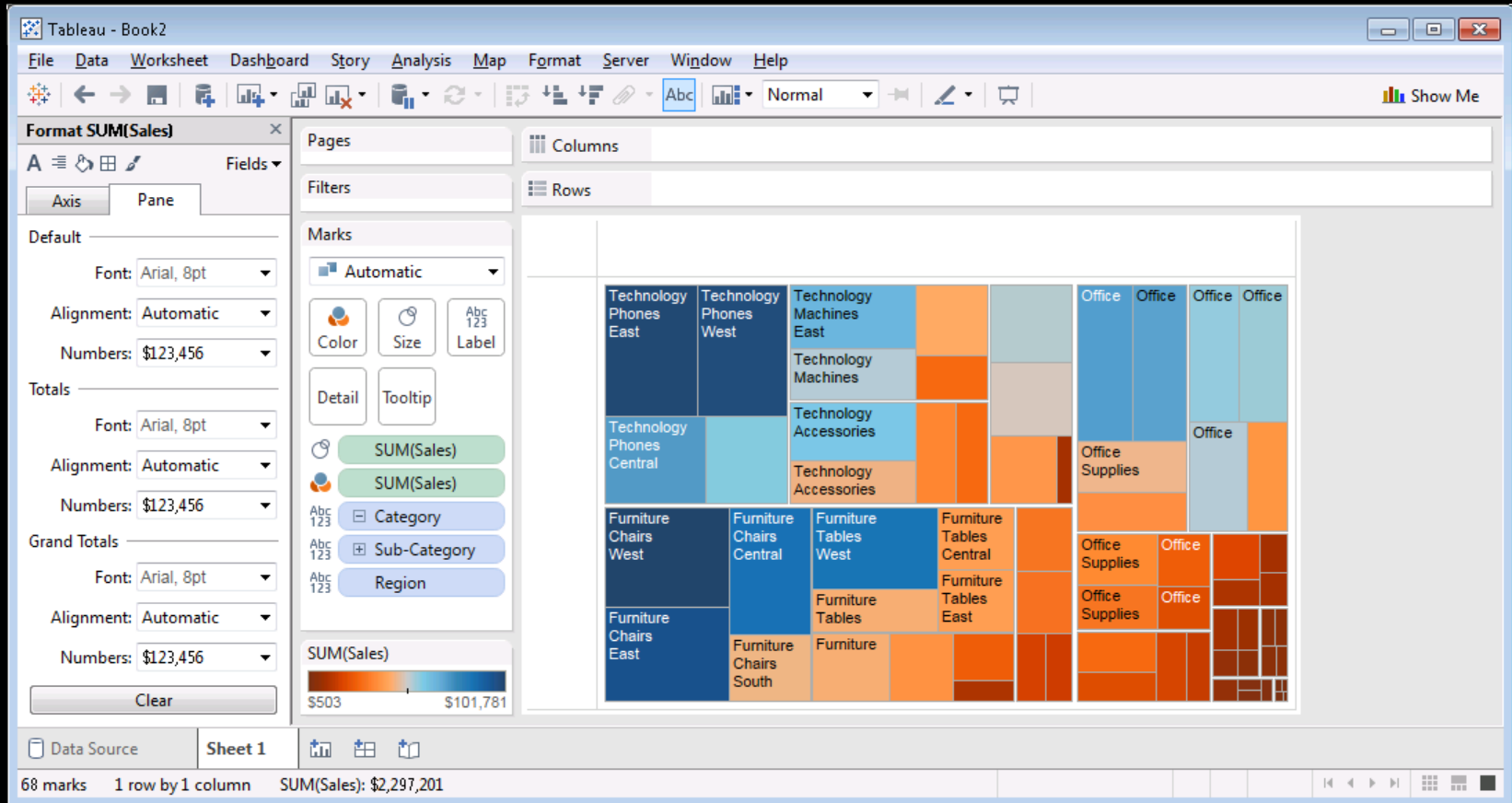
Maps



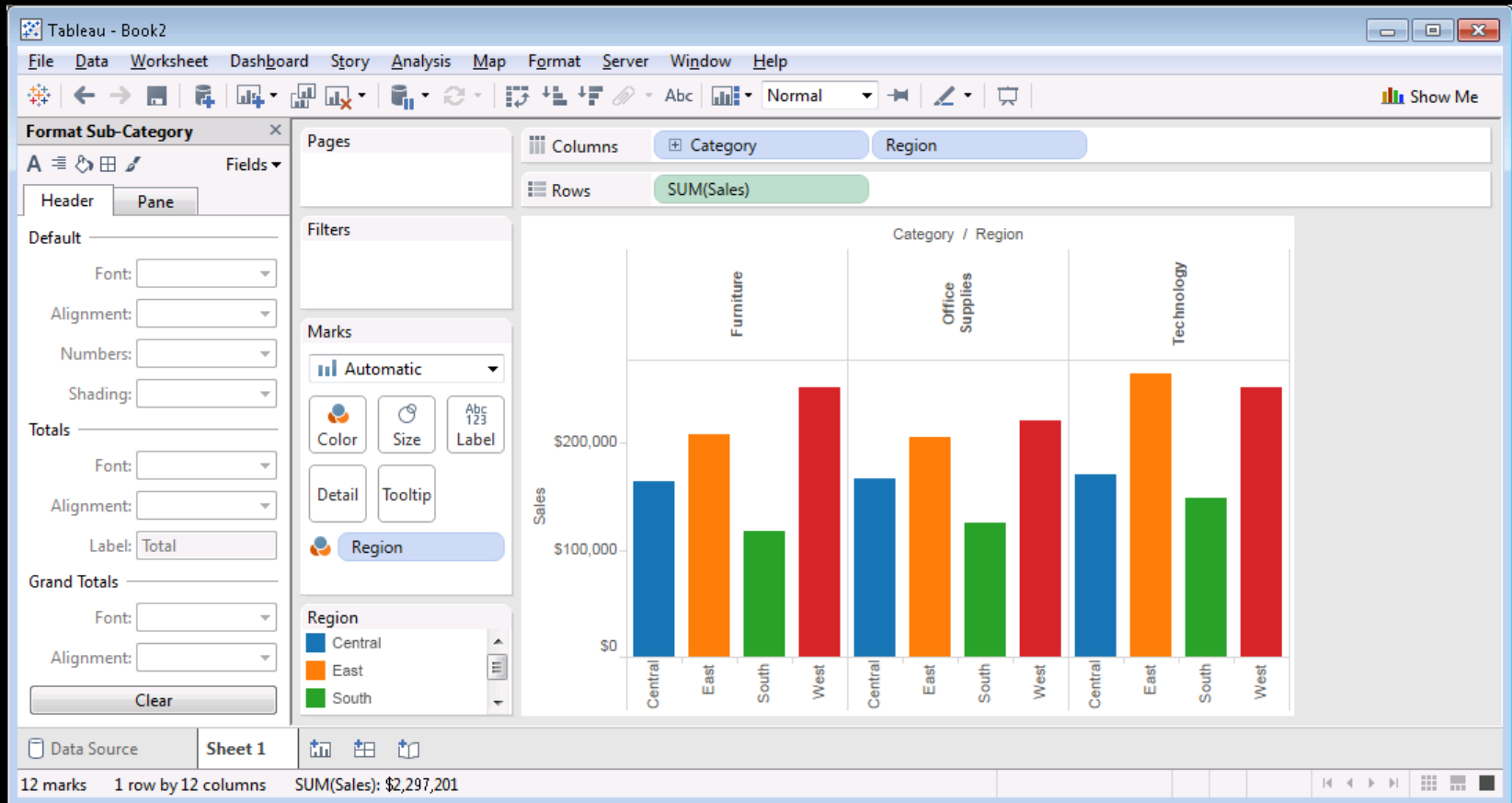
Heat Map



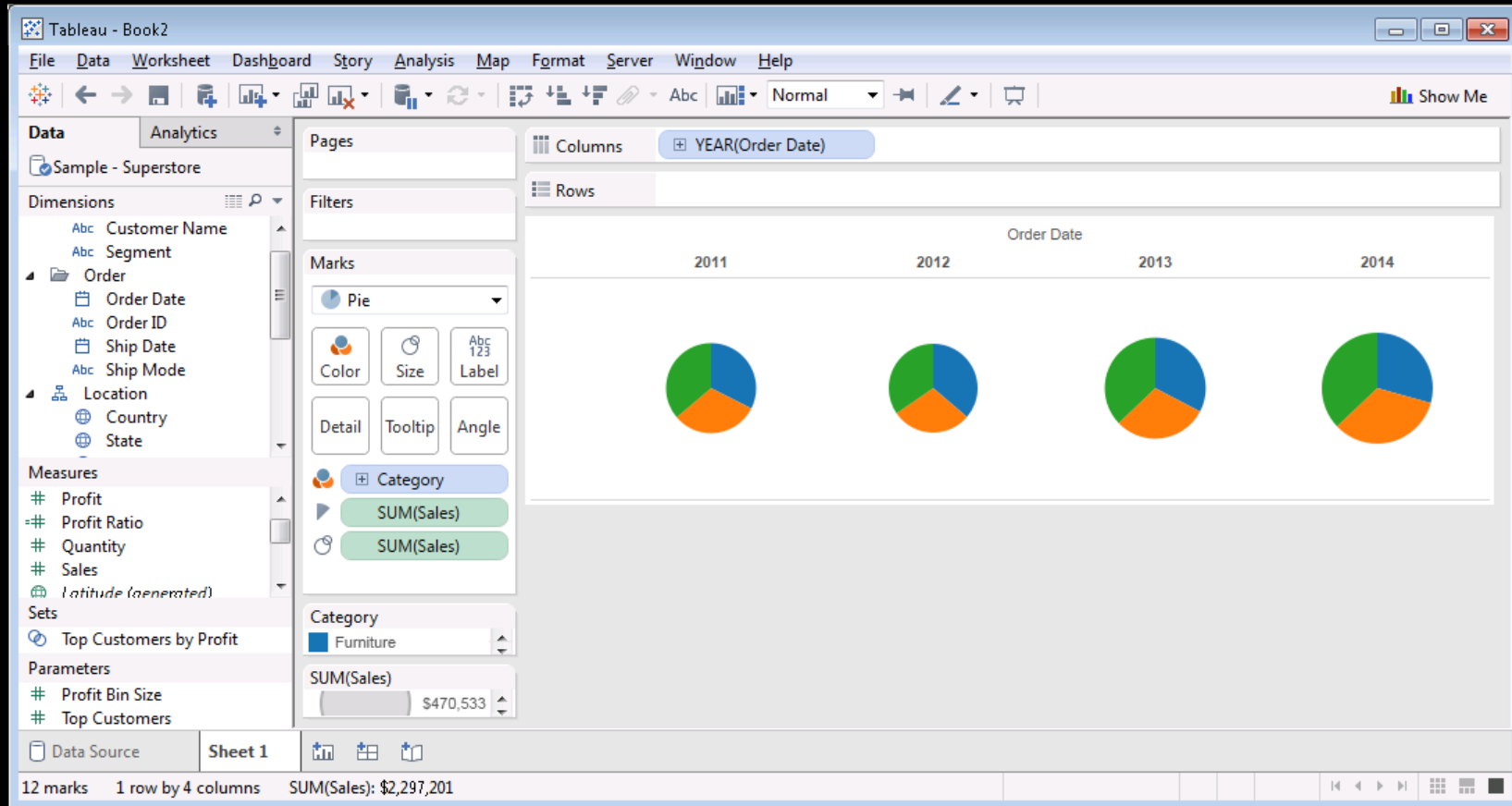
Tree map



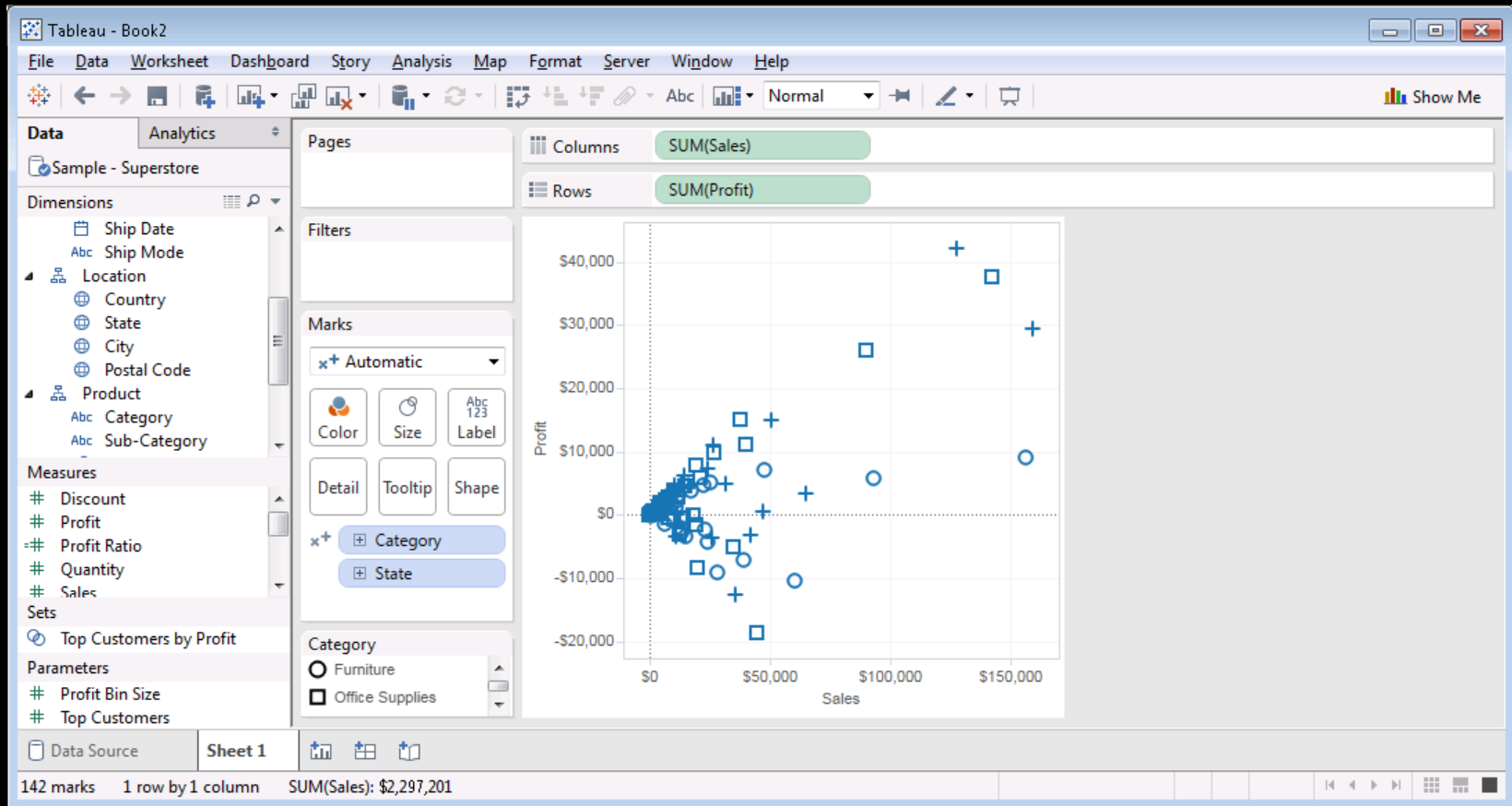
Bar Charts



Pie Charts



Scatter Plots

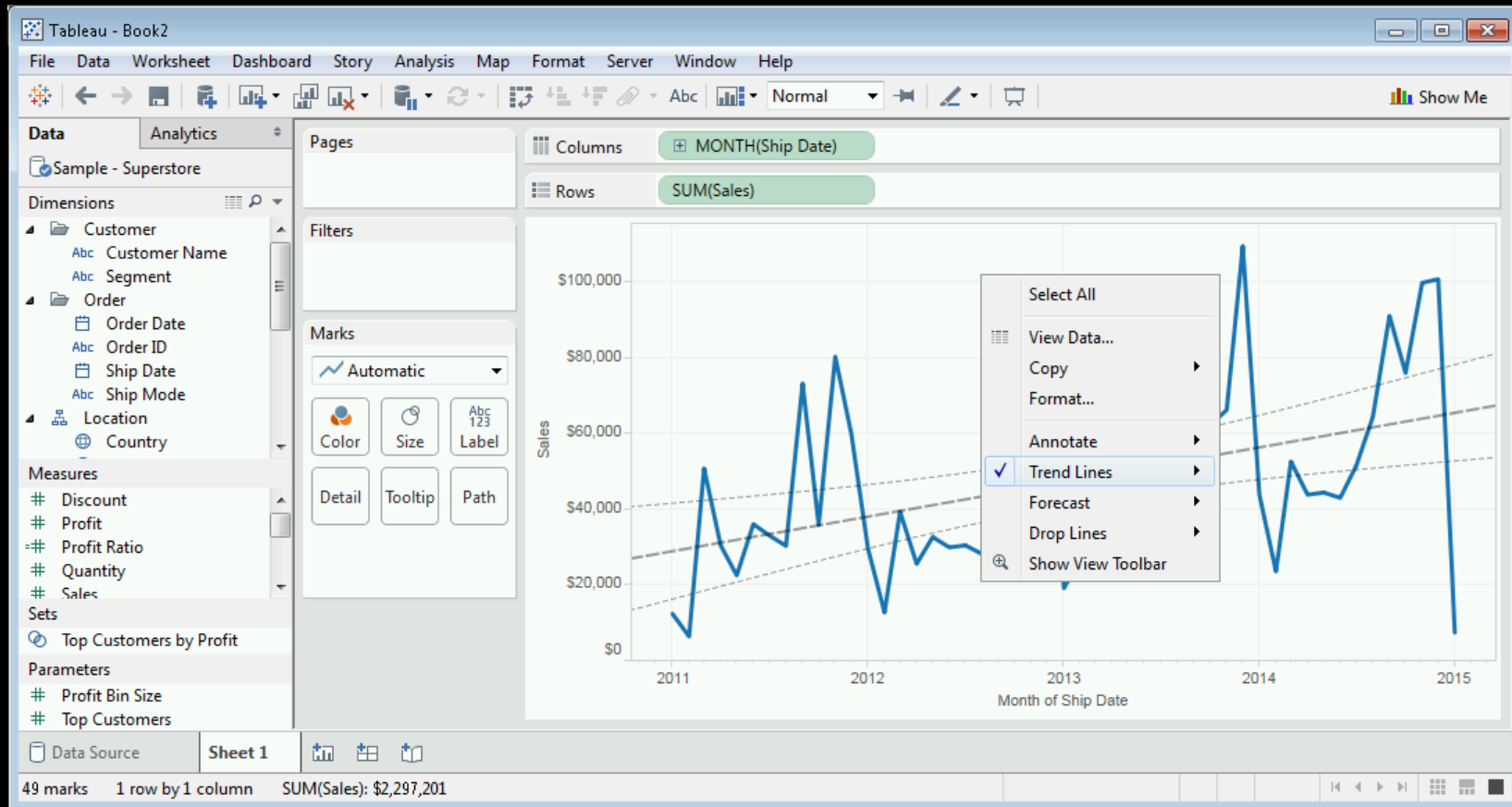


Trend Lines and Reference Lines

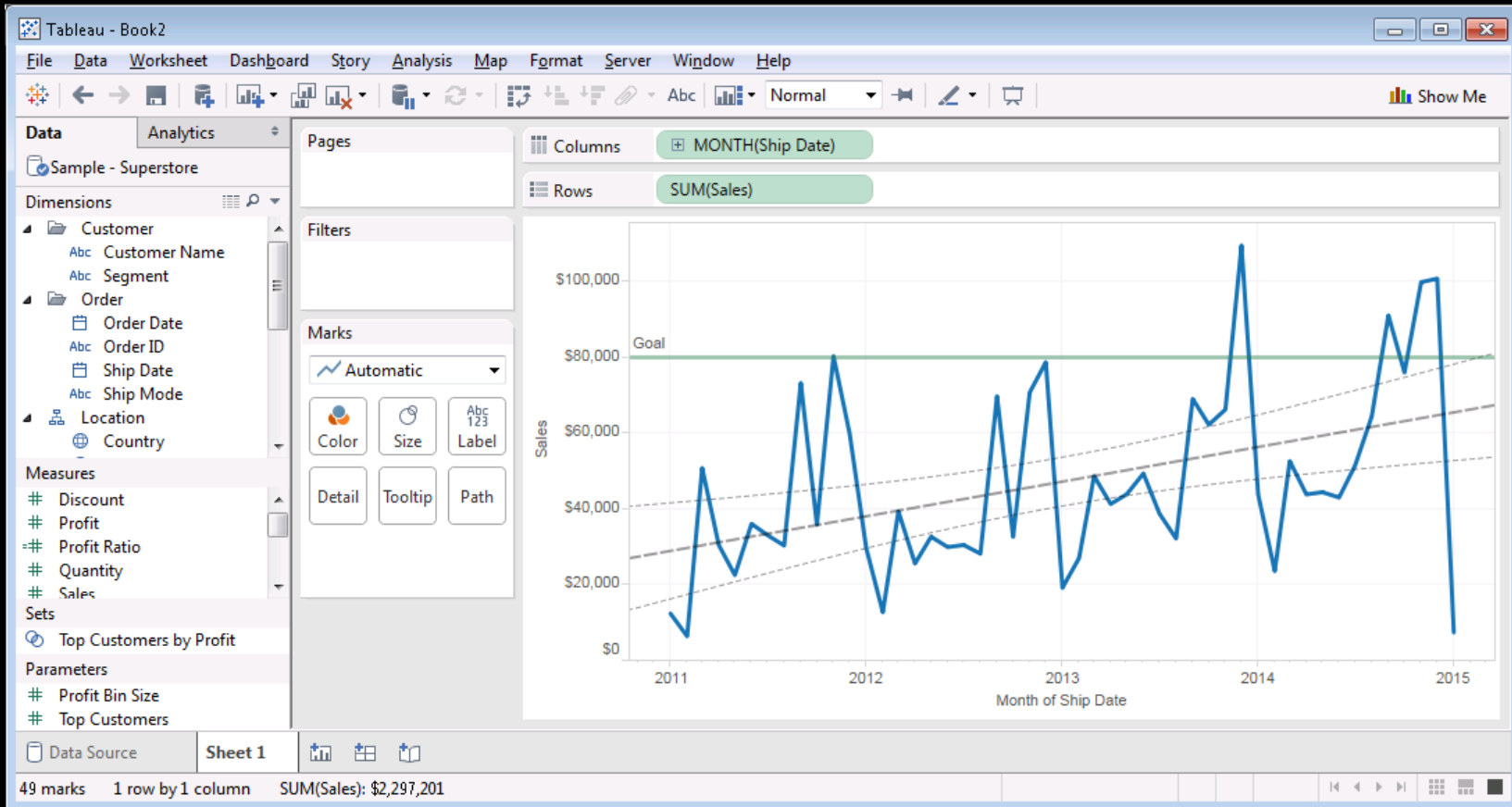
Trend lines show patterns in data using a line of best fit.

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

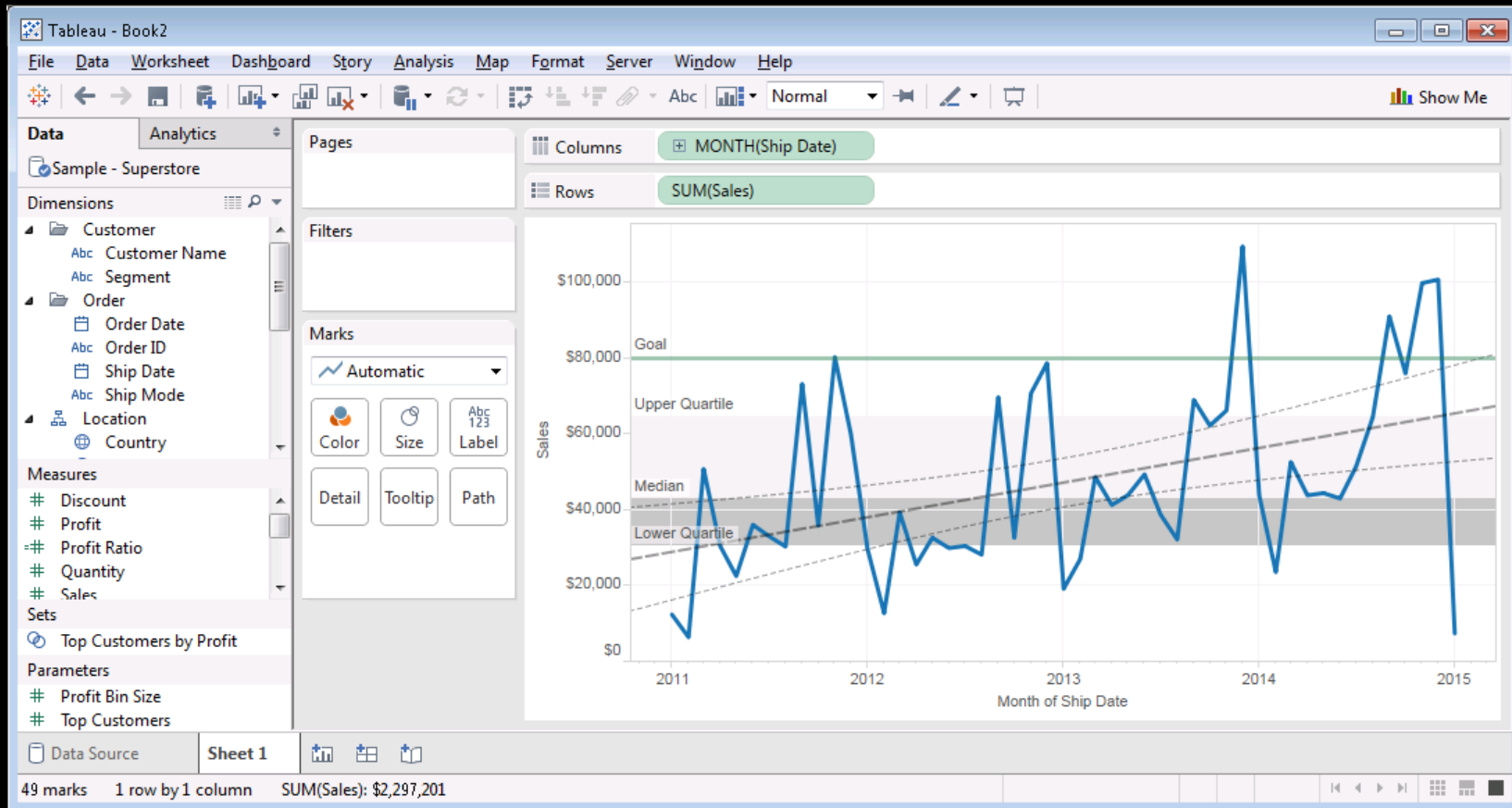
Adding Trend Lines



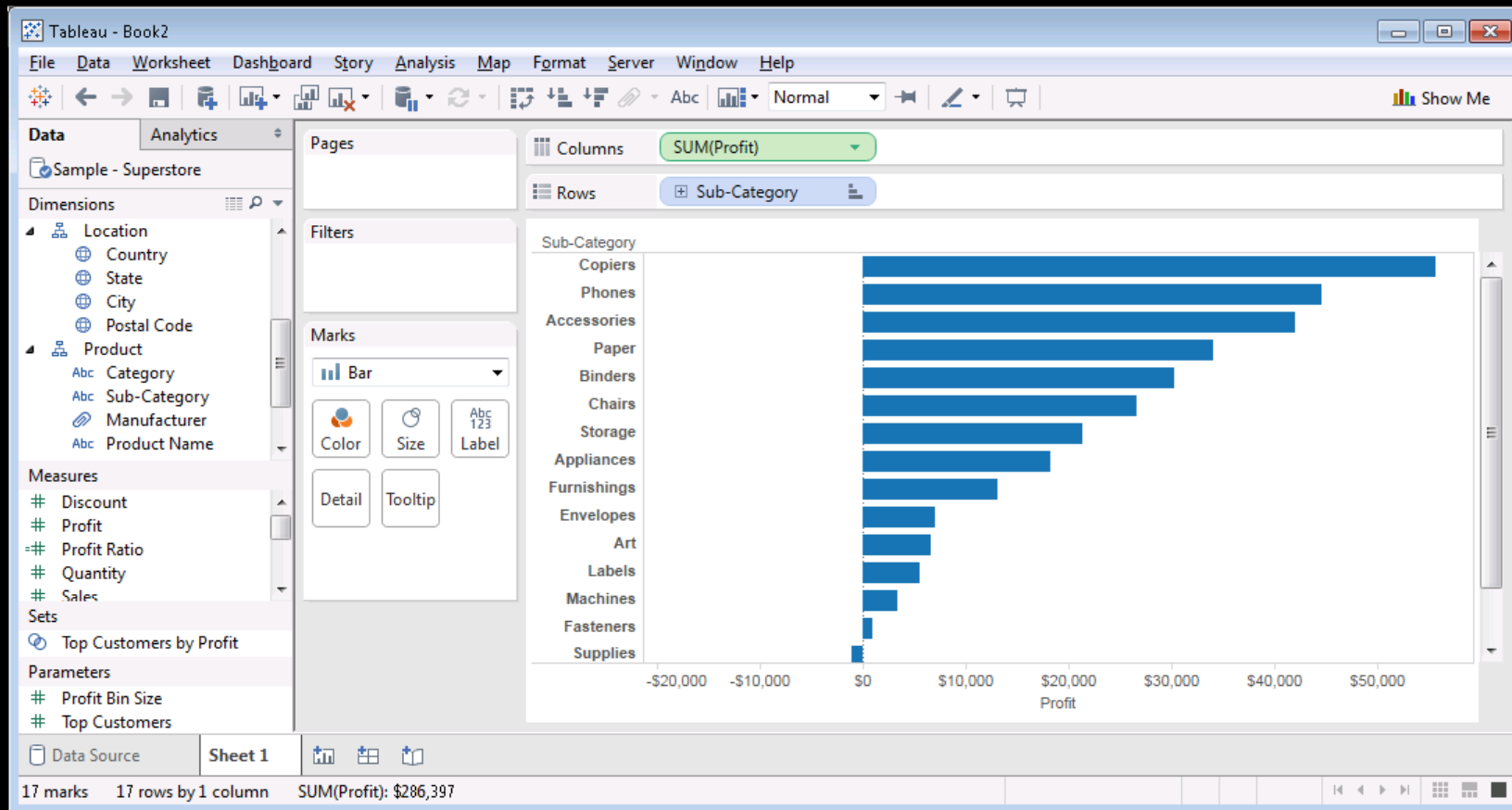
Trend Lines and Reference Lines



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

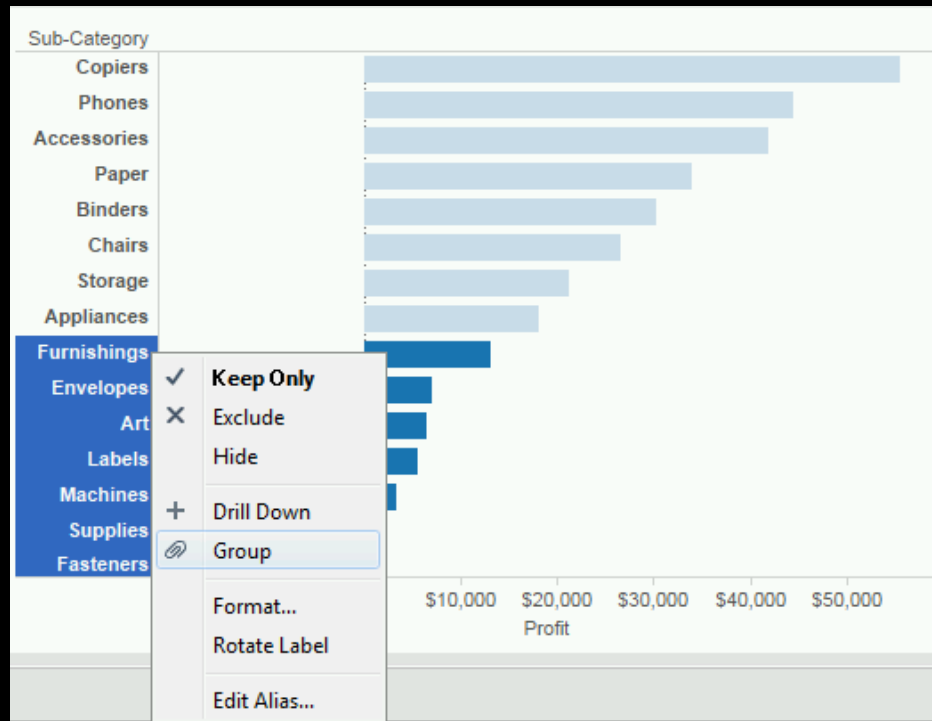
There are multiple ways to define filters:

- 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.

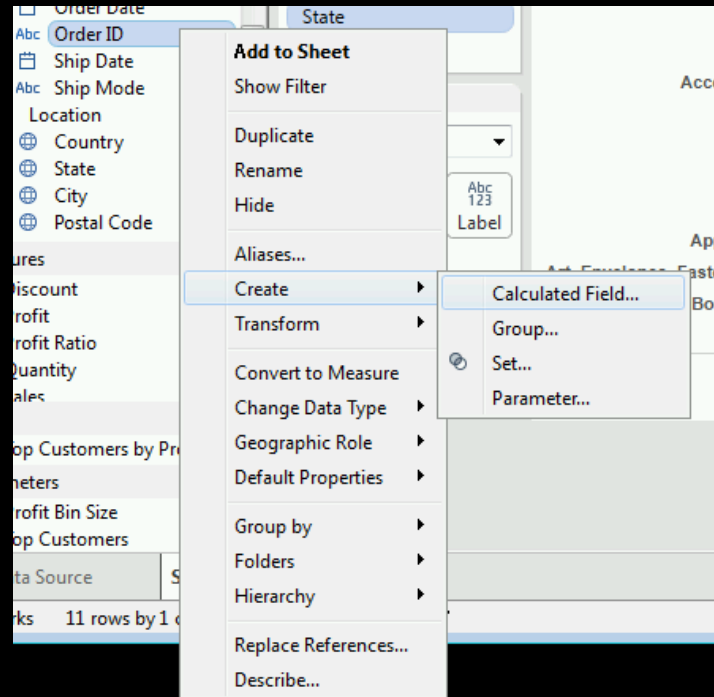
- Multi-select elements then in pop-up menu select Group



Calculations

Calculated fields are performed on data source when possible.

Table calculations are performed locally in Tableau.



Creating a Calculated Field

The screenshot shows a dialog box for creating a calculated field. The title bar reads "Sample Calculation" with a close button. The main area contains the formula `UPPER ([Order ID])`. Below the formula, a status message says "The calculation is valid." At the bottom left are "Apply" and "OK" buttons. On the right, a dropdown menu is open, showing a list of functions: SUM, TAN, THEN, TODAY, TOTAL, TRIM, UPPER (highlighted), USERDOMAIN, USERNAME, and VAR. To the right of the dropdown, the field name "Order ID" and its data type "Data type: String" are displayed. A "Describe..." button is located at the bottom right of the dialog.

Sample Calculation

`UPPER ([Order ID])`

The calculation is valid.

Apply OK

All

Enter search text

- SUM
- TAN
- THEN
- TODAY
- TOTAL
- TRIM
- UPPER
- USERDOMAIN
- USERNAME
- VAR

Order ID

Data type: String

Describe...

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.

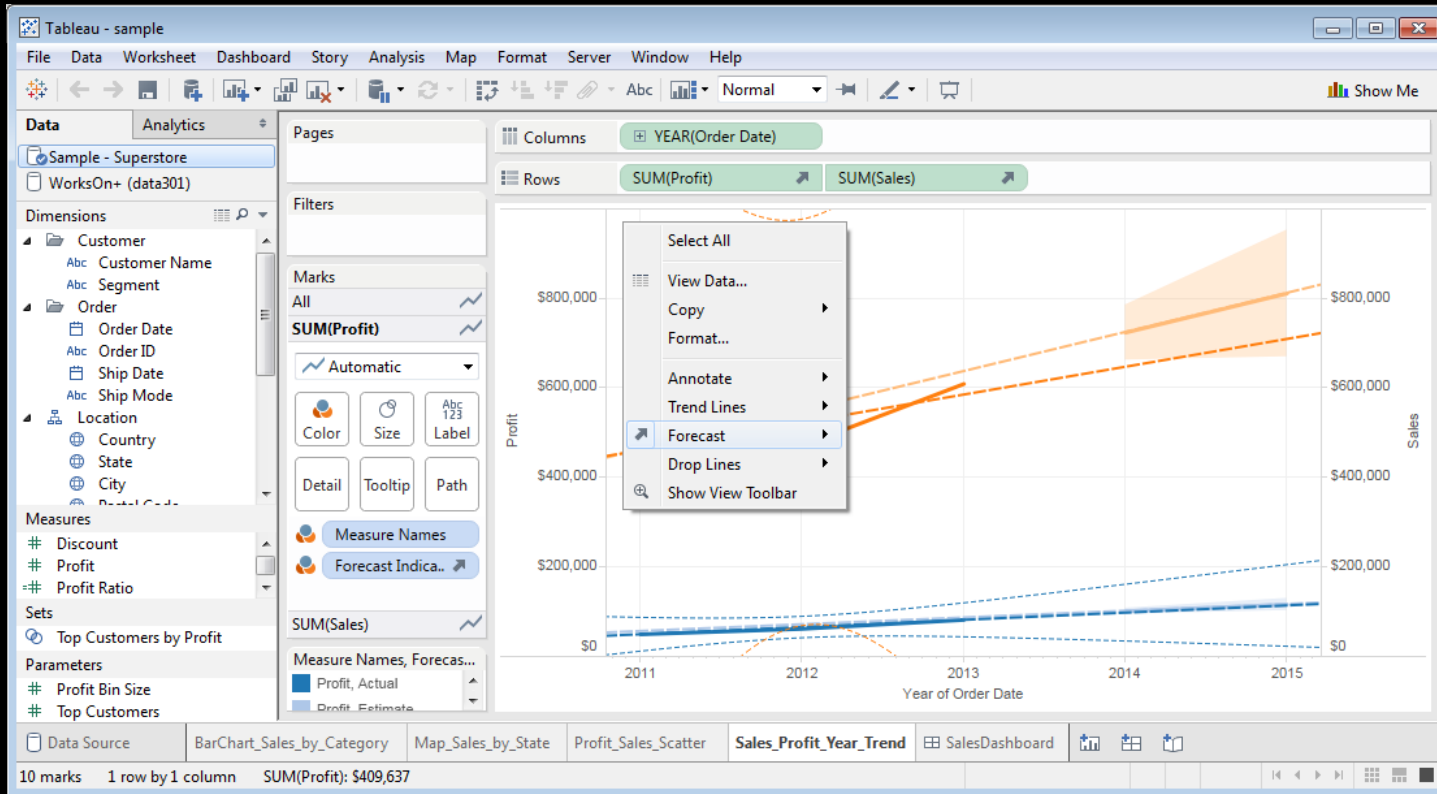


Tableau Charts Question

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:

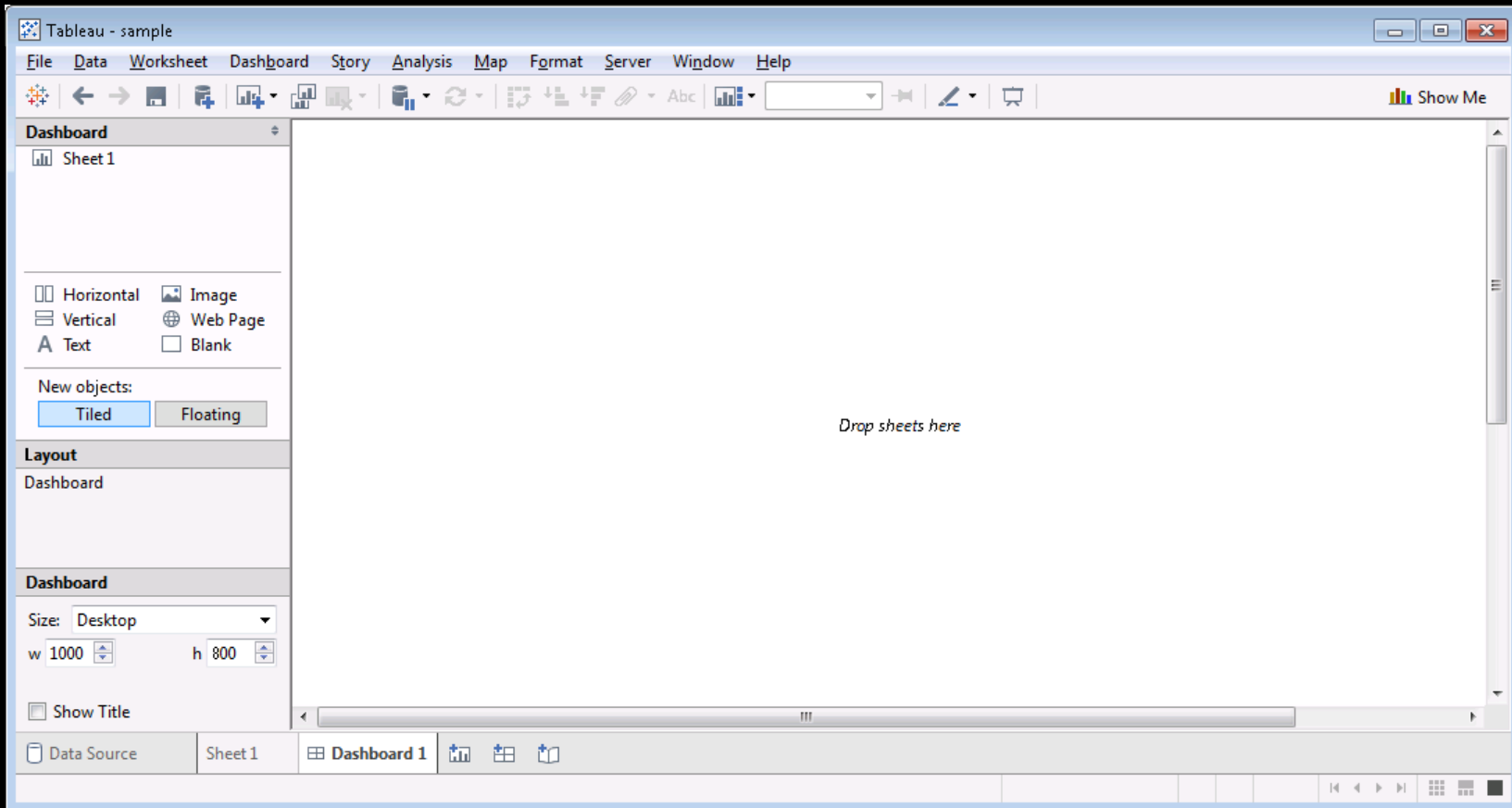
- line chart (with forecast and trend line)
- bar chart (with filters and sorting)
- pie chart (with a parameter)
- heat map (with grouping)
- scatter plot (with a calculated field)
- histogram
- circle view

Dashboards

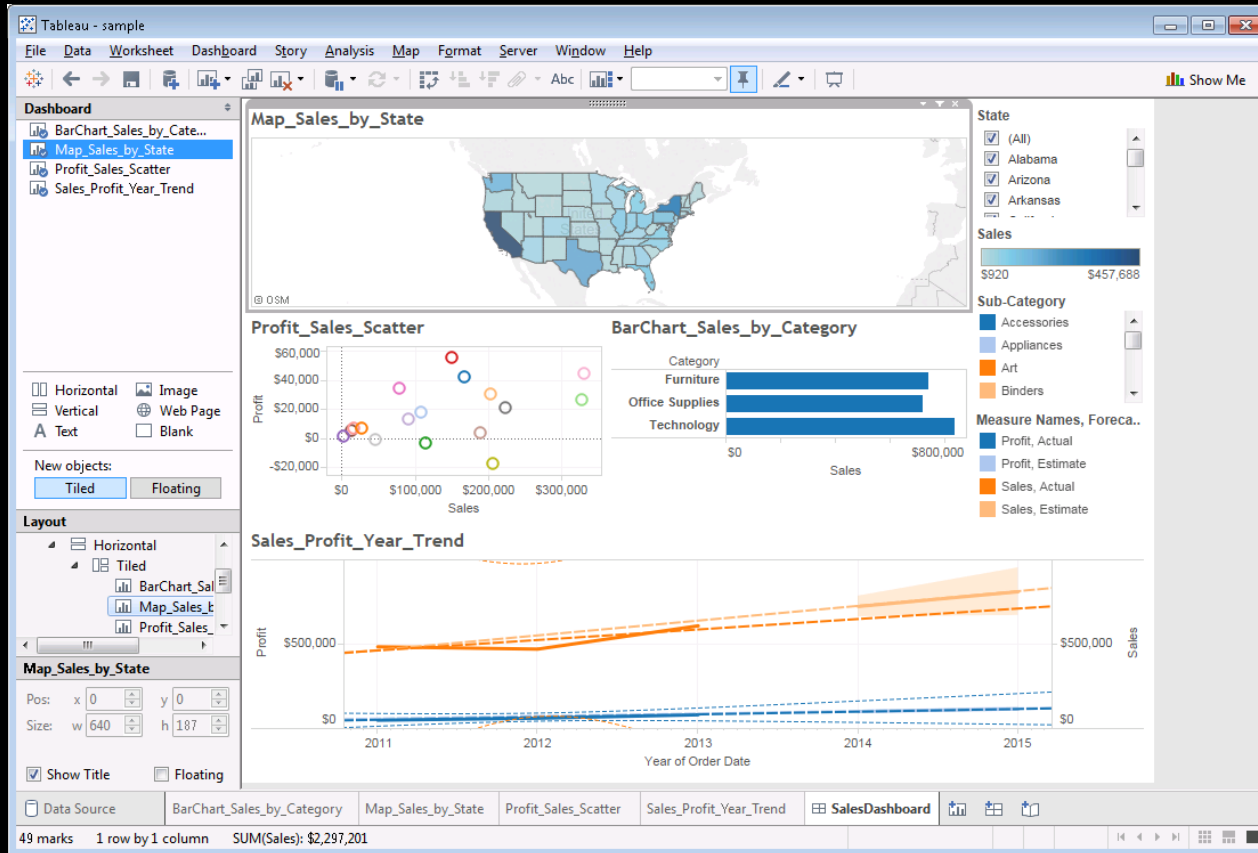
A *dashboard* consists of multiple sheets organized to make information and its relationships more understandable.

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