

Data Analysis Fundamentals with Tableau

Ben Sullins
Bensullins.com
ben@bensullins.com



pluralsight 
hardcore developer training

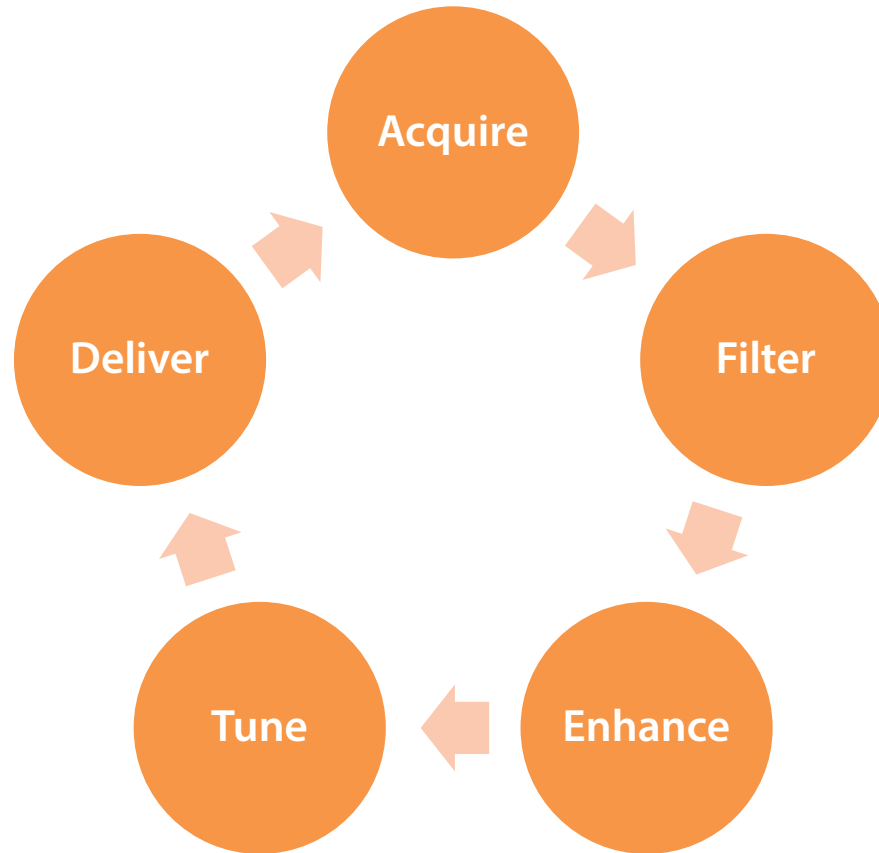
Course Outline

- Introduction to Tableau
- Getting dirty with your data
- Show me the data!
- Adding the intelligence factor
- Painting the whole picture
- Sharing your work

Overview

- Visual analysis
- Visual perception
- Tableau's product family
- Connecting to your data
- Data terminology
- View terminology

Introduction to Visual Analysis

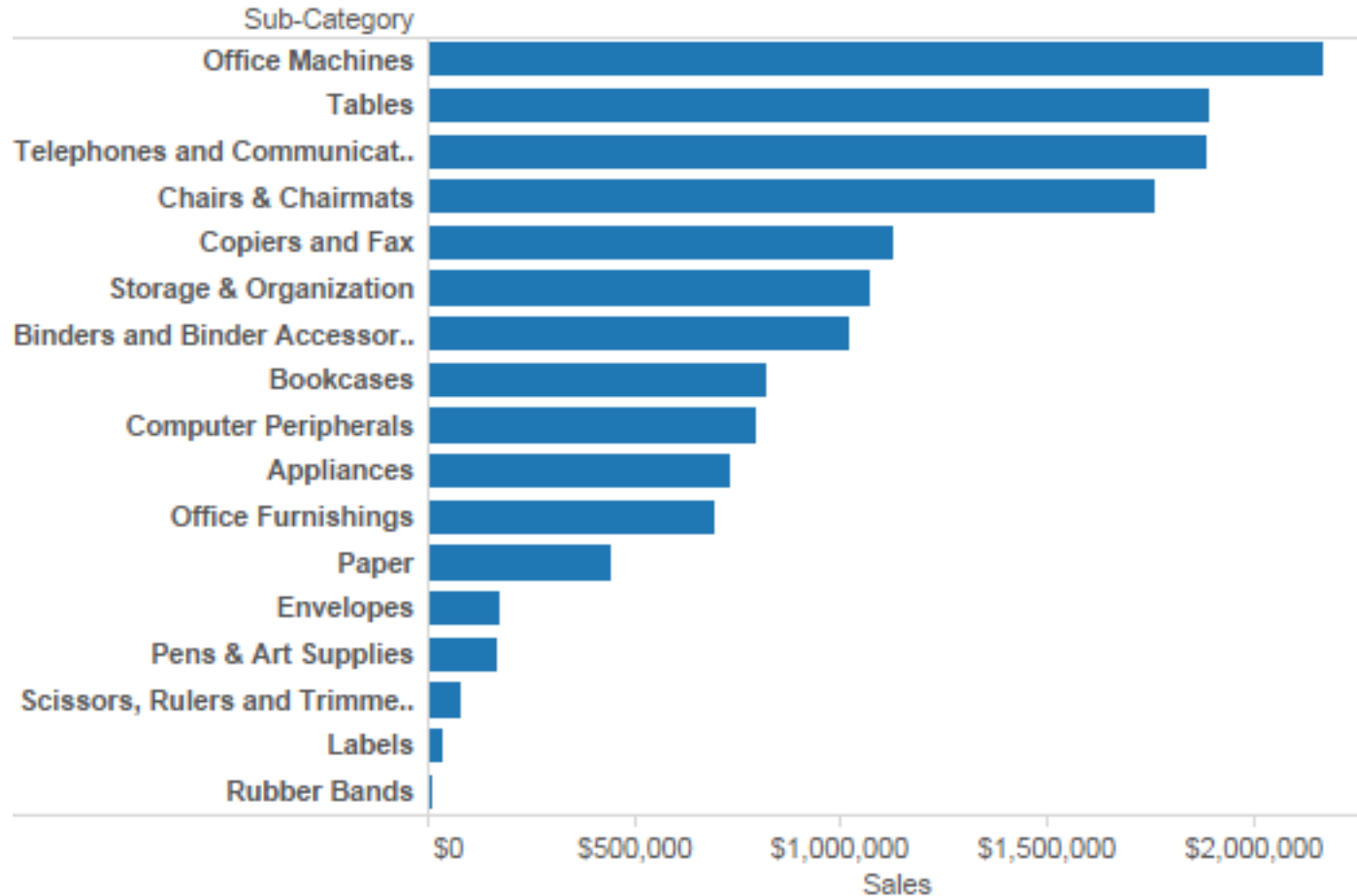


“Visual analysis means exploring data visually.
A story unfolds as
you navigate from one visual summary into
another.”

Hanrahan, Stolte, Mackinlay. “Visual Analysis for Everyone.” *Tableau White paper* (2007).

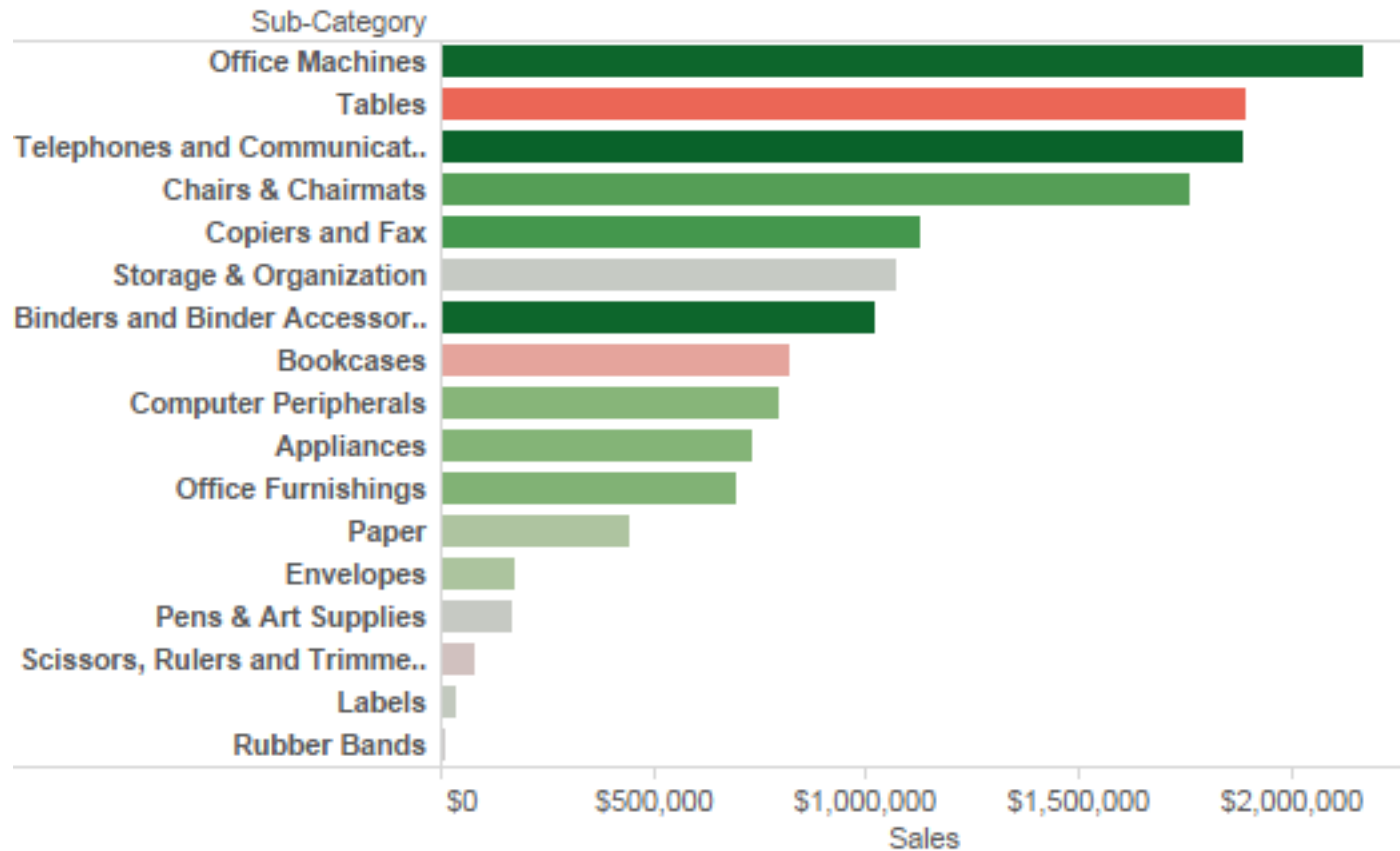
“Visual analysis means exploring data visually...

Sales by Sub Category



... a story unfolds as
you navigate from one visual summary into another."

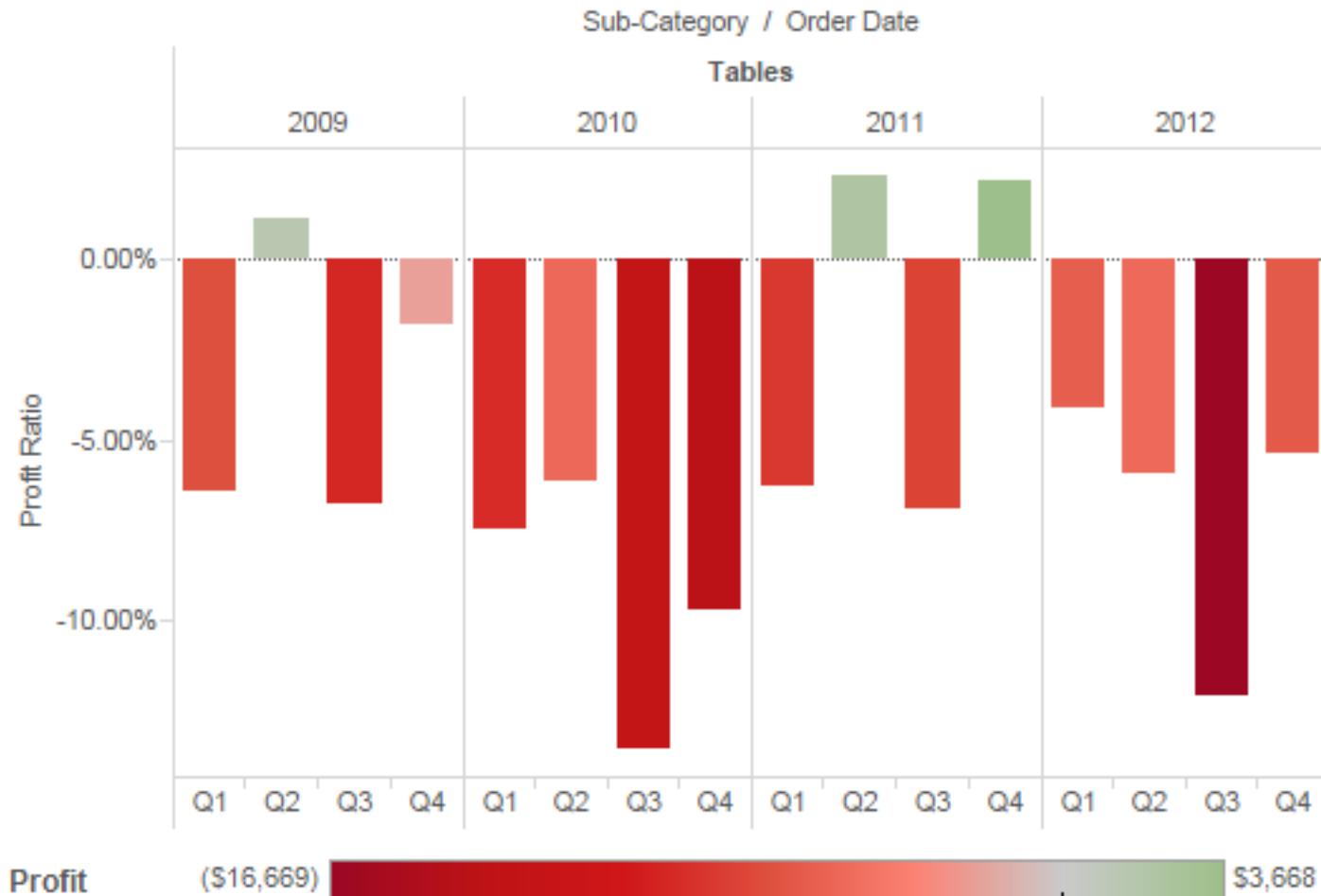
Sales by Sub Category



Profit (\$99,063) \$316,952

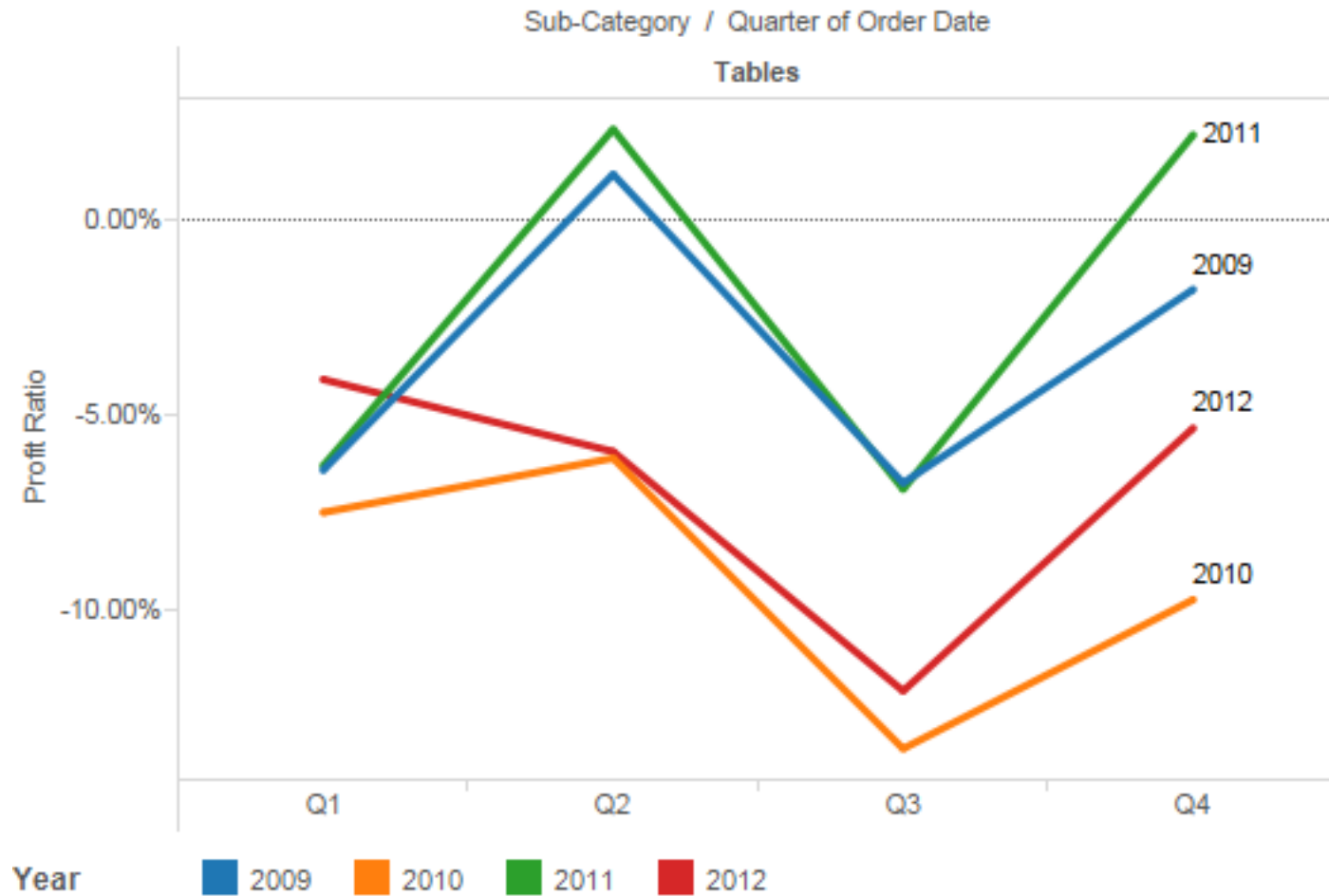
Instantly change *what* you are looking at

Sales by Sub Category



Instantly change the *way* you are looking at it

Sales by Sub Category



“Simplicity is the essence of
clear communication”

- Steve Jobs

Count how many times the number “5” appears

987349790275647902894728624092406037070570279072
803208029007302501270237008374082078720272007083
247802602703793775709707377970667462097094702780
927979709723097230979592750927279798734972608027

**Now count the number of time “5” appears in the
same set of numbers**

98734979027**5**647902894728624092406037070**5**70279072
803208029007302**5**01270237008374082078720272007083
24780260270379377**5**709707377970667462097094702780
927979709723097230979**5**927**5**0927279798734972608027

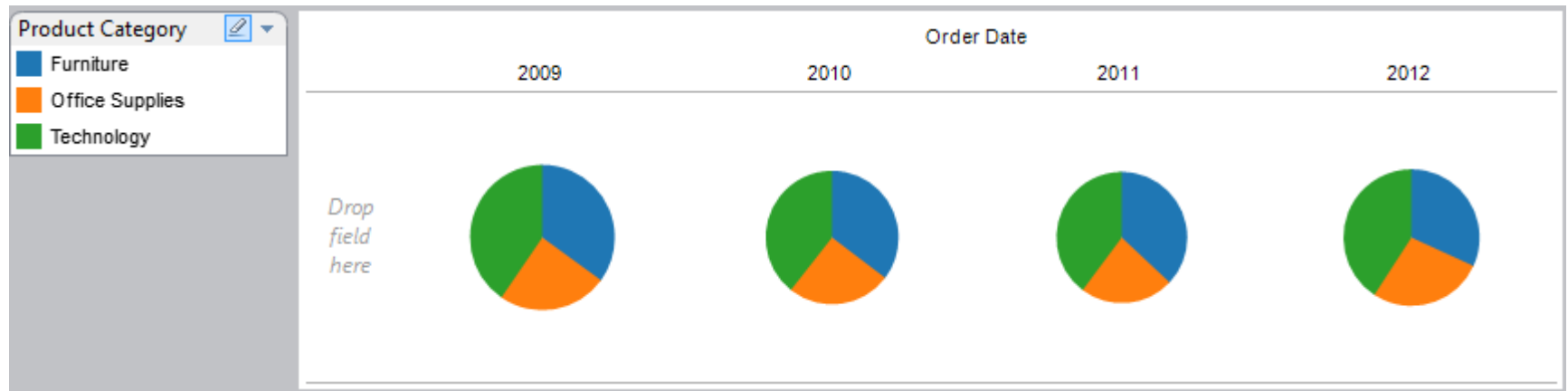
Attentive Processing

987349790275647902894728624092406037070570279072
803208029007302501270237008374082078720272007083
247802602703793775709707377970667462097094702780
927979709723097230979592750927279798734972608027

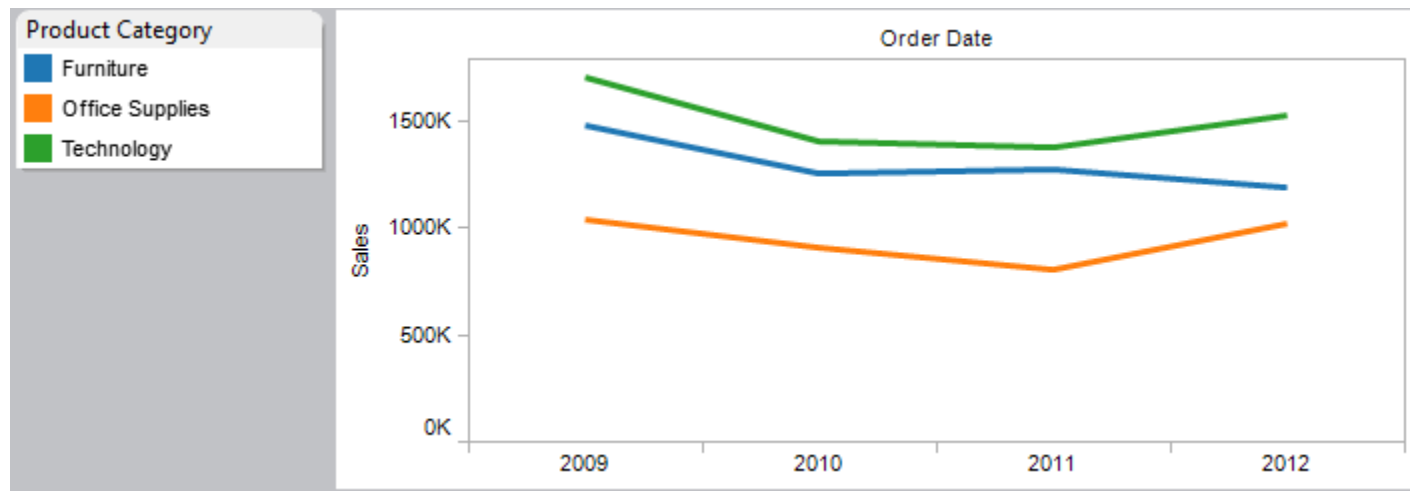
Pre-attentive Processing

98734979027**5**647902894728624092406037070**5**70279072
803208029007302**5**01270237008374082078720272007083
24780260270379377**5**709707377970667462097094702780
927979709723097230979**5**927**5**0927279798734972608027

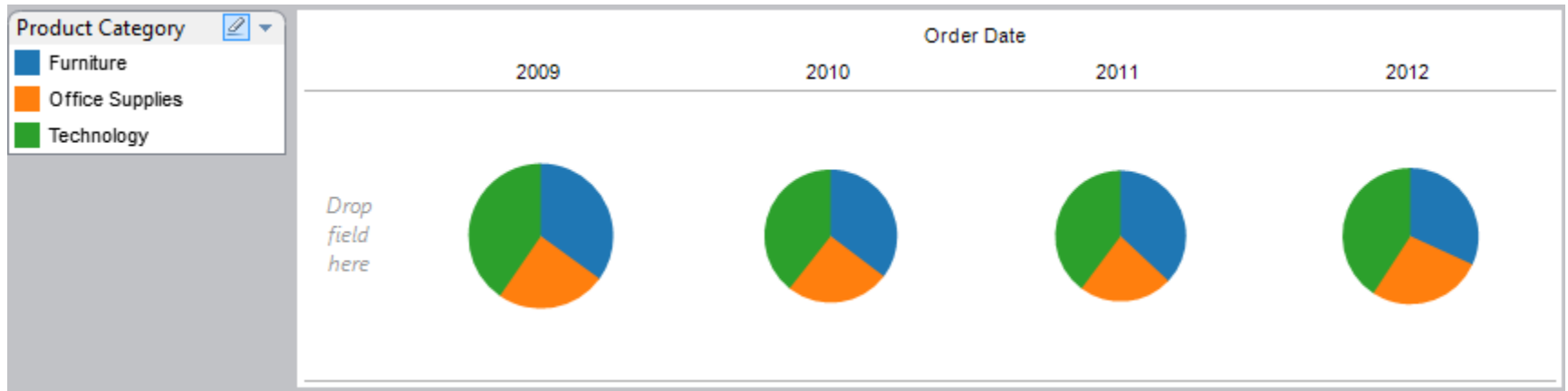
Which product categories are trending up?



Which product categories are trending up?



Attentive



Pre-attentive

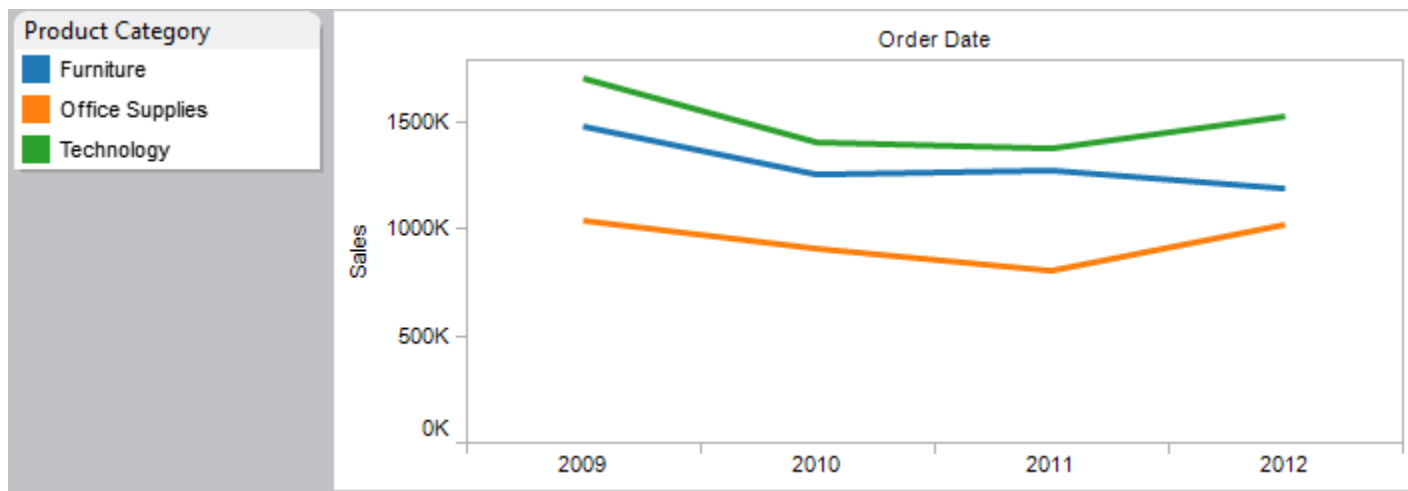


Tableau Product Family

Tableau Desktop

- Author/Create
- Visualize



Tableau Reader

- Free Desktop Client
- Only packaged workbooks
- No data Security



Tableau Public

- Free hosted solution
- No security
- Data source & size limitations



Tableau Server

- Self hosted
- Scalable & Secure
- Enterprise Solution



Tableau Online

- Paid SaaS solution
- Data source limitations
- Named User Licensing



Connecting to Data

File based sources

Big Data sources

Connect to Data

In a file

- Tableau Data Extract
- Microsoft Access
- Microsoft Excel
- Text File
- Import from Workbook

On a server

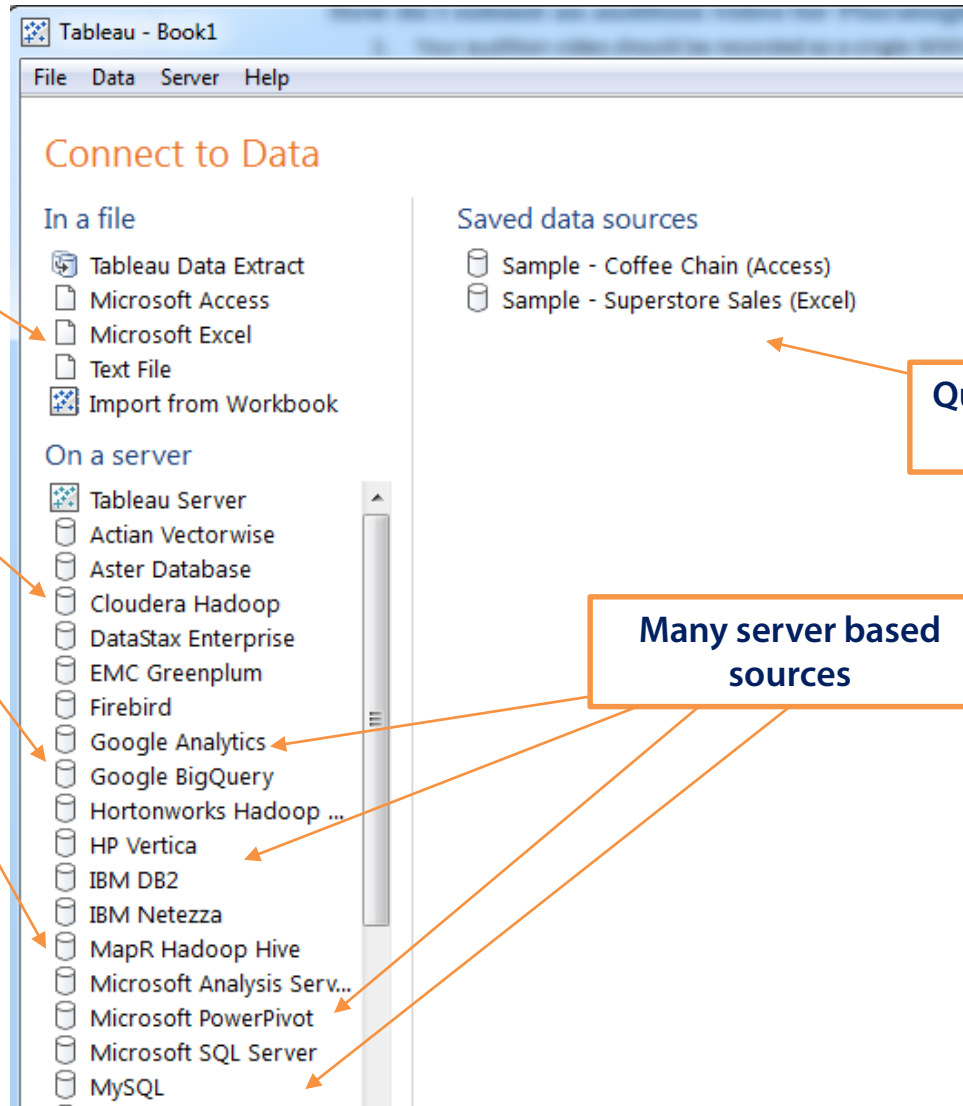
- Tableau Server
- Action Vectorwise
- Aster Database
- Cloudera Hadoop
- DataStax Enterprise
- EMC Greenplum
- Firebird
- Google Analytics
- Google BigQuery
- Hortonworks Hadoop ...
- HP Vertica
- IBM DB2
- IBM Netezza
- MapR Hadoop Hive
- Microsoft Analysis Serv...
- Microsoft PowerPivot
- Microsoft SQL Server
- MySQL

Saved data sources

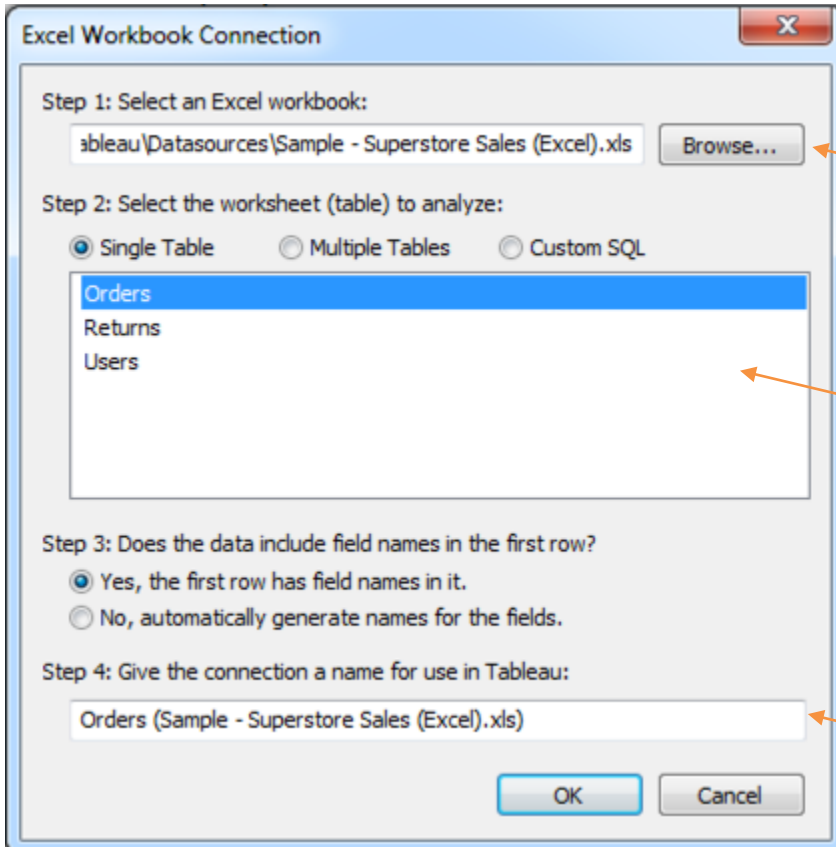
- Sample - Coffee Chain (Access)
- Sample - Superstore Sales (Excel)

Quick Access to Saved Data Sources

Many server based sources



Setting up the Connection – Single Table



The image shows the 'Excel Workbook Connection' dialog box in Tableau. It has four steps: Step 1: Select an Excel workbook; Step 2: Select the worksheet (table) to analyze; Step 3: Does the data include field names in the first row?; Step 4: Give the connection a name for use in Tableau. Annotations with orange boxes and arrows point to specific elements: 'Connection String' points to the 'Browse...' button in Step 1; 'Choose a single table' points to the 'Orders' table in the list in Step 2; 'Name the connection' points to the text box in Step 4.

Excel Workbook Connection

Step 1: Select an Excel workbook:

Tableau\Datasources\Sample - Superstore Sales (Excel).xls Browse...

Step 2: Select the worksheet (table) to analyze:

☒ Single Table ☐ Multiple Tables ☐ Custom SQL

Orders
Returns
Users

Step 3: Does the data include field names in the first row?

☒ Yes, the first row has field names in it.
☐ No, automatically generate names for the fields.

Step 4: Give the connection a name for use in Tableau:

Orders (Sample - Superstore Sales (Excel).xls)

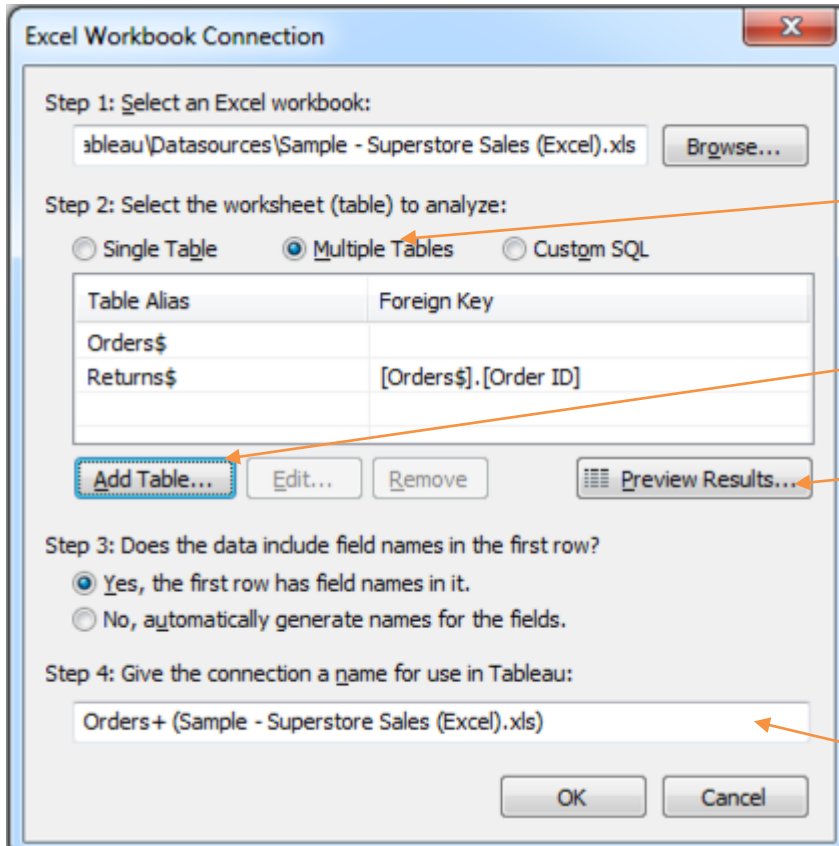
OK Cancel

Connection String

Choose a single table

Name the connection

Setting up the Connection – Multiple Tables



The image shows the 'Excel Workbook Connection' dialog box in Tableau. It is divided into four steps. Step 1: 'Select an Excel workbook:' with a text field containing 'tableau\Datasources\Sample - Superstore Sales (Excel).xls' and a 'Browse...' button. Step 2: 'Select the worksheet (table) to analyze:' with three radio buttons: 'Single Table', 'Multiple Tables' (which is selected), and 'Custom SQL'. Below these is a table with two columns: 'Table Alias' and 'Foreign Key'. The table contains two rows: 'Orders\$' and 'Returns\$'. The 'Returns\$' row has '[Orders\$].[Order ID]' in the 'Foreign Key' column. Below the table are three buttons: 'Add Table...', 'Edit...', and 'Remove'. To the right of these buttons is a 'Preview Results...' button. Step 3: 'Does the data include field names in the first row?' with two radio buttons: 'Yes, the first row has field names in it.' (selected) and 'No, automatically generate names for the fields.' Step 4: 'Give the connection a name for use in Tableau:' with a text field containing 'Orders+ (Sample - Superstore Sales (Excel).xls)'. At the bottom are 'OK' and 'Cancel' buttons. Four orange arrows point from text boxes on the right to specific elements: 'Choose Multiple Tables' points to the 'Multiple Tables' radio button; 'Add Table' points to the 'Add Table...' button; 'Preview Results' points to the 'Preview Results...' button; and 'Name the connection' points to the text field in Step 4.

Excel Workbook Connection

Step 1: Select an Excel workbook:

tableau\Datasources\Sample - Superstore Sales (Excel).xls Browse...

Step 2: Select the worksheet (table) to analyze:

☐ Single Table ☒ Multiple Tables ☐ Custom SQL

Table Alias	Foreign Key
Orders\$	
Returns\$	[Orders\$].[Order ID]

Add Table... Edit... Remove Preview Results...

Step 3: Does the data include field names in the first row?

☒ Yes, the first row has field names in it.
☐ No, automatically generate names for the fields.

Step 4: Give the connection a name for use in Tableau:

Orders+ (Sample - Superstore Sales (Excel).xls)

OK Cancel

Choose Multiple Tables

Add Table

Preview Results

Name the connection

Setting up the Connection – Custom SQL

Excel Workbook Connection

Step 1: Select an Excel workbook:
\\bleau\Datasources\Sample - Superstore Sales (Excel).xls Browse...

Step 2: Select the worksheet (table) to analyze:
☐ Single Table ☐ Multiple Tables ☒ Custom SQL ...

```
SELECT [Orders$].[City] AS [City],  
[Orders$].[Customer Name] AS [Customer Name],  
[Orders$].[Customer Segment] AS [Customer Segment],  
[Orders$].[Discount] AS [Discount],  
[Orders$].[Order Date] AS [Order Date],  
[Orders$].[Order ID] AS [Order ID],  
[Orders$].[Order Priority] AS [Order Priority]
```

Step 3: Does the data include field names in the first row?
☒ Yes, the first row has field names in it.
☐ No, automatically generate names for the fields.

Step 4: Give the connection a name for use in Tableau:
Custom SQL (Sample - Superstore Sales (Excel).xls)

OK Cancel

Choose Custom SQL

Click the ellipse (...)

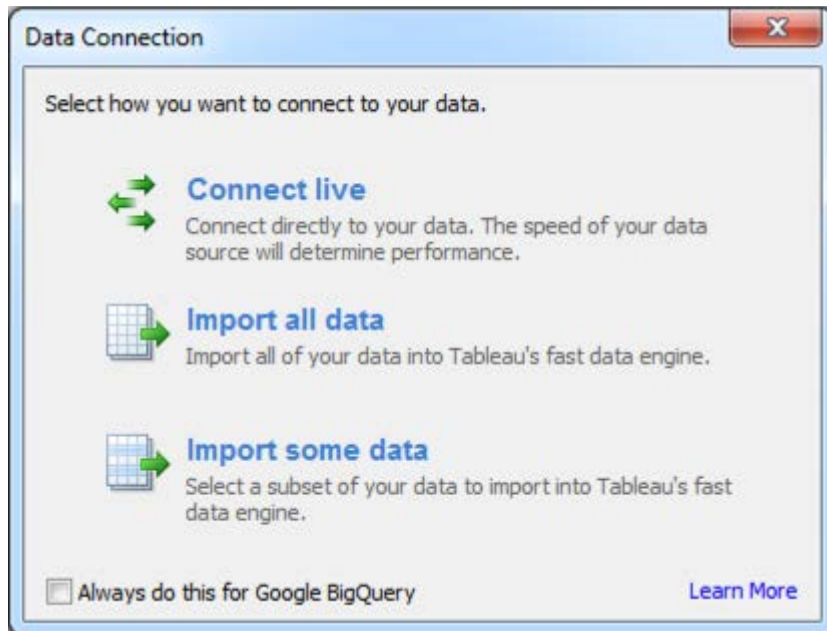
Name the connection

Enter SQL

Edit Custom SQL

```
SELECT [Orders$].[City] AS [City],  
[Orders$].[Customer Name] AS [Customer Name],  
[Orders$].[Customer Segment] AS [Customer Segment],  
[Orders$].[Discount] AS [Discount],  
[Orders$].[Order Date] AS [Order Date],  
[Orders$].[Order ID] AS [Order ID],  
[Orders$].[Order Priority] AS [Order Priority],  
[Orders$].[Order Quantity] AS [Order Quantity],  
[Orders$].[Product Base Margin] AS [Product Base Margin],  
[Orders$].[Product Category] AS [Product Category],  
[Orders$].[Product Container] AS [Product Container],  
[Orders$].[Product Name] AS [Product Name],  
[Orders$].[Product Sub-Category] AS [Product Sub-Category],  
[Orders$].[Profit] AS [Profit],  
[Orders$].[Region] AS [Region],  
[Orders$].[Row ID] AS [Row ID],  
[Orders$].[Sales] AS [Sales],  
[Orders$].[Ship Date] AS [Ship Date],  
[Orders$].[Ship Mode] AS [Ship Mode],  
[Orders$].[Shipping Cost] AS [Shipping Cost],  
[Orders$].[State] AS [State],  
[Orders$].[Unit Price] AS [Unit Price]
```

Data Connection Options



Real-time results directly from data source but could be slow depending on source.

All data is imported to Tableau's own Columnar format called a Tableau Data Extract (.tde)

Some data is imported and you have more options for future refreshes of the data source. If I'm importing data I often choose this option.

Importing Some Data

The screenshot shows the 'Extract Data' dialog box with the following sections and options:

- Specify how much data to extract:**
 - Filters (Optional):** A table with two columns: 'Filter' and 'Details'. The first row is selected, showing 'corpus_date' in the filter column and 'includes values greater than or equal to 1,000' in the details column. Below the table are buttons for 'Add...', 'Edit...', and 'Remove'.
 - Aggregation:** Includes a checkbox for 'Aggregate data for visible dimensions' and a 'Roll up dates to' dropdown menu.
 - Number of Rows:** Includes radio buttons for 'All rows' (selected) and 'Sample'. Under 'All rows', there is a checked 'Incremental refresh' checkbox, a dropdown for 'Identify new rows using column:' set to 'corpus_date', a status bar saying 'All rows will be added.', and a 'Full Refresh' button.
 - At the bottom, there are radio buttons for 'Top:' and 'Sample:', each followed by a text input field and the word 'rows'.
- Buttons:** 'History...', 'Hide All Unused Fields', 'Extract', and 'Cancel' are located at the bottom of the dialog.

Filter down to
relevant data
set

Aggregate for increased
performance

Setup Incremental data refreshes using
dates or integers

Pull a sample data set only

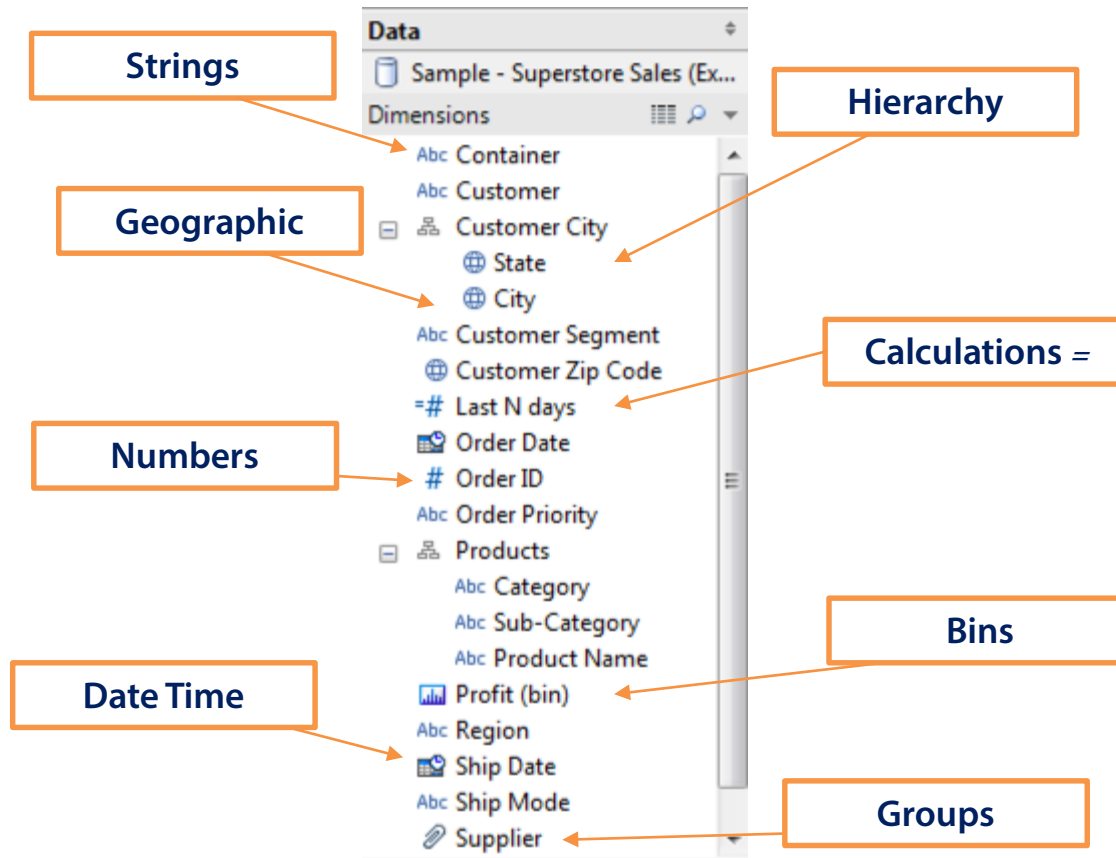
The Data Pane

The diagram illustrates the Tableau Data Pane with several annotations pointing to specific sections:

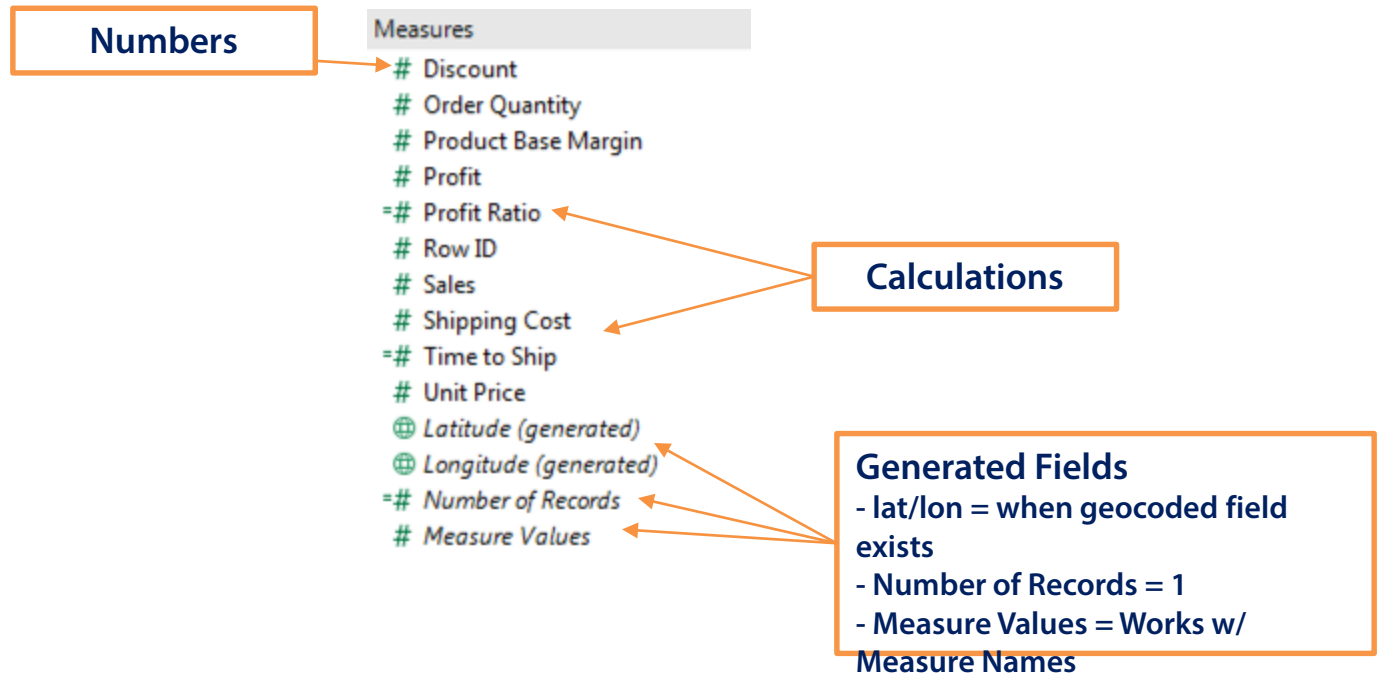
- Our Connections (no limit to how many)**: Points to the **Data** section, which lists the data source: **Sample - Superstore Sales (Ex...**
- Dimensions (context / level of detail)**: Points to the **Dimensions** section, which lists attributes like **Container**, **Customer**, **Customer City**, **State**, **City**, **Customer Segment**, **Customer Zip Code**, **Last N days**, **Order Date**, **Order ID**, **Order Priority**, **Products**, and **Category**.
- Hierarchy (drill-down attributes)**: Points to the **Customer City** and **State** items in the Dimensions list.
- A Calculation (equal sign indicator)**: Points to the **Last N days** item in the Dimensions list, which is marked with an equals sign.
- Measures (aggregates)**: Points to the **Measures** section, which lists aggregate functions like **Discount**, **Order Quantity**, **Product Base Margin**, **Profit**, **Profit Ratio**, **Row ID**, **Sales**, and **Shipping Cost**.
- A Calculation (equal sign indicator)**: Points to the **Profit Ratio** item in the Measures list, which is marked with an equals sign.

The Data Pane interface also includes a **Pages** shelf, **Filters**, **Marks** (with **Color** and **Detail** buttons), and a **Sets** section containing **Top Customers by Profit**. The **Parameters** section at the bottom includes **Profit Bin Size**. The bottom right corner shows **Sheet 1**.

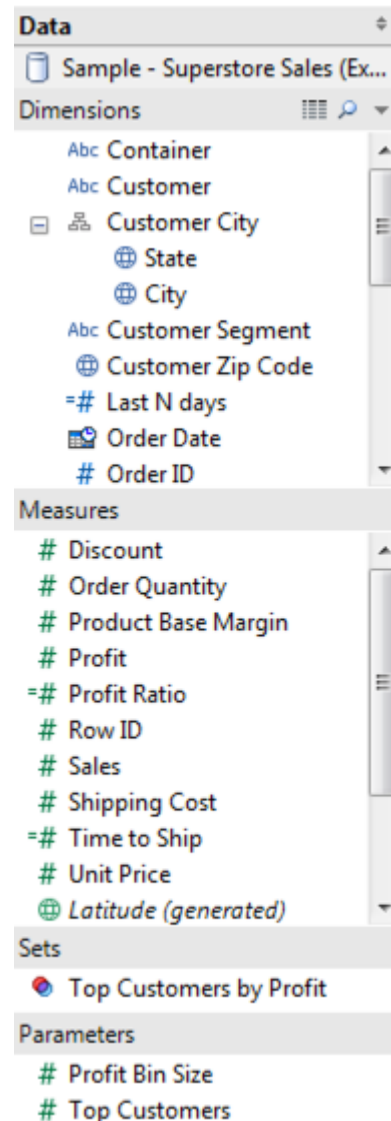
Data Terminology - Dimensions



Data Terminology - Measures



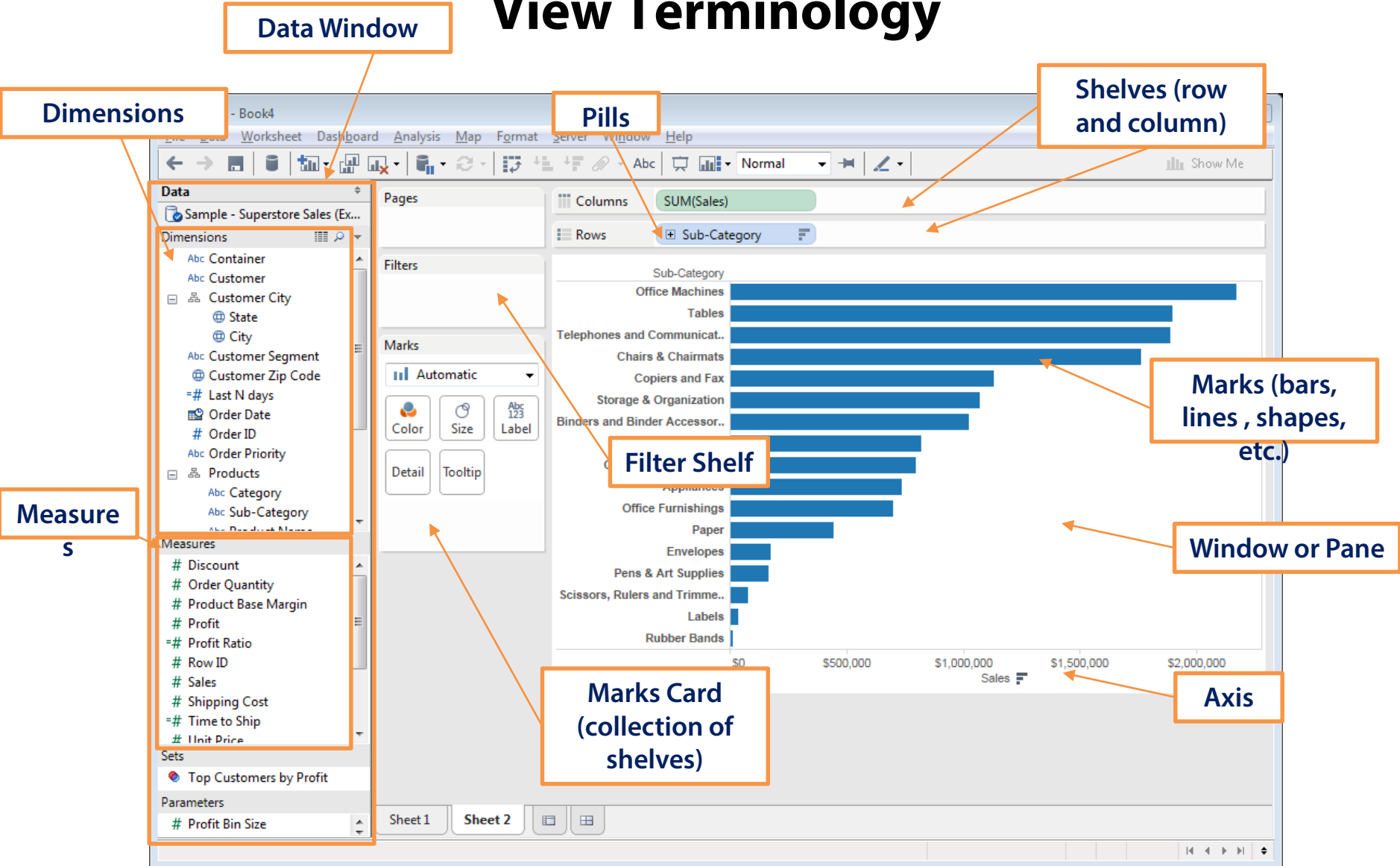
Data Terminology – Sets & Parameters



Sets
- collections of
dimension members

Parameters
- Independent variables
that allow user input

View Terminology



Summary

- Visual analysis
- Visual perception
- Tableau's product family
- Connecting to your data
- Data terminology
- View terminology

Getting dirty with your data

Ben Sullins
Bensullins.com
ben@bensullins.com



pluralsight 
hardcore developer training

Overview

- **Connecting to your data**
- **Joins and Blending**
- **Filtering**
- **Sorting**
- **Hierarchies**
- **Grouping**
- **Sets**
- **Data Types**
- **Dates**
- **Measure Names / Values**

Connecting

Joins and Blending

Filtering

Sorting

Hierarchies

Grouping

Sets

Data Types

Date Values

Measure

Names / Values

Show me the data!

Ben Sullins
Bensullins.com
ben@bensullins.com

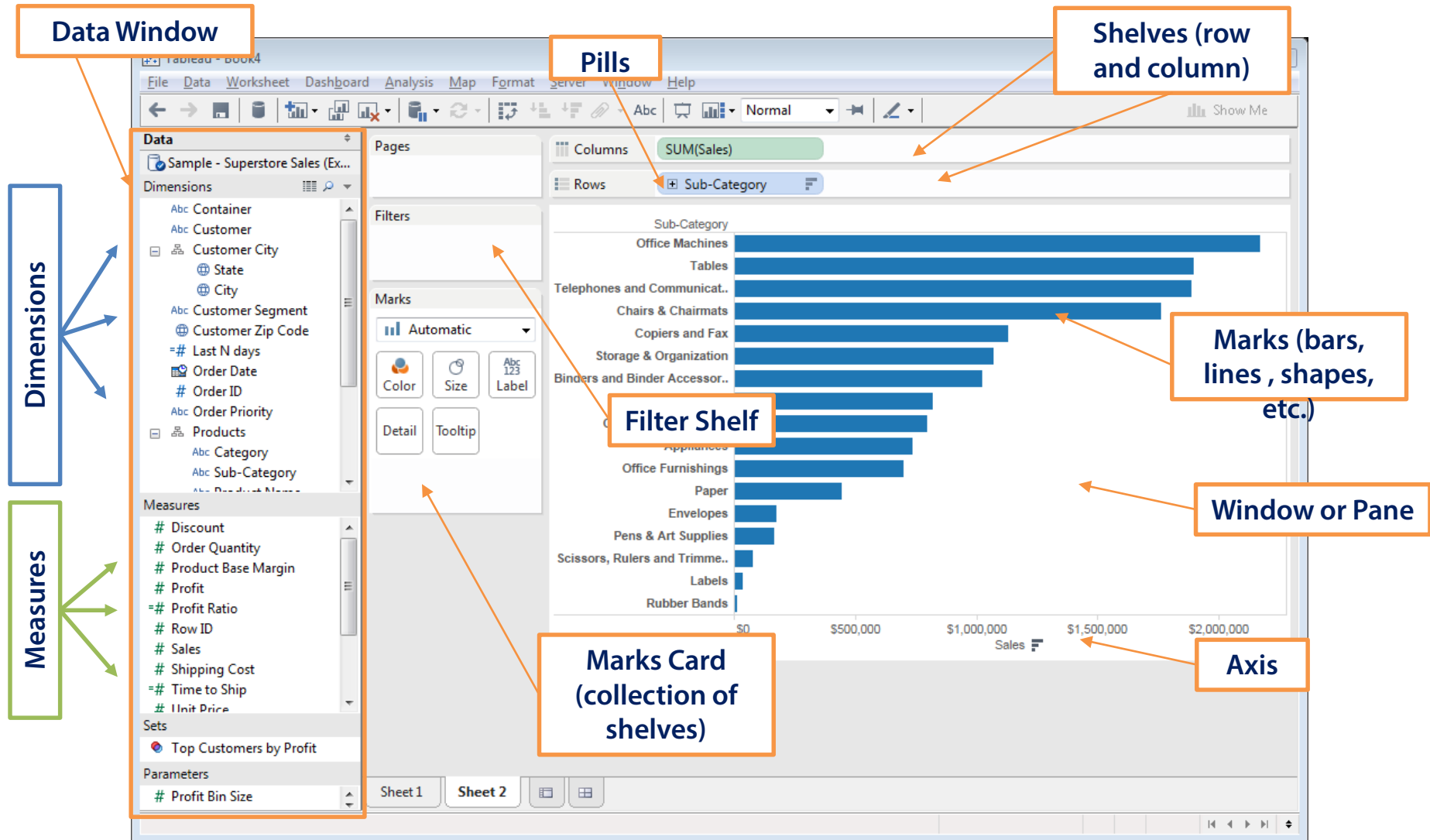


pluralsight 
hardcore developer training

Overview

- Terminology Review
- Basic Charting
- Mapping
- Formatting
- Advanced Charting

Terminology / User Interface



Basic Charting

Mapping

Formatting

Advanced Charting

Summary

- Terminology Review
- Basic Charting
- Mapping
- Formatting
- Advanced Charting

The Intelligence Factor

Ben Sullins
Bensullins.com
ben@bensullins.com



pluralsight 
hardcore developer training

Overview

- **Basic Calculations**
- **Table Calculations**
- **Parameters**
- **Stats Functions**

Basic Calculations

Table Calculations

Parameters

Stats Functions

Summary

- **Basic Calculations**
- **Table Calculations**
- **Parameters**
- **Stats Functions**

Painting the whole picture

Ben Sullins
Bensullins.com
ben@bensullins.com



pluralsight 
hardcore developer training

Overview

- **Dashboards Defined**
- **Basic Dashboarding**
- **Formatting**
- **Actions**

Dashboards Defined

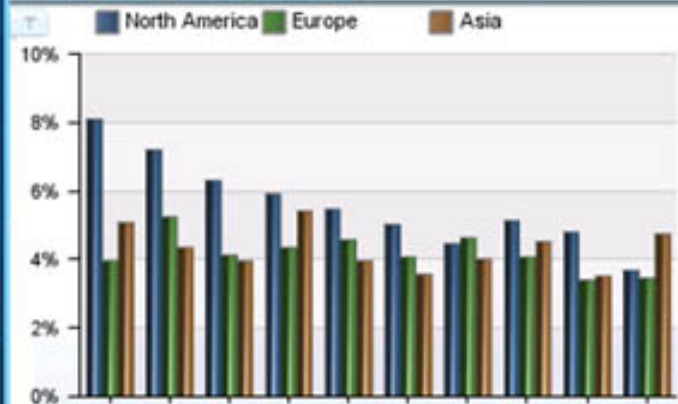
Visual display
of the
most important information needed to
achieve one or more objectives
which
fits entirely on a single screen
so it can be
monitored at a glance

Examples

Regional Performance

Win/Loss Trend	Region	KPP	Rep Days	Win / Loss %	Win / Loss Target
	North America		1869	37.06%	
	Europe		609	33.03%	
	Asia		2168	10.81%	

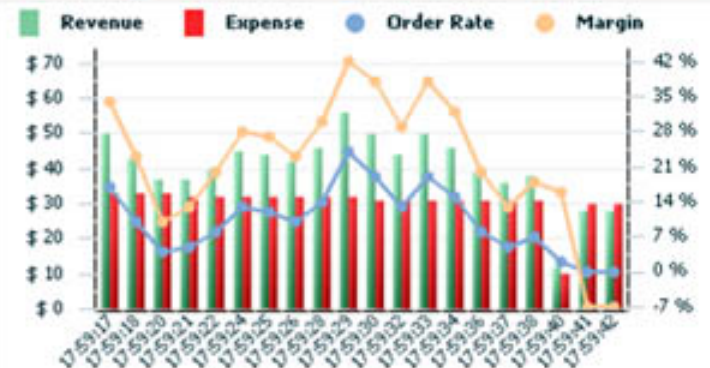
Regional Market Share



Product Performance

Customer Region	Product Line	Growth	Win / Loss %	Shipping Cost
Asia	Alpha	6%	6.6%	6,964,426.54
	Charger	7%	20.1%	2,424,143.07
	Nova	7%	22.8%	2,168,947.17
Asia - Summary		7%		
Europe	Alpha	21%	29.0%	1,302,856.25
	Charger	26%	24.5%	1,487,554.56
	Nova	29%	41.0%	700,060.55
Europe - Summary		25%		

Advertising Performance



Time Period

2008-Q3 2008-Q4 2009-Q1 2009-Q2 2009-Q3 2009-Q4 2010-Q1 2010-Q2 2010-Q3 2010-Q4

2008-Q3 - 2010-Q4

Future Chips
INTERNATIONAL



BUSINESS PERSPECTIVE

Select View

Corporate

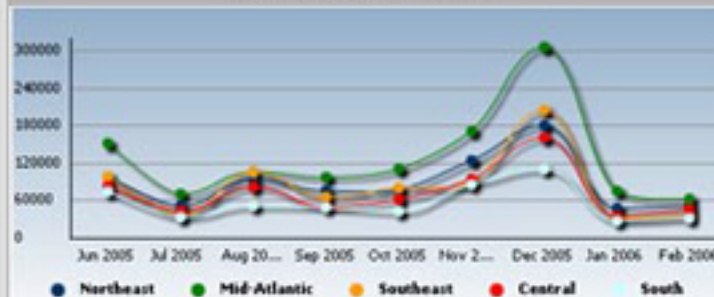
Regional

Corporate Revenue

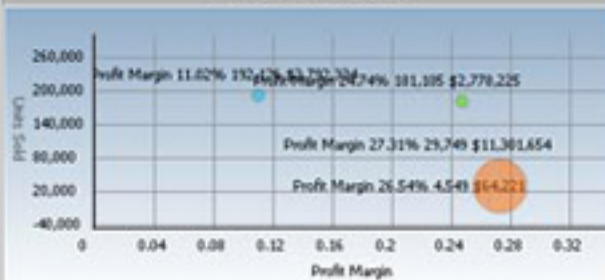


Revenue in millions of dollars

Regional Performance

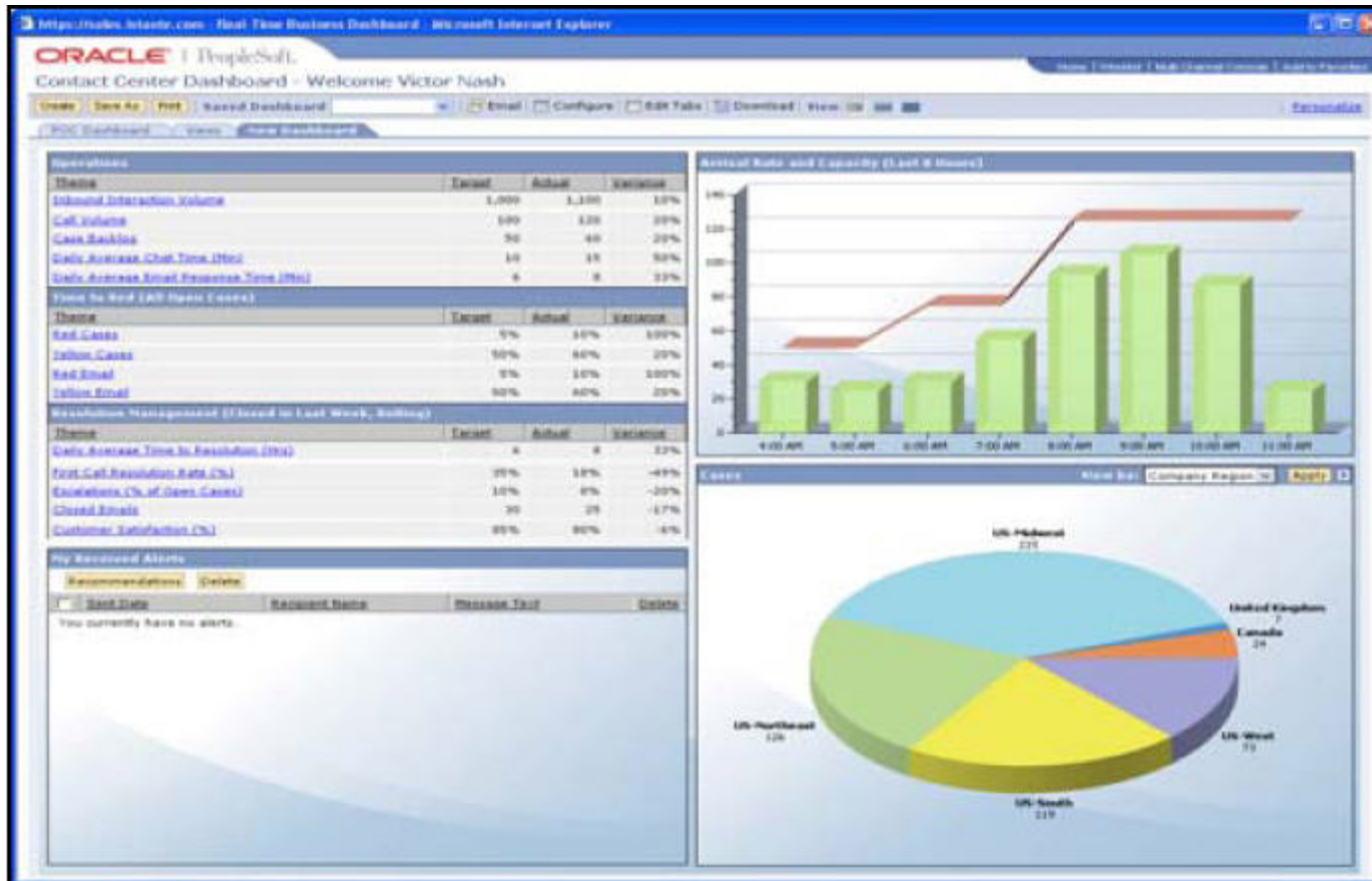


Category Analysis

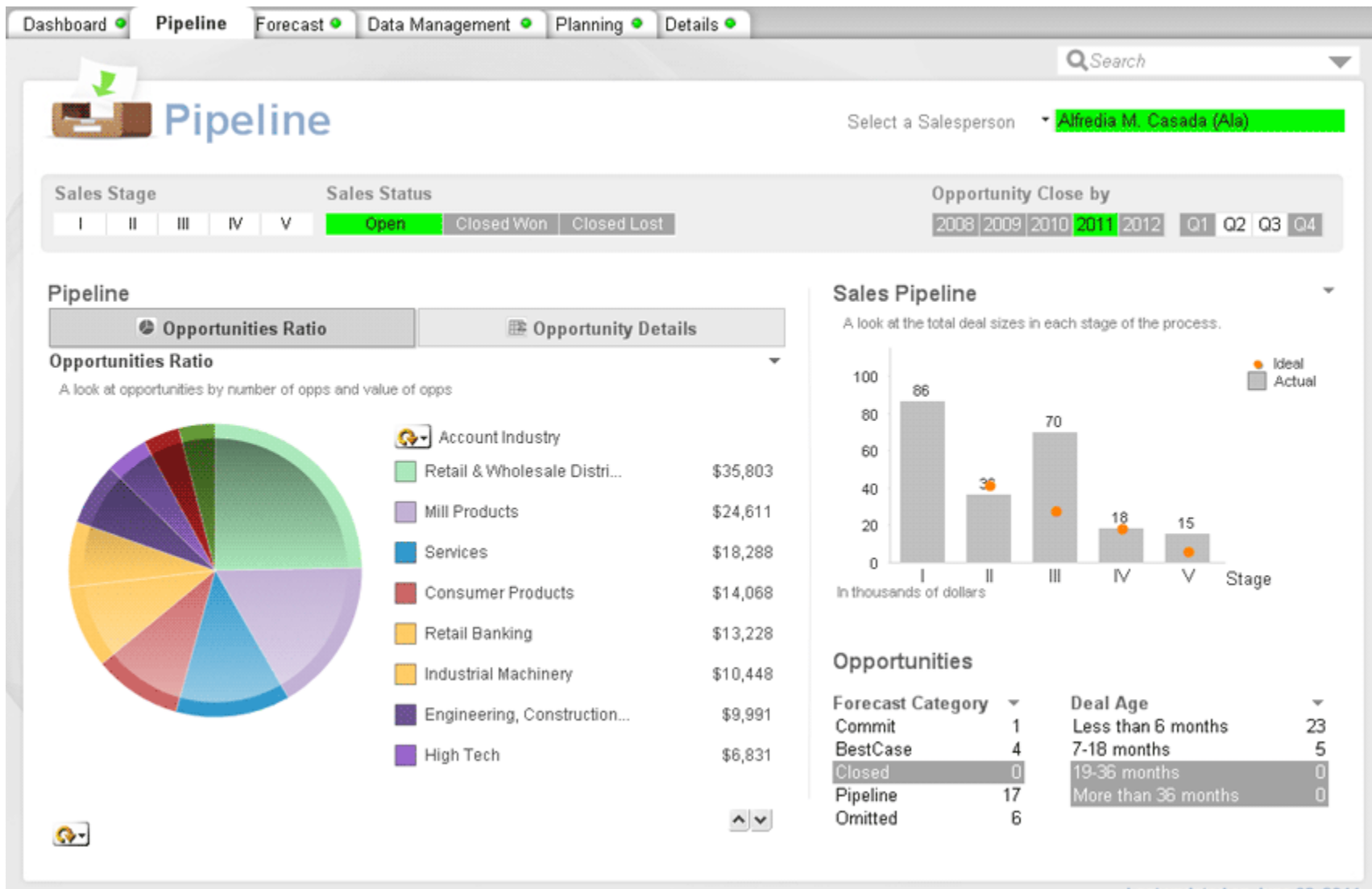


Subcategory Analysis

Books	Electronics	Movies	Music
Subcategory	Revenue Forecast		Units Sold
Business	\$14,146		1,041
Art & Architecture	\$13,341		682
Literature	\$12,606		1,706
Science & Technology	\$11,423		341
Sports & Health	\$6,739		552
Books - Miscellaneous	\$2,035		227







QlickView

U.S. Forest Fire Hot Spots, 2002-2012

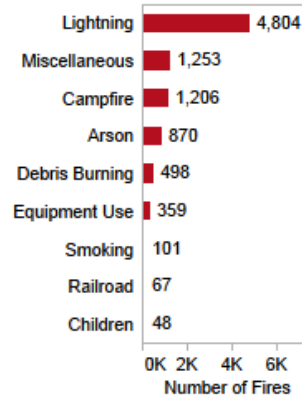


10 years of U.S. forest fire data reveals that most fires burn less than a quarter of an acre, occur in the western part of the country during the summer months and are caused by lightning.

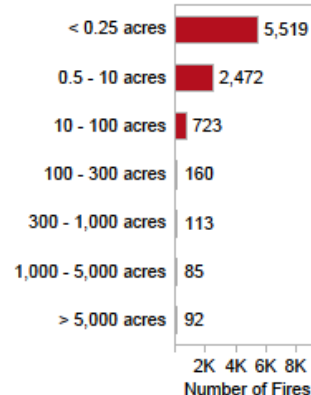
Clicking on a bar filters all other views.

Select a Year:
2007

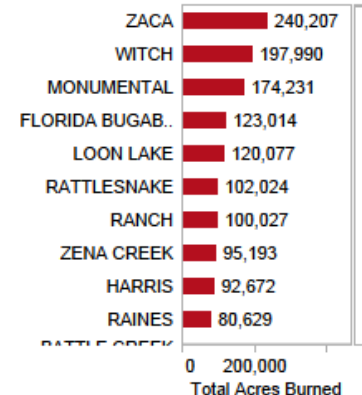
What was the cause?



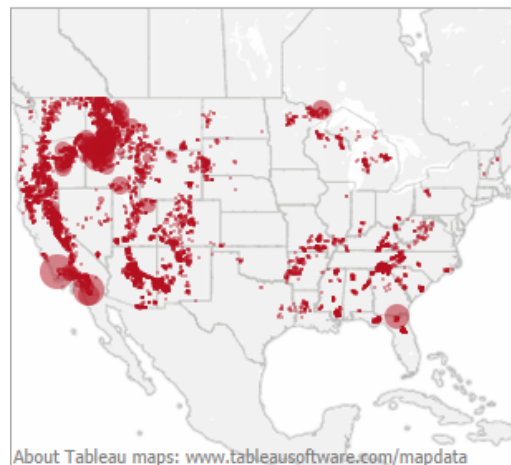
How big is the fire?



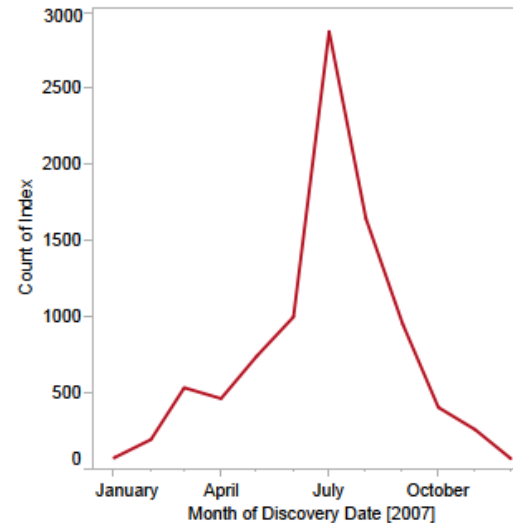
Largest fires by location



Where do they occur?



What time of the year do they occur?



Source: <https://fam.nwgc.gov/fam-web/>

Now

Network

ERP

Data Warehouse

Web Site

Email

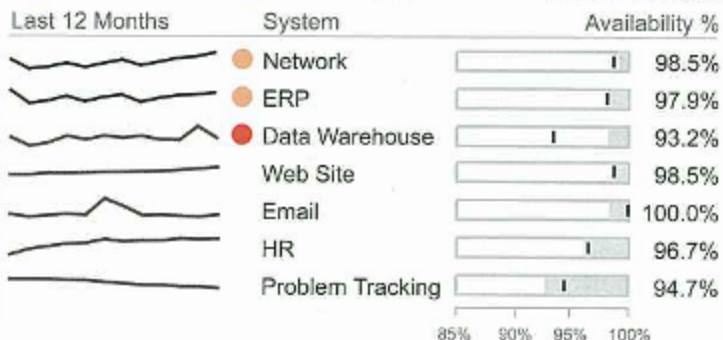
CIO Dashboard

(As of December 19, 2004, 3:35 PM)

Help

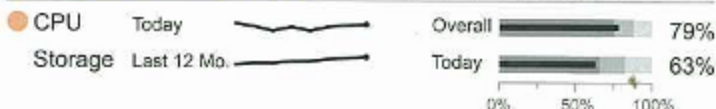
System Availability (Last 30 days)

(■ Actual; ■ Acceptable)

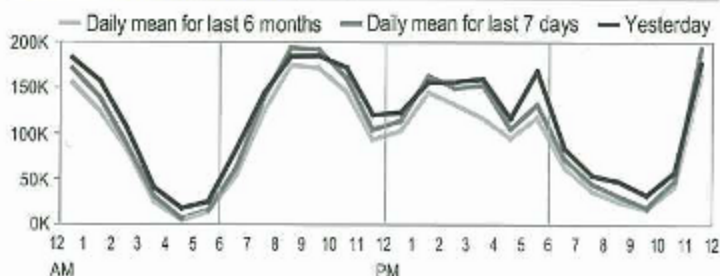


Hardware % of Capacity

(■ Actual; ■ Good; ■ Excessive; ■ Critical)

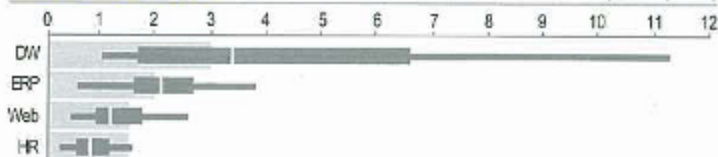


Daily Network Traffic (Kilobytes)



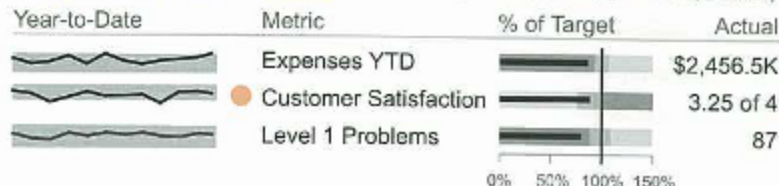
Response Time (Distribution in seconds)

(■ Acceptable)



Key Non-System Metrics

(■ Actual; ■ Good; ■ Excessive; ■ Critical)



Major Project Milestones

Project	Milestone	Due Date	Days Until/ Past Due
ERP Upgrade	Full system test	01/10/06	22
● Add services data to DW	ETL coding	12/15/05	-3
● Upgrade mainframe OS	Prepare plan	12/23/05	4
Disaster recovery site	Install hardware	01/08/06	20
● Budgeting system	Hire team	12/06/05	-13
Web site face-lift	Move into production	01/05/05	17

5 Top Projects in the Queue (Sorted by priority)

Project	Status	Funding Approved	Sched. Start
1 Professional services module	Pending available staff	X	05/10/06
2 Upgrade MS Office	Cost-benefit analysis		02/15/06
3 Failover for ERP	Preparing proposal		06/02/06
4 Upgrade data warehouse HW	Evaluating options	X	04/15/06
5 Executive dashboard	Vendor assessment		07/01/06

Critical Events (Next 14 Days)

Event	Group Responsible	Date
Full system maintenance outage from 9-11 PM	G. Jones	12/21/05
Present hardware upgrade proposal to CEO	Self/M. Smith	12/22/05
Tom visiting from Asia office	Self	12/23/05
Prepare quarterly financial for public announcement	Self	01/04/06
Present revised information strategy to steering comm.	J. Kane	01/06/05

Basic Dashboarding

Formatting

Actions

Summary

- **Dashboards Defined**
- **Basic Dashboarding**
- **Formatting**
- **Actions**

Sharing your work

Ben Sullins
Bensullins.com
ben@bensullins.com



pluralsight 
hardcore developer training

Overview

- Overview of Sharing Options
- Tableau Reader
- Tableau Public
- Tableau Server
- Tableau Online

Tableau Sharing Options

Tableau Desktop

- Author/Create
- Visualize



Tableau Reader

- Free Desktop Client
- Only packaged workbooks
- No data Security



Tableau Public

- Free hosted solution
- No security
- Data source & size limitations



Tableau Server

- Self hosted
- Scalable & Secure
- Enterprise Solution



Tableau Online

- Paid SaaS solution
- Data source limitations
- Named User Licensing



Tableau Reader

Tableau Public

Tableau Server

Tableau Online

Summary

- Overview of Sharing Options
- Tableau Reader
- Tableau Public
- Tableau Server
- Tableau Online