Let’s start with a small visual puzzle

How many 9s in the grid below?

Is it easier to count now?

Let’s solve some brainstorming puzzle

We have 15 mins to do it
Just to get a feel of some
parameters which we will discuss later

Einstein puzzle

- There are 5 houses in 5 different colors. In each house lives a person with a different nationality. The 5 owners drink a certain type of beverage, smoke a certain brand of cigar, and keep a certain pet. No owners have the same pet, smoke the same brand of cigar, or drink the same beverage.“

- **The question is:** Who owns the fish?

The constraints are -

- The Brit lives in the red house.
- The Swede keeps dogs as pets.
- The Dane drinks tea.
- The green house is on the left of the white house.
- The green homeowner drinks coffee.
- The person who smokes Pall Mall rears birds.
- The owner of the yellow house smokes Dunhill.
- The man living in the center house drinks milk.
- The Norwegian lives in the first house.
- The man who smokes Blend lives next to the one who keeps cats.
- The man who keeps the horse lives next to the man who smokes Dunhill.
- The owner who smokes Blue master drinks beer.
- The German smokes prince.
- The Norwegian lives next to the blue house.
- The man who smokes Blend has a neighbor who drinks water.

Where did we face the difficulty?

Solution to the question

	House 1	House 2	House 3	House 4	House 5
Color	Yellow	Blue	Red	Green	White
Person	Norwegian	Dane	Brit	**GERMAN**	Swede
Drink	Water	Tea	Milk	Coffee	Beer
Smoke	Dunhill	Blends	Pall Mall	Prince	Blue-masters
Pet	Cats	Horse	Birds	**FISH**	Dogs

Difficulty lies in thinking in 6 directions

Why did we face the difficulty ?

- The previous problem had 6 dimensions –
- First 5 can be guessed easily
  - Position is also a Dimension
  - What about Difficulty level of Data with 150 Dimensions

Why data visualization?

- Human visual cortex dominates our perception

- Accelerates the identification of hidden patterns in data
  - “A picture is worth a thousand words”

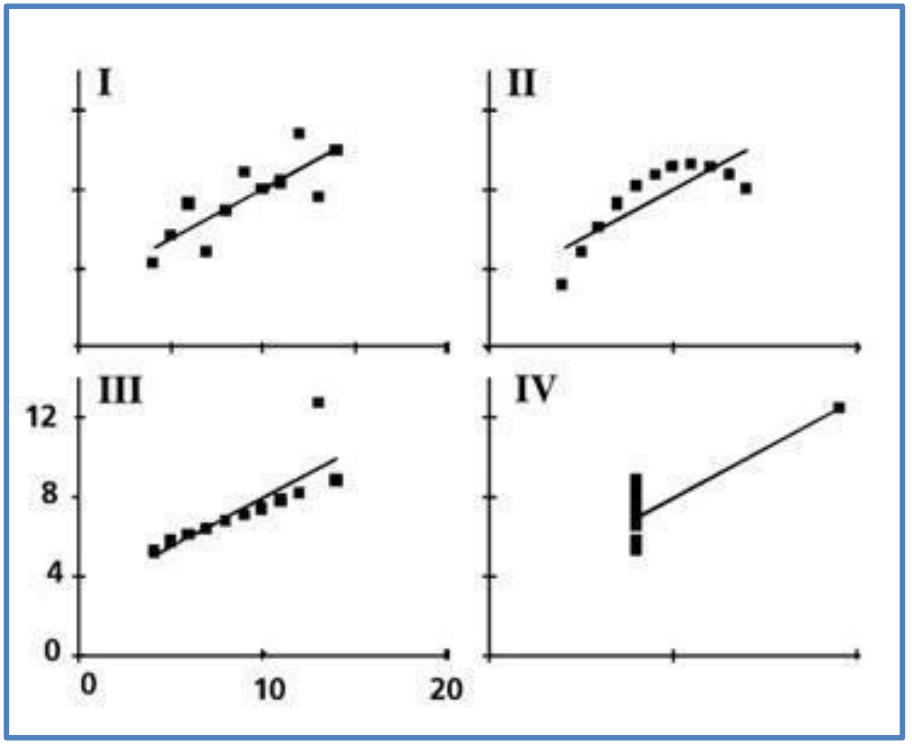
Why is it important?

- Human beings are tremendously influenced by sensory perceptions…
- The way that we learn, grow, understand, and adapt is based on our ability to view, perceive, and conceptualize thoughts and ideas…
- The power to visualize and graphically represent results, ideas, solutions, and problems in multiple dimensions, as well as to manipulate data and virtually collaborate with others, is the next big revolution in technology.

Anscombe's Quartet

Anscombe's Quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89



- This graphic is from the statistician FJ Anscombe to demonstrate why it is important to graph data before analyzing it.
- The quartet consists of four sets of data that have identical simple statistical properties.
- They are, however, very different when graphed. The key take-away here is that graphics do not simply represent the data in visual form, they reveal what the data means.

What is Data Visualization?

- "...to convey information through visual representations."
- "...produces (interactive) visual representations of abstract data to reinforce human cognition; thus enabling the viewer to gain knowledge about the internal structure of the data and causal relationships in it."
- "…involve detection, measurement, and comparison, and is enhanced via interactive techniques and providing the information from multiple views and with multiple techniques"

Visualization Goals

- Answer questions (or discover them)
- Make decisions
- See data in context
- Support graphical calculation
- Find patterns
- Present argument or tell a story
- Inspire

Three Functions of Visualization

- **Record:** store information
- **Analyze:** support reasoning about information
- **Communicate:** convey information to others

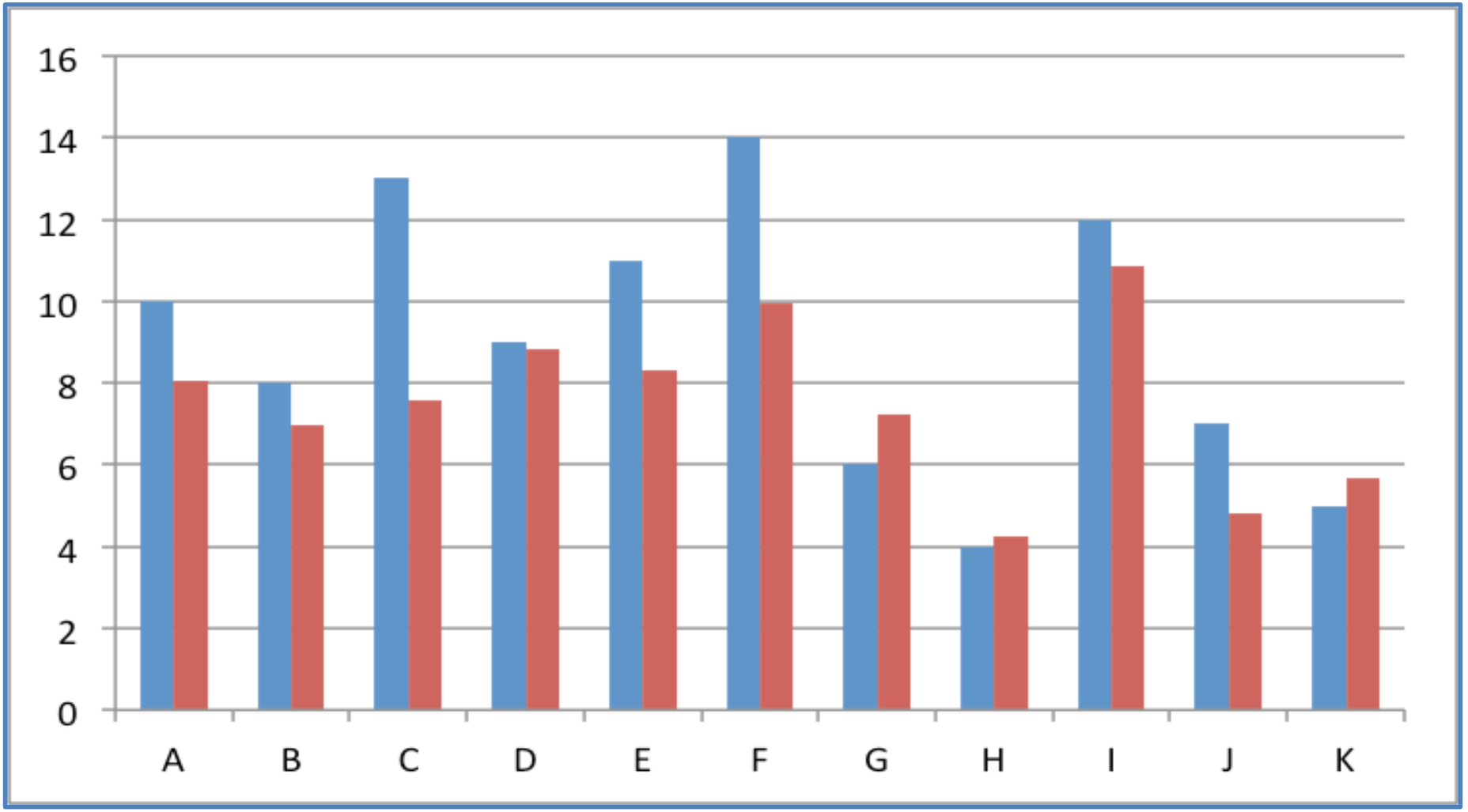
Objectives of Data Visualization

- Two basic types:
  - **Exploration:** find a story the data is telling you
  - **Explanation:** tell a story to an audience

- Represents large quantities of data coherently and efficiently
- Helps the user to discern relationships in the data
- Does not distort what data has to say
- Takes into account your audience's expectations

Types of visualizations

Bar Chart

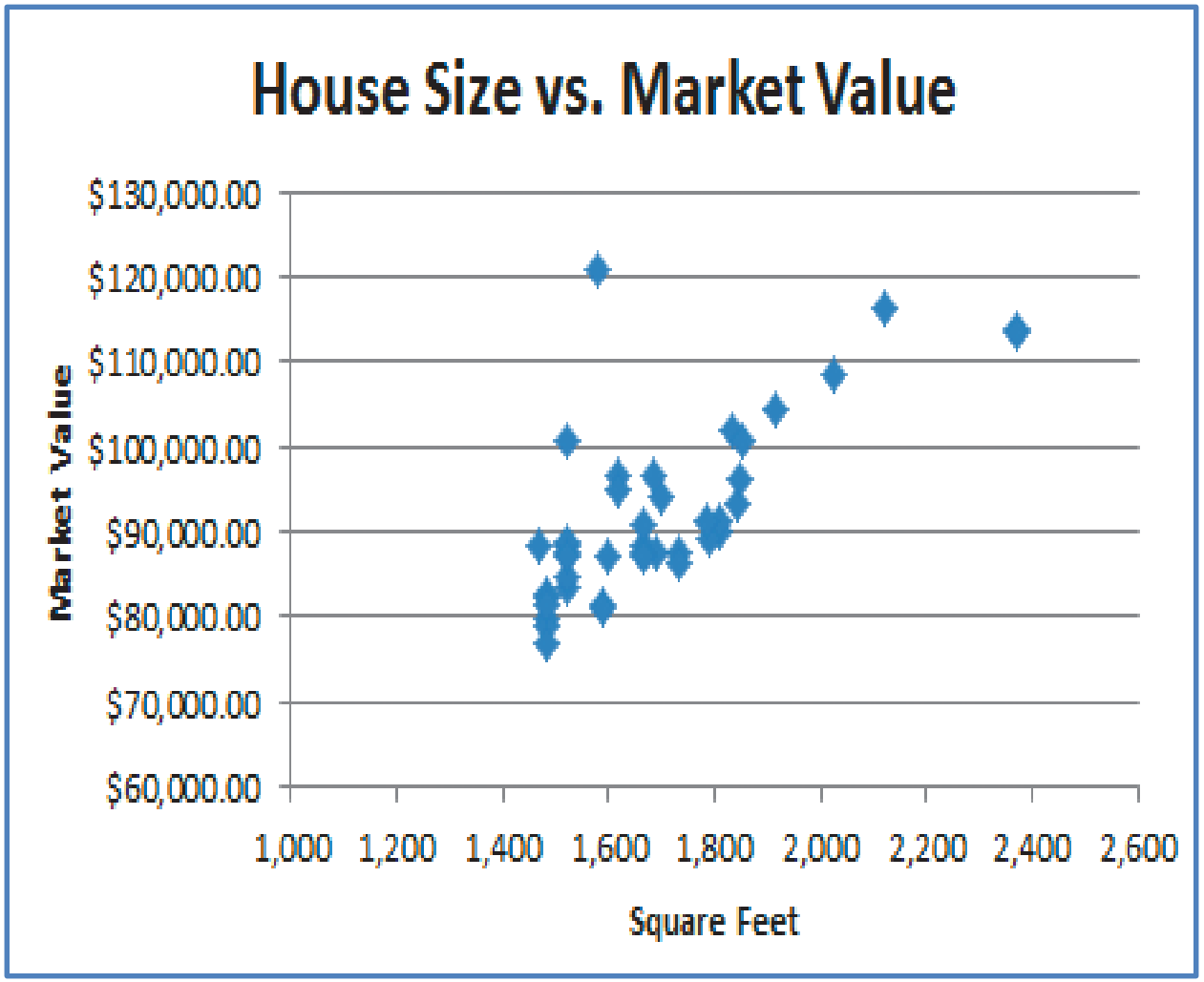


Sample bar chart from Excel

Line Chart

Sample Line chart from Excel based on US Exports data

Scatter Chart



- Suggests correlation between two variables.
- Correlations may be positive (rising), negative (falling), or null (uncorrelated).
- A line of best fit (alternatively called 'trendline') can be drawn.
- Ability to show nonlinear relationships between variables.

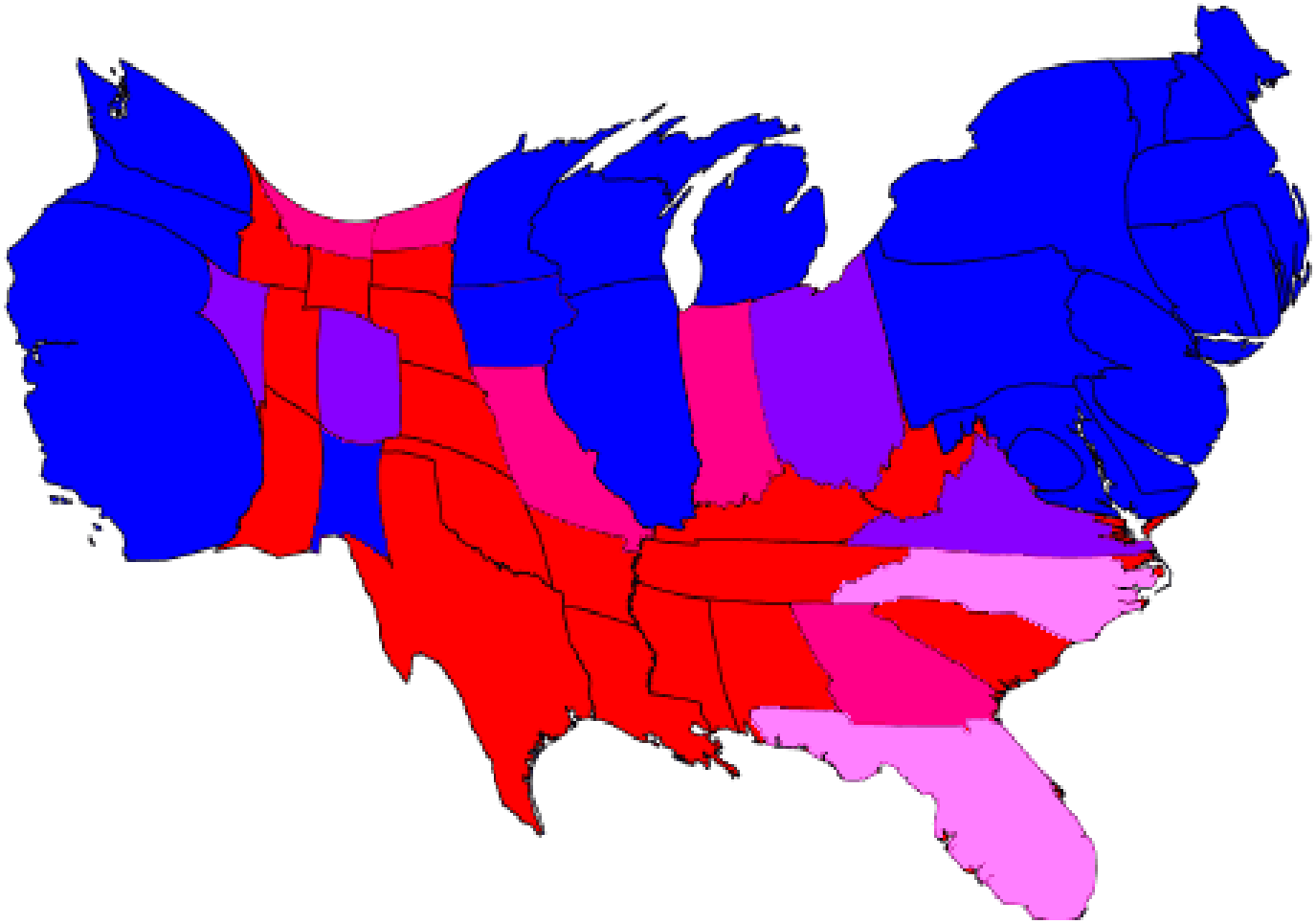
Sample Scatter chart from Excel based on real estate data

Bubble Chart

- Bubble chart is a technique in which a set of numeric quantities is represented by closely packed circles whose areas are proportional to the quantities

Sample Bubble chart from Excel comparing stock characteristics

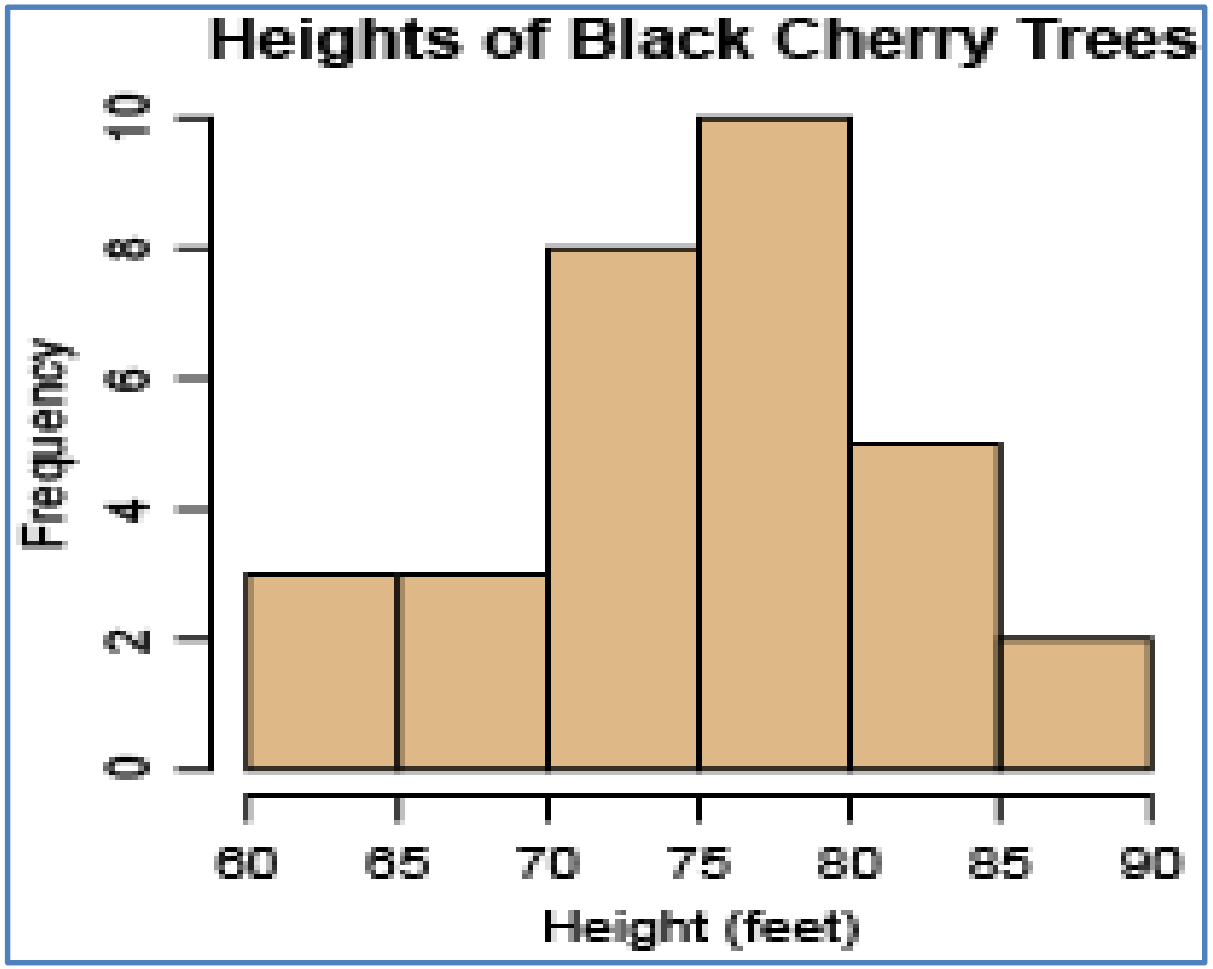
Cartogram



- A cartogram is a map in which some thematic mapping variable – such as travel time, population, or Gross National Product – is substituted for land area or distance

Cartogram of election results based on data from FiveThirtyEight.com

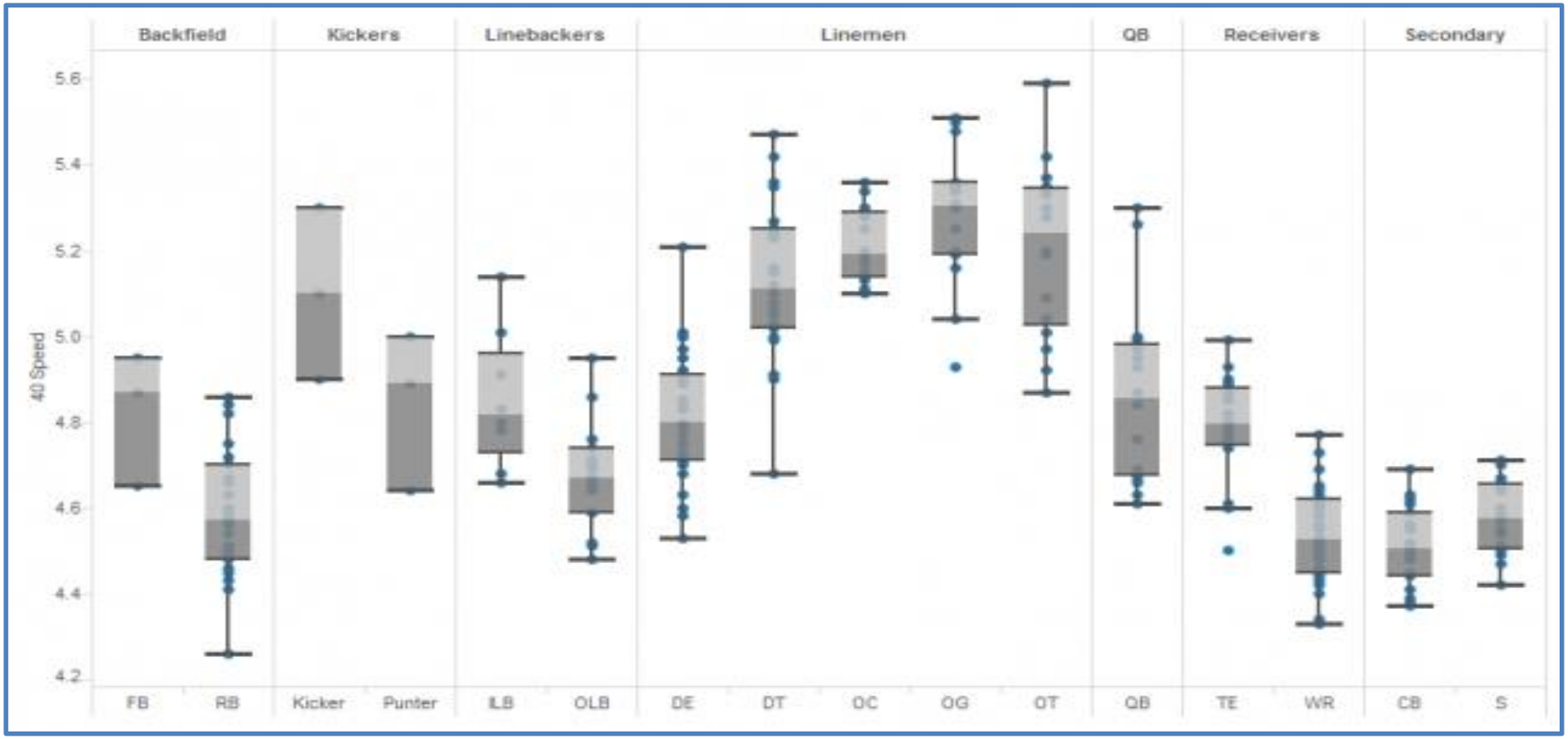
Histogram



- Shows entire distribution of one particular variable.
- Each column's height is determined by the count of the number of items which fall into the **bin.**
- Bin size is a variable you can play with: wider is more smooth, while smaller bins can yield erratic plots.

Cartogram of election results based on data from FiveThirtyEight.com

Box and whisker plot



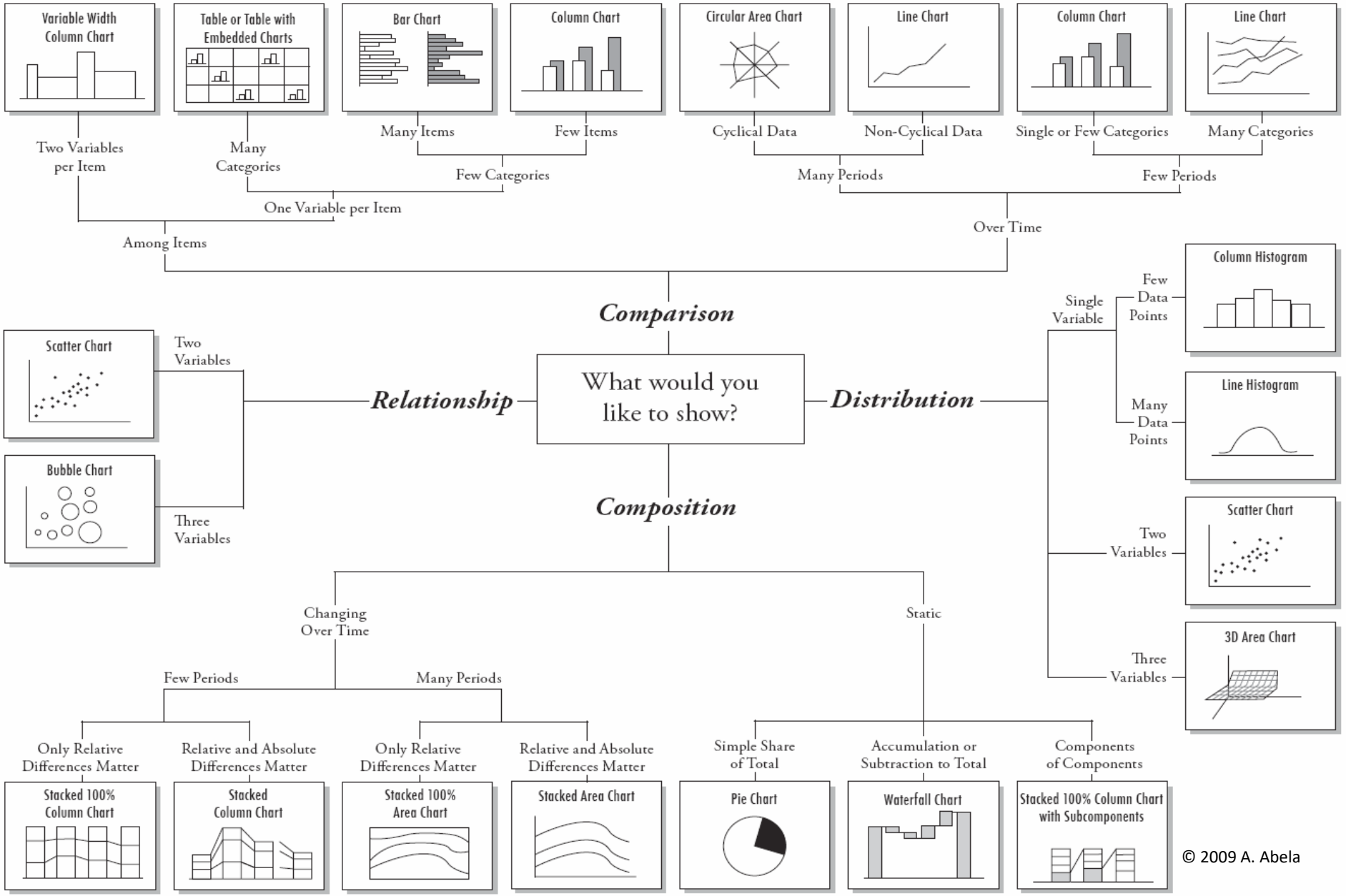
- Display differences between subpopulations in your data.
- Furthest lines are min/max.
- Box shows 25th to 75th percentiles.
- Thick line shows the 50th percentile (the median).

Waterfall chart

- A waterfall chart can be used for analytical purposes, especially for understanding or explaining the gradual transition in the quantitative value of an entity which is subjected to increment or decrement.

Waterfall created in tableau to gauge company's profit

Forming Analytical Patterns Out of Preattentive Attributes



And many more ways to visualize data

Common Data Visualization Issues

- Inappropriate display media i.e. pie chart, 3D charts
- Variety for the sake of variety
- Too much information
- Inconsistent ordering and placement
- Encoding quantitative data inaccurately
- Poorly designed display choices that use noisy fill patterns, line styles, or saturated / bright colors

KDD (Knowledge Discovery in Databases) Process for Data Visualization

- Selection
  - Obtain data from all of sources
- Preprocessing
  - After selecting the data, clean it to make sure it is consistent
- Transformation
  - After preprocessing the data, analyze the format/amount of data
- Data Mining
  - Once the data is in a useable format/content, apply various algorithms based upon the results trying to be achieved
- Interpretation/Evaluation (Visualization)
  - Finally, present the results of the data mining step to the user, so that the results can be used to solve the business need at hand

It is important to consider that the KDD process is useless if the results are not understandable

Data Visualization Best Practices

- **Fulfill the purpose -** Define the purpose of your visualization and you can focus on achieving that objective
- **Consider your audience -** Knowing who will be using your visualization is just as critical as knowing its purpose. You should consider your audience and their degree of familiarity with the subject, the underlying data, and their comfort level with charts and graphs.
- **Provide context -** If your audience doesn't know the source of the data, when it was collected, or how it relates to them, it will lack meaning and impact. Ensure that your visualization contains supporting information that provides answers to all the questions that your audience might have.

Data Visualization Best Practices

- **Show the numbers -** One of the benefits of visualization over spreadsheets is showing a lot more information in less space and without having to show every number. Nonetheless, the numbers are still important and should be shown for all critical components of the viz.
- **Present the most important information first -** Make sure you present the most critical information first. This applies both to displaying the overall layout as well as to sorting data that is shown within each element of the viz.
- **Pay Attention to Aesthetics -** Overall aesthetics are also very important for good visualizations. Make sure the overall presentation of your visualization is neat and crisp – with effective use of color, aligned components, and consistent use of fonts, size, and placement.

Data visualization tools

There are many tools in the Market

- D3
- Qlikview
- Tableau
- Spot fire
- Many more

Tableau?

- Tableau is a rapid BI software
- Tableau is fast becoming the number one tool for many data visualization professionals
  - Straight Forward
  - Easy
  - For the non-programmer that sees data visualization as an important part of his job, Tableau is probably the tool for you.
  - Tableau is a visual analysis solution that allows people to explore and analyze data with simple drag-and-drop operations

Prerequisites

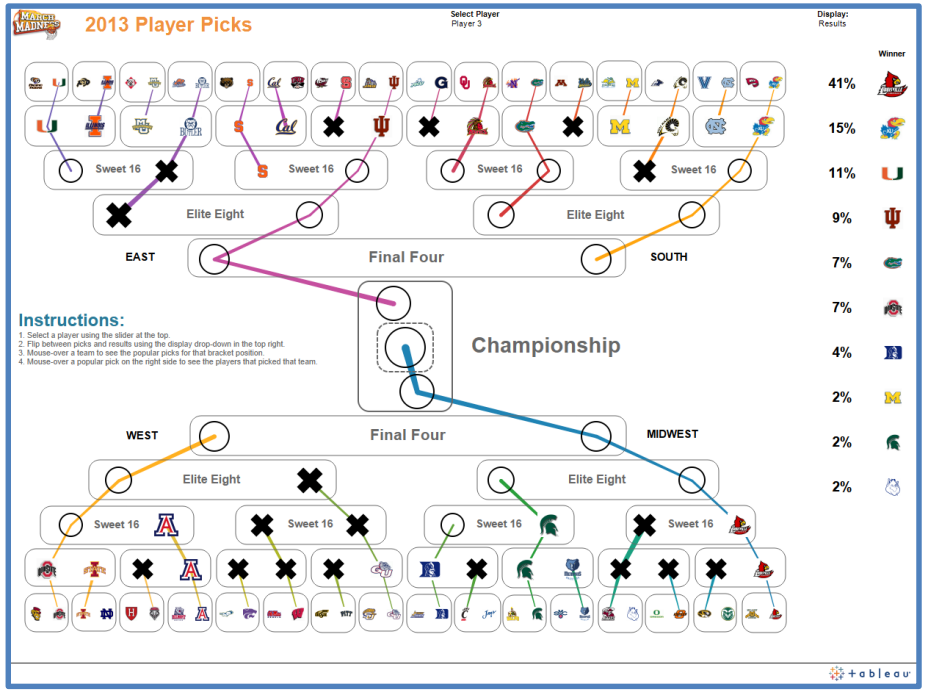
Bit of database knowledge

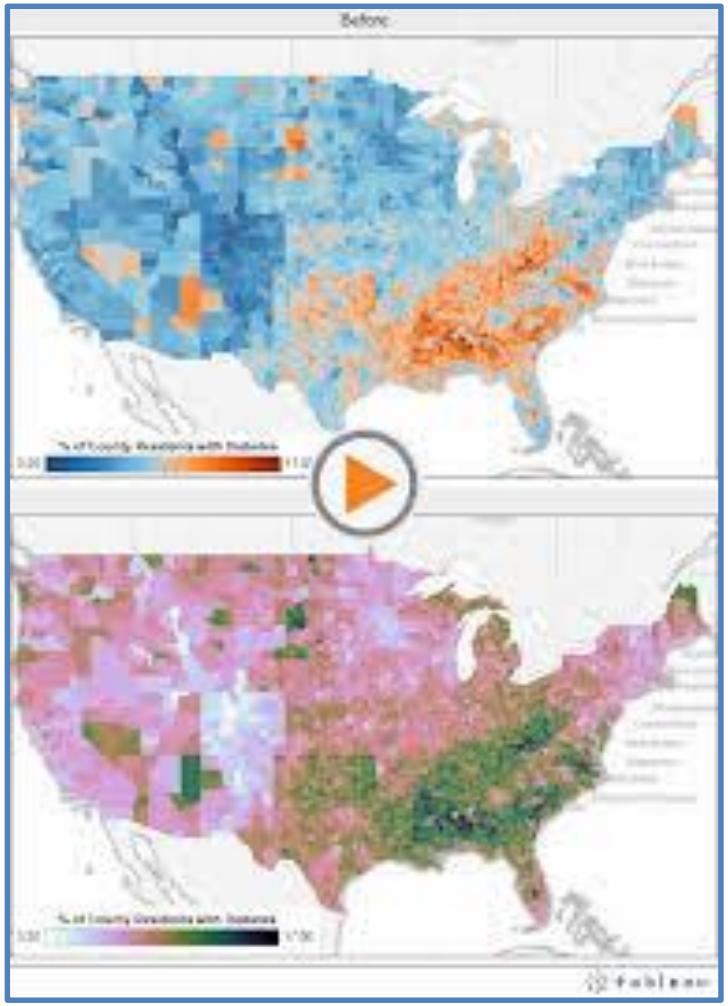
Basic idea of visualizations and dashboards

Windows OS with minimum 2 GB RAM

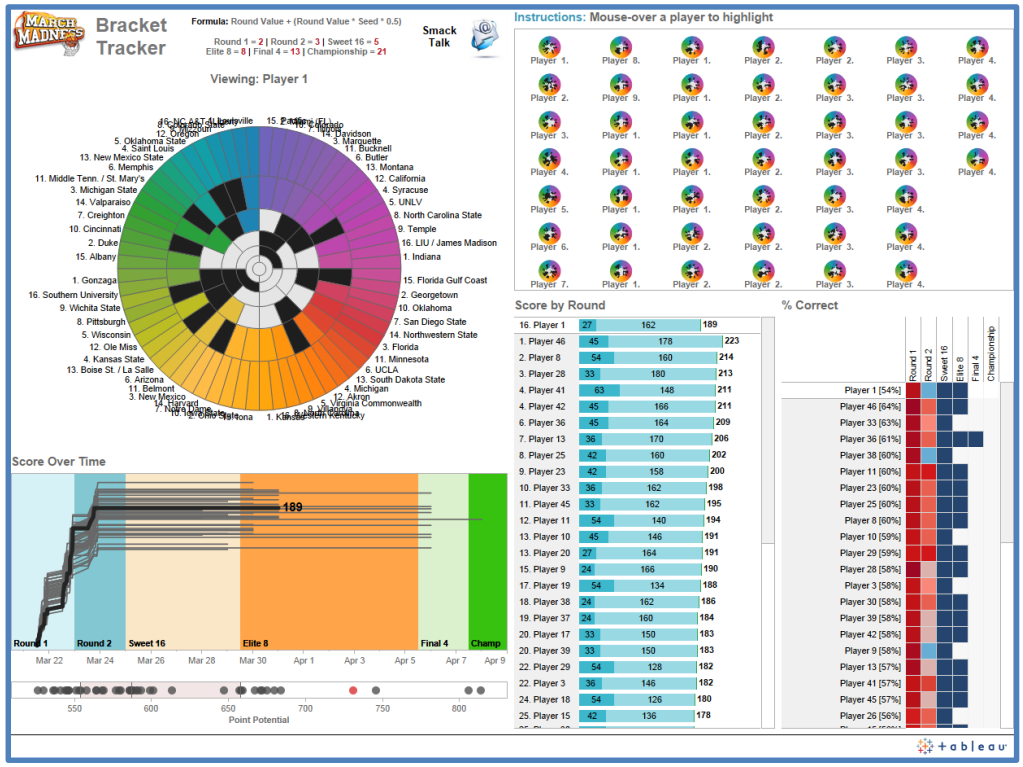
Latest version of Tableau Public or Tableau Desktop Trial version

Sample Tableau Dashboards





Sample Tableau Dashboards



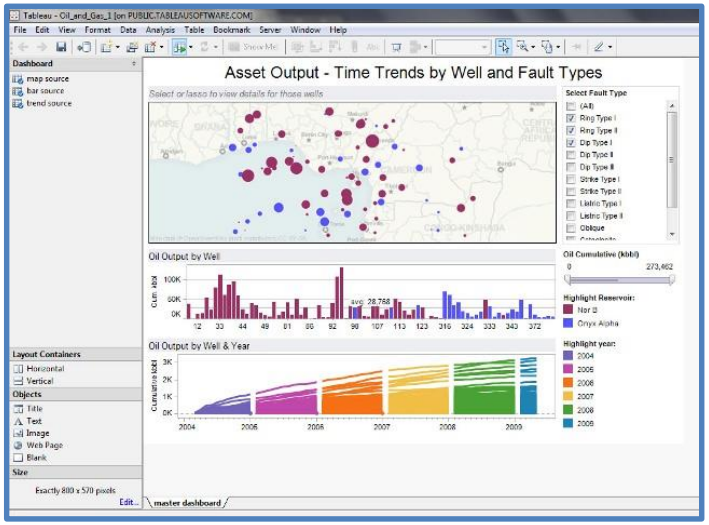


Tableau Products

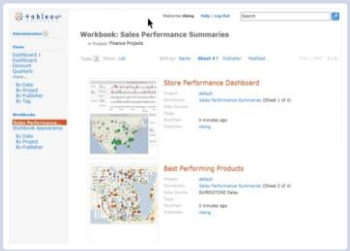
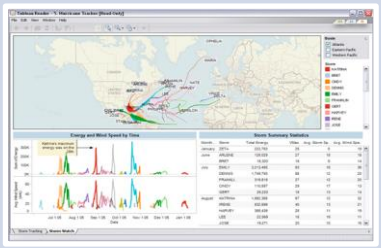
Tableau Desktop	Tableau Server Share - Web	Tableau Reader Share - Local	Tableau Public Share - Everyone
			
+ ad hoc analytics, dashboards, reports, graphs + explore, visualize, and analyze your data + create dashboards to consolidate multiple views + deliver interactive data experiences	+ business intelligence solution scales to organizations of all sizes + share visual analytics with anyone with a web browser + publish interactive analytics or dashboards + secure information and manage metadata + collaborate with others	+ share visualizations & dashboards on the desktop + filter, sort, and page through the views + “Acrobat for Data” + free download	+ create and publish interactive visualizations and dashboards + embed in websites and blogs + free download and free hosting service

Tableau Timeline

PIXAR

STANFORD COMPUTER SCIENCE

VISIO

parc Palo Alto Research Center

•Defense grant to Stanford

• Stanford spin-out

• First patents; invention of **VizQL™**

• Launched Tableau Desktop

Hyperion

• Hyperion global OEM deal

1997 - 2005

PC MAGAZINE Best of the Year 2005

• Awarded **"Product of the Year"** for 1st time by PC Magazine

• Launched *Beta* Tableau Server product

2006

SQL Server EDITOR'S CHOICE 2007 GOLD

• Launched Tableau Server product

• **Chuck Geschke (Chairman of Adobe)** joins Board of Directors

2007

2008 SIIA //CODiE// WINNER

• 1st Annual Customer Conference

(sold out)

• Tableau S receives a Stevie

THE STEVIES

• Global OEM deal (expansion) signed with Oracle

ORACLE

2008

Inc.500

• Inc. 500 list

• 2nd Annual Customer Conference 67% increase in attendance

• Deloitte Fast 500

Deloitte. 2009 Technology Fast 500

• SQL Server Magazine –

Best BI & eporting Tool

SQL Server 2009 Editors' Best GOLD

2009

SOFTWARE 500

Gartner

• 1st Year Magic Quadrant Challenger

• Launched Tableau Public

tableau public

AMERICAN BUSINESS AWARDS STEVIES 2010 WINNER

2010 SIIA //CODiE// WINNER

SQL Server 2010 Community Choice GOLD

SQL Server 2010 Editors' Best GOLD

SQL Server 2010 Community Choice SILVER

SQL Server 2010 Community Choice BRONZE

2010

Gartner

• 2nd Year Magic Quadrant Challenger

• Launched Tableau Digital

The Seattle Times | Business / Technology

Times Watch

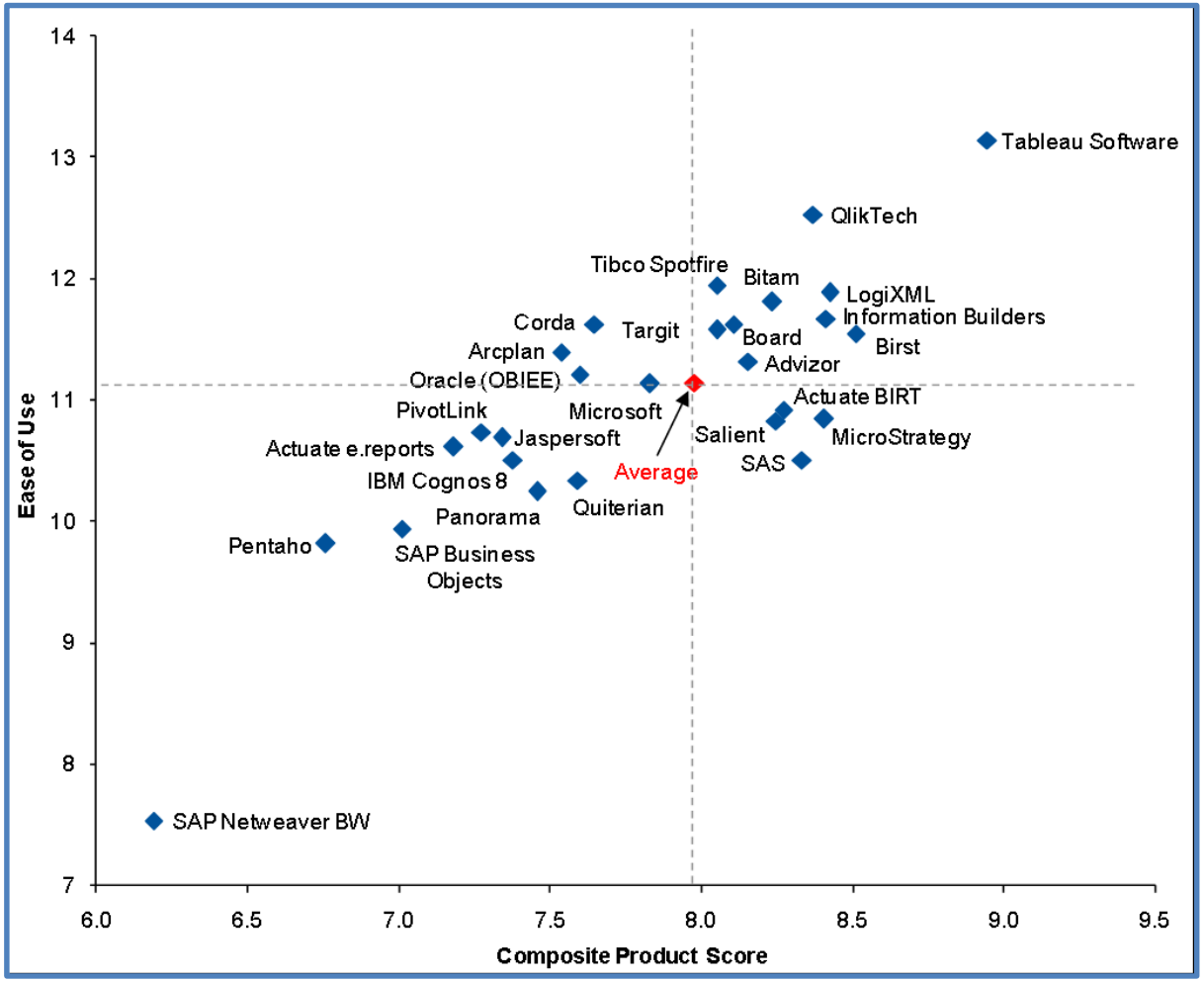
The Big Picture: The state's gains seem minuscule

2011

For [**Data Visualization**]

Overall Product Score vs. Ease of Use

- Find Tableau in the upper right corner
- Overall, Tableau has the thirteen capabilities that Gartner says a BI tool should have better than anyone else
- Overall, it is the easiest to use



Install tableau

1) Open http://www.tableausoftware.com/
2) Download the free trial version
3) Give your Email Id and Register
4) This is free Trial for 14 Days

Sample Workbooks

- When Tableau is Installed , you will get 4 pre-installed datasets to work on

1) Coffee chain
2) Super store English
3) Super store Subset
4) World Bank Indicators

View

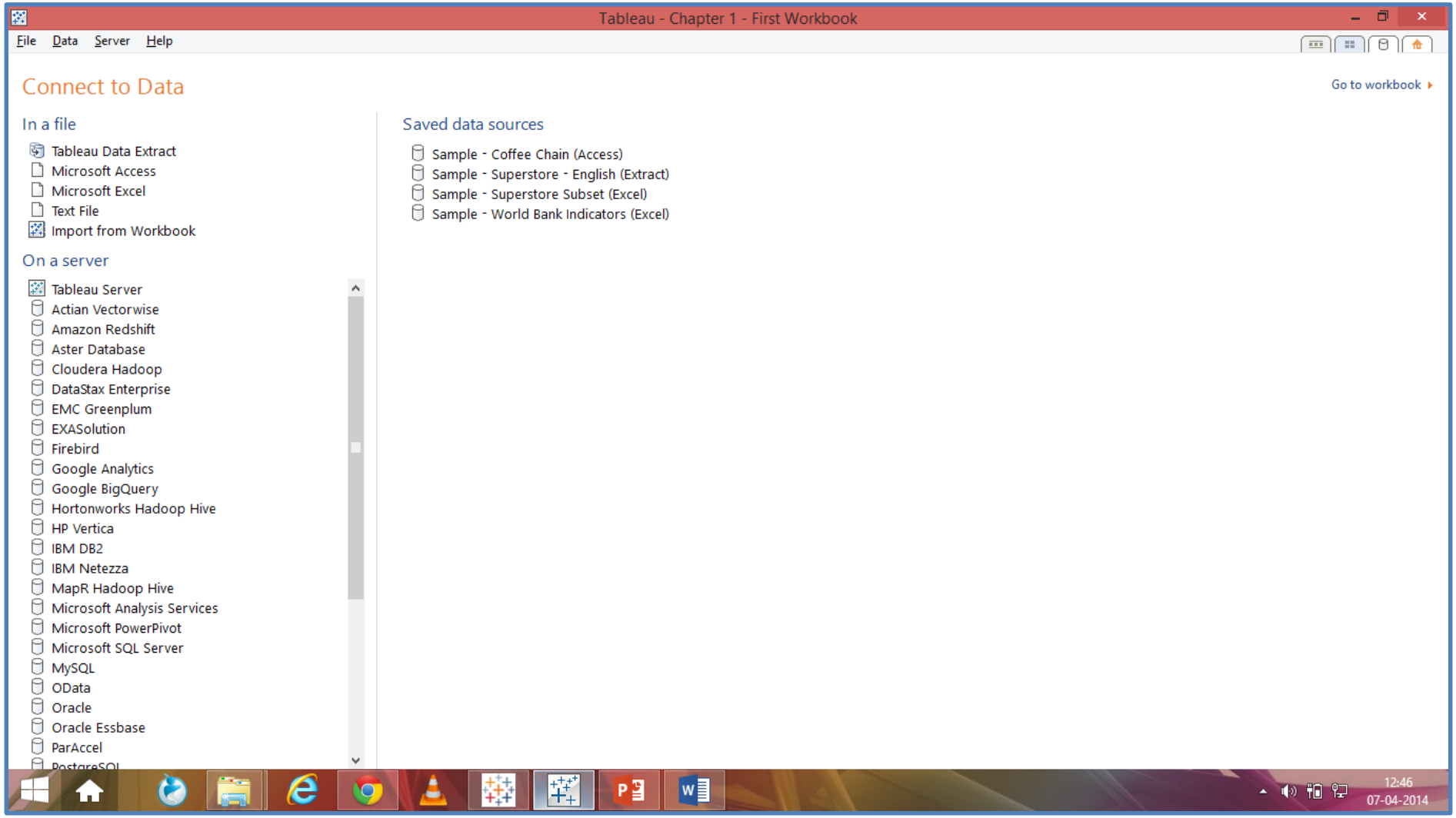


Tableau - Chapter 1 - First Workbook
File Data Server Help
Connect to Data
Go to workbook
In a file
Tableau Data Extract
Microsoft Access
Microsoft Excel
Text File
Import from Workbook
On a server
Tableau Server
Actian Vectorwise
Amazon Redshift
Aster Database
Cloudera Hadoop
DataStax Enterprise
EMC Greenplum
EXASolution
Firebird
Google Analytics
Google BigQuery
Hortonworks Hadoop Hive
HP Vertica
IBM DB2
IBM Netezza
MapR Hadoop Hive
Microsoft Analysis Services
Microsoft PowerPivot
Microsoft SQL Server
MySQL
OData
Oracle
Oracle Essbase
ParAccel
Saved data sources
Sample - Coffee Chain (Access)
Sample - Superstore - English (Extract)
Sample - Superstore Subset (Excel)
Sample - World Bank Indicators (Excel)
12:46
07-04-2014

Tableau Files and File types

- Workbooks
  - Tableau workbook files have the **.twb** file extension and are marked with the workbook icon. Workbooks hold one or more worksheets and dashboards.
- Bookmarks
  - Tableau bookmark files have the **.tbm** file extension and are marked with the bookmark icon. Bookmarks contain a single worksheet and are an easy way to quickly share your work.
- Packaged Workbooks
  - Tableau packaged workbooks have the **.twbx** file extension and are marked with the packaged workbook icon. Packaged workbooks contain a workbook along with any supporting local file data sources and background images. This format is the best way to package your work for sharing with others who don't have access to the data.

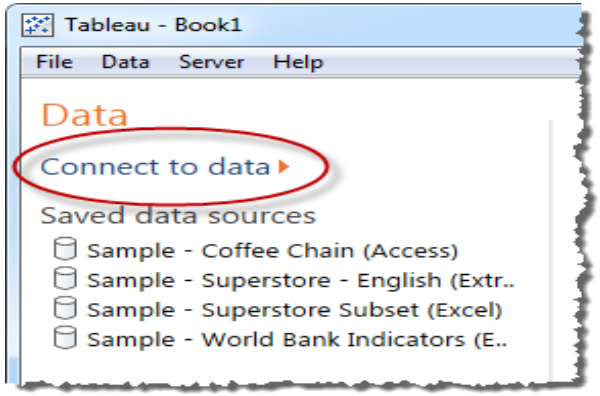
Tableau Files and File types

- Data Extract Files
  - Tableau data extract files have the **.tde** file extension and are marked with the extract icon. Extract files are a local copy of a subset or entire data source that you can use to share data, work offline, and improve database performance.
- Data Connection Files
  - Tableau data connection files have the **.tds** file extension and are marked with the data connection icon. Data connection files are shortcuts for quickly connecting to data sources that you use often.

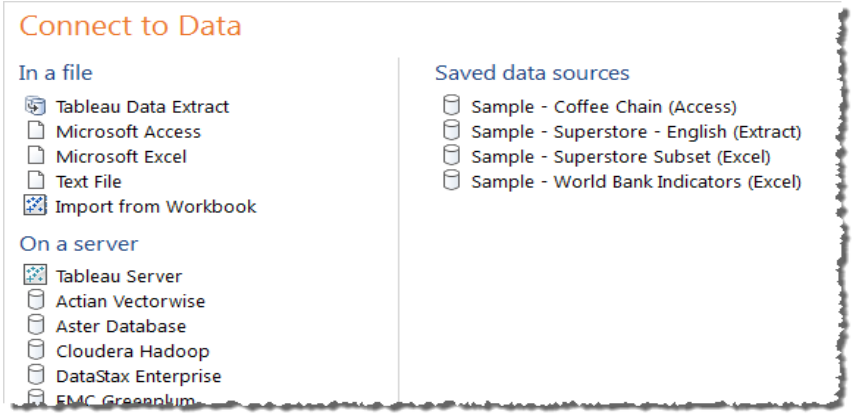
Connect to data

- To build views of your data, you must first connect Tableau to a data source. You can connect to any supported data source with the Connect to Data dialog box.

1) Select Data > Connect to Data or press Ctrl + D on your keyboard. You can also select the Open Data option on the start page.



2) On the Connect to Data page, select the type of data you want to connect to.

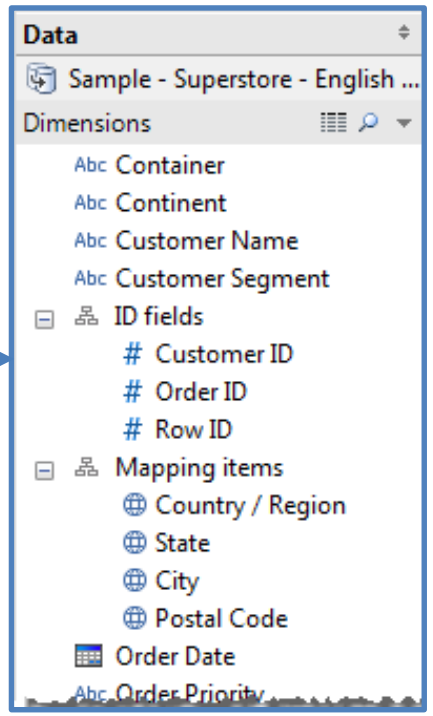


Connect to data (Cont'd…)

3) A data source-specific dialog box opens that allows you to complete the connection process.

- Another way to connect to data is to import from a workbook. A workbook can contain multiple connections to different data sources. To import a connection from a workbook click the Import from Workbook button at the bottom of the Select Saved Connection tab in the Connect to Data dialog box.

- After the connection is established, the data source fields display on the left side of the workbook in the Data window.



Dimensions and Measures

Dimensions

- Dimensions typically produce headers when added to the rows or columns shelves in the view. By default, Tableau treats any field containing qualitative, categorical information as a dimension. This includes, for instance, any field with text or dates values.
- For instance, you might calculate the Sum of “Sales” for every “State”. In this case the State field is acting as a dimension because you want to aggregate sales for each state.

Measures

- Measures typically produce axes when added to the rows or columns shelves. By default, Tableau treats any field containing numeric (quantitative) information as a measure.
- For instance, you might calculate the Sum of “Sales” for every “State”. In this case, the Sales field is acting as a measure because you want to aggregate the field for each state.

Data Types

Icon	Description
Abc	Text values
	Date values
	Date & Time values
#	Numerical values
T/F	Boolean values (relational only)
	Geographic values (used with maps)

- Sometimes Tableau may identify a field with a data type that is incorrect.
- For example, a field that contains dates may be identified as an integer rather than a date.
- You can change the data type in Tableau by right-clicking the field in the Data window, selecting Change Data Type, and then selecting the appropriate data type.

Rows and Columns Shelves

- The Columns shelf creates the columns of a table, while the Rows shelf creates the rows of a table. You can place an unlimited number of fields on these shelves.
- When you place a dimension on the Rows or Columns shelf, headers for the members of that dimension are created. When you place a measure on the Rows or Columns shelf, quantitative axes for that measure are created. As you build up your data view with more fields, additional headers and axes are included in the table and you get an increasingly detailed picture of your data.
- In the view shown below, the members of the Customer Segment dimension are displayed as column headers, while the Profit measure is displayed as a vertical quantitative axis.

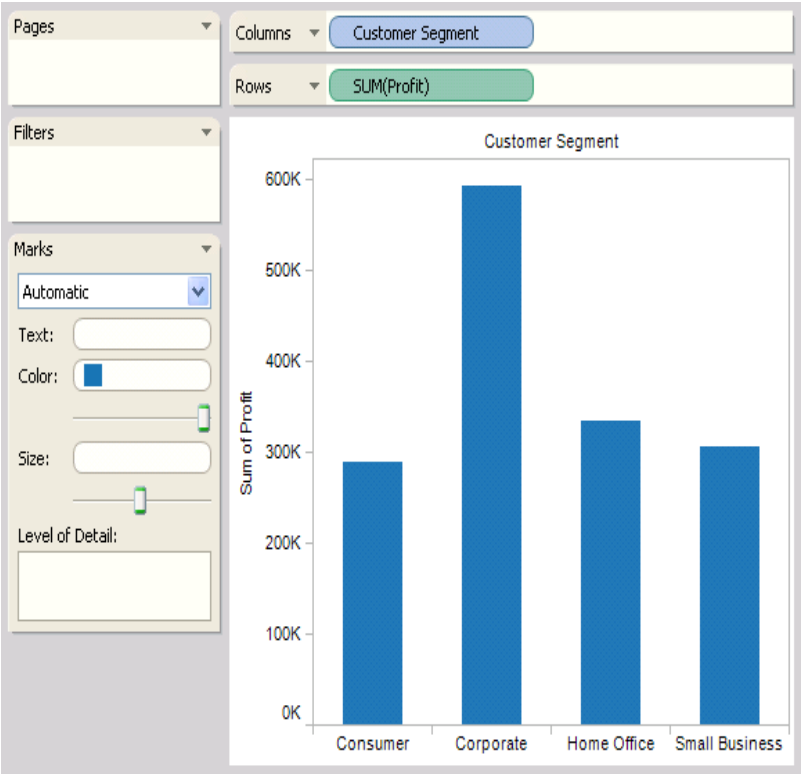
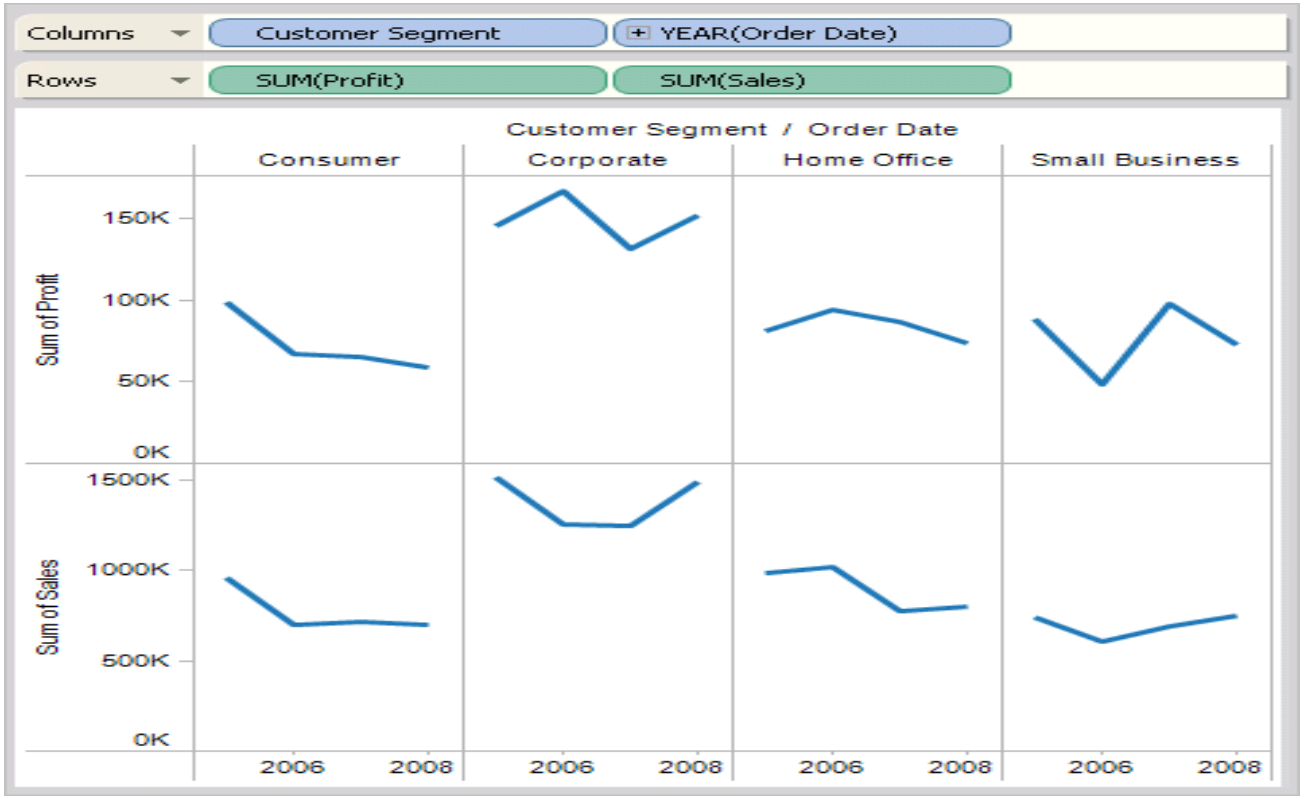


Tableau displays data using marks, where every mark corresponds to a row (or a group of rows) in your data source.

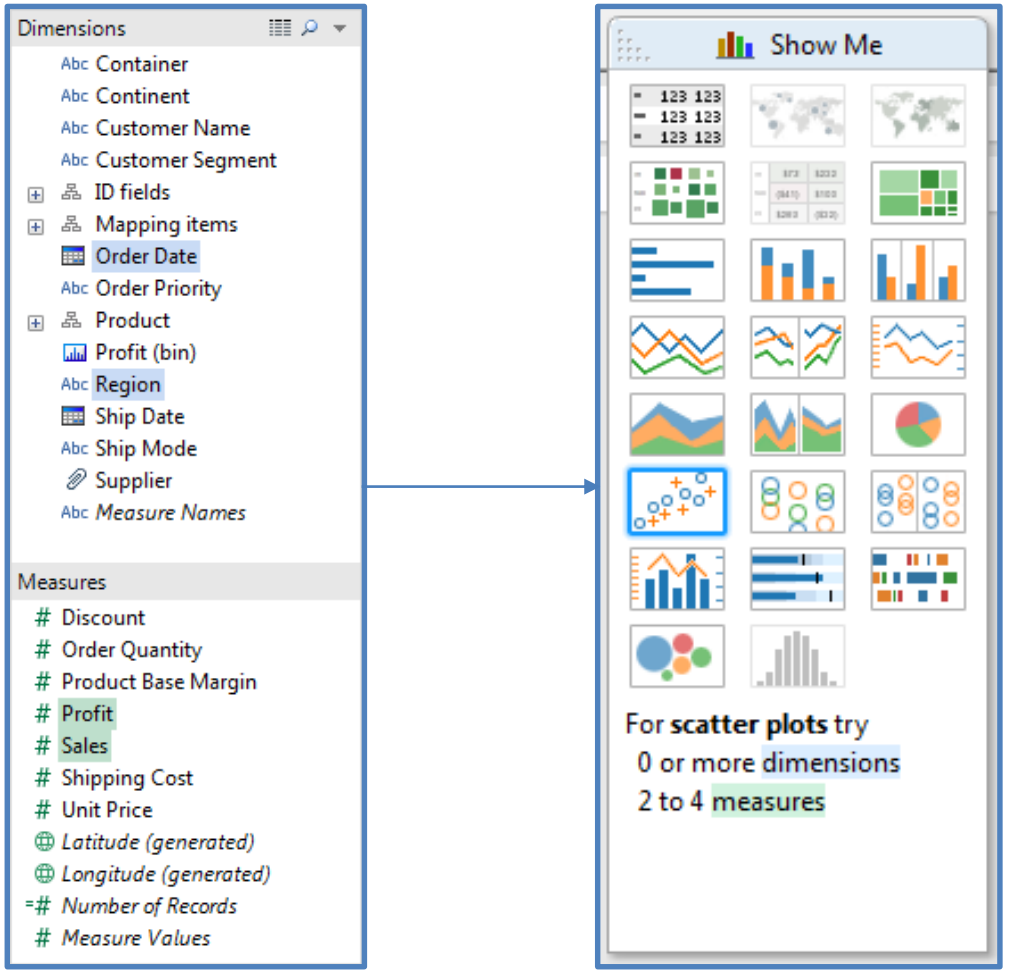
Rows and Columns Shelves (Cont'd…)

- Adding more fields to the Rows and Columns shelves adds more rows, columns, and panes to the table.



Show me: Charts

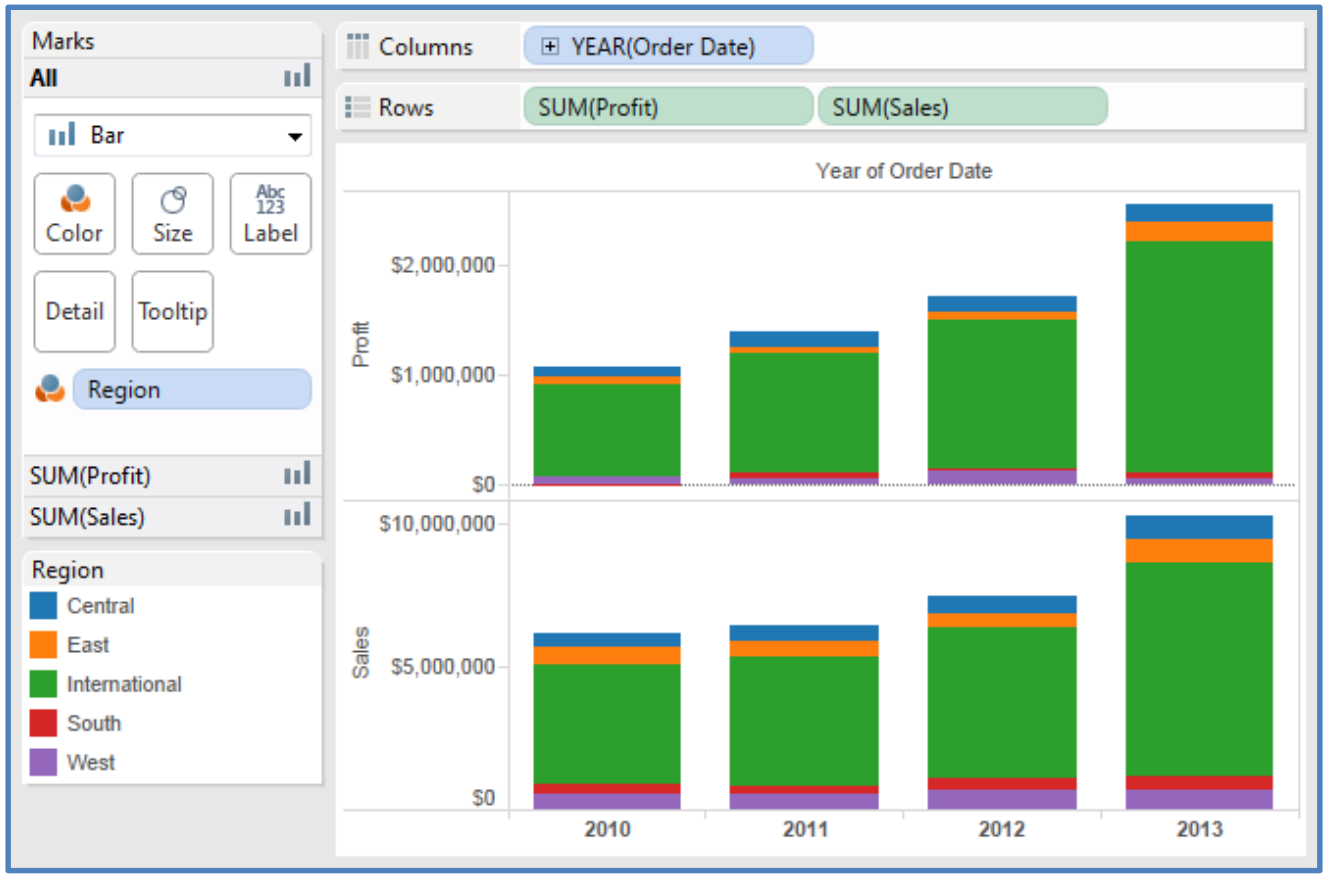
- Show Me creates a view based on the fields already used in the view and any fields you've selected in the Data window. Open Show Me by clicking Show Me on the toolbar.
- When you use Show Me simply select fields you want to analyze in the Data window and then select the type of view you want to create. Tableau automatically evaluates the selected fields and gives you the option of several types of views that would be appropriate for those fields.



Any view type that is not gray will generate a view of your data.

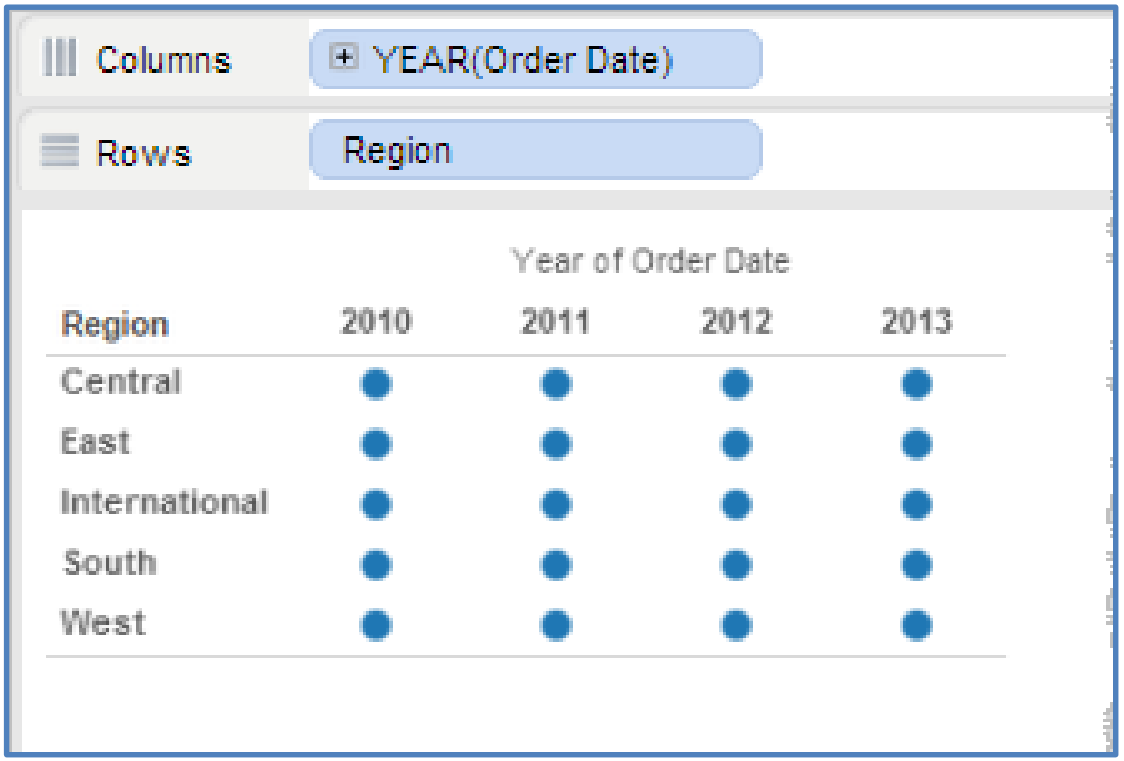
Show Me: Charts (Cont'd…)

- View the Result. Tableau automatically creates a view of the data.



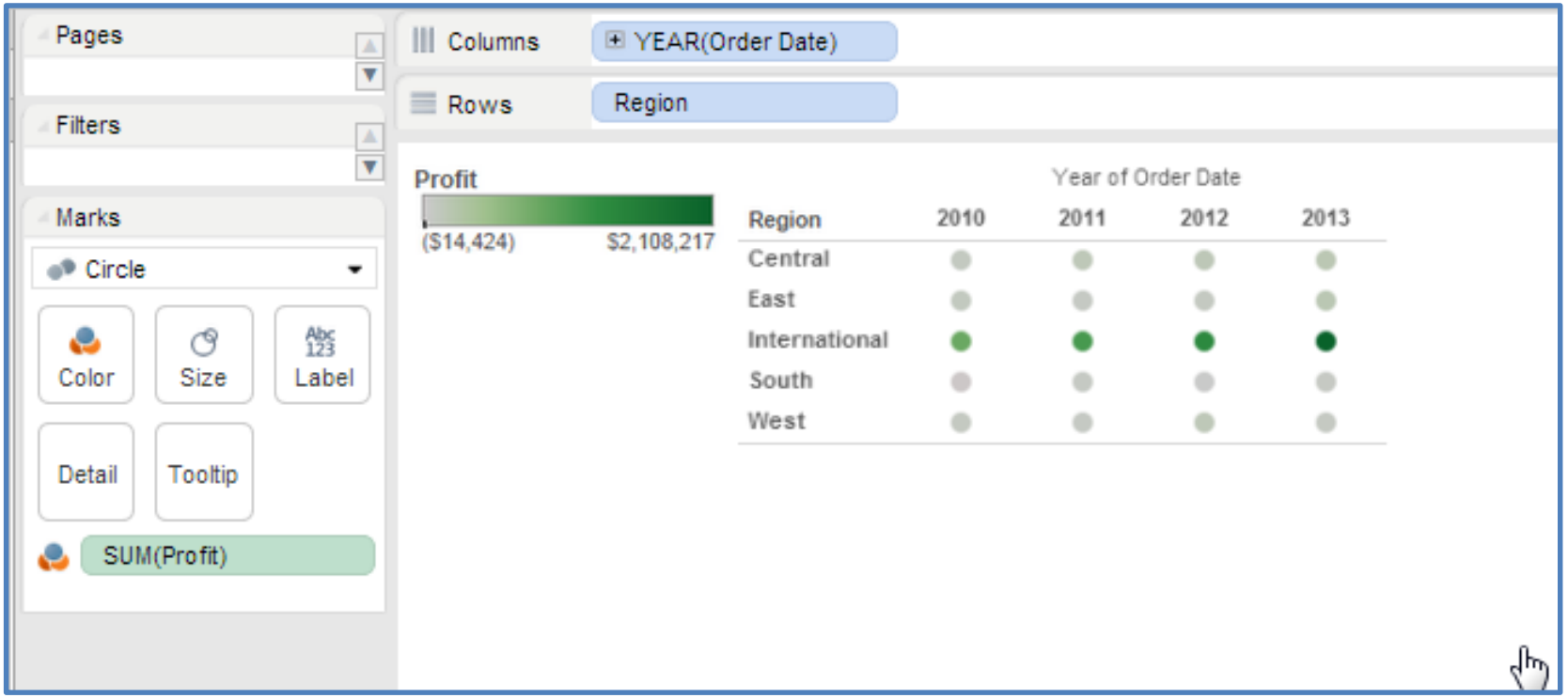
The Marks Card

- When you drag fields to the view, the resulting data points are displayed using marks. Each mark represents the intersection of all of the dimensions in the view. For example, in a view with Region and Year dimensions, there is a mark for every combination of those two field--for example, East 2011, East 2012, West 2011, West 2012, etc.:



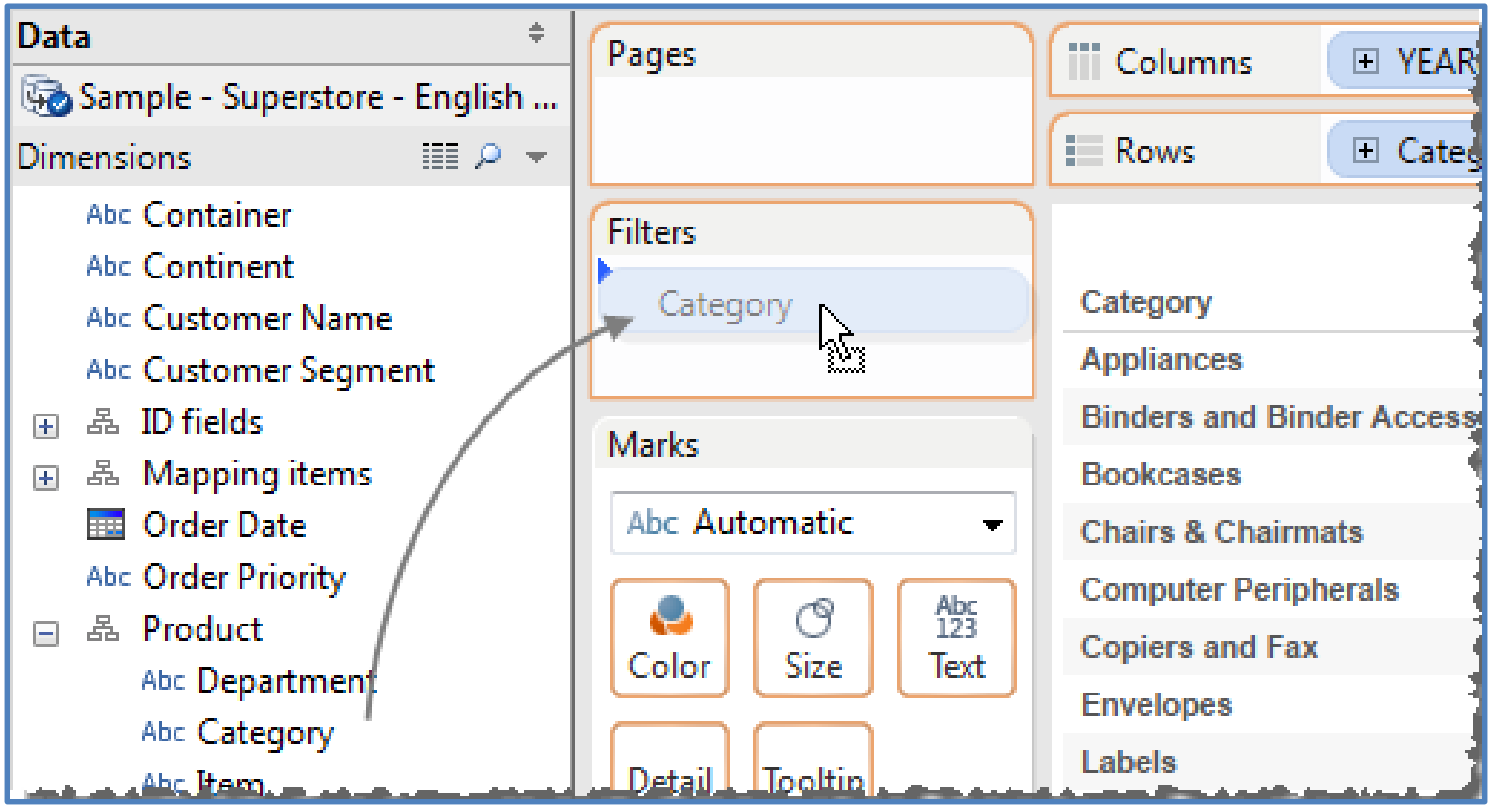
The Marks Card (Cont'd…)

- Marks can be displayed in many different ways including lines, shapes, bars, maps, and so on. You can show additional information about your data using mark properties such as color, size, shape, and labels. The type of mark you use and the mark properties are controlled by the Marks card. Drag fields to the Marks card to show more data. For example, you can enhance the view above by dragging Profit to Color. With this additional information, it becomes apparent that the International region is consistently more profitable than other regions:



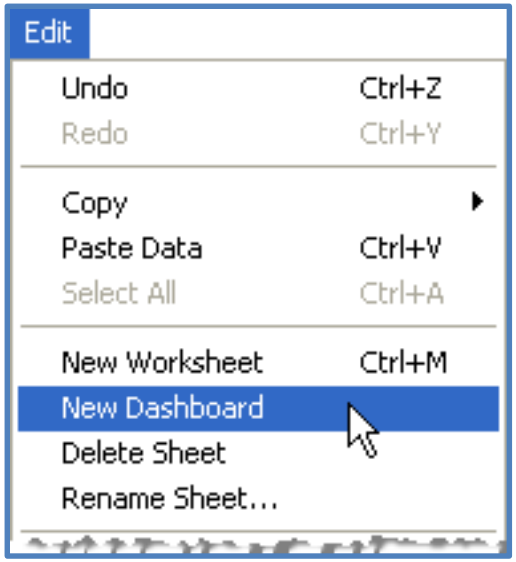
Filters

- **Basic Categorical Filters**
  - Drag a field from the Data window to the Filters shelf. You can also right-click a field on any shelf and select Filter.



Dashboard

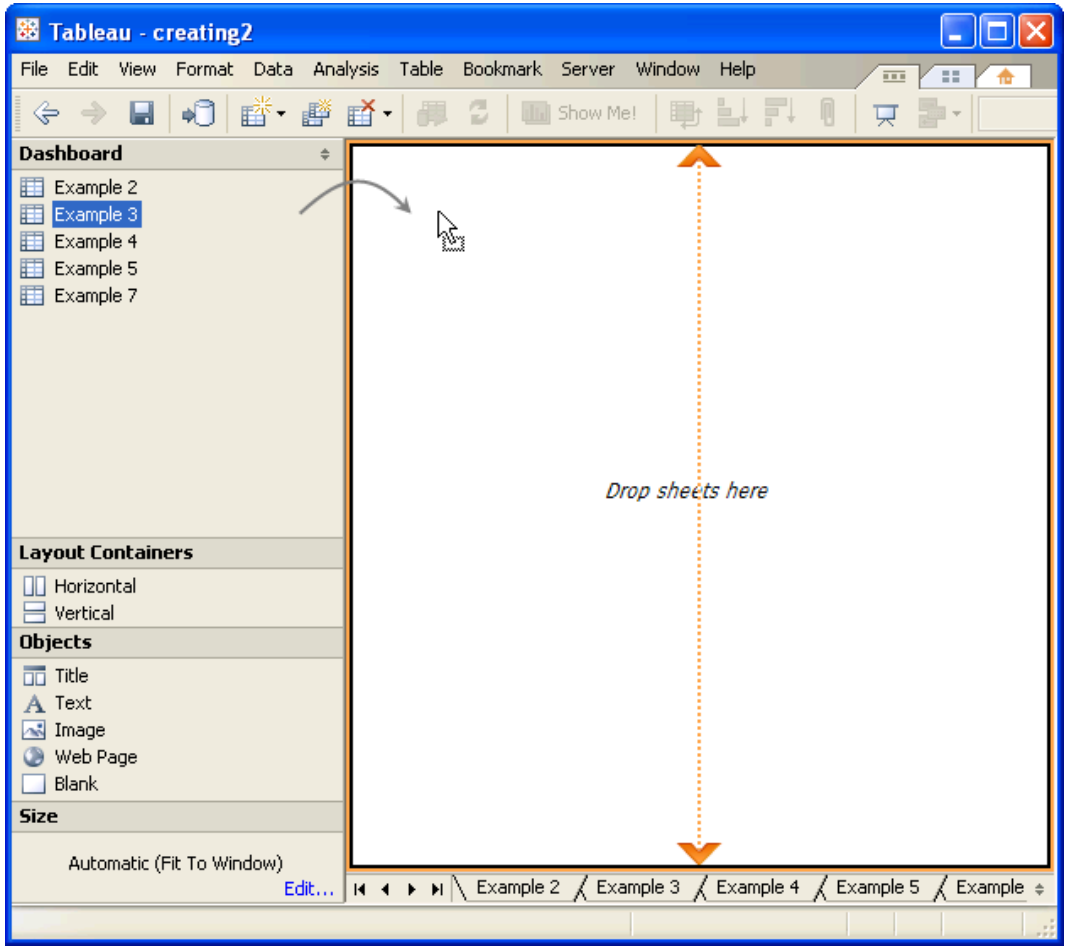
- **Create Dashboard**
  - Select Edit > New Dashboard



- **Adding Views to a Dashboard**
- When you open a dashboard the Dashboard window replaces the Data window on the left side of the workbook. The Dashboard window lists the worksheets that are currently in the workbook. As you create new worksheets, the Dashboard window updates so you always have all worksheets available when adding to a dashboard.

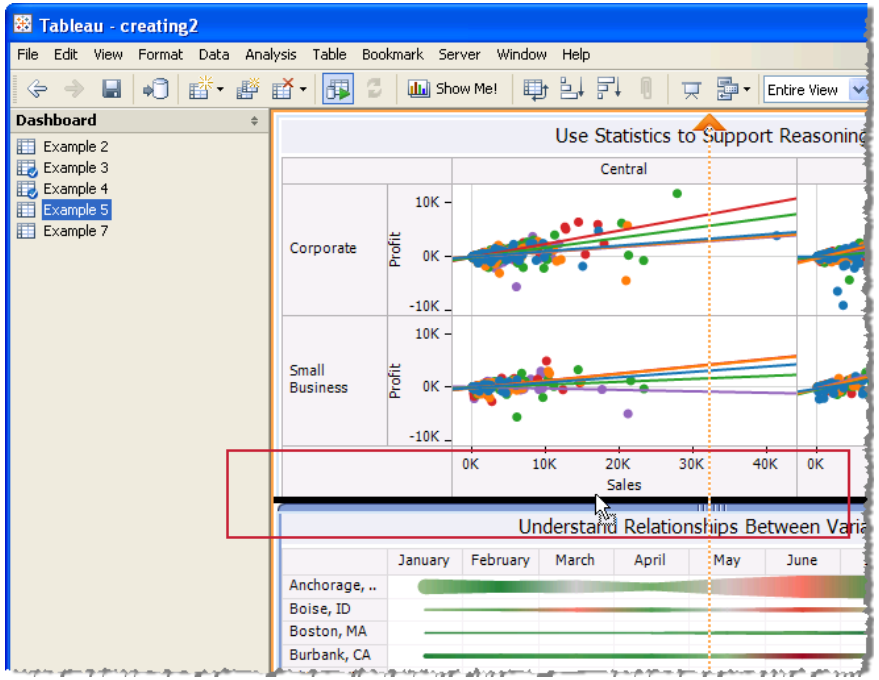
Dashboard (Cont'd…)

- To add a view to a dashboard, click and drag a worksheet from the Dashboard window to the dashboard on the right:

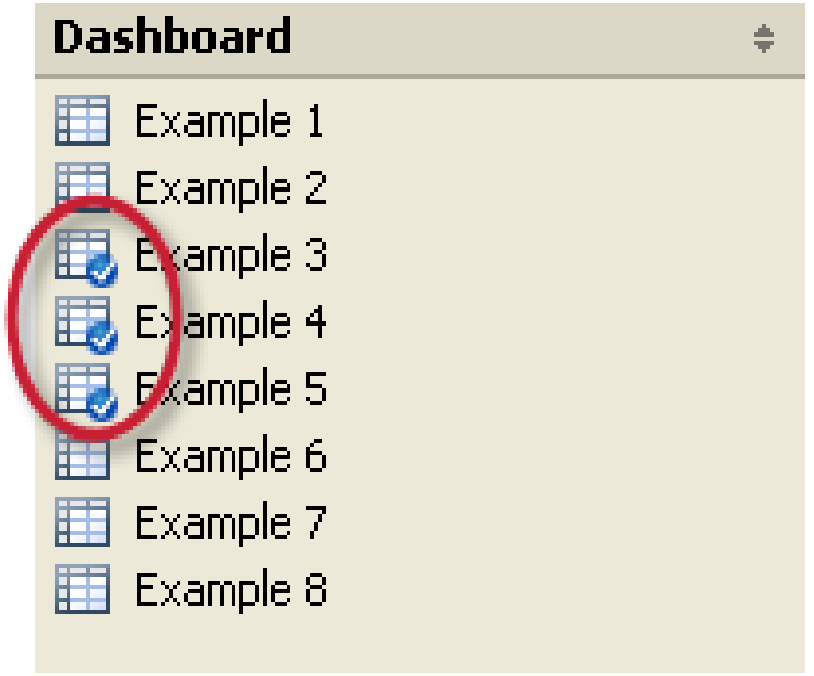


Dashboard (Cont'd…)

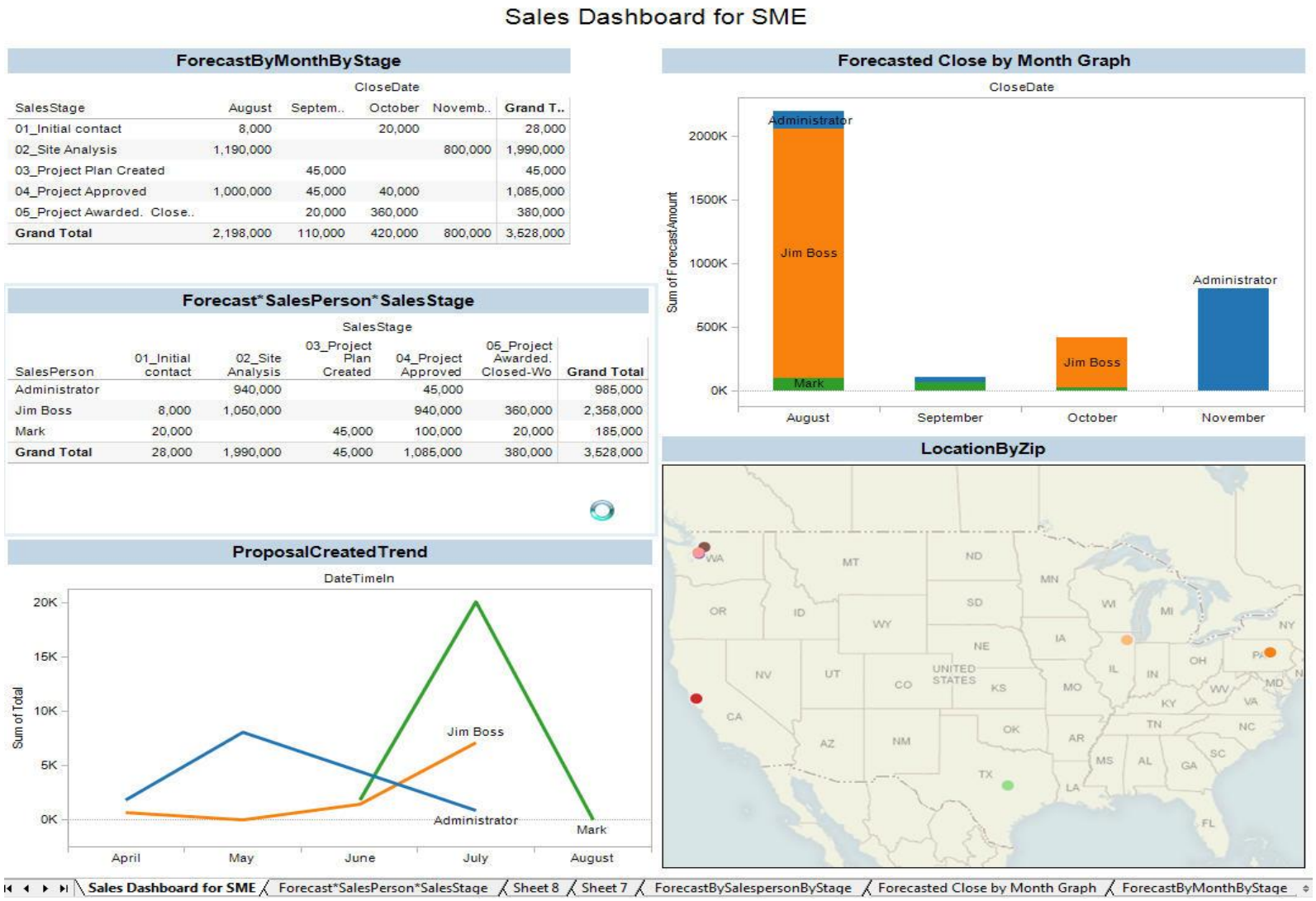
- Continue to drag as many of the worksheets to the dashboard as you like. Notice as you drag worksheets around over the dashboard a black bar displays, indicating the various places you can drop it.



- After a view is added to the dashboard, the worksheet is marked with a check mark in the Dashboard window. Also, any legends or quick filters that are turned on for the sheet are automatically added to the dashboard.



Dashboard: Example



Advanced Topics

- Data blending
- Advanced Mapping Techniques
- Data Calculations
- Filtering Conditions and Measures
- Trend lines, Residuals and Forecasting

References

- www.tableausoftware.com
- http://www.slideshare.net/parthacharya/getting-started-with-tableau? (Parth Archarya)
- http://www.slideshare.net/visualisingdata/the-8-hats-of-data-visualisation (Andy Kirk)
- http://www.slideshare.net/thompsonkaren/data-visualization-theory (university of Idaho)
- http://www.slideshare.net