

Assignment 2 – Introduction to Tableau

Professor John Sokol | Due 2/5

Tableau Introduction:

- Tableau background as a company
- Pros and limitations of Tableau
- Importing data
- When you first open Tableau
- Key terminology
- Tableau data types

Tableau background:

Tableau Software is an American computer software company headquartered in Seattle, WA. It generates interactive data visualization products focused on business intelligence. The company was established at Stanford University's Department of Computer Science between 1997 and 2002. Tableau is traded publicly on the NYSE under the ticker DATA.

The different products that Tableau offers are the following:

Tableau Desktop

- A data visualization application to facilitate you to examine virtually any kind of structured data and generate highly interactive, beautiful graphs, dashboards, and reports within minutes. Once a quick installation, you can tie to virtually any data source from spreadsheets to data warehouses and display information in several graphic perspectives. Designed to be easy to utilize, you'll be working more rapidly than ever before.

Tableau Server

- Enterprise level online server to upload and store company/organization Tableau workbooks.

Tableau Public:

- Individual online server to upload and store your Tableau workbooks in an online medium, on your Tableau Public profile. 10 GB of space and able to personalize your profile.
- <https://public.tableau.com/profile/john.sokol#!/>

Tableau Reader

- The user can only read and view Tableau files; unable to edit the workbook.

Pros:

- Interactivity
 - Tell a story with your data that your users can interact with.
 - Ability to “drill down”, or to filter from summary to detail level data in real time.
- High Performance
 - Fast and efficient querying (accessing data from the data source), especially from “flat files” such as Excel spreadsheets and csv files.
- Growing market
 - Tableau is somewhat fresh in the BI marketplace, but its market share is increasing. It has huge user base, which consists of big companies that utilize Tableau on a daily basis to support their business Intelligence initiatives.
- Rich online community support
 - If you’re stumped, Google it. It’s likely the Tableau community has a solution.
- Easy to share
 - Present your dashboard.
 - Upload the dashboard to your Tableau Public profile.

Limitations:

- Not an operational reporting tool
 - Text boxes are not visualizations but are commonly found in many dashboards. From industry experience in healthcare, I find business administration wants to see the raw data that comprises the visualization (Nonetheless, there is a Tableau solution).
- Expensive
 - \$35/month per user, but only able to connect to Google sheets and Excel. To connect to all data sources, such as SQL databases, \$70/month per user. Multiple users can be exorbitant overhead for a business.
- Occasionally crashes abruptly. Save your workbooks often.

Saving workbooks:

- Tableau regular workbook
 - Contains all visualizations, data connection information, but none of the actual data is stored.
- Tableau *packaged* workbook

- All visualizations, data connection information, and the data

****For this class, save all workbooks, especially assignment submissions as *packaged* workbooks.****

Importing data:

Since we'll be working with flat files in this course, there are two options to import flat files into Tableau.

- Excel file
 - Imports Excel files
- Text file
 - Imports tsv, csv files

I emphasize these options are firm; you need to import Excel files using the Excel option and csv files using the text file option.

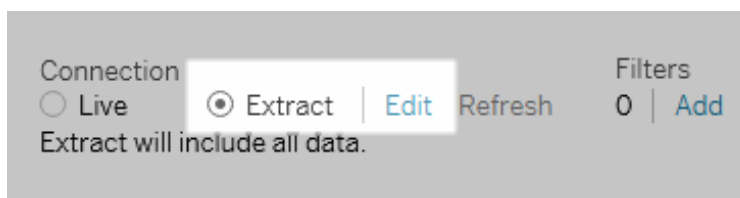
Data connections:

Live:

- A live connection will query the underlying data in the datasource/database. Only refreshing the workbook will return the latest results.
- Live connections are appropriate when connecting to server databases.

Extract:

- Tableau pulls the data from the original source and stores it in a TDE file.
- Since we're working with flat files, switch the data connection to extract at the top right of the data sources tab for the best workbook performance.


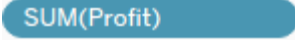


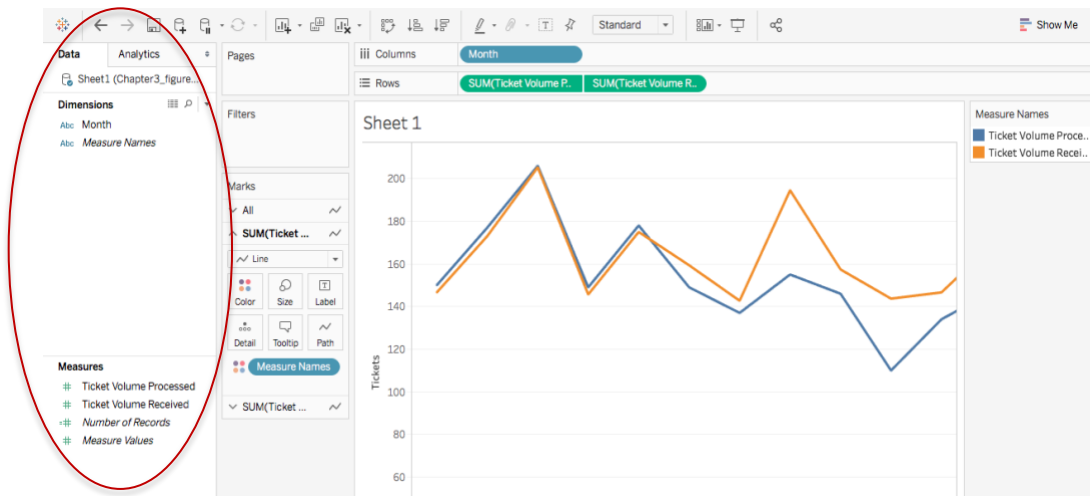
- Tableau is now querying and getting results from the extract. Data can be retrieved from the source again to refresh the extract. Thus, each extract is a snapshot of the data source at the time of the latest refresh. Extracts offer the benefit of being portable and extremely efficient.

When you first open Tableau to create a single visualization in a worksheet:

1. Import your dataset (either Excel or csv).
2. Check the Data Source tab; ensure all data fields are assigned the correct data type.
3. Choose appropriate data connection.
4. Drag the necessary data fields to the columns or rows shelf to begin creating your visualization.
5. Modify size, color, details, tooltips, and visualization formatting.

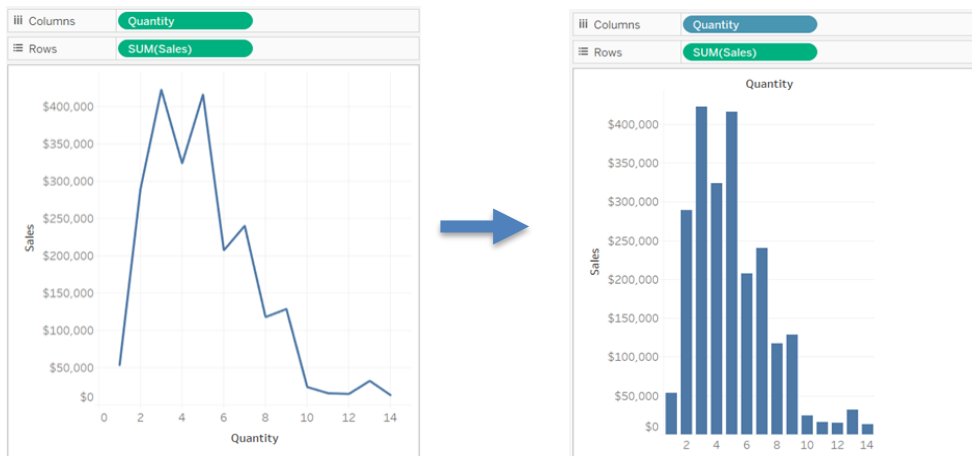
Key Terminology:

- Data field
 - A column where data is stored
- Measures
 - Quantitative (numbers) values that math functions work with
 - Measures are continuous, meaning they form an unbroken whole, without interruption. Continuous values form a continuum on a certain interval.
 - Measures CAN be aggregated
 - Aggregate function is a function where the values of multiple rows in a single data field are grouped together to yield a single value.
 - The most frequently used aggregate functions are common math functions such as sum, median, average, maximum, minimum, etc. For example, the average of 10, 20, and 30 is 20. So 20 is an aggregate of the 3 listed values.
 - Measures are green pills  they are continuously infinite. Generally, continuous fields add axes to the worksheet view.
- Dimensions
 - Qualitative (not numbers) values
 - Region, employee name, locations, dates, etc
 - Dimensions can NOT be aggregated
 - Dimensions are blue pills  they are discrete, finite values. Generally, discrete fields add headers to the view.



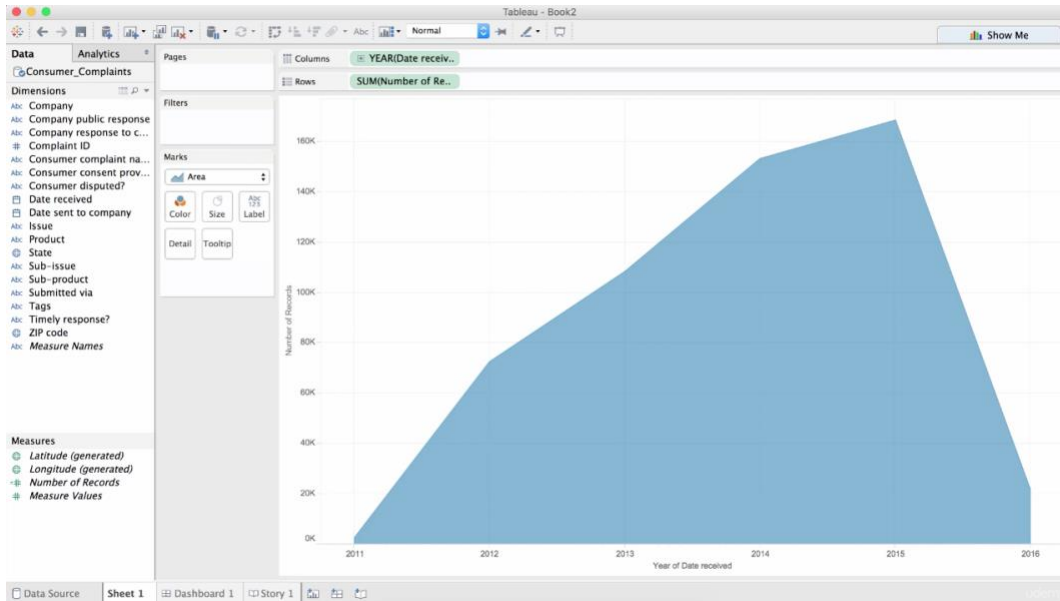
Above: Dimensions and Measure are listed to the left

- Notice Tableau automatically changes a line graph to a bar graph when the “quantity” value changes from a continuous measure (green) to a discrete dimension (blue) value because the quantity data field as a dimension treats the integers of 1 – 14 as categories, not real numbers.

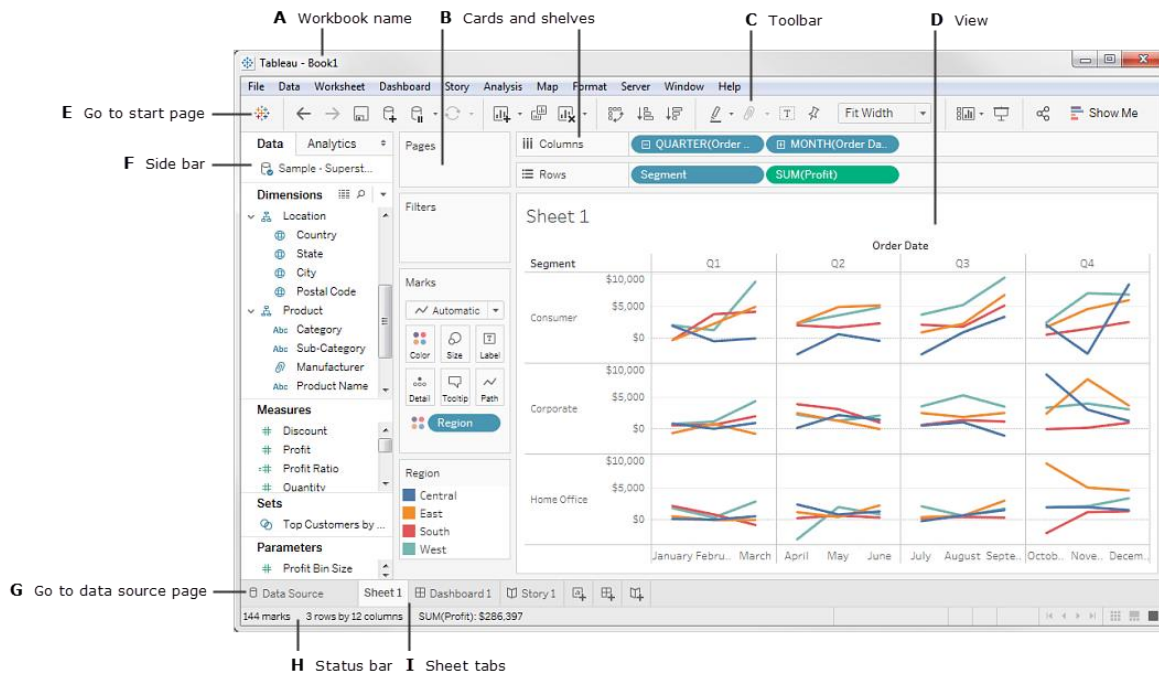


Worksheet

- A space to create a single visualization, or “Viz” in Tableau parlance (see below).



Worksheet mapping:



Dashboard

- A space to combine multiple worksheets in one view (see below).

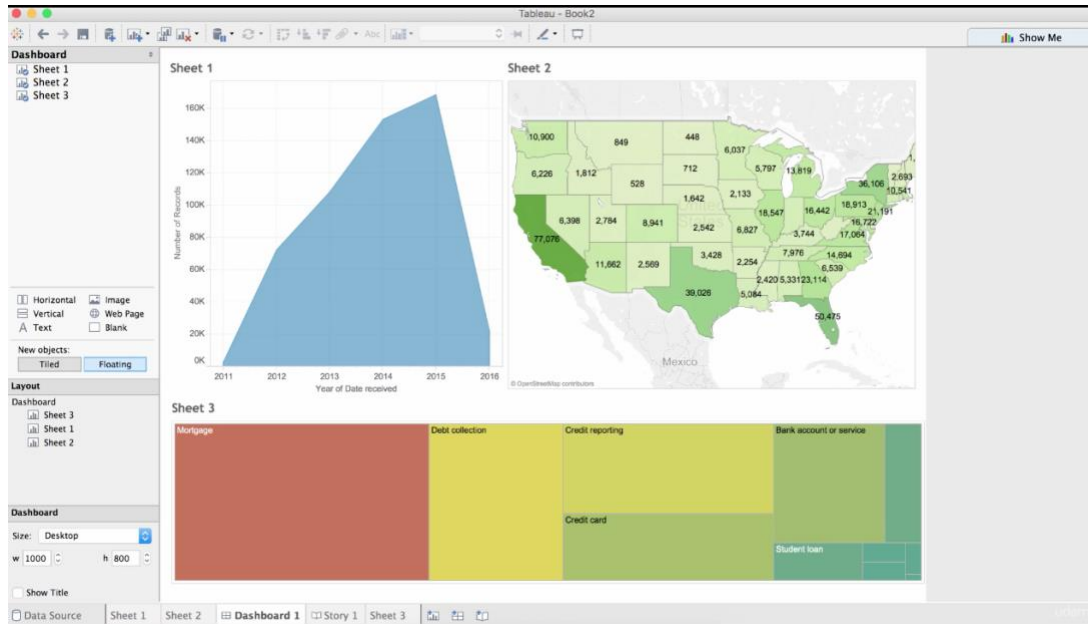


Tableau Data Types:

Data Type	Description	Example	Icon
String	Any sequence of zero or more characters. They are enclosed within single quotes.	'Hello' 'Quoted' 'quote'	Abc
Number (integer or float)	Either integers or floating points. Floats are decimal, non-integer values.	Int – 3 Float – 142.58	#
Date	Date value	9/10/2018	📅
Date/Time	Date and time value	9/10/2018 10:01:09 AM	🕒
Geographic Values	Longitude and latitude	39.6494670, 74.1772330	🌐
Boolean	Logical values	True/False	=T F

Corresponding data type identification:

String ↓	Date ↓	Integer ↓	Integer ↓	Boolean ↓
----------	--------	-----------	-----------	-----------

Abc Sheet1 Location	📅 Sheet1 Reporting Date	# Sheet1 Tickets Processed	# Sheet1 Tickets Received	=T F Calculation TRUE/FALSE
Manahawkin	1/31/2010	541	763	True
Egg Harbor Township	1/31/2010	314	498	False
Linwood	1/31/2010	393	690	False
Hammonton	1/31/2010	287	650	False
Ventor	1/31/2010	479	810	False
Cape May Court House	1/31/2010	232	545	False
Atlantic City	1/31/2010	834	1,453	False
Marlton	1/31/2010	134	231	False

Geographic (notice the globe icon):

🌐 Sheet1 Mail City	🌐 Sheet1 Mail State	🌐 Sheet1 Mail Zip	🌐 Sheet1 Latitude	🌐 Sheet1 Longitude
Beach Haven	NJ	08008	39.6494670	-74.1772330
Mullica Hill	NJ	08062	39.6515790	-74.1729540
Lanoka Harbor	NJ	08734	39.6428970	-74.1804160
Lanoka Harbor	NJ	08734	39.6428970	-74.1804160
Ship Bottom	NJ	08008	39.6428970	-74.1804160
Princeton	NJ	08540	39.6474880	-74.1861780
Philadelphia	PA	19102	39.6428970	-74.1804160
Essex Fells	NJ	07021	39.6463660	-74.1847570
Robinsville	NJ	08691	39.6436110	-74.1791620
Spray Beach	NJ	08008	39.6428970	-74.1804160
Peekskill	NY	10566	39.6447030	-74.1851490
Ship Bottom	NJ	08008	39.6428970	-74.1804160
Upper Saddle River	NJ	07458	39.6414270	-74.1837500
Huntingdon Valley	PA	19006	39.6391340	-74.1873320

Instructions:

- Import the IT tickets dataset into Tableau, then replicate my worksheet I created in class and in my online video tutorial.
- In the same workbook, import the 'Life Expectancy Final' dataset. Create a line graph depicting the life expectancy of one country. Reference the 'Changing Data Sources' video to learn about changing data sources in one workbook.

Deliverables:

- Submit a single Tableau *packaged* workbook of:
 - Line graph worksheet of the IT tickets dataset created in class
 - Line graph worksheet of Sokol heart rate 12-25-19 to 1/25/19
 - The line graph should only include data from the last week of the month; 1/18/19 to 1/25/19
 - Bar graph worksheet of Denver arrest rates; same in class Excel graph but made in Tableau
 - Crime and traffic bars
 - Use calculated fields with functions SUM() and COUNT()
- Half to one-page write-up of your thoughts about Tableau. Some thoughts to consider:
 - Challenges using the software?
 - Questions for me about Tableau, including anything you want to learn in particular and if Tableau is capable of a certain task.

I hope this assignment helps you become more comfortable with the Tableau interface.