U.S. General Services Administration

D2D Beginner Tableau Training

January 10th and January 11th

presented by Walter Mehra
D2D Team Member

Pre-Class Activities

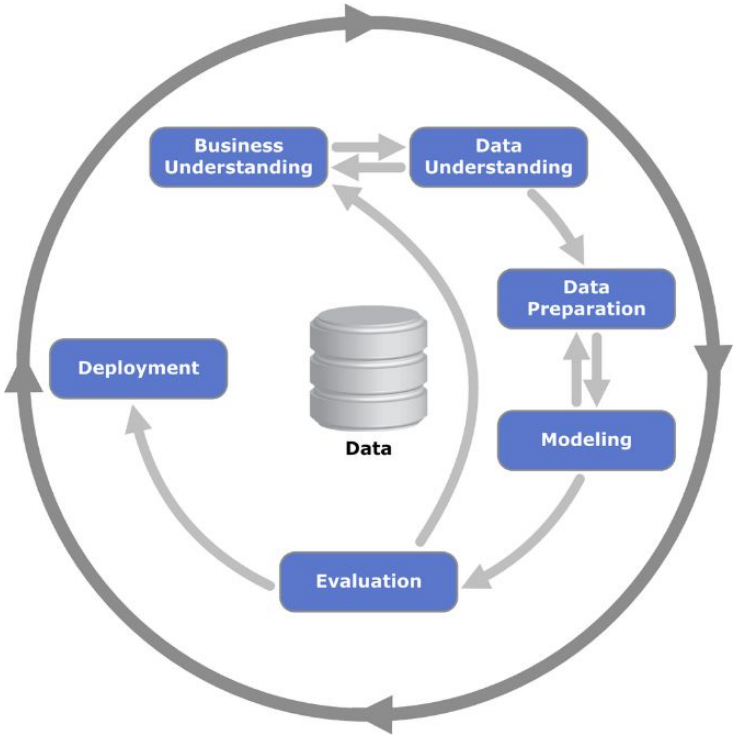
- Please make sure you do the following prior to the training session:

  - Ensure that Tableau Desktop (version 9.3) or higher (D2D is on 10.3.4) is installed on your computer and you have a valid license (or are on Tableau granted two-week trial license);
  - Review the Basic Data Concepts document
  - Download the dataset used for this training to your computer's desktop

Objectives

- Tableau UI (User Interface)
- Loading and Joining Data
- Creating Analytics and Dashboards

Data Analytics and Visualization Steps (CRISP-DM)



Retrieved from: https://en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining#/media/File:CRISP-DM_Process_Diagram.png

- **Business Understanding** of the project objectives and requirements from a business perspective
- **Data Understanding** phase starts with initial data collection and proceeds with activities that enable you to become familiar with the data, identify data quality problems, discover first insights into the data, etc.
- **Data Preparation** phase covers all activities needed to construct the final dataset including tasks such as table, record, and attribute selection, as well as transformation and cleaning of data
- **Modeling** phase covers the build out of analytics, whether descriptive, predictive, or prescriptive
- **Evaluation** phase requires thorough evaluation and review, to be certain the analysis properly achieves the business objectives. A key objective is to determine if there is some important business issue that has not been sufficiently considered. At the end of this phase, a decision on the use of the data mining results should be reached.
- **Deployment** can vary based on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process across the enterprise

Type of Analytics

According to Gartner Research:

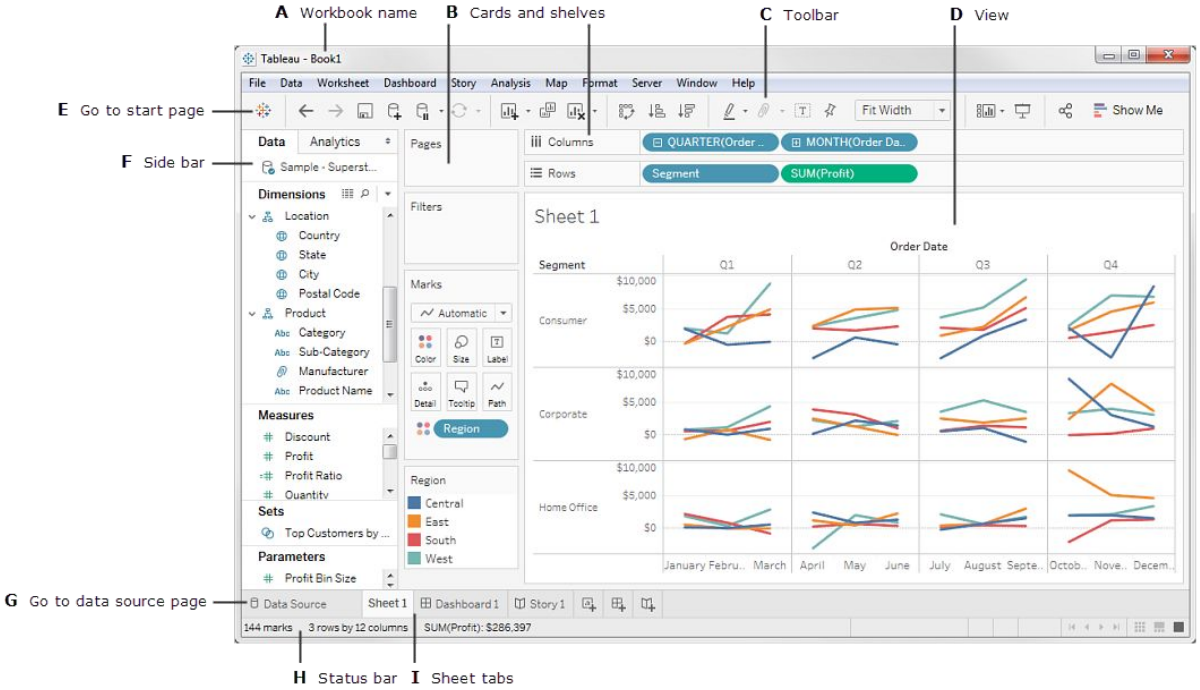
- Descriptive Analytics - is the examination of data or content, usually manually performed, to answer the question "**What happened**?" (or What is happening?), characterized by traditional business intelligence (BI) and visualizations such as pie charts, bar charts, line graphs, tables, or generated narratives.

- Predictive Analytics - advanced analytics which examines data or content to answer the question "**What is going to happen**?" or more precisely, "What is likely to happen?", and is characterized by techniques such as regression analysis, forecasting, multivariate statistics, pattern matching, predictive modeling, and forecasting.

- Prescriptive Analytics - advanced analytics which examines data or content to answer the question "**What should be done?" or "What can we do to make _____ happen**?", and is characterized by techniques such as graph analysis, simulation, complex event processing, neural networks, recommendation engines, heuristics, and machine learning.

http://www.gartner.com/it-glossary/

The Tableau Workspace



A. Workbooks and Sheets name. For more details on workbooks and sheets.
B. Shelves and Cards - Drag fields to the cards and shelves in the workspace to add data to your view.
C. Toolbar - Use the toolbar to access commands and analysis and navigation tools.
D. Parts of the View - This is the workspace where you create your data visualizations.
E. Goes to the start page. For more information, see Start Page.
F. The Side Bar - The side bar provides two panes: the Data pane and the Analytics pane. For more information, see The Side Bar.
G. Go to the data source page. For more information, see Data Source Page.
H. Status Bar - Displays information about the current view.
I. Sheet tabs - Tabs represent each sheet in your workbook. This can include worksheets, dashboards and stories. For more information, see Sheets.

For Reference: The Tableau Workspace

Tableau Terminology

- Tableau Files - there are six type of files that Tableau creates:
  - .twb - Tableau Workbook (does not include/package the data with the workbook)
  - .twbx - Packaged Tableau Workbook which contains data and can contain external files such as images, etc.; think of it as a zip file full of everything you need to visualize the data
  - There are four other Tableau file, you can read about them on Tableau's website: http://onlinehelp.tableau.com/current/pro/online/mac/en-us/environ_filesandfolders.html

- Sheets are part of Tableau Workbooks, there are three types: Worksheet, Dashboard, and Story

Data Terminology

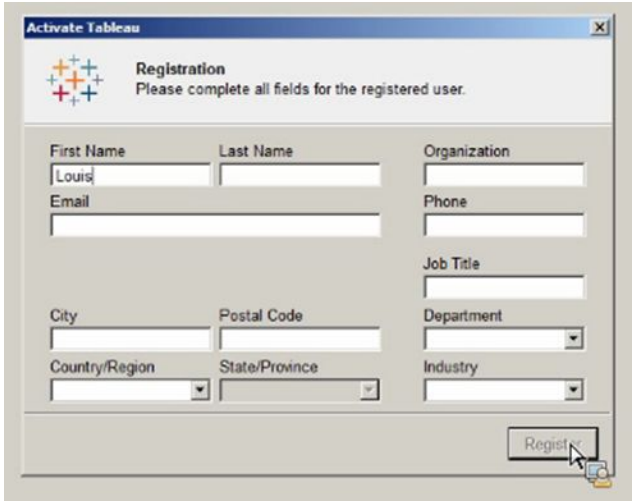
Data Type	Descriptions/Details	Examples
Number (Decimal) or Number (Whole)	These are either integers or floating points. If a variable can take on any value between two specified values, it is called a continuous variable; otherwise, it is called a discrete variable	3 or 3.14159265359 (Continuous)
Date or Date and Time	Tableau recognizes dates in almost all formats, these values are typically used for time series or trend analysis	11/28/2017 or 11/28/2017 1:00 PM
Boolean	They are logical values	True or False
String	Any sequence of characters. They are enclosed within single quotes. The quote itself can be included in a string by writing it twice.	GSA, Budget Activity, Lease, NAICS
Geographic Role	*Not a datatype, but does clarify the desired outcome for the data field* Identifier for Tableau to facilitate map building included roles are Airport, Area Code, CBSA/MSA, City, Congressional District, Country / Region, County, Latitude, Longitude, NUTS Europe, State/Province, Zip Code/Postcode	KIAD (Airport), 202 (Area Code), VA (State/Province), etc.

Tableau will categorize your data fields into one of two buckets:

- *Dimension* is something you categorize with (e.g., color of a shirt)
- *Measure* is something you do math with (e.g., the number of white shirts)

Tableau Desktop Version & Registration

- Check that Tableau Desktop is installed on your workstation. Version should ideally be the current GSA version, 10.3.4.
- Ensure Tableau Desktop is registered in one of two ways.
  - New users: open the software and register for the free 2-week trial.
  - Current users: no action needed; you should already have a registered version of the software.



Using Tableau (Data Sources)

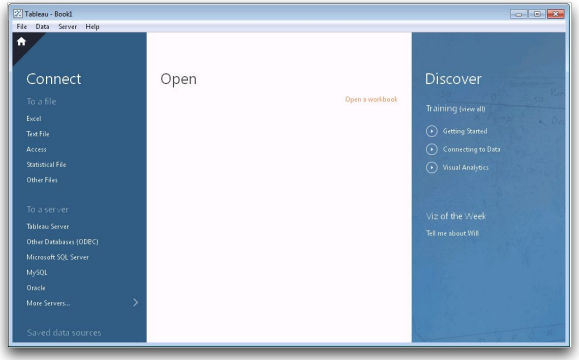
- Using the **Discover** pane on the Tableau Desktop home page is a great resource to find assistance for the various Tableau capabilities and sample dashboards

- The **Connect** and **Open** pane are the primary panes to work from, one will provide you the ability to connect to new data sources while the other will enable you to open workbooks you've recently worked on

- Dataset(s): REXUS Building Data
- Question: What are each of the three panes on the Tableau Desktop home page used for?

Steps:

1. Open Tableau
2. Browse the different sections of the Start Page:
    1. Connect
    2. Open
    3. Discover
3. Connect to our data source, Excel



Using Tableau (Data Sources cont'd)

- Data Source Page provides tools to change data and metadata:
  - Joins
  - Changing field names
  - Changing data types
  - Filters

- Extract vs Live Connection:
  - Extract will generally be faster but not provide a real-time connection Live will provide a real-time connection to your data but may be limited by performance or features

- Using your Tableau Repository or Tableau Server, multiple workbooks can connect to the same Tableau data source

❑ Dataset(s): REXUS Building Data

❑ Question: What are each of the three panes on the Tableau Desktop home page used for?

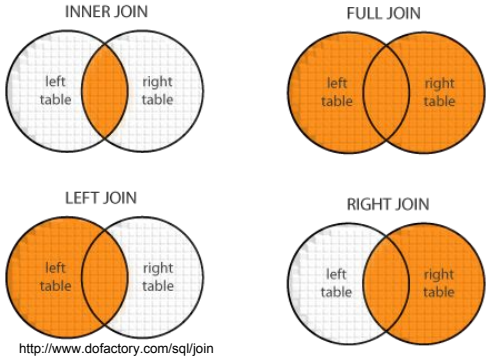
Steps:

1. Drag sheet datagovbldgrexus to the window
2. Select connection type of **Extract**
3. Additional updates to the data include:
   i. Adding filters
   ii. Adding additional data to join
4. Review other options, such as number of records to display, changing data types, etc.
5. Click on **Sheet 1** to begin analysis

Data Joining vs Data Blending

Data Join

- Only join data within the same database (fixed in Tableau 10.0 as a cross-database join)
- More flexibility in joining data, you can easily identify a left, right, inner, and full joins
- Join is done once at the data source and used for every everything while a blend is specific



http://www.dofactory.com/sql/join

Data Blend

- Join data across databases (for versions below 10.0)
- When you use data blending to combine data, a query is sent to the database for each data source that is used on the sheet.
- Always mimics a left join, emulation of other types of joins is possible but a bit more work:
- Primary/Driver data source (left-side of join) is defined by the first variable in a dataset dragged into the worksheet:

Table A (Primary/Driver)

ID	Name	Height
1	Arya	5' 0"
3	Cersei	5' 5"
5	Daenerys	5' 2"
7	Jamie	6' 2"
9	Jon	5' 10"
11	Tyrion	4' 5"

Table B (Secondary)

ID	Name	Weight (lbs)
2	Joffrey	123.5
4	Samwell	231
5	Daenerys	115
7	Jamie	194
9	Jon	168
12	Sansa	141

Left Joined Table

ID	Name	Height	Weight
1	Arya	5' 0"	N/A
3	Cersei	5' 5"	N/A
5	Daenerys	5' 2"	115
7	Jamie	6' 2"	194
9	Jon	5' 10"	168
11	Tyrion	4' 5"	N/A

SELECT a.ID, a. Name, a.Height, b.Weight
FROM Table_A a
LEFT JOIN Table_B b ON Table_A.ID = Table_B.ID;

Data Blending

Joins vs. Blends

- Joins are more efficient when data is in the same database and when there is high-cardinality (lots of unique numbers)] in the joining fields
- Blends are useful when there is a one-to-many relationship in the data and you want to summarize it more efficiently (inclusion of * for the one-to-many relationship to show multiple values), also known as granularity of the data
- Joins occur once while blends call the database each time you make a change to your worksheet

Primary vs. Secondary Datasources

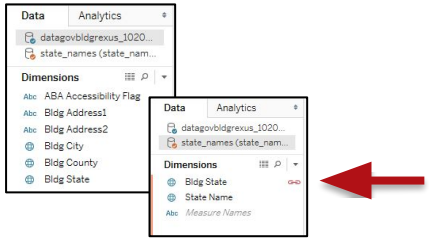
- Primary datasource (left-side of join) is defined by the first variable in a dataset dragged into the worksheet
- Tableau auto selects the relationship between datasources when they have similar names
- ***Data → Edit Relationships to define or edit relationships***

❑ Dataset(s): REXUS Building Data, State Names

❑ Question: How many records are there per state?

Steps:

- Open the REXUS data and drag sheet datagovbldgrexus to the window
- Rename Sheet 1 to **Blend**
- Add a New Data Source; import the State Names dataset and switch back to the **Blend** worksheet
- Drag **Bldg State** from the REXUS dataset to the rows shelf and **Number of Records** to columns shelf
- Switch to the State Names dataset and drag **State Names** from the State Names dataset to the filter shelf, filter the results to only include California, Texas, Florida, and New York



Cross-Database Joins

Joins vs. Blends

- Joins are more efficient when data is in the same database and when there is high-cardinality (lots of unique numbers)] in the joining fields
- Blends are useful when there is a one-to-many relationship in the data and you want to summarize it more efficiently (inclusion of * for the one-to-many relationship to show multiple values), also known as granularity of the data
- Joins occur once while blends call the database each time you make a change to your worksheet

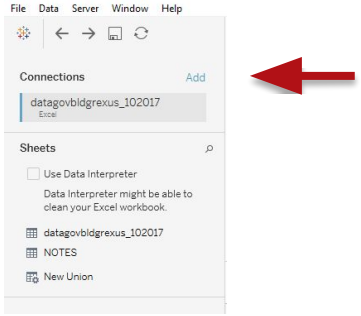
Primary vs. Secondary Datasources

- Primary datasource (left-side of join) is defined by the first variable in a dataset dragged into the worksheet
- Tableau auto selects the relationship between datasources when they have similar names
- **Data → Edit Relationships to define or edit relationships**

- ❑ Dataset(s): REXUS Building Data, State Names
- ❑ Question: How many records are there per state?

Steps:

1. Open up Tableau Desktop 10 (if you don't have it, just follow along)
2. Create a new connection for our first dataset (REXUS)
3. Click on Add, then let's connect to the second dataset (State Names)
4. **Sheet1** from the State Names should have automatically populated and joined to our first dataset, if not, drag it into the canvas
5. Review the data for a left, right, inner, and outer join



Shaping Your Data for Tableau

Crosstabs Not Ideal for Tableau

Agency	***2016***	***2017***	***2018***
GSA	103	50	125
Dept. of Interior	88	100	75

Shape Your Data in Columns

Agency	***Year***	***Value***
GSA	2016	103
GSA	2017	50
GSA	2018	125
Dept. of Interior	2016	88
Dept. of Interior	2017	100
Dept. of Interior	2018	75

- Normalize your data into the First Normal Form (1NF)
  - Remove redundant columns (e.g., Year)
- Remove titles from your data tables if you're using Excel or Google Sheets
- Every column is a data field/variable, so if you want to create a filter on Year, it should all be in one column
- Don't include Totals from your source data, let Tableau generate your totals

End of Part 1

- Any questions?
- Anything you'd like covered for next time?

Analysis – Data Exploration (Exercise 1)

- Dimensions contain qualitative values (such as names, dates, or geographical data).
- Measures contain numeric, quantitative values that you can measure. You can apply calculations to them and aggregate them.
- When a measure or dimension is green, it is continuous. Its values are treated as an infinite range. Generally, continuous fields add axes to the view.
- When a measure or dimension is blue, it is discrete. Its values are treated as finite. Generally, discrete fields add headers to the view.
- Worth noting how each field type affects the visualization in a view and the graph options (i.e., Show Me)

- ❑ Dataset(s): REXUS Building Data
- ❑ Question: What are the differences between Tableau Field Types (Continuous and Discrete Dimensions and Measures)?

Steps:

1. Create a new Tableau workbook and import in the REXUS dataset
2. Rename the worksheet **Field Types**
3. Right click on the **Region Code** dimension and select **Describe…**
4. In the Describe Field box, click on **Load** to load the set of valid values (only first 20, if more than 20) for the field, review results and exit
5. Drag **Region Code** dimension to the rows shelf
6. Drag the **Bldg ANSI Usable** to the columns shelf
7. In the columns shelf, right click on **Region Code** and change the field type to **Continuous**
8. In the columns shelf, right click on **Region Code** and change the field type to **Measure** and calculation to **Count (Distinct)**
   - Change the aggregation to **Count, Sum, Average**
9. Change the field type to a **Discrete** from **Continuous**

Analysis - Summarize (Exercise 2)

- There are two data roles in Tableau:
  - Dimension are for categorizing
  - Measure are for calculations

- Know the Tableau-generated fields, including:
  - Measure Names
  - Measure Values
  - Number of Records

- Use “Show Me” to change the graph type

- Three ways to create a view:
  - Click & drag a field
  - Double-click a field
  - Highlight field(s) then click “show me” to see Tableau’s “best guess” graphs

❑ Dataset(s): REXUS Building Data

❑ Question: What are the number of properties per State?

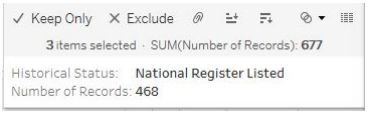
Steps:

1. Create a new worksheet, title it **Summary**
2. Drag **Bldg State** field to the rows shelf
3. Drag the **Number of Records** field to the columns shelf
4. Click on **Show Me** to change your visualization from a bar graph to a crosstab

Analysis - Grouping (Exercise 3a)

- Creating groups is one way to combine multiple responses into one and enable aggregation that isn't readily available in the underlying data

- Marks tooltip:



- ❑ Dataset(s): REXUS Building Data
- ❑ Question: What percent of GSA's owned building inventory is historic?

Steps:

1. Create a new worksheet, title it **Historical Status**
2. Drag **Historical Status** to the columns shelf
3. Drag **Number of Records** to the row shelf
4. Create a table calculation on Number of Records by right-clicking on the **Number of Records**, selecting Quick Table Calculation then select Percent of Total
5. In the graph, multi-select (drag across the categories or use the control key on your keyboard and select each one at a time):
   i. National Historic Landmark
   ii. National Register Listed
   iii. National Register Eligible
6. Hover over the selected group until the tool tip appears with multiple options, select the paper clip icon (Group)
7. From the marks card, drag the newly created grouped field (Historical Status (group)) and replace the **Historical Status** pill in the rows shelf (hold the control key on your keyboard to copy the grouped field as-is)
8. Drag the **Number of Records** field from the columns shelf to the label control on the marks card

Analysis - Grouping (Exercise 3b)

- Creating groups is one way to combine multiple responses into one and enable aggregation that isn't readily available in the underlying data

- Dataset(s): REXUS Building Data
- Question: How many buildings are there per Region?

Steps:

1. Create a new worksheet, name it **Regions**
2. Drag **Bldg State** to the view, a geographic map should auto-populate, click on **Show Me** and change it to a filled map
3. Drag **Region Code** into the labels marks control
4. Select each region and group them (by Bldg State) using the marks tool tip (see guidance in Exercise 3a example)
5. Drag **Number of Records** to the color marks control card
6. Remove **Bldg State** from the details marks control card

Analysis – 508 Compliance (Exercise 4)

- The filter shelf controls all view-based filter functionality
- The only other filters in Tableau are those within the data import process.
- Different types of filters are available for different data types
- The ColorBrewer site is another resource for identifying an optimal color palette that is color blind safe

- ❑ Dataset(s): REXUS Building Data
- ❑ Question: What tools can we leverage for 508 Compliance in Tableau?

Steps:

1. Navigate back to the **Historical Status** tab
2. Add title
   i. E.g., descriptive title “Percent of Historical Buildings”
   ii. E.g., question you’d like viewers to ask, such as “Does historic status vary among leased versus owned properties?”
3. Rename y-axis **% of Properties**
4. Drag **Bldg State** to the Filter Shelf
5. Enable Quick filters
6. Change view fit from **Normal** to **Entire View**
7. Change color scheme to **Color Blind 10**
8. Display and review the Summary Card
9. Display and review the Caption Card
10. Review how to enhance dashboards with Tool Tips

Analysis – Viz Controls (Exercise 5)

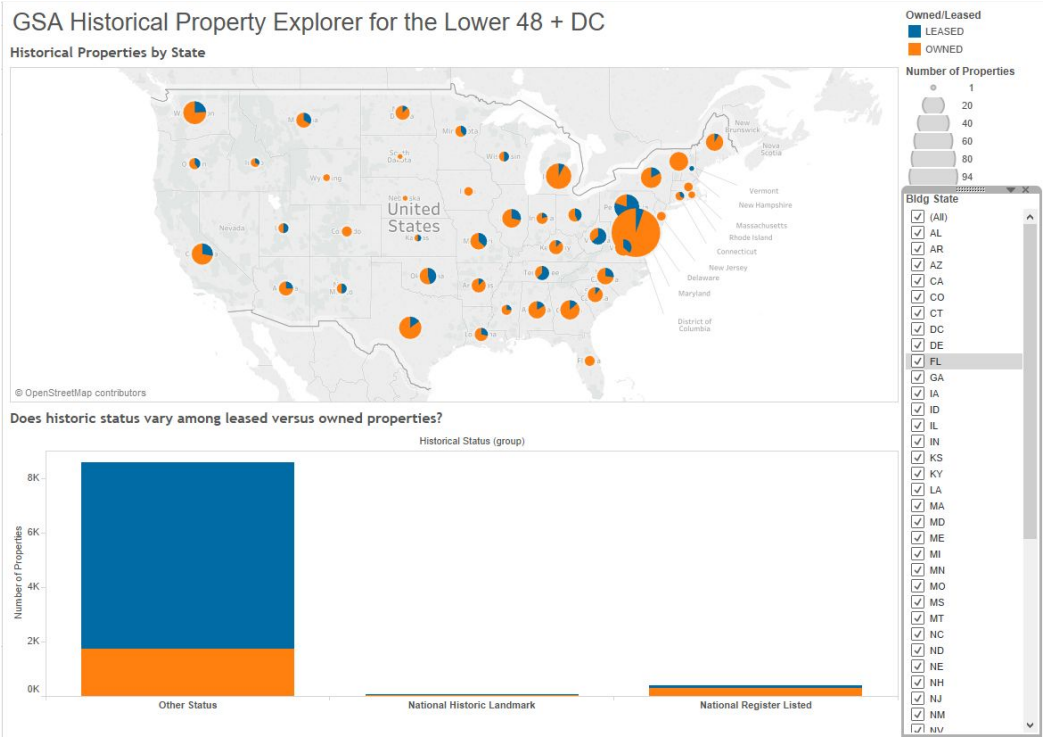
- Sketch your visualization before you start creating it or risk aimlessness
- Use your variables correctly, remember from earlier:
  - Dimension are for categorizing
  - Measure are for calculations
- You can use the search box on your map to zoom into specific locations:
  - Continent
  - Country
  - State or province
  - County
  - City
  - Postcode

- ❑ Dataset(s): REXUS Building Data
- ❑ Question: Which states contain more historic properties than others?

Steps:

1. Double-click **Bldg State**
2. Drag **Historical Status (group)** to the color control
3. Drag **Number of Records** to the **Size** control and into the view
4. Change marks type from **Automatic** to **Pie** (you may want to change the color of the pie)
5. Focus map on United States (continental). To do this, create Set "Lower 48 + DC" that excludes 7 areas: AK, AS, GU, HI, MP, PR, and VI. Drag to filter shelf.
6. Drag **Bldg State** to the filter box then right-click on the pill in the filter box and select **Show Filter**
7. Increase the size of the Pie marks
8. Edit pie legend to read Number of Properties
9. Give the worksheet a title **Historical Properties by State**

Exercise 6 – Creating a Dashboard



Dashboard (Exercise 6)

- Use vertical and horizontal containers to better arrange your objects.

- Dataset(s): REXUS Building Data
- Question: Which states contain more historic properties than others?

Steps:

1. Create a new sheet, select Dashboard
2. Add a Vertical Container
3. Drag in a Text Box for a title
4. Write title in text box “GSA Historical Property Explorer for the Lower 48 + DC”
5. Drag “Exercise #4” sheet below Text Box
6. Drag “Exercise #2” sheet below “Exercise #4” sheet
7. Apply the “State” filter to “All Worksheets using this Data Source”

Dashboard - Notes (Exercise 7)

- Do not go overboard! Think of the notes tab as a map rather than an encyclopedia.
  - Pithy > Wordy
  - Sufficient > 100% Comprehensive

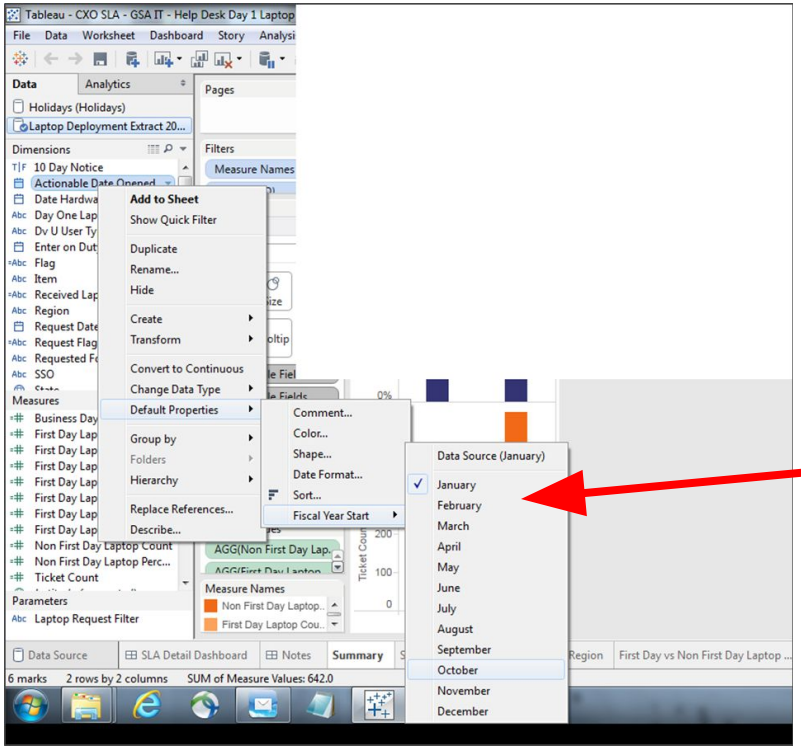
- Dataset(s): REXUS Building Data
- Question: None

Steps:

1. Create a new dashboard tab called “Notes.”
2. Write necessary documentation to do the following:
   i. interpret the dashboard
   ii. identify source systems
   iii. reproduce the calculations

Fiscal Year Start

Right-click on a date field



Hover until you get to Fiscal Year Start > Month. Choose desired month, e.g. October.

Tableau Official & Unofficial Resources

Official Tableau Resources

- Online Help: http://onlinehelp.tableau.com/v9.0/pro/online/windows/en-us/help.htm
- Quick Start Guide: http://onlinehelp.tableau.com/v9.0/pro/online/windows/en-us/help.htm
- Training, videos, webinars, white papers, events: http://www.tableau.com/learn
- Tableau Keyboard Shortcuts - http://onlinehelp.tableau.com/current/pro/desktop/en-us/shortcut.html
- Tableau User Groups -- these exist across the country (for Washington, DC go to http://community.tableau.com/groups/washington-dc)
- **Tableau Public Gallery, lots of interesting-looking examples: https://public.tableau.com/s/gallery**
- Whitepaper “Designing Efficient Workbooks.” http://www.tableau.com/learn/whitepapers/designing-efficient-workbooks

Unofficial Tableau-related Sites

- http://www.dataplusscience.com/TableauReferenceGuide/
- http://vizpainter.com/
- http://drawingwithnumbers.artisart.org/
- https://3danim8.wordpress.com/


GSA