U.S. General Services Administration

D2D Intermediate Tableau Training

February 13th and February 15th

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

- Analysis on the City Pairs Program (2015 Data)
  - Loading and Joining Data
  - Working with Calculations
  - Creating Analytics and Dashboards

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)

Importing Data
(Clean / Enhance our Data)

- You can **union** your data to combine two or more tables by appending values (rows) from one table to another. Note, the tables must come from the same connection.
    - Try adding the Source sheet to your union. What happens?
    - The **Wildcard (automatic)** union method enables you to combine data more easily; think of combining multiple fiscal years or

- **Hiding** a field simply means that when you create/refresh the extract, those hidden columns are not included (performance optimizing step).

❑ Dataset(s): CPP – Awards Info

Steps:

1. Open Tableau
2. In the Connect pane, select **Excel**
3. Browse to the CPP – Awards Info Excel workbook and open it
4. Drag the option **New Union** to the Canvas (i.e., the area that has a label over it 'Drag sheets here')
5. Select **Specific (manual)** as the option
6. Drag *CPP Data Source 02012018* into the Union pop-up window
7. Drag *CPP Data Source 02132018* into the Union pop-up window
8. Review the newly created columns (Sheet and Table Name) then **Hide** them
9. Rename our union to CPP Data Source

Question - Tableau made some assumptions about our data. What assumptions has Tableau done incorrectly thus far?

Importing Data
(Clean / Enhance our Data)

- You can use **split** or **custom split** options in Tableau to separate the values based on a separator or a repeated pattern of values present in each row of the field.
    - A string field can be split automatically based on a common separator that Tableau detects in the field (in our example, it's a hyphen).
    - A custom split enables you to define precisely how the split should occur.
- String functions allow you to manipulate string data (i.e. data made of text). Note, processing strings is a process heavy step, you should avoid, if possible (performance optimizing consideration).
- You can embed Tableau functions within other functions to achieve the desired outcome.

❑ Dataset(s): CPP – Awards Info

Steps:

1. In the Data Grid convert **AWARD_YEAR** to a date data type
2. We will split the Origin and Destination location fields to provide us more insight (think Text to Columns in Excel)
3. Select the drop-down button for the **ORIGIN_CITY_STATE_AIRPORT**, select **Split**
4. Rename the two new split fields **Origin_City_State and Origin_Airport**, respectively, hide the original field (ORIGIN_CITY_STATE_AIRPORT)

Question – Does the field Origin_City_State have any geographic role in Tableau (i.e., can we map it)?

5. Select the drop-down button for **Origin_City_State**, select **Create Calculated Field…**, label the new field **Origin_State** and type in the formula:
   RIGHT([ORIGIN_CITY_STATE],2)
6. Select the drop-down button for **Origin_City_State**, select **Create Calculated Field…**, label the new field **Origin_City** and type in the formula:
   LEFT([ORIGIN_CITY_STATE],LEN([ORIGIN_CITY_STATE])-2)
7. Hide both **ORIGIN_CITY_STATE_AIRPORT** and **ORIGIN_CITY_STATE**
8. Repeat steps 3 – 6 for the **DESTINATION_CITY_STATE_AIRPORT** field

Importing Data
(Clean / Enhance our Data)

- Calculated fields are NOT materialized into the data source, therefore you can not use the calculated fields to create joins.
    - In Tableau 10.2 and greater, you can use the Create Join Calculation…to define a custom join field.

Join Type	Result
Inner 	the result is a table that contains values that have matches in both tables
Left 	the result is a table that contains all values from the left table and corresponding matches from the right table. When a value in the left table doesn't have a corresponding match in the right table, you see a null value in the data grid.
Right 	the result is a table that contains all values from the right table and corresponding matches from the left table. When a value in the right table doesn't have a corresponding match in the left table, you see a null value in the data grid.
Full outer 	the result is a table that contains all values from both tables. When a value from either table doesn't have a match with the other table, you see a null value in the data grid.
Union	Though union is not a type of join, union is another method for combining two or more tables by appending rows of data from one table to another. Ideally, the tables that you union have the same number of fields, and those fields have matching names and data types.

❑ Dataset(s): CPP – Awards Info, Location Details, Airlines, and Fares

Steps:

1. In the Left Pane, next to Connections, select the **Add** option to perform a cross-database join
2. Browse to the CPP – Location Details Excel workbook and open it
3. Drag the **Origin City Info** sheet to the Canvas to capture additional airport location details
4. Under the Data Source side of the join window, select **Create Join Calculation…**and enter the following formula to extract the three-digit airport abbreviation

    RIGHT([ORIGIN_CITY_STATE_AIRPORT],3)
5. Under the Origin City Info side of the join window, select **Origin Airport Abbrev**, change the join to a **Left** join
6. Repeat steps 3 – 5 for the **Destination City Info** sheet
7. Add the remaining two databases to our join; browse to the CPP – Airlines and CPP – Fares Excel workbooks and open and join them to the CPP Data Sources dataset
8. Switch the Data Grid to a **Metadata Grid,** remove duplicate values and clean up variable names (i.e., proper or camel case, etc.)
9. Create an extract filter and *remove* any records that have a **Destination Country** outside of the USA, save your extract then save your workbook as a packaged worked (call it CPP Analysis)

Question – What type of join or joins do we need to use for the remaining workbooks? Why?

End of Part 1

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

Calculations in Tableau

- Basic calculations:
  - Broad set of built-in functions to create new variables which don't exist in the dataset
    - *e.g., I want to sum all my records, take the absolute value of my measure, if-then logical test, return string size, etc.*
  - Create new fields by using the "Create calculated field…" tool

- Table calculations:
  - Table calculations are transformations to the values of a measure in your View
    - *e.g., I want to calculate percent of total, the index, running average, etc.*
  - Common functions include INDEX(), RANK_PERCENTILE(), RANK(), etc
  - *http://onlinehelp.tableau.com/current/pro/desktop/en-us/calculations_tablecalculations_understanding_addressing.html*

- Level of Detail (LOD) Expressions:
  - Enable varying level of details in your View
    - *e.g., I want to compare the average expenditures by all agencies (Avg(Expense) in your Viz) to the aggregated per agency expenditure average (create an LOD Expression/Calc: {INCLUDE [AGENCY]: SUM([Expenses])}*
  - Curly braces are used in a calculation to indicate that the calculation should be independent of what is going on in the View
  - Scoping keywords for LOD Expressions are FIXED, INCLUDE, EXCLUDE (they affect the calculation but aren't required) but do affect how your data is displayed (FIXED does not impact your Viz, or, consider dimensions in your Viz, while INCLUDE/EXCLUDE do)

Outliers
(Basic Set of Dimensions)

- A box plot shows a minimum (-1.5 * IQR), First Quartile, Median, Third Quartile, and Maximum (+1.5 * IQR)
    - Great way to quickly identify outliers in your data
    - Definition of outlier is not standard across all datasets, what is considered an outlier in one dataset may not be an outlier in another

- Visualization of analytics can ease the consumption of vast quantities of data

- ❑ Dataset(s): CPP – Awards Info, Location Details, Airlines, and Fares
- ❑ Question: What are the outliers for each destination city?

Steps:

1. Drag **Destination City Name** to the Columns shelf
2. Drag **Unrestricted Coach Fares (YCA)** to the Rows shelf, change the measure aggregation to Average from Sum
3. Drag **Origin Airport Abbrev** to the Rows shelf
4. Change the graph to a box-and-whisker plot
5. Swap rows and columns
6. Click on the Size Marks card and increase the size of the marks just a bit
7. Right-click on any of the box plot minimum/maximum line and click on Edit . . .
8. Click on Hide underlying marks (except outliers)
9. Change the Fill to a welcoming color, make any other edits that you think will enhance the viz and click on Ok

Questions – Do we have any outliers? If so, what city? What considerations do you have to make in your analysis of any outliers?

Outliers
(Additional Dimensions)

- A box plot shows a minimum (-1.5 * IQR), First Quartile, Median, Third Quartile, and Maximum (+1.5 * IQR)
    - Great way to quickly identify outliers in your data
    - Definition of outlier is not standard across all datasets, what is considered an outlier in one dataset may not be an outlier in another

- Visualization of analytics can ease the consumption of vast quantities of data

- ❑ Dataset(s): CPP – Awards Info, Location Details, Airlines, and Fares
- ❑ Question: What are the most expensive Unrestricted Coach Fares (YCA) for destinations for DCA, ORD

Steps:

1. Drag **Destination City Name** to the Rows shelf
2. Drag the **Federal Employees Passenger Count** to the Size Marks card
3. Drag the **Unrestricted Coach Fares (YCA)** to the Color Marks card, change the aggregation to Average from Sum
4. Drag the **Origin Airport Abbrev** to the Filters box
5. Change the graph to a Tree Map
6. Drag the Destination State to the Label Marks card
7. Edit the Unrestricted Coach Fares (YCA) to a Palette that is intuitive for reviewing high/low (e.g., Red-Green Diverging)

Question – What are we looking for in this analysis? What will cause us to dig deeper into the set fare amount if we're planning for subsequent years?

Aggregation (Basic Calcs)

- Basic calculations allow you to transform values or members at the data source level of detail (a row-level calculation) or at the visualization level of detail (an aggregate calculation)

- What is the ATTR (Attribute) function?

  Shows value if values are unique, otherwise displays an asterisk (indicates more than one non-unique entry in the data)

- More than one way to completely logical calculations (e.g., CASE, IIF [simple Boolean tests], ISNULL, IFNULL, etc.)

- Logic functions are executed in the manner they are written; try to write your logic so most cases are executed in the first statement/line

❑ Dataset(s): CPP – Awards Info, Location Details, Airlines, and Fares

Steps:

1. Create a new worksheet, call it Aggregations
2. Create a new calculated field called **City Pair** (alternatively we can create a combined field)

   [Origin Airport Abbrev] + " ✈ " + [Destination Airport Abbrev]
3. Create a three new calculated fields:
   - SUM Calc 1: *sum([Unrestricted Coach Fares (YCA)]/[Number of Records])*
   - *SUM Calc 2: sum([Unrestricted Coach Fares (YCA)])/sum([Number of Records])*
   - *SUM Eval: if [Sum Calc 1] > [Sum Calc 2] then "Calc 1 > Calc 2" else "Calc 1 <= Calc 2" end*
4. Drag **City Pair, Unrestricted Coach Fares (YCA), Number of Records, Sum Calc 1, and Sum Calc 2** to the rows shelf
5. Convert the four measures to Discrete
6. Drag **Sum Eval** to the rows shelf
7. Drag **Sum Eval** to the Filters shelf, select “Calc 1 > Calc 2”

Top N Calculation
(Table Calcs and Nested Sorting)

- The Index function returns the index of the current row in the partition, without any sorting with regard to value
- Combine fields to create a cross product of members from different dimensions. You would combine dimensions if you want to encode a data view using multiple dimensions.
- Use the Index function to find Top N, it will return the index of the current row in the partition, ***without any sorting with regard to value***.
- Control-drag to “copy move”; remember calculation performed in a measure / dimension reflects the data in the Viz (e.g., Using Index from the Data Pane to filter vs. the Index we used with a table calc.)

❑ Dataset(s): CPP – Awards Info, Location Details, Airlines, and Fares

Steps:

1. Create a new worksheet, rename it to Table Calc
2. Drag **Origin Airport Abbrev** and **Destination City Name** to the Rows shelf
3. Drag **Unrestricted Coach Fares (YCA)** to the Text marks card (or in the Viz), change measure aggregation to Average from Sum
4. Sort Ascending
5. Create a new combined field **Origin Airport Abbrev** and **Destination City Name**
6. Drag the new combined field to the Rows shelf, just before Destination City Name
7. Right-click on the combined field and select Sort
8. Change the Sort by to Field, and select **Unrestricted Coach Fares (YCA)** and Aggregation to Average
9. Right-click on the combined field and select Show Header
10. Create a calculated field, call it **Index**, and set it to call the function **Index()**
11. Drag **Index** to the Rows shelf, convert it to discrete
12. Right-click on **Index** and select Edit Table Calculation…
13. Select Specific Dimensions then select **Origin Airport Abbrev** and **Destination City Name**
14. Change the value of At the level to Destination City Name and the value of Restarting every to Origin Airport Abbrev
15. Change the Sort order to Custom Sort, select **Unrestricted Coach Fares (YCA)** and Average then drag Index to the Filters box

Top N Calculation
(Basic Calcs and Parameter)

- Alternative to creating the Top N PAX Calc:
  - IIF([Index] <=[Top PAX],"Display", "Don't Display")
  - Won't handle NULLs well

- Alternative to creating the City Pair calculation:
  - Create a combined field the Origin and Destination Airport Abbrev fields
  - Drag Origin and Destination fields each to the rows shelf

- Parameters:
  - Add more functionality for the user
  - Require accompanying calculated fields to be of any use
  - With calculated fields can implement complex logic
  - Can be used to filter across data sources; the only other real option there is to use filter actions
  - Run faster if their values are integers; choose numeric values for parameters, and have them "Display As" something else

❑ Dataset(s): CPP – Awards Info, Location Details, Airlines, and Fares

Steps:

1. Create a new worksheet, rename it to Top PAX
2. Create a new Parameter, call it **Top PAX,** set it to an integer type with a default value of 1
3. Create a calculated field called **Top N PAX Calc**

   if [Index] <=[Top PAX] then "Display" else "Don't Display" end
4. Drag **City Pair** to the Rows shelf
5. Drag **Federal Employees Passenger Count** to the Text marks card, change the measure type to Average from Sum
6. Drag **Index** to the Rows shelf, change it to Discrete
7. Sort the data in descending order
8. Drag the **Top N PAX calculation** to the Filters shelf
9. Right-click on the parameter and select **Show Parameter Control**
10. Replace the number "1" to switch among the top travelled city pairs

Resolving Duplicates
(Level of Detail Calc)

- Just like basic calculations, LOD calculations allow you to compute values at the data source level and the visualization level. However, LOD calculations give you even more control on the level of granularity you want to compute. They can be performed at a more granular level (INCLUDE), a less granular level (EXCLUDE), or an entirely independent level (FIXED) with respect to the granularity of the visualization.

- LOD Expressions can be set to:
    - FIXED (not impacted by the dimensions in the Viz)
    - INCLUDE (includes dimensions within expression AND in the Viz)
    - EXCLUDE (exclude dimensions in the Viz from the calculation)

Tableau Resource on LOD Expressions

❑ Dataset(s): CPP – Awards Info, Location Details, Airlines, and Fares

Steps:

1. Create a new worksheet, rename it to LOD De-Dupe
2. Drag **Origin Airport Abbrev** and **Destination Airport Abbrev** to the Rows shelf
3. Drag the **Number of Records** to the Text marks card (or drop it into the Viz)
4. Drag **Unrestricted Coach Fares (YCA)** two times to the Rows shelf
5. Change both pills to Discrete then change the second pill to Average from Sum
6. Create a new calculated field called **LOD (De-Dupe Unrestricted Coach Fares)**

    { FIXED [Origin Airport Abbrev],[Destination Airport Abbrev] :min([Unrestricted Coach Fares (YCA)])}
7. Drag **LOD (Unrestricted Coach Fares)** to the Rows shelf and convert it to Discrete

Duplicate the worksheet and rename it LOD Exclude

1. Create a new calculated field called **LOD (Exclude Destination Airport)**

    { EXCLUDE [Destination Airport Abbrev]:AVG([Unrestricted Coach Fares (YCA)])}
2. Create a new calculation call it **Fares Above or Below Average**

    [Unrestricted Coach Fares (YCA)]/ [LOD (Exclude Destination Airport)]
3. Remove the third and fourth pill in the Rows shelf
4. Drag **LOD (Exclude Destination Airport)** and **Fares Above or Below Average** to the Rows shelf, convert them to discrete
5. Right-click on **Fares Above or Below Average** and click on Show Filter

Accessibility

- The wider and more visible the dashboard, the more important accessibility becomes Nightmare scenario:
  - Someone needs to use your dashboard to do a job;
  - They can't because of accessibility issues and have no recourse;
  - Job doesn't get done, frustration or worse ensues.

- Don't hesitate to contact 508 Accessibility POCs if needed. D2D Team can also assist up to a point.

❑ Dataset(s): Applies to Any Dataset

❑ Question: How to improve 508 compliance?

Steps

1. Make the raw data visible / available. Some options:
   A. Right click sheet title > duplicate as crosstab, make crosstab available.
   B. Click Data menu > view data > select all > copy > paste in spreadsheet.

   C. **Highly Recommended:** unless data sensitivity/confidentiality issues, link to a downloadable text or Excel dataset on a screen-readable website
2. **Highly Recommended:** Add screen-readable disclaimer & contact information
3. Other measures
   1. Use Colorblind-friendly color palette
   2. Use Auto-captions
   3. Design visualizations to appear and/or operate in similar ways; seek/consider *standards*

Outside Tableau Resources*

Official Tableau Resources

1. Online Help: http://onlinehelp.tableau.com/v9.0/pro/online/windows/en-us/help.htm
2. Quick Start Guide: http://onlinehelp.tableau.com/v9.0/pro/online/windows/en-us/help.htm
3. Training, videos, webinars, whitepapers, events: http://www.tableau.com/learn
4. Tableau User Groups -- these exist across the country. For example, here is the one for Washington, DC: http://community.tableau.com/groups/washington-dc
5. Tableau Public Gallery, lots of interesting-looking examples: https://public.tableau.com/s/gallery
6. **Whitepaper "Designing Efficient Workbooks." http://www.tableau.com/learn/whitepapers/designing-efficient-workbooks**

Unofficial Tableau-related Sites

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


GSA