Hands on Training: Intro to Tableau

# About this Data

The [data set](http://data.worldbank.org/data-catalog/crowd-sourced-price-collection) is from the World Bank, collected as part of a pilot study on the feasibility of crowd-sourced price data collection. Price data was collected by non-professionals for thirty specific food items across several months in eight countries.

The Excel file contains geographic data (city and country), time data (date of observation), product information (name, quantity) and price (standardized quantities, prices standardized to USD).

# Agenda

1. Format of the Training
2. Connect to Data & Data Cleanup
3. Overview of Tableau Basics and Concepts
4. Mapping the Data
5. Analysis
6. (Optional) LOD Expressions Exercise
7. Story Points
8. Dashboards in Story Points

|  |
| --- |
| Format of the Training Each section has a guiding statement or question  Overview:  Provides basic directions on the purpose of the section and what the desired outcome is. For those who want to dive in and try it on their own, this may be enough guidance to begin.  Detailed Steps:  Provides step-by-step level directions on how to complete the section. |

|  |
| --- |
| Connect to Data Connect to an Excel file and join two tables  Overview:  Open the file “Crowd Sourced Grocery Prices” and cleanup the data  Detailed Steps:   1. Open Tableau Desktop 9.0 (Tableau 9.0 Beta if applicable) 2. On the Connect Pane, click Excel 3. Navigate to the file on your machine (desktop) and open it 4. Click and drag out the sheet named *Observations* 5. View the data in the preview pane below 6. Find the data field called Location. Split this field into City and Country (we’ll want to analyze these separately)    1. Click the drop down menu next to Location and select Split    2. Change the Geographic Role for Location – Split 1 to be City by clicking the =Abc Icon    3. Change the Geographic Role for Location – Split 2 to be Country by clicking the =Abc Icon    4. Rename the column headers by editing the metadata      * 1. Rename “Location – Split 1” to be City   2. Rename “Location – Split 2” to be Country  1. Click “Sheet 1” |

|  |
| --- |
| Overview of Tableau Basics and Concepts Concepts to go over   1. A good way to approach data analysis is to **start by asking questions of the data**, then planning how to answer the questions with Tableau. Often the questions will change as you begin working with the data but it’s helpful to have a guiding question when you start. 2. Fields are broken up into Dimensions and Measures.    * **Dimensions** (blue) are categorical fields. They are the labels in a visualization, the buckets that data falls into such as locations, product names, etc.    * **Measures** (green) are quantitative fields. They are the axes in a visualization, the numbers that can be analyzed, such as price and counts of records. 3. Tableau creates some fields that can be used in a visualization that do not exist in the original data set.    * If the data set contains geographic fields, such as country or city, Tableau searches an internal database and generates **Latitude** and **Longitude** fields. This enables the geographic data to be plotted on a map    * **Number of Records** is a simple count of rows in the data set 4. **Show Me** can be accessed in the upper right corner of the screen. With field(s) selected, Show Me offers one-click options for chart types 5. The Tableau canvas    * Data window – purple – drag fields from here to bring them into the view, either data fields or, if you’re on the Analytics tab, items like reference lines and box plots.    * Shelves – blue – areas where fields can be placed to control exactly how they appear in the view    * Canvas – green – where the visualization is built. Fields can be placed directly here as well as on shelves    * Toolbar and ribbon – orange – additional controls and menu options, including Undo and Clear Sheet     Take the next few minutes and explore the data – ask a question and try to answer it, play with dragging fields to different places or using Show Me to get a feel for how the software behaves. |

|  |
| --- |
| Mapping the Data What does our data look like globally?  Overview:   1. Where is data coming in from?    1. Plot countries as a map 2. Which country has the most expensive average price?    1. Color by **Price**, changing the default aggregation to average 3. Make a filled map of countries   Detailed Steps:  Where is data coming in from?   1. Clear the sheet  (in the ribbon) or create a new sheet  (at the bottom) 2. Drag **Country** from the data window to the canvas    1. If you get anything other than a map, undo and try dropping the field **Country** into the large, bottom right rectangle that says “Drop field here”.       1. Alternatively, double click on the field name to bring **Country** out as a map    2. Note that the generated Latitude and Longitude were automatically plotted on the Rows and Columns shelves     Which country has the most expensive average price?   1. Drag **Price** from Measures to the Color shelf (button)    1. This should create a filled map colored on a gradient of “SUM(Price)”    2. Note that the default aggregation is SUM 2. Right click (on a Mac, control + click or two-finger click) on the pill on the Marks Card that says “SUM(Price)”    1. Click on the option “Measure(Sum)”    2. Click on “Average” to change the aggregation from Sum to Average 3. Hover over the darkest green country to see which country has the highest average price     Make a filled map of countries   1. Clear the sheet 2. Double click on **Country** to plot the map 3. Drag **Country** from the data window to the color shelf 4. Select the dropdown on the Marks card and change from “Automatic” to “Filled Map” 5. Right click on the sheet tab and Rename the sheet “Countries” and create a new sheet     If desired, try creating additional maps to answer further question. What cities are represented in the data? Are there an equal numbers of observations from all locations? |
| Analysis Questions for analysis: How much variation is there in product prices across each country?  Overview:   1. If the analysis relates to variation, what visualization type makes the most sense?   Detailed Steps:  How much variation is there in product prices per country?   1. Decide which chart type to use. Here, we’ll use Bar Charts. 2. Holding down the Control key (Command on a Mac) click to select the fields: **Product** **Name**, **Price** 3. With those two fields selected, click on the Show Me tab 4. Select the Horizontal Bar Chart and click the Show Me tab again to close it 5. Double click on the “SUM(**Price**)” pill on the Columns shelf    1. Change the aggregation from a SUM to a MEDIAN by replacing SUM with MEDIAN    2. Hit Enter 6. Hover over the words “Median Price” on the X axis until the Sort icon appears. Click the icon to sort 7. Drag **Country** from the Rows shelf to the Columns shelf in front of **Price**. 8. Drag a new copy of **Country** from Dimensions to the Color shelf 9. Right click on the sheet tab and rename the sheet “Price Variation by Country”   Analysis:  Here, we can get a sense of overall prices across countries. Nigeria has more tall bars, indicating it has higher prices, than other countries. What other patterns do you see? What insight do you get?   1. Click to the Analytics pane 2. Drag Average Line and drop on Pane to add an average reference line per pane 3. Click on a reference line and select Edit 4. Choose to label the line with the Value 5. Click OK 6. Play around with this view. Try selecting multiple bars. What happens to the Average line? 7. Right click on the sheet tab and rename the sheet “Price Variation by Country”     If time remains, explore other analyses of the data. Can you create a boxplot of the data to see what products are outliers across countries? What quantities were looked at for each product? How might you best compare prices for a single product across countries? |

|  |
| --- |
| Optional LOD Expression Exercise Which City has the greatest maximum price? Which Country has the highest average of the maximum price?  Overview:   1. Brief explanation of creating a Calculated Field 2. Creating and using an LOD Expression   Setup/Background for LOD Expressions – No need to follow along in this section  A key part of using calculations in Tableau is understanding the aggregation and/or granularity of your data. When placing dimensions/measures on the canvas, Tableau will automatically aggregate the data to the most granular dimension placed on the view.  For example, in this particular dataset, **City** is more granular than **Country**. If we start with a view of only **Country** and **MAX(Price)**, each row of data (price of one item in one city) will be grouped by **Country** and then the maximum price is shown on the view.  *Note: the default aggregation of SUM has been changed to a MAX, similarly to how the SUM(****Price****) was changed to AVG(****Price****) earlier in the training.*    If we place **City** next to **Country** on the Rows Shelf, the length of the bars becomes the maximum price within each **City/Country** combination (i.e. the data is more granular than the previous view).    Let’s say we’d like to take the maximum price per city and average that for each country. We want to use the information about each city’s maximum price without having to make our view granular to the level of city.  LOD Expressions allow you to explicitly define the level of aggregation on the view—we’ll explicitly define the level of detail to be **City**. The syntax for using an LOD Expression is as follows:  {INCLUDE [City] : MAX([Price])}   * {}: Wrap your LOD Expression in curly braces * *INCLUDE:* This syntax uses the *INCLUDE* keyword, which will include the following dimension in the calculation without displaying it on the view.   + Note: there are two other keywords (*EXCLUDE* and *FIXED*)—each of which will incorporate the level of detail differently. * [City]: Place a dimension after the keyword. This dimension will be the level of detail used in the calculation.   + Note: you may place multiple dimensions here. Simply separate each with a comma. * : MAX([Price]): the colon signifies the beginning of the aggregation. The aggregation (MAX([Price])) will be computed for each **City**.   The value resulting from this LOD Expression (on the right in the image below) will give us the average of the maximum **Price** for each city—which is computing at the level of **Country**, because of how we built the view, but also INCLUDING the level of **City**, as set in the LOD Expression. On the other hand, if we had just placed the AVG**(Price)** on the view (on the left in the image below), the value would be the average **Price** for each country, since that is the granularity in the view.    Follow along the remainder of the training to see how to use this LOD Expression in the view to answer our initial question.  Detailed Steps:   1. Create a Calculated Field    * Right click in the Data Pane to Create a Calculated Field. If you click on a field instead of white space, your menu may look slightly different. If you don’t see Create calculated field, look for a Create menu, expand that, then click Calculated field.      1. The Calculation Editor – for a more thorough discussion of Calculations, refer to the [Online Help](http://kb.tableau.com/).    * Calculation name – orange – give your calculation a name    * Calculation editor canvas – green – areas where calculations can be built. These calculations may include, but are not limited to using fields, functions, and other operators.    * Functions – blue – using functions, you may compute values such as a sum, average, and minimum, to just name a few. You can use the menu to drill down into specific types of functions (date, string, number, etc.).    * Function description – purple – additional information using the function     This remainder of this LOD Expression exercise goes a bit beyond true introductory topics, which is why this is optional, but it’s a game changing 9.0 feature and you’ll want to at least have a basic knowledge.   1. Create the LOD Expression    * Name your calculation “Max Price by City”    * Define the calculation as: {INCLUDE [City] : MAX([Price]) }    * Select OK      1. Drag **Country** to the Rows Shelf 2. Drag a new copy of **Country** from Dimensions to the Color shelf 3. Drag **Max Price by City** to the Columns Shelf 4. Double click on the “SUM(**Max Price by City**)” pill on the Columns shelf    * Change the aggregation from a SUM to a AVERAGE by replacing SUM with AVG    * Hit Enter   From our previous view, we know that there is a large variation in the prices of each **Product Name**. Because of this, the current view may be somewhat misleading as it only accounts for the variation in prices among the cities. Let’s filter the data to show only one **Product Name** at a time.   1. Right click on **Product Name** anywhere you see it and select “Show Quick Filter”    * Click on the caret in the upper right corner of the filter to bring up the menu and select “Single Value (Dropdown)” 2. Click through a few different **Product Names**. Does Kenya have the highest average of maximum price for each **Product Name**?     If you’d like to dive deeper into LOD Expressions, please stay tuned for the LOD Expressions whitepaper.  Explore a few other views using the LOD Expression we created. What other insights can you find? |

|  |
| --- |
| Story Use Story Points to present your findings on the data set  Overview:  Create a Story using the visualizations created above to tell the story of the data. Stories can have captions, floating descriptions, and are fully interactive. Filter selections can be saved (updated) or duplicated as new points.  Detailed Steps:  I. Create a story and add a point with description   1. Create a blank story by clicking on the tab at the bottom 2. If necessary, use the sizing menu in the bottom left corner to adjust the size of the story to fit your screen (laptop or automatic) 3. Click and drag out **Countries** (Map) 4. Click in the navigator box to add a caption for the map    1. “Data was collected from 8 countries around the world” 5. Click and drag out the Description option from the left pane to add a comment about the map    1. “Prices for specific food items were recorded for several months via crowd sourced data collection” 6. Click the Story menu -> Format    1. Change the Navigator format to have Black shading and Segoe UI, 12pt font     II. Add another point and modify the visualization   1. Double click **Price Variation by Country** 2. Click in the navigator box to add a caption    1. “Prices of the food products varied between countries” 3. Go back to the sheet for **Price Variation by Country** 4. In the ribbon, use the drop down to change the fit from “Normal” to “Entire View”    1. If you want to maximize screen real estate, you can remove the color legend. Click on the caret on the Countries color legend and select “Hide Card” 5. Click back on the story to verify there are no longer scroll bars on the story    1. Note: most changes to a visualization must be made on the underlying sheet, not in the Story     III. Showcase interactivity and saving a specific state of a visualization   1. On the current point (“Prices of the food products varied between countries”),    1. Hold down the control key (command key on a Mac) and click on several bars that don’t seem to fit the overall trends 2. Click the button “Save as a New Point”    1. This duplicates the point – the visualization is the same but with those selections still highlighted    2. If you want to change which bars you’ve selected, you can select other bars and click “Update” above the navigator box to save those modifications 3. Click in the navigator box to add a caption    1. “Which products were more expensive varied by country”    2. If the entire text doesn’t display on the story point, drag the story point size to increase the width. Note: this will increase the height of all story points on this story     Rename the story to be “Crowd Sourced Grocery Prices”. Now let’s go back and make a couple more visualizations to put into our Story. |
| Dashboards and Stories Questions for analysis: Are there price fluctuations or have prices held fairly steady?  Overview:   1. If the analysis relates to variation, what visualization type makes the most sense? 2. Note: instead of choosing a specific country or product, allow the end-user to choose by using quick filters   Detailed Steps:  Are there price fluctuations or have prices held fairly steady?   1. Create a new sheet 2. Drag **Obs Date** to the Columns shelf 3. Right click on the pill and select the option “Week Number” 4. Drag **Price** to the Rows shelf 5. Double click on the “SUM(**Price**)” pill on the Rows shelf    1. Change the aggregation from a SUM to a Average by replacing SUM with AVG    2. Hit Enter 6. Drag **Country** to Color 7. Right click on **Country** anywhere you see it and select “Show Quick Filter”    1. If desired, click on the caret in the upper right corner of the filter to bring up the menu and select “Multiple Value (Dropdown)” 8. Drag **Product Name** to Detail 9. Right click on **Product Name** anywhere you see it and select “Show Quick Filter”    1. Click on the caret in the upper right corner of the filter to bring up the menu and select “Single Value (Dropdown)” 10. Right click on the sheet tab and rename the sheet “Timeline of Price Fluctuation”   Analysis:  Clicking through the **Product Name** filter, what patterns emerge? What products have a more or less price variation? Which products are the most stable, within or across countries? Are there any countries that deviate from the overall pattern?   Now let’s build a dashboard to bring two elements together Overview:   1. Make a dashboard with a timeline and map   Detailed Steps:   1. Create a blank dashboard by clicking on the tab at the bottom 2. Fit the size of the dashboard to “Fit to Crowd Sourced Grocery Prices” 3. Drag out **Countries** 4. Drag out **Timeline of Price Fluctuation** to the bottom 5. Click on the map to bring up its border    1. At the top right corner, click the caret to bring up the menu    2. Select “Use as Filter” to set the map as a filter for the timeline 6. Right click on the title “Countries” above the map and chose “Hide title” 7. If desired, close the Country color legend by clicking to bring up the border and clicking the X. 8. If desired, close the Country quick filter by clicking to bring up the border and clicking the X. 9. Double click on the tab to rename the sheet **Price Fluctuation** |

|  |
| --- |
| Back on the Story IV. Add another point and finish the story   1. Double click **Price Fluctuation** to bring it out to the story 2. Click in the navigator box to add a caption    1. “Explore the dashboard to see how prices changed over time” 3. Click and drag out the Description to add a caption    1. “Click on a country to see just the records for that country”    2. “Click on a product (or “All”) to change the timeline”      1. Double click “Story Title” to add a title to the story    1. “Variation of Food Prices” |

|  |
| --- |
| If you made any other sheets, continue adding new points. Experiment with updating points or creating new points with filter or highlighted selections.What story can you find in the data? |