INTRODUCTION TO TABLEAU FOR HEALTH AND POLICY

Class 1: Introduction to Tableau

Douglas Joubert, NIH Library

Version: April 2017
# Table of Contents

About the Data.............................................................................................................................................. 4
  County Level Variables........................................................................................................................ 4
  Population counts ................................................................................................................................. 4
  Poverty counts ...................................................................................................................................... 4
  SNAP participant counts ....................................................................................................................... 4
  SNAP participant benefits ..................................................................................................................... 5
  Ratios .................................................................................................................................................... 5

Introduction to Tableau ................................................................................................................................ 5
  What is Tableau? ....................................................................................................................................... 5
  Tableau Versions ....................................................................................................................................... 5
  Introducing Tableau file types .................................................................................................................. 6

Managing Data Sources .................................................................................................................................. 6
  Connecting to a data source ..................................................................................................................... 6
  Joining related data sources ................................................................................................................... 11
  Reading loosely structured Excel files ..................................................................................................... 17

Exploring the Tableau Interface .................................................................................................................. 22
  Dimensions and Measures...................................................................................................................... 22
    Dimensions.......................................................................................................................................... 23
    Measures............................................................................................................................................. 25
  Menu and Toolbar................................................................................................................................... 27
  Data Pane................................................................................................................................................ 28
  Analytics Pane......................................................................................................................................... 29
  Shelves and Cards ................................................................................................................................... 30
    Pages Shelf .......................................................................................................................................... 30
    Filters Shelf.......................................................................................................................................... 30
    Marks Card .......................................................................................................................................... 38

Managing Tableau Data .............................................................................................................................. 38
  Adding, duplicating, and renaming worksheets ..................................................................................... 40

Saving Your Work ........................................................................................................................................ 47

Exporting Data ............................................................................................................................................. 47

Exporting as an Image................................................................................................................................... 48
About the Data

In this session, I will be imported USDA data, on their SNAP program.

The County data worksheet contains 46 variables.

The data includes the following variables (filtered for the years 2006 – 2010):

**County Level Variables**

- FIPSTXT: FIPS code in text format with leading zeros.
- COFIPS: County code portion of total FIPS code (last three digits), not unique across States.
- CODEF: Source of county definition: BEA_ONLY for Bureau of Economic Analysis, CEN_ONLY for Census Bureau, and CEN_BEA when both Census Bureau and BEA use the same definition.
- **CONAME**: County name and two-digit State identifier.
- URBCODE: 2003 Rural-Urban Continuum Code per ERS. For more information about the rural-urban code, see 2003 Rural-Urban Continuum Codes.
- METRO: Metro/non-metro code per U. S. Census Bureau. For the metropolitan/non-metropolitan codes used in the data set, see Rural Definitions: Data Documentation and Methods.

**Population counts**

Area estimates of population counts are obtained from the U.S. Census Bureau's data, which includes a description of the methods used to create the estimates.

- **POPxx**: Estimated population count, with the year for the data indicated by the last two digits of the variable name.

**Poverty counts**

The Small Area Income & Poverty Estimates Program of the U.S. Census Bureau provides area estimates of the number of persons in households\(^1\) with total income below 100 percent of the Federal poverty level (FPL)\(^2\).

- **NUMPOVxx**: Estimated total number of poor people, with the year for the data indicated by the last two digits of the variable name. Data are provided for 1999, 2000, 2006, 2007, and 2010.

**SNAP participant counts**

Participation counts at the county level are provided primarily by the Small Area Income & Poverty Estimates (SAIPE) program of the U.S. Census Bureau.

- **PRGNUMxx**: Estimated number of active SNAP/Food Stamp Program (FSP) participants. The last two digits (xx) indicate the year for the data.

---

\(^1\) Income" above refers to "modified adjusted gross income" (MAGI). For most people, it is the same or very similar to "adjusted gross income" (AGI). MAGI isn't a number on your tax return.

\(^2\) In 2017, this is 12,060 for individuals or $24,600 for a family of 4.
• **PRGxx_yy**: Percentage change in the number of SNAP/FSP participants between the two years represented by "xx" and "yy" in the variable name.

**SNAP participant benefits**
The Regional Economic Information System, Bureau of Economic Analysis of the U.S. Department of Commerce provides estimates of total annual benefits issued by area. State and county estimates are based on county tabulations of the distributed benefits from FNS, USDA.

• **PRGBENxx**: Estimated total annual amount of SNAP/FSP benefits (in thousands) issued to State or county residents in a year. The last two digits indicate the calendar year for the data.

**Ratios**
A number of variables in the file compare participant counts or program benefits to an underlying population group (computing the values in green).

• **PRG_POPxx**: The ratio of program participants to area population, with the last two digits of the variable name indicating the data year.
• **PRG_POVxx**: The ratio of program participants to area count of poor persons, with the last two digits of the variable name indicating the data year.
• **AMB_POPxx**: The ratio of total average monthly program benefits issued to all participants in an area to that area's total population count, with the last two digits of the variable name indicating the data year.
• **AMB_POVxx**: The ratio of total average monthly program benefits issued to all participants in an area to that area's count of poor persons, with the last two digits of the variable name indicating the data year.
• **AMB_PARxx**: The ratio of total average monthly program benefits issued to all participants in an area to that area's count of program participants, with the last two digits of the variable name indicating the data year.

**Introduction to Tableau**

**What is Tableau?**
Tableau is a tool that allows you to easily connect to data, then visualize and create interactive, sharable dashboards.

**Tableau Versions**
There are many different versions of Tableau, all with slightly different capabilities.

There is a completely free version called [Tableau Public](#), which includes a Tableau Public interface app that you download. This version lets you do almost everything the other versions can do.

However, with the free version you can only connect to Excel spreadsheets, text files, and Microsoft access files. Also, all Tableau Public analysis is stored online and can be viewed by anyone.

There are also paid Tableau desktop versions. These come in two flavors:

1. **Personal Desktop**
2. Professional Desktop

Both of these versions allow you to save data and workbooks to your desktop and provide ways to share your work with others online.

The NIH Library has a license to Tableau Professional Desktop, which is installed on our Data Sciences workstation.

Introducing Tableau file types
Tableau has a number of different file types:

1. The first is the Tableau workbook, which ends in the .twb file extension. The workbook stores a visualization without including the source data.
2. The second is the Tableau data source, or .tds file. This file stores the server address, password, and other information required to access a data source.
3. The next file type is the Tableau bookmark, which has a .tbm extension. Tableau bookmarks are a connection to a worksheet in another Tableau workbook. Creating these connections lets you reference other workbooks from within a file so you do not have to recreate the visualization in your current file.
4. Next, is the Tableau data extract, or .tde file. These files store Tableau data as a filtered and aggregated extract. You might not get the individual results you are looking for, but you will have a look at an aggregated view of your data.
5. Finally, there is the Tableau packaged workbook. These files have a .twbx extension, and they store extracted data and visualizations for viewing in Tableau, or Tableau Reader.

Managing Data Sources
Connecting to a data source
Worksheet: SNAP_just_county

When you are ready to create a visualization within Tableau, you need to connect to an outside data source.

In this section, I will show you how to connect to an Excel workbook. However, the skills that you learn here, will also apply to other types of data sources as well.

When you fire up Tableau, you will see the start screen [Figure 1].
Figure 1: Tableau Start Screen.

The section on the left is where you connect to data sources. You can see that Tableau can connect to different files, different servers, or databases.

In the middle section of the start screen are previously used data sets, or example workbooks.

On the right are training resources.

1. To connect to an Excel worksheet, simply click on the Excel link, and navigate to the file that you want to import [Figure 2].
2. For this example, I am going to open the **SNAP just_county** file. The screen you will see is the data import screen [Figure 3].

3. Drag the **just_county** sheet to the **data preview** section of the import screen [Figure 3].

4. Starting from the top of the import screen, I can:
   a. Click on the data icon near the name of the worksheet, or click **Add** to add a new data source, on the Connections panel [Figure 3].
   b. Choose a live or extracted view of the data.
   c. Filter the data by selecting **Filter>Add>Add>** List of variable names.
   d. Search for sheets by clicking on the **Search** icon.
5. There are two dropdown menus over each variable name.

6. With the first options, I can correct any errors in the field types. For example the FIPS codes are being "recognized" as a number, when it should be represented as a text field.

7. Using the second option, I can rename/describe the variable. These options can also be changed after you import the data.
   
a. Choose **Describe>Load** variable.

b. As you can see from Figure 4, there are 3,196 members (counties) in this file.
8. Let us go ahead and change the values for the FIPS codes, to string. Because I am using the live view of the data, a query is executed against the original data file.

9. Click on the tab labels as **Sheet 1** to switch to what I call the **workbook** view [Figure 5]. You can always switch back to the data view by clicking on the Data Source tab.
Joining related data sources

1. Sometimes it might be necessary to grab data from different sources and join them using a common field (primary key).
2. The primary key in this data set is the FIPS code. This means that both the just_grants and snap_benefits tables have FIPS fields.
3. Click on the Data Sources tab near the bottom left.
4. There are two ways to create a union. The first is to double-click on the New Union icon. The second is to drag the New Union icon to the sheets workspace.
5. Double-click on the snap_benefits table to add it to the workspace.
6. If you perform a join and you see something like Figure 6, it means that Tableau could not figure out which fields to join.
7. If you look closely at Figure 6, you can see that the FIPS column/field in the grants table is still formatted as a number.
Figure 6: A join mismatch.

8. Scroll over to the FIPS (snap!benefits) column and covert this value to a string. Once this is done, the join will work [Figure 7].
9. If you click on the **Join** icon, the join parameters will pop up [Figure 8]. As you can see in Figure 8, there are two types of joins possible:
   a. **Inner join**: Only matches within both tables.
   b. **Left-join**: All the values in the left table (just county) and all matching values in the right table (snap_benefits). Members with no matches will show up as **null** on the right.

---

**Figure 7**: Performing a join in Tableau.
Figure 8: Join properties.

10. You can filter each of the data sources (look for other variables) or change the type of "connection expression" via the join popup menu [Figure 9]. This is useful because sometimes tables will have the same data in a field...but the column name might be different.
11. Click on the **Sheet 1** tab.

12. Figure 10 is displaying the joined data. As you can see in Figure 10, the data panel is now divided into two separate sections: *(just_county and snap_benefits)*.
13. If I wanted to, I could rename each one of the variables [Figure 11].
14. If you click on the little blue half circle [Figure 11], you can change the variable back to its default name.

Figure 10: Data sheet with the joined data.
Figure 11: Renaming variables in Tableau.

Reading loosely structured Excel files
Worksheet: acf_data_tab2016
Worksheet: ag_ruraldefinitions_national

Tableau has the ability to "correct" the formatting for some types of data. If you look at the table in Figure 12, you can see that it is not formatted like the previous Excel file.

The Data Interpreter tool allows you to tweak irregularly structured data so that it can be imported into Tableau. It works with simple files like the ones shown in Figure 12.
Figure 12: Formatted Excel table.

However, for more challenging files [Figure 13], Tableau is less successful. This is because this worksheet has been formatted for print.
Figure 13: USDA data file with complex formatting.

1. Open up Tableau, and connect to the acf_data_tab2016 worksheet.
2. As you can see, there are 5 worksheets in the workbook.
3. Drag the nonpivot (2) table to the sheets workspace. If Tableau cannot interpret the data, the Data Interpreter icon will appear [Figure 14].
Figure 14: Data Interpreter.

4. Click **Use Data Interpreter**.

5. Figure 15 shows the results of the reformatted data. If you notice, the merged columns at the top have been collapsed and the columns have been renamed.
6. Click on Review Results to see how Tableau parsed the data [Figure 16].
Figure 16: Color-coded Data Interpreter report.

Exploring the Tableau Interface

Worksheet: SNAP just_county data

I want to briefly explore the Tableau interface so that you have a better idea of what you are looking at when you add data to the workspace.

If you are not already connected to this data, click on the Data Source tab.

1. Add the just_county and snap_benefits data to the worksheet.
2. Make sure the join works.
3. Double-click on Sheet 1 and rename it to usda_snap.

Dimensions and Measures

Worksheet: SNAP just_county data

When you connect to a data source, Tableau assigns each field in the data source as playing one of two possible data roles: (1) dimensions or (2) measures.

If we think of Tableau data in terms of variables, the dimensions are discrete/categorical variables, and the measures are continuous variables.
Dimensions
Tableau considers discrete/categorical information (for example, fields where the values are strings or Boolean values) as dimensions.

1. Double-click on the **County Name** dimension, or drag it to the Rows shelf [Figure 17].

![Figure 17: Representing a dimension (Grantee) in Tableau.](image)

Catagorical data

2. Tableau treats any field you drag on to a shelf as a "pill."
3. A pills background color indicates whether it is **discrete (blue)** or **continuous (green)**.
4. You can change the field to a measure just by clicking the field and choosing **Measure** from the dropdown menu.
5. Change the **County_Name** variable to a measure by clicking on the dropdown menu and then choosing **Grantee>Measure>Count** [Figure 18].
6. As illustrated in Figure 19, the view will now contain a **continuous axis** instead of column or row headers, and the field's background will become green.

7. Also, CNT has been added to the variable name in the shelf.

8. If I hover over the bar in the graph, the count of counties (3,196) is displayed.
Figure 19: County variable as a measure.

Your options are limited when converting a dimension to a measure. For the **County name** variable, you can have either the count or the unique count, meaning you cannot have a min/max for a categorical variable.

Change **County** back to a dimension.

**Measures**

Tableau assigns any fields that contain quantitative, numerical information (that is, fields where the values are numbers) to the **Measures** area in the **Data** pane.

1. Rename the **PRGBEN06** to **Program Benefits 06**.
2. Move the **Program Benefits 06 variable** to the **Columns** shelf [Figure 20].

As you can see in Figure 20, when you drag a **Measures** variable to **Rows** or **Columns**, Tableau creates a **continuous axis**.
Figure 20: Adding a measure to a graph.

3. If you need to change any properties of a variable, you can also use the variable drop-down menu (Figure 21).
Figure 21: Converting a variable to a measure.

Just remember that the methods used to represent your variables (discrete or continuous) needs to be informed by your research question, or by the type of analysis you are performing.

Menu and Toolbar

At the top of the screen are a series of menu options and the toolbar. We will be diving into these functions later, I simply wanted to give you a quick overview of the interface.

You can also access training resources from the Help menu.

The Toolbar is located right below the Menu bar, and gives you access to common functions.

Important icons on the main menu are [Figure 22]:

1. Undo
2. Clear sheet
3. Menu>Analysis
4. Menu>Format
Figure 22: Common Tableau analysis/formatting tools.

Data Pane
At the far left of the screen are two panels: (1) the Data panel and (2) the Analytics panel.

The Data Pane lists all of the dimensions and measures in your workbook.

If you click on the View Data icon [Figure 23], you can open up the data associated with the worksheet.
Figure 23: Viewing the data from the Data Panel.

You can also use the panel to select values and copy them to Excel.

Ctrl+mouse to select individual rows.

Shift+mouse to select multiple rows.

Analytics Pane

The Analytics pane provides quick and easy access to common analytic features in Tableau.

For example, you can drag reference lines, box plots, trend lines forecasts, and other items into your view from the Analytics pane [Figure 24].

As I previously mentioned, only the appropriate options will be highlighted in dark grey [Figure 24].

Double-click on the Analytics option to add it to the chart. Or, you can drag it to the chart to choose custom options.
Figure 24: Analytics pane.

**Shelves and Cards**

**Worksheet: SNAP_just_county data**

If not already on the worksheet, move the **Program Benefits 06** variable to the **Columns** shelf, and the **County Name** dimension, or drag it to the **Rows** shelf.

**Pages Shelf**

The **Pages** shelf lets you break a view into a series of pages so you can better analyze how a specific field affects the rest of the data in a view.

When you place a dimension on the **Pages** shelf, you are adding a new row for each member in the dimension.

The **Pages** shelf creates a set of pages with a different view on each page. Each view is based on a member of the field you placed on the **Pages** shelf. You can easily flip through the views and compare them on a common axis.

**Filters Shelf**

The **Filters** shelf allows you to specify which data to include and exclude.

1. Add the **State Cd** dimension to the **Filters** shelf, and choose **AL**, **LA**, and **MS** as the states to filter [Figure 25].
2. Choose **Apply**, and then **OK**.
Figure 25: Adding a variable to the Filters shelf – choose the attribute.

3. The worksheet now displays only the Program Benefits 06 in Alabama, Louisiana, and Mississippi [Figure 26].
4. Data options are accessed by clicking anywhere on the x-axis [Figure 27].
5. There are three options:
   a. Sort ascending
   b. Sort descending
   c. View data
6. Click on the View Data icon.
7. There are two data views: (1) summary view, and (2) full data view [Figure 28].
8. You can export the data from either of these tabs.
Figure 28: Data view options.

9. Close the Data View popup box.

Want to make the filter more dynamic? No problem.

1. Click **Show Filter** from the Filter pulldown menu [Figure 29].
2. This adds a filter to the worksheet [Figure 30], which gives you the ability to quickly filter the data.
3. If you hover over the filter, a dropdown menu will appear in the upper left hand side of the filter [Figure 31]. This menu provides you with a number of options for formatting the filter.

4. Change the filter to **Multiple Values (dropdown)** [Figure 31].
5. As you can see in Figure 32, this gives us more room on the worksheet.
Figure 32: Multiple Values filter in Tableau.

**Marks Card**

The Marks card is where you control mark properties. It contains the mark type selector along with controls for Color, Size, Label, Detail, Tooltip, Shape, and Angle. The controls that are available depend on the mark type.

We will explore Marks cards in the second webinar.

**Managing Tableau Data**

1. Select Clear Sheet, to clear the Tableau workspace.
2. Double-click on the **County Name** variable to add it to the Rows shelf [Figure 33].
3. Double-click on all of the **Prgnum** variables: **Prgnum06**, **Prgnum07**, and **Prgnum10** [Figure 33].

As a reminder, these variables estimate the total annual amount of SNAP/FSP program participants by state or county residents in a year.

As you can see in Figure 33, when you add multiple measures, the variable is automatically added to the Filters card/shelf.
4. Change each of the *Prgnum* variables, so that they are easier to understand (Program Numbers 2006, 2007, and 2010) [Figure 34].

As you can see in Figure 34, each time I change the variable name...the Measure Values and the Column Headers are updated in the worksheet.
Adding, duplicating, and renaming worksheets

I have already shown you how to rename a worksheet.

However, now I would like to explore more worksheet options.

1. First, if you hover over a worksheet, Tableau will give you a preview of the data or the graph.
2. If you right-click on a worksheet, you will see all of the available worksheet options [Figure 35].
3. Let us go ahead and duplicate this worksheet.
4. Rename the worksheet `snap_chart`.
5. Open up the **Show Me** tab [Figure 36].
6. Choose **Horizontal Bars** as the option [Figure 36].
7. We are going to cover visualization in another webinar. However, I wanted to point out a couple of things about the **Show Me** tab.
   a. Horizontal Bar is highlighted as a recommended chart.
   b. A visualization tip is provided at the bottom of the **Show Me** tab. In this example, we need at least 1 measure to make a horizontal bar chart.
   c. The only charts that are displayed are the ones that are appropriate for your data.
8. Figure 37 displays the reformatted data.
9. While we are on this screen, I would like to show you two very useful tools. The first is the *Swap* function (Ctrl+w) [Figure 38].

10. The options in the *Show/Hide* card menu (near the *Fit* tool) allow you to turn on/off various cards [Figure 38]. You are probably noticing that Tableau gives you many options for exploring the same functions.

11. As a reminder, I can always right-click near the bottom of the worksheet, to reset all of the cards.
I just showed how to turn off various options (like the cards/shelfs), now I would like to show you how to turn on particular elements.

By default, the title, caption, and summary statistics are turned off. However, you can turn them on by going to the Worksheet menu. Title and caption are self-explanatory and I do not use them that often. However, I would like to show you the **Show Summary** and the **View Toolbar** tools.

1. Select **Worksheet> Show Summary** [Figure 39].
Figure 39: Displaying summary statistics for a worksheet.

2. Figure 40 shows the summary statistics for each of the three quantitative variables.
Figure 40: Summary statistics for the SNAP data.

3. You can change the display (types of statistics displayed) for this tab by clicking on the small arrow near the right top of the tab [Figure 41].
Figure 41: Changing the summary stats that are displayed.

4. Click **Hide Card** to remove this tab [Figure 41].

**Saving Your Work**

Tableau provides three ways for you to save your work:

1. **Workbooks**: Saves all open worksheets.
2. **Packaged Workbooks**: Saves the workbook along with all referenced local file data sources and images into a single file.
3. **Bookmarks**: Saves the current worksheet.

You can share workbooks and bookmarks, provided users can access the relevant data sources. If users do not have access to the data you used in your analysis, you can save your work as a packaged workbook.

Custom fields such as binned measures, calculated fields, groups, and sets are saved with workbooks and bookmarks

**Exporting Data**

There are three methods for exporting workbook data:

1. Exporting as an Image
2. Exporting as Data
3. Exporting as Crosstab to Excel

Exporting as an Image
The export image command saves the current view as an image file. You can export to an image file with the following three steps.

1. Select Worksheet > Export > Image [Figure 42].
2. In the Copy Image dialog box, select the contents you want to include in the image and the legend layout (if the view contains a legend).
3. Click Save.
4. In the Save Image dialog box, navigate to where you want to save the image file and type a file name into the text box. Select a file format from the Save as type drop-down menu.
5. Click Save.

![Figure 42: Exporting a worksheet as an image.](image)

You can also publish one or more views to PDF by selecting File > Print to PDF.

When printing a sheet, filters in the view are not included. To show filters, create a dashboard containing the sheet and print the dashboard to PDF.

Exporting as Data
Select Worksheet > Export > Data.

Creates a connection to an Access database.
Exporting to Excel (as crosstab)

Select **Worksheet > Export > Crosstab to Excel**.

Tableau automatically pastes a crosstab version of the current view into a new Excel workbook. This option automatically opens a new instance of the Excel application.

As you can see in Figure 43, this process is similar to using the Copy feature available from the Data View pop-up menu.

![Figure 43: Exporting summary data to Excel.](image)