

## MIS0855: Data Science

### In-Class Exercise on Feb. 2-4 – Getting Familiar with Tableau

**Objective:** Learn the basics of Tableau by visualizing a sample data set

#### Learning Outcomes:

- Create textual and graphical visualizations in Tableau.
- Combine data from multiple tables to create visualizations.
- Create calculated fields to categorize data.
- Create a dashboard to view multiple visualizations at once.

This tutorial will take you through the basics of Tableau. You'll be working with the USDA Food Access Research Atlas from the US Department of Agriculture. This is an extensive set of data, by US county, regarding access to grocery stores, convenience stores, restaurants, and socioeconomic and health data.

This data set was accessed through data.gov and hosted on the US Department of Agriculture web site (see <http://catalog.data.gov/dataset/food-access-research-atlas>).

Make sure you are working on a computer with Tableau Desktop installed. The classroom laptops have Tableau installed on them. If you're working with your own computer, make sure you've followed the instructions on the Community Site to install Tableau.

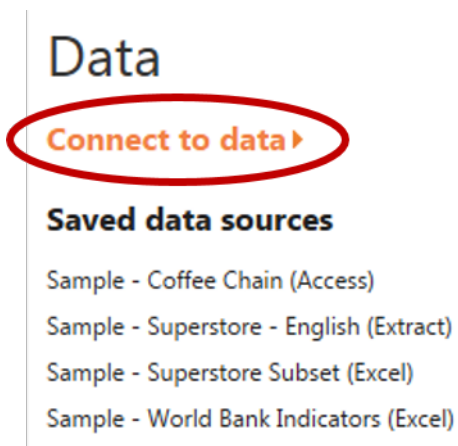
#### Part 1: Download the data file and start Tableau.

- 1) Go to the Community Site and look for the post with this in-class exercise.
- 2) Right-click on the link to the data set (FoodAtlas.xlsx) and save it to your computer.  
Remember where you saved it!

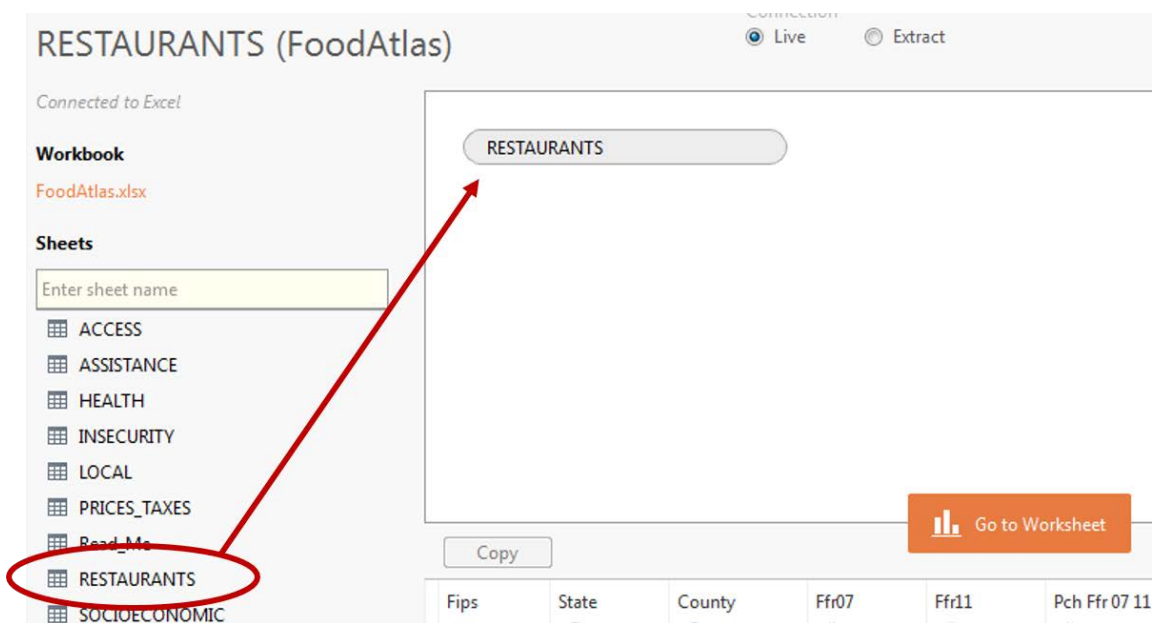
You can open the file in Excel and take a look through it. You'll see that there are a number of worksheets. Each one is a separate table. The Data Dictionary is in the tab "Variable\_List"

- 3) Start Tableau Desktop. If you're using Windows 7, find it on the Start Menu. If you're using Windows 8, it will be an icon on the Start screen, or you can just search for "Tableau" from the Start screen.

- 4) Click on “Connect to data”



- 5) Click on “Microsoft Excel” under “In a file.”
- 6) Navigate to the location where your data file (FoodAtlas.xlsx) is stored and select it.
- 7) You’ll see a list of Excel worksheets at the left side of your screen. These are all the sheets contained within the workbook. Drag the RESTAURANTS sheet to the workspace:



- 8) Wait for the data to be loaded at the bottom and click Go to Worksheet.

## Part 2: Working with Data: Rows, Columns, and Marks

1) Drag the “State” Dimension to the Rows shelf. You’ll see this:

The screenshot shows the Tableau interface with the 'State' dimension on the Rows shelf. The Marks card is set to 'Automatic'. The data table is as follows:

State	
AK	Abc
AL	Abc
AR	Abc
AZ	Abc
CA	Abc
CO	Abc
CT	Abc
DC	Abc
DE	Abc
FL	Abc
GA	Abc
HI	Abc
IA	Abc

2) Now drag “Ffrpth11” (the number of fast food restaurants per 1000 residents as of 2011) measure to the Text area within Marks (how values will be displayed). You’ll now see this:

The screenshot shows the Tableau interface with the 'Ffrpth11' measure on the Text area of the Marks shelf. The data table is as follows:

State	
AK	8.5
AL	37.5
AR	37.4
AZ	8.2
CA	37.3
CO	45.8
CT	4.3
DC	1.2
DE	2.0
FL	33.2
GA	84.7
HI	3.6
IA	48.5

- 3) Notice that Tableau guesses you want to add together all the values for each state (SUM). There are many entries in this table for each state – one for each county. But since our measure is a ratio, we rather want to average these numbers, not add them.

So right-click SUM(Ffrpth11) and select Measure > Average. You'll now see the average of all the values for that state:

The screenshot shows the Tableau interface with the following components:

- Columns:** Empty
- Rows:** State
- Marks:** AVG(Ffrpth11)
- Table:** A table with two columns: State and AVG(Ffrpth11). The data is as follows:

State	AVG(Ffrpth11)
AK	0.3546
AL	0.5598
AR	0.4989
AZ	0.5450
CA	0.6424
CO	0.7048
CT	0.5347
DC	1.2476
DE	0.6656
FL	0.4959
GA	0.5326
HI	0.7157

- 4) Then drag the “County” dimension to the Rows shelf, next to State. You'll now see a breakdown by county. If you see a warning dialog, click “Add all members.”

Scroll down to PA and find Philadelphia County. We seem to have a high ratio of fast food restaurants.

The screenshot shows the Tableau interface with the following components:

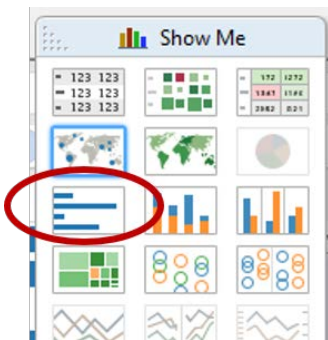
- Columns:** Empty
- Rows:** State, County
- Marks:** AVG(Ffrpth11)
- Table:** A table with three columns: State, County, and AVG(Ffrpth11). The data is as follows:

State	County	AVG(Ffrpth11)
PA	Northampton	0.853
	Northumberland	0.486
	Perry	0.456
	Philadelphia	0.851
	Pike	0.246
	Potter	0.229
	Schuylkill	0.481
	Snyder	0.578
	Somerset	0.439
	Sullivan	0.772
Susquehanna	0.324	
Tioga	0.377	

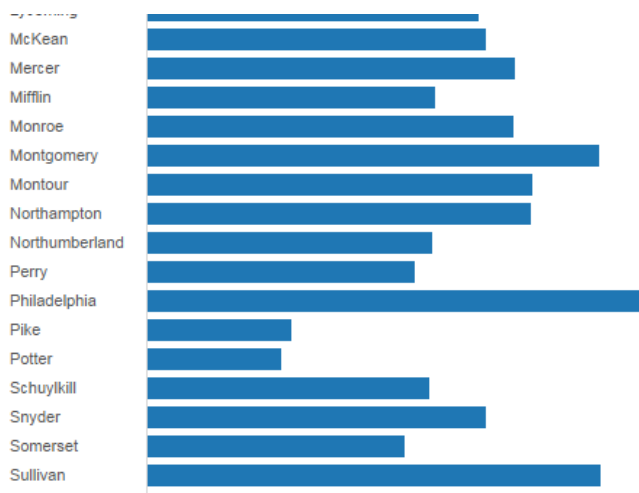
### Part 3: Creating Charts and Maps

So let's really see how Philadelphia compares with the rest of the state. First, we'll filter our data to only look at Pennsylvania.

- 1) Right-click "State" in the Rows shelf, then Filter...
- 2) In the Filter[State] dialog, click the "None" box, then scroll down and select only PA. Then click "OK."
- 3) Nothing much has changed, except that now we only see data for PA counties. And all the data is difficult to compare. It's difficult to tell if Philadelphia County really has the highest number of fast food restaurants per 1000 residents in the state. So now let's try some visualizations to make the data easier to see.
- 4) In the "Show Me" tab in the upper right corner, click the bar chart icon (third row, first column):

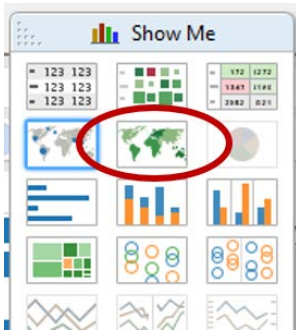


- 5) Now you'll see a bar chart, like this (you'll also notice that AVG(Ffrpth11) has moved from Marks area to the Columns shelf):

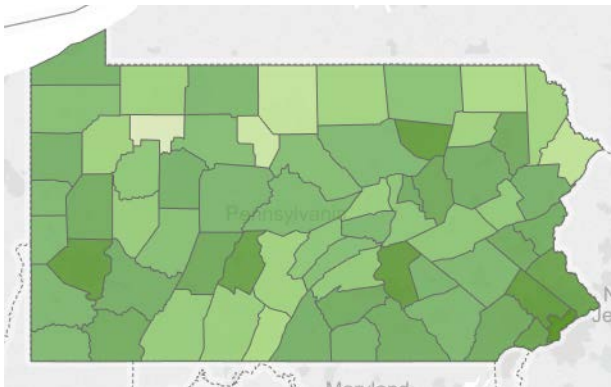


Scroll down the list and you'll see that Philadelphia really is the champion when it comes to the most fast food restaurants per 1000 residents. But it's still cumbersome to navigate such a complex bar chart.

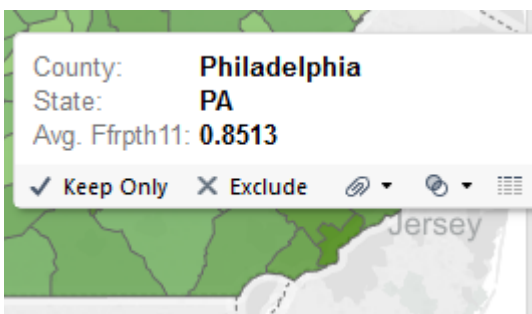
6) So now click on the filled map icon in the "Show Me" tab (second row, second column):



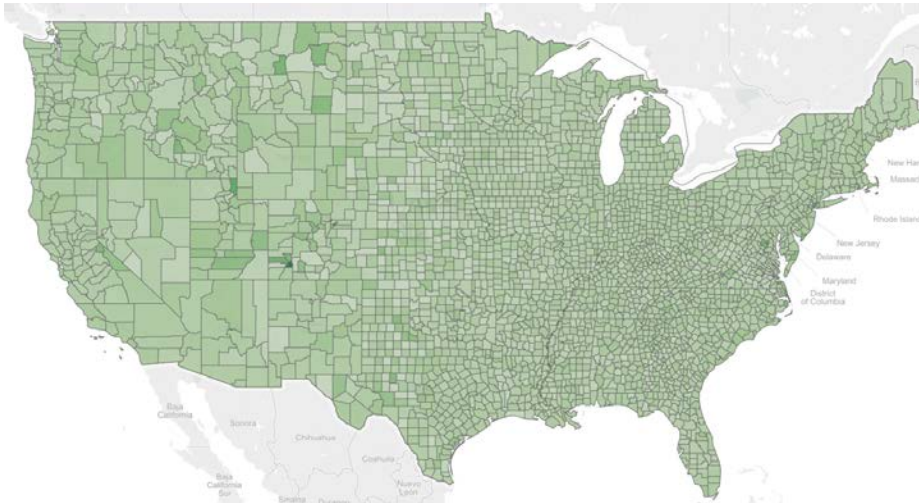
and after a while, you'll see a shaded map of Pennsylvania by county, where the darker shaded areas mean higher values. Notice that it also changed the values in Rows and Columns to Longitude and Latitude, and State, County, and AVG(Ffpth11) are now in the Marks area.



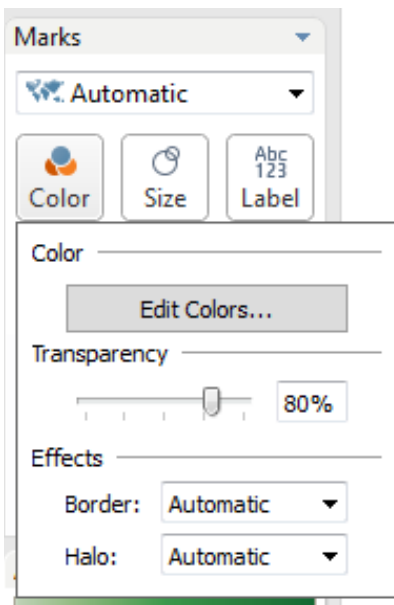
From this, Philadelphia clearly is among the highest values for fast food restaurant density in the state. Just to be sure, you can run your mouse over the other similarly-shaded counties and see the exact values:



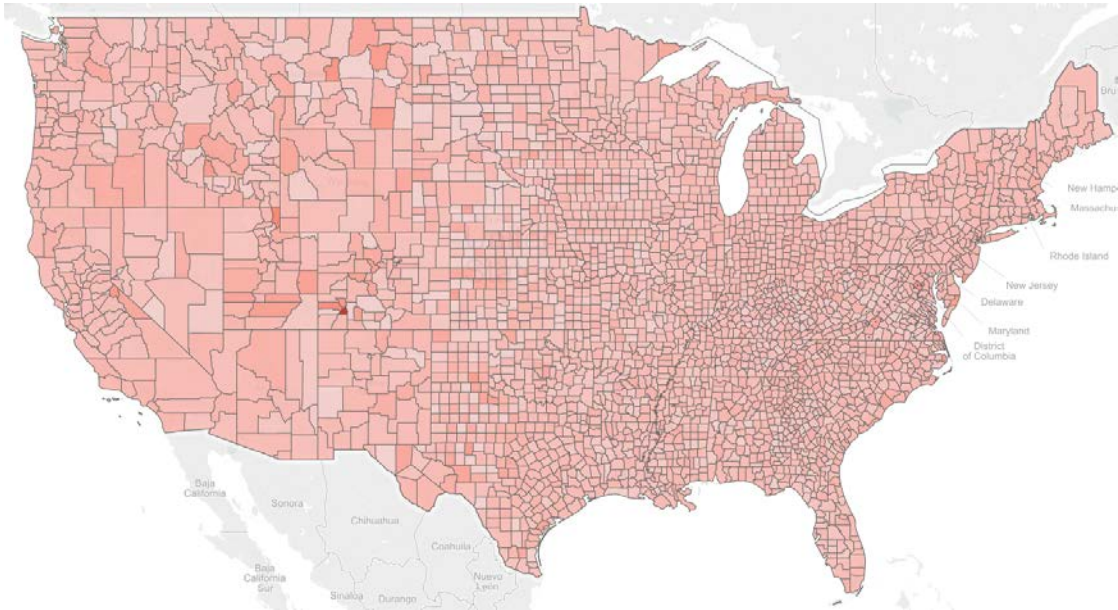
- 7) Now let's create a map of the entire continental United States. Right-click on "State" (now under the Marks area) and select "Filter..."
- 8) Click "All" and then unselect AK, HI, and PR, so that the map will show 48 states in the continental U.S. You'll now see a map like this:



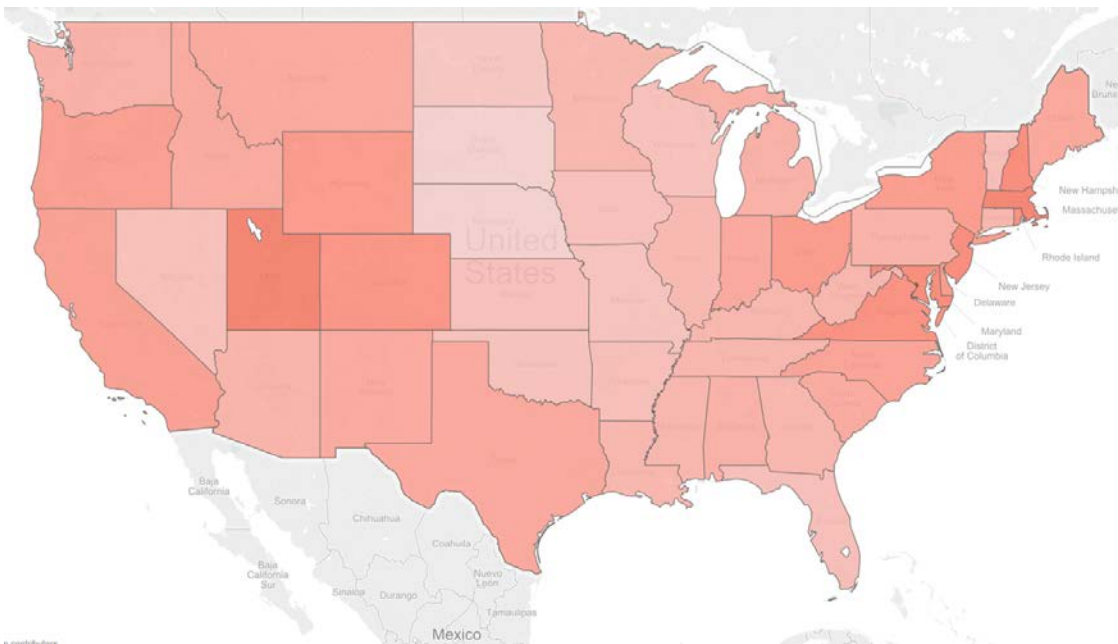
- 9) Let's change the color to make it easier to read. Click "Color" under Marks and then "Edit Colors..."



10) Choose “Area Red” and click OK. You’ll now see this map:



11) Now let’s view the map by states, not counties. Right-click “County” within the Marks area and select “Remove.” You’ll now see this:



12) Finally, rename the worksheet you’ve created by double-clicking on the “Sheet 1” tab at the bottom of the Tableau window (or right-click and choose “Rename Sheet”) and typing “Fast Food Density by State”



13) This would be a good time to save your Tableau workbook (File/Save). Save the file in the same folder as you put the Excel spreadsheet.

#### Part 4: Analyzing Data across Multiple Tables (Worksheets)

Sometimes you'll have related data in more than one table (or in the case of Excel, in more than one worksheet). Tableau will let you combine that data using something called a "join."

1) Right-click on the RESTAURANTS entry under "Data" at the top left of the screen.



Select Edit Data Source...

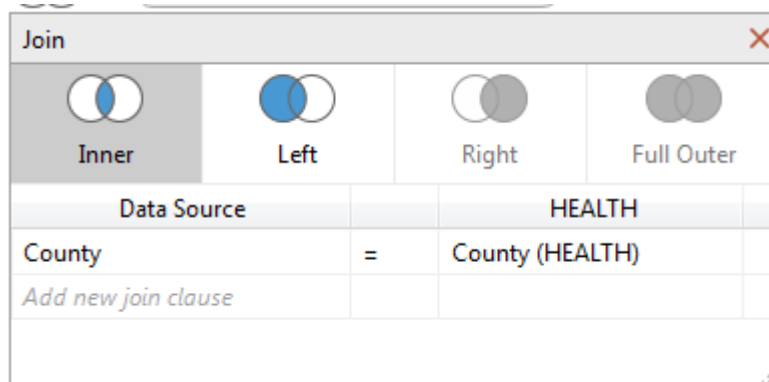


2) Drag the HEALTH worksheet to the whitespace area next to RESTAURANTS. The HEALTH table has data like obesity rates and the number of recreation facilities, also recorded by county.

Now click the join symbol:




3) When you do, you'll see a dialog that looks like this:



It guesses (correctly) that you want to connect County in RESTAURANTS with County in HEALTH. In other words, it wants to associate data across those two tables based on county.

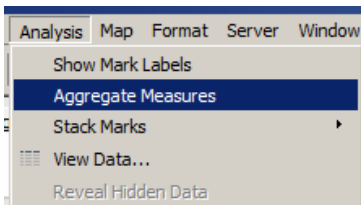
- 4) When you can see the data, click “Go to Worksheet.” You’ll see new Dimensions and Measures labeled HEALTH\$ to reflect data from the new table! The data from before is listed under RESTAURANTS\$.



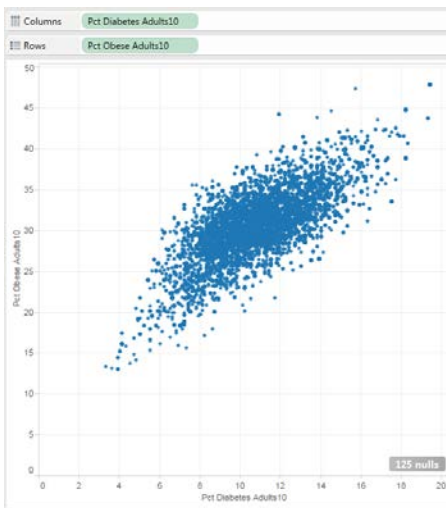
- 5) Create a new Tableau worksheet by clicking on the (  ) icon at the bottom of the screen.

- 6) Let’s look for a relationship between Diabetes rates and Obesity rates. Under measures, drag “Pct Obese Adults10” (the percentage of obese adults in 2010, by county) to the Rows shelf, and drag “Pct Diabetes Adults10” (the percentage of adults with diabetes in 2010, by county) to the Columns shelf.

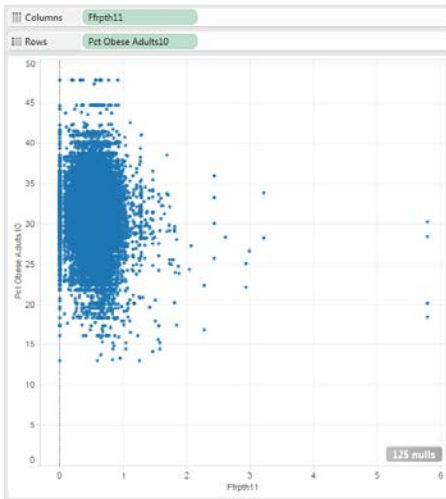
- 7) Now go to the Analysis menu and de-select “Aggregate Measures.” This tells Tableau to plot every data point (county) separately, and not add them together or average them into one number.



You’ll see a scatterplot that clearly shows a positive relationship between obesity and diabetes:



- 8) Now let’s combine data across tables. Right-click “Pct Diabetes Adults10” from Columns and click Remove. Drag “Ffrpth11” to Columns Remember, that’s the number of fast food restaurants per 1000 residents from our original RESTAURANTS table. Now you get this scatterplot:



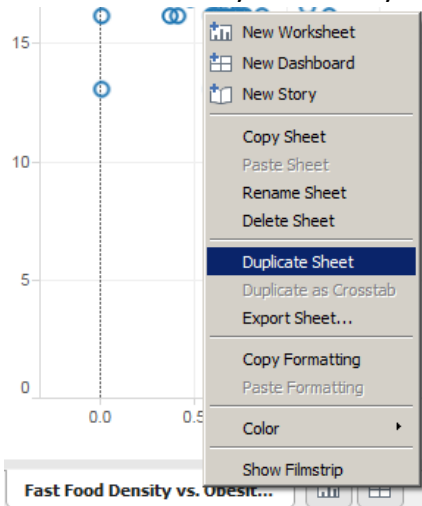
This gives us the somewhat surprising result that the density of fast food restaurants seem to have no relationship with obesity rates!

- 9) Change the name of this worksheet to “Fast Food Density vs. Obesity Rates”

### Part 5: Create a calculated field

You can create additional data fields based on the ones that are already there. This can be useful if you want to create additional categorizations, or if you need to perform new calculations.

- 1) First, let’s create a new worksheet, based on the one we just created. Right click on the “Fast Food Density vs. Obesity Rates” tab and select “Duplicate Sheet.”



“Sheet 3” will be created.

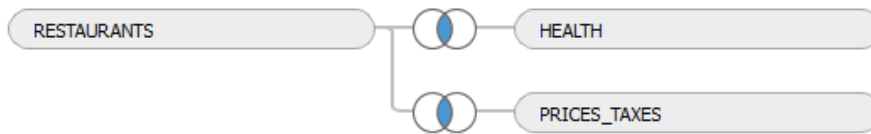
2) Now remove Ffrpth11 from the Columns shelf and replace it with “Pct Diabetes Adults10.” (Drag this to Columns.) You’ll see the same scatter plot in Step 7 of the last section.

3) Right-click on the RESTAURANTS entry under “Data” at the top left of the screen and choose “Edit Data Source...”

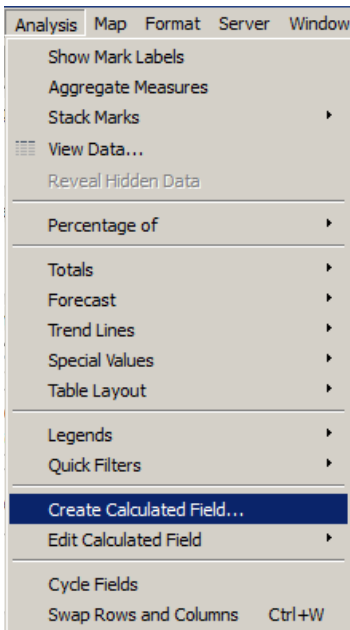


4) Drag the PRICES\_TAXES sheet to the whitespace below HEALTH and RESTAURANTS. This sheet has information about the relative prices of milk, soda, and tax rates attached to junk food.

5) You’ll see it created the join for you. Click “Go to Worksheet”



6) Go to the Analysis Menu and select “Create Calculated Field.”

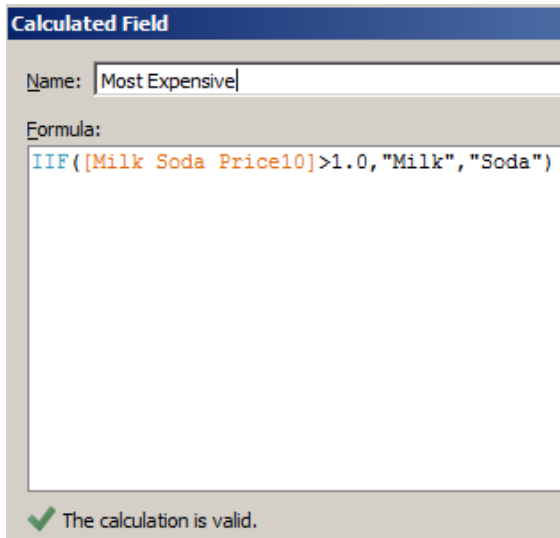



- 7) We'd like to know if the relative price of milk to soda has any relationship with obesity and diabetes rates. We have a field in PRICES\_TAXES called "Milk Soda Price10," which is the price of milk divided by the price of soda, by county, in 2010.

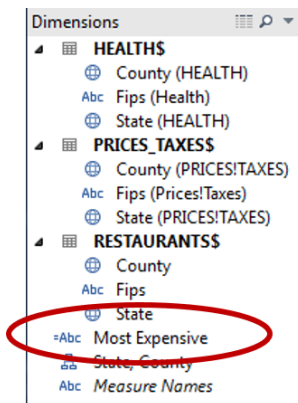
In the dialog box, type "Most Expensive" for Name, and the following formula for Formula:

```
IIF([Milk Soda Price10]>1.0,"Milk","Soda")
```

**NOTE: IIF is NOT a typo. Make sure you've got two "I"s in there!**

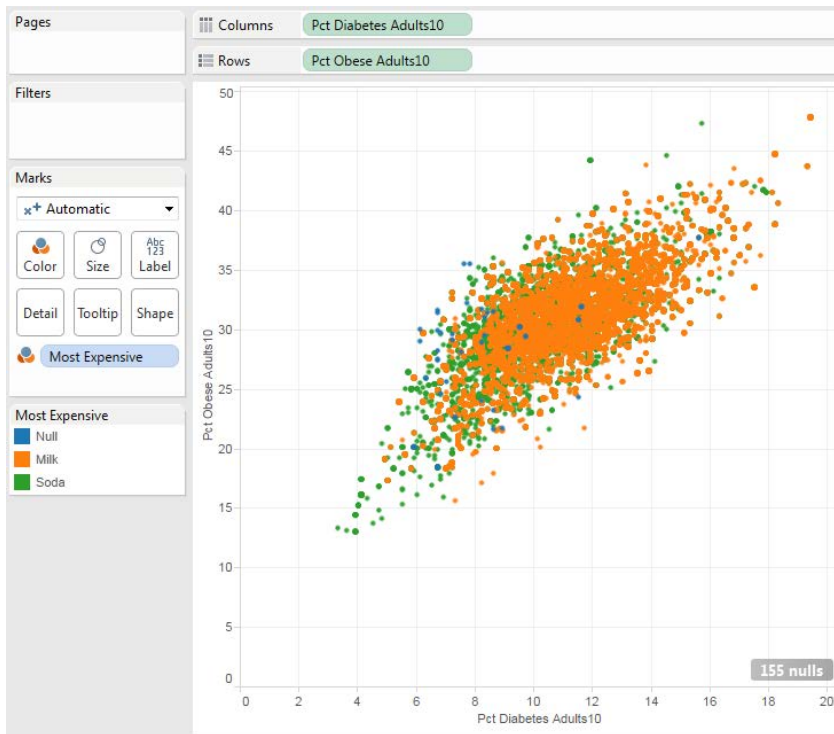


- 8) If you typed it correctly, you'll see  The calculation is valid. under the Formula text box. Then click "OK". You'll see the new field appear under Dimensions:



- 9) Drag your new "Most Expensive" dimension to the "Color" button of the Marks area.

10) It will now color code your scatterplot:



Orange dots represent counties where Milk is more expensive, and Green dots represent counties where Soda is more expensive. The blue dots represent counties where this figure wasn't available, so let's get rid of that.

11) Right-click on "Most Expensive" under the Marks area and select "Filter..." Unselect "Null" and click "OK."

12) The plot will now look a little cleaner:



13) We learn a few interesting things from this plot:

In most counties, milk is more expensive than soda.

Where milk is more expensive, obesity and diabetes rates tend to be higher. The orange dots tend to be at the upper right of the plot.



Where soda is more expensive, obesity and diabetes rates tend to be lower. The green dots tend to be at the lower left of the plot.

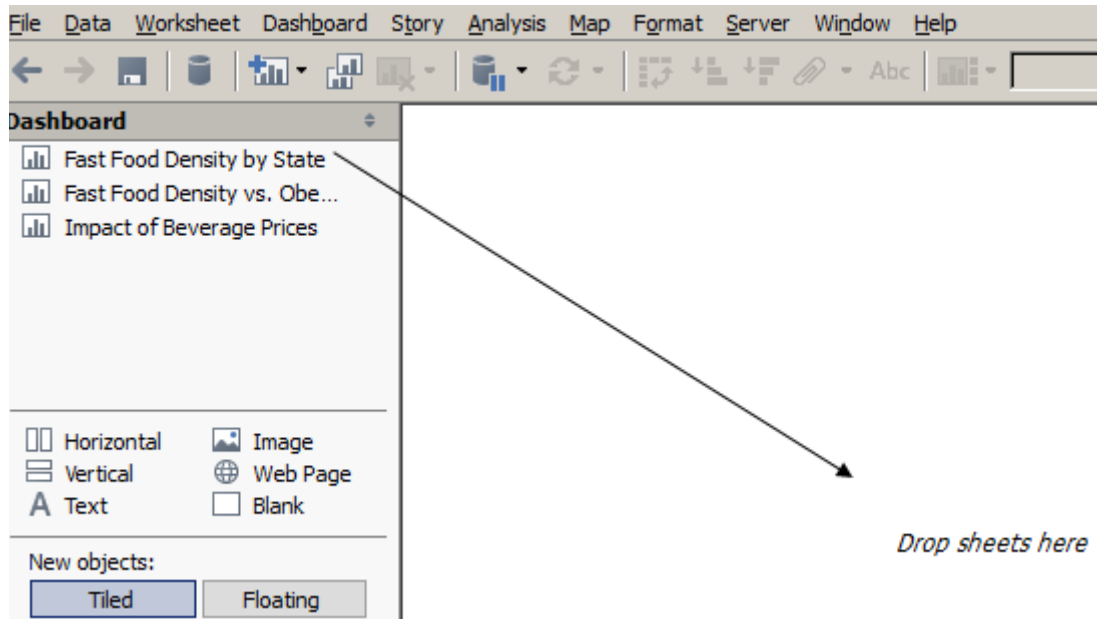
14) Rename "Sheet 3" to "Impact of Beverage Prices"

15) Save the Tableau workbook.

## Part 6: Create a dashboard

A dashboard is a collection of worksheets, arranged so that they are easy to view all at once.

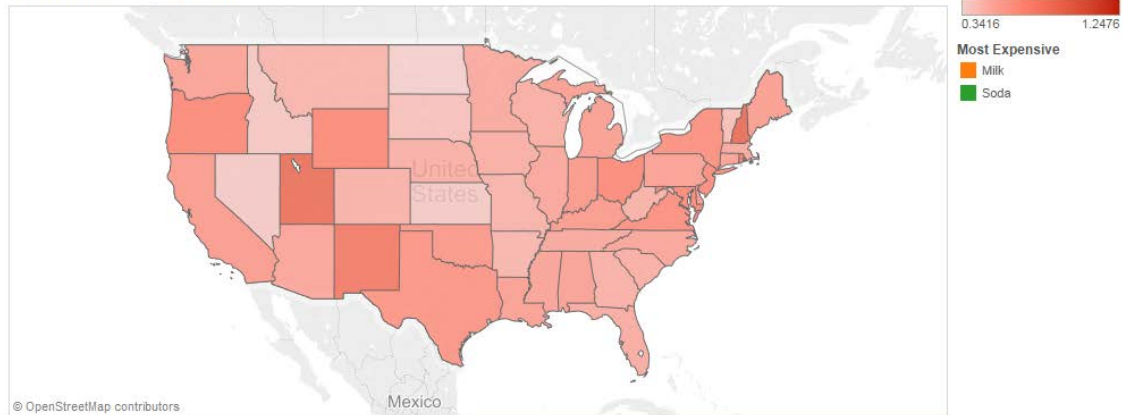
- 1) Click on the  tab at the bottom of the screen to create a new dashboard. Not this  for a new worksheet.
- 2) Drag the “Fast Food Density by State” worksheet (under Dashboard) to the “Drop sheets here” whitespace.



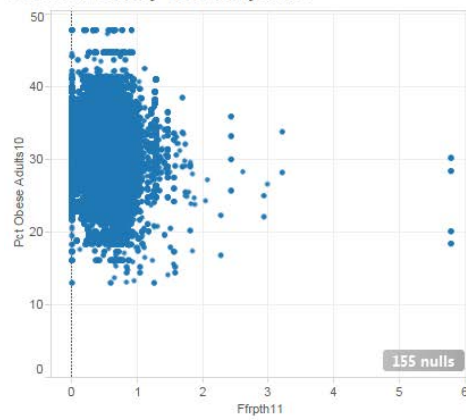
- 3) Now drag the “Fast Food Density vs. Obesity Rates” worksheet to the bottom half of the workspace (the map). When you’re in the right spot, you’ll see the bottom half of the map shaded light gray.
- 4) Now drag the “Impact of Beverage Prices” worksheet to the bottom right corner of the workspace (the right side of the scatterplot). It will look like this:



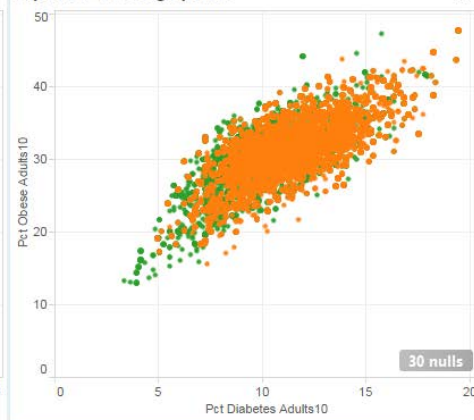
Fast Food Density by State



Fast Food Density vs. Obesity Rates



Impact of beverage prices



5) Rename the “Dashboard 1” tab to whatever you’d like and then save your workbook.

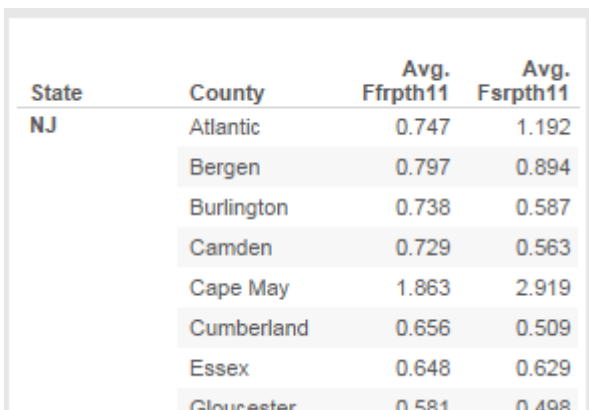
**Part 7: Now try it on your own!**

- 1) Create a new Tableau worksheet. Create a filled map showing 2010 adult obesity rates by county. Exclude Alaska (AK), Hawaii (HI), and Puerto Rico (PR).

**HINT: Remember to average percentages, not sum them.**

- 2) Create a new Tableau worksheet. Create a text table that compares the percentage of fast food restaurants per 1000 people (Ffrpth11) with the percentage of full-service restaurants per 1000 people (Fsrpth11). Limit your results to counties in New Jersey.

The table should look something like this:



State	County	Avg. Ffrpth11	Avg. Fsrpth11
NJ	Atlantic	0.747	1.192
	Bergen	0.797	0.894
	Burlington	0.738	0.587
	Camden	0.729	0.563
	Cape May	1.863	2.919
	Cumberland	0.656	0.509
	Essex	0.648	0.629
	Gloucester	0.581	0.498

**HINT: Drag your dimensions into Rows and your measures to columns. Then change the chart type to “text tables.”**

- 3) Modify the table you created in #2 to color code the text in the table. Counties where there are more fast food restaurants than full service restaurants should be one color, and counties where there are more full service restaurants than fast food restaurants should be another color.

**HINT: You’ll need to create calculated field to compare AVG(Ffrpth11) to AVG(Fsrpth11).**