



INTRODUCTION TO TABLEAU PUBLIC (10.4/10.8)

Making Charts and Maps

ABSTRACT

This lesson covers several data organizing, analysis, and visualization techniques. It works with data from the US Census bureau.

Ayanda Masilela

CSSCR Workshops (2018)

TO USE THIS DOCUMENT

The document is arranged in sections, which can be overviewed in the Table of Contents. It is subsequently broken down as an outline.

Descriptions and purposes are listed in black

Instructions for execution are highlighted in purple

Questions are highlighted in burnt orange

YOU WILL NEED

- Tableau Public Desktop 10.4, 10.8
- 2018.1 may have some minor differences

TABLE OF CONTENTS

3.....	Getting Started
3...	Connecting to Datasets
3...	Setting Data Types
5...	Joining Datasets
5.....	Working in Sheets
5...	Bar Charts, Scatter Plots, and Pie Charts
5...	Bar Charts
6...	Scatter Plots
7...	Creating Pie Charts with Categorical Data
7...	Creating Pie Charts with Numerical Data in Columns
9...	Creating a Map (+1 additional layer)
10.....	Creating a Story

3

GETTING STARTED

1. Connecting to Datasets

- a. Tableau uses the language of “connection” rather than “opening” because edits and manipulations made within Tableau do not modify the source file
- b. Click “Connect to File” and then “Text File”
 - i. In the new window, click the drop menu at the bottom and select “All Files *.*”
 - ii. Navigate to the Com_S0802_165yr.csv spreadsheet
 1. .csv stands for “Comma Separated Value”
 2. Essentially, it is a series of words and data entries separated by commas
 3. It is a highly simplified file type
 4. Excel spreadsheets can also be imported in a similar way
 - iii. Tableau integrates .csv files easily
 1. If all went well, you should see a series of labels along the white top of a spreadsheet, and rows of data represented in alternating stripes within the spreadsheet
 2. Grab the scroll bar at the bottom of the spreadsheet, and scroll to the right – make sure that everything looks good
 - iv. In the gray column to the left that says “Columns”, click the “Add” link
 1. Add PopD.csv just like you did the first dataset
 - a. When it is done, look to your left in the gray column that says “Files”
 - i. This is a listing of all the data sources that are available for querying and arrangement in this instance of Tableau
 - b. Double-click Com_S0802_165yr, and reduce the name to “S0802” – from here, the dataset will be referred by this reduced title

2. Setting data types

- a. If the S0802 dataset isn’t already in the blank white area above the spreadsheet, double-click it in the file list to add it
 - i. Look at the white bar above the spreadsheet that contains the column headers
 1. **Abc** means “string” or text
 2. **#** means “number” – can be decimal or whole
 3. The calendar means “date”
 4. (not shown) The clock and calendar mean “date and time”
 5. (not shown)The **T|F** means “Boolean”
 - ii. Drag the S0820 box out of the top white area
 - iii. Drag PopD.csv into the
 1. State and County have globes
 - a. Tableau is able to read some geographic information
 - b. Click on the globe that says “County” and pan to the bottom of the list where it says “Geographic Role”
 - i. “County” should be highlighted

- c. Do the same for state
 - i. The more geographic information we have, the easier the task will be for Tableau to correctly place our data
 - 2. But didn't S0820 also have geographic data?
 - a. Yes, but it was not formatted in a way that Tableau can read it
 - i. Scales of geographic data (country→state→county→city) must be stored in separate columns
 - ii. S0820 stores all of this information in one column, so Tableau treats it as a category, but nothing more
- 3. Joining Datasets
 - a. S0802 is a census dataset that was downloaded with the American Fact Finder
 - i. It has been greatly reduced for the purposes of this demonstration – nearly 800 columns for analysis
 - b. The population data in PopD.csv was derived from the US Census DP03 dataset
 - c. The goal for the next stage is to include overall population data, as well as taking a few measures of population density
 - i. We'll need to combine the datasets in order to make this possible
 - ii. This is called a **Join**
 - d. Pan over the list of datasets, and click on the miniature spreadsheet next to the file name of S0802. Do the same for PopD
 - i. Do you see any similar columns across which we can conduct a new Union?
 - e. Types of Joins
 - i. **Inner:** the resulting table keeps all matching values on both sides – it trims data that does not have matches
 - ii. **Left:** Keeps all records on the left side, regardless of whether it has matches
 - 1. Unmatched left-side results receive null values
 - iii. **Right:** Keeps all records on the right side, regardless of whether it has matches
 - 1. Unmatched right-side results receive null values
 - iv. **Full outer:** Keeps all values from both sides
 - f. A **Join** is not a **Union**
 - i. A union melds datasets and increases the number of case entries. Ideally, the melding datasets will have a similar structure and number of columns
 - 1. We will do this at another time
 - g. Click and drag S0802 to the blank space above the spreadsheet
 - h. Next, click and drag PopD.csv
 - i. A box with blue and white overlapping circle will appear
 - ii. **Inner Join** is already selected
 - iii. Click "Add New Join Clause" on the left side, and select "GEOid2"
 - iv. Click the right side, and select "Fips"

1. Once you have selected the two, the spreadsheet window should automatically populate with the newly joined datasets
2. Scroll to the right to make sure the data matches

WORKING IN SHEETS

1. Click “Sheet 1”, which is highlighted in orange
 - a. You will see a large white canvas with several “Drop field here” indicators
 - b. Double-click the Sheet 1 tab, and name it “Bar Charts”
 - c. Look to the left at the white column with a “Data” tab on top
 - i. This is an outline of all your data
 1. Categorical and ordinal data are sorted under **Dimensions**
 2. Continuous data are sorted under **Measures**
 - d. At the top of the screen, click Analysis → Create Calculated Field
 - i. This will open a new window
 - ii. Click where it says “Calculation 1” and type in “Pop Density”
 - iii. Look to the left in the column that says **Measures**
 1. Click and hold “Population” – it will turn into a green pill
 - a. Drag the pill into the calculation window
 - b. Type in the “/” division operator
 - c. Click and hold Sq Mi and drag it into the calculation window
 2. There should be a message at the bottom of the window that the calculation is valid – click “OK”
2. Bar Charts, Scatter Plots, and Pie Charts
 - a. Tableau’s engine attempts to read your data, and will recommend specific visualization types
 - i. As you pan over the different panels, observe the data recommendations at the bottom of the window
 - b. At the top right, click **Show Me**, and observe how your options for representation change
 - c. Drag “4-wheel Alone” and “Pop Density” into the **Rows** bar at the top of the window
 - d. Drag “County” into the **Columns** bar at the top of the window
 - e. You can pan over the specific bars to get more information about the actual values
 - i. Which county has the highest population density, and what is that number?

 - ii. Which county has the highest count of commuters who drive to work in their own car? _____
 - iii. Why may this not be the most descriptive measure? How can we look at this measure differently? _____
 - f. Go to Analysis → Create Calculated Field
 - i. Repeat the steps as in part 1-d, using “4-wheel alone” as the numerator and “Workers 16+” as the denominator

- ii. Set the title to “%Solo_Drivers”
 - iii. Add the new “%Solo_Drivers” measure to the **Rows** bar
 - iv. How do the results of this measure differ from the initial two bar charts?
-

- g. Add a new sheet and name it “Regional Summations”
 - i. From the **Dimensions** pane drag “Counties” into the **Columns** row
 - ii. Drag “Carpool %Female” and “Carpool %Male” into the rows column
 - 1. The resulting table displays a summation by region
 - 2. There are multiple counties within each region
 - iii. But there’s a problem – it is taking the sum total of percentages, which doesn’t make sense for this analysis
 - 1. Mouse over to the right side of the green “Carpool %Female” pill and click the drop arrow
 - a. Pan down to **Measure** → **Average**
 - b. The chart will now display the mean percentage of female people who carpool per region
 - c. You can change the measure to appropriately match the arrangement of your data
- h. At the bottom-left side of the window, click “New Worksheet”
 - i. Rename it “Scatterplot”
 - ii. Drag “County” into the **Marks** box next to the white **Measures** column
 - iii. Go to Analysis → Create Calculated Field
 - 1. Repeat the steps as in part 1-d, using “Workers 16+” as the numerator and “Sq_Mi as the denominator”
 - 2. Name the newly calculated field “Worker Density”
 - 3. Repeat these same steps with “Carpool” as the numerator and “Wordkers 16+” as the denominator
 - 4. Name the new field “%Carpool”
 - iv. Drag “%Carpool” into the **Columns** bar and “Worker Density” into the **Rows** bar
 - v. Drag “County” into the **Marks** pane
 - vi. If Tableau does not automatically assign **Scatterplot**, activate the **Show Me** window and click the “Scatter Plots” panel
 - vii. Go to Analysis → Trend Lines → and activate **Show Trend Lines**
 - viii. Pan over the trend line. What does the P-value indicate? _____

- i. Creating Pie Charts with Categorical Data
 - i. Tableau is most capable of making pie charts with categorical data. We’ll use the regional data to demonstrate this section
 - ii. Make a new sheet, and name it “Cat Pie Chart”
 - iii. In the **Marks** pane, click the drop menu at the top and select **Pie**
 - iv. From the **Dimensions** pane, drag “Region” into the **Color** box

- v. Drag “Worker Density” into the **Angle** box
 - vi. Click **CTRL+SHIFT+B** a few times to zoom in if you need
 - vii. Which region has the highest density of working age people? The lowest?
-

j. Creating Pie Charts with Numerical Data in columns

- i. Given our data structure, Tableau does not perform well with the creation of proportional pie charts
- ii. Nonetheless, we can walk through these steps and get an idea of how a pie chart can be structured given these limitations
- iii. Open a new sheet and name it “Num Pie Chart”
- iv. In the box labeled **Marks**, click the drop menu item that says “Pie”
 - 1. In the box labeled **Dimensions**, click and drag “Measure Names” into the **Color** box inside the **Marks** box
 - a. It may ask you if you are sure that you want to use all names. Select **Add All Members**
 - 2. In the box labeled **Measures**, scroll to the bottom, select “Measure Values”, and drag it to the **Angle** box inside the **Marks** box
 - 3. You should now have a single pie chart with the totals of all 39 counties
 - 4. Tableau automatically populates all numerically valued entries, so we’ll have to select out the ones we want
 - a. In the **Measure Values** column, keep the following entries
 - i. 4-wheel alone
 - ii. Carpool
 - iii. Pubtrans
 - b. Delete all other entries
 - c. Hold **CTRL** and drag the blue “Measure Names” pill in the **Marks** box into the Label box
 - d. What is the most common form of transportation? _____
 - e. What is the total number of people who use public transit?

- 5. Click and drag “County” from the **Dimensions** box
 - a. You should now see a scrollable list of pie charts for the 39 counties
 - b. Click **CTRL+SHIFT+B** a few times to zoom in
 - c. Hold **CTRL** and drag the blue “Measure Names” pill in the **Marks** box into the Label box
 - d. You should now see the pie chart sections labeled with the categories listed in *Section 4a*
 - e. These are not so intuitive or attractive, so let’s change them – Right click, and click **Rename** in the drop menu
 - i. Change “4-Whell alone” to “Drove Alone”

- ii. Change “Carpool” to “Carpooled”
 - iii. Change “Pubtrans” to “Public Transit”
 - f. If you are dissatisfied with the color scheme that was automatically assigned, click the **Color** box inside the **Marks** box, and click **Edit Colors**
 - i. Which two counties have the greatest proportion of public transportation users? _____
 - g. Note that even though the display is proportional, when you pan over the pie slices, Tableau does not return a percentage
 - h. This function can however be adapted by inputting genuine percentages as your fields. This will be demonstrated in the following section
- 3. Creating a map (+1 additional layer)
 - a. Tableau is able to generate maps, as well as displays with multiple layers
 - b. Click “New Worksheet” and name it “%Solo Drivers”
 - i. From the **Dimensions** box, drag “County” into the large empty square at the middle of the sheet
 - 1. The map will probably generate a few point locations – these are county centroids
 - 2. In the **Show Me** menu, the map with circles (symbol map) of differing sizes may be selected
 - a. Click the map next to it that doesn’t have circles, but does have varying colors (filled map)
 - 3. It is possible that some counties are missing
 - a. This is normal, given that there are multiple counties that share names throughout the US
 - 4. Drag “State” and “Country” into the empty square, as well
 - a. The missing counties should now populate the map
 - 5. Hold the “Shift” key, which will give you an arrow indicator to pan around the map
 - ii. From the **Measures** box, drag “%Solo Drivers” into the **Color** box in the **Marks** menu
 - 1. The map should now display colors that fluctuate through varying shades of blue
 - 2. If you would like to use different colors click the **Color** box and then click the **Edit Colors** button
 - 3. Click the **Palette** drop menu – this shows you the pre-made options for color sequences
 - 4. Click the blue box on the left, and select a shade of green
 - 5. In the **Edit Colors** menu, click **Reversed**

- a. The greener the county, the lower the proportion of solo drivers are present
 - b. Click **Apply** and see if you like the color scheme
 - c. Click **OK**
 - d. You can click on an individual county to observe any of the data that is featured in the **Marks** box
 - e. Which counties appear to have the greatest percentage of solo drivers? The lowest percentage? _____
-
6. Now hold the “CTRL” key and click the green “Latitude” pill in the **Rows** bar at the top of the screen, and drag it to the right to duplicate it
 - a. You should now see two choropleth displays
 7. In the **Marks** box at the bottom, click the second “Latitude” entry, and remove the “%Solo drivers” pill
 8. In the **Show Me** box, make sure that “symbol map” is selected
 9. Drag “Measure Names” into the **Color** box
 10. Drag “Measure Values” into the **Angle** box
 11. Set the colors to a combination that won’t get lost in the color of the basemap
 - a. I did salmon and apricot
 12. In the right green “Latitude” pill, right-click and select **Dual Axis**
 - a. This will place your pie charts over their corresponding counties
 - b. They might be pretty small, so select the **Size** box in the **Marks** pane and use the slider to adjust the size of the pie charts to something legible
 - c. List a few counties where a higher proportion of solo drivers are men: _____
 - d. What proportion of solo drivers are women in Jefferson County? _____

10

CREATING A STORY

1. At the bottom window where you were creating new sheets, click on the book with the “+” sign and **Add New Story**
 - a. On the left, you should see a listing of all the charts and maps we made in the previous exercises
 - b. Double click the text that says “Story 1” at the top of the screen, and give your story an intuitive title
 - c. You can add a caption below the title to give the reader more context about your project
2. Drag “Bar Charts” into the blank box
3. Drag “Regional Summations” into the area to the right of the **Caption Box**
 - a. You should see two arrows indicating that it will be placed to the right of the first page
4. Click the **Layout** tab
 - a. Here you can choose how you want to set up your navigation
 - i. Captions – add text to describe each story page
 - ii. Numbers – list pages by page number
 - iii. Dots – each dot represents a page
 - iv. Arrows Only – shows page #/# and navigates with arrow buttons only
 - b. So far it doesn’t look like either page allows us much space to add descriptions about the context of the project, so click the **Blank** button in the **New Story Point** pane
 - i. Click **Drag to Add Text** and drag the text box to the top of the page
 1. Fill the text box with a brief introduction to the project
 2. Add another text box that describes the data source
 - c. Drag all of the other tables to the top, and style your navigation bar as you please
 - i. Add descriptions about each page
 - ii. Add descriptions about findings on each page
 - iii. When you are satisfied, save your project to Tableau
 1. Tableau saves your work in its online interface
 2. Since we are using the Public version, all of your projects will be available to other users