

TUTORIAL

# Get Started with Tableau Desktop

Learn how to connect to data, create data visualizations, present your findings, and share your insights with others.

This tutorial walks you through features and functions of Tableau Desktop. As you work through this tutorial, you will create three views in a Tableau workbook. The steps you'll take is based on a Dillard's Live data set from the University of Arkansas Enterprise System group in the Information Systems Department.

**Dataset Overview:**

Table Name	Time Frame	Rows	Attributes	Size (GB)
CUSTOMER	n/a	83,513,612	7	8.037
DEPARTMENT	n/a	938	6	0.001
SKU_STORE	n/a	520,148,909	4	19.969
SKU	n/a	29,594,797	10	3.239
STORE	n/a	411	6	0.001
TRANSACTION	10/17/2016 to 7/20/2019	425,589,589	15	42.340
<b>10 TABLES</b>			<b>48</b>	<b>137.593</b>

Platform data is currently available: Yes

The dataset has a total of 1,058,848,256 records and 48 attributes distributed among 6 tables.

This tutorial shows you how to:

1. Connect directly to the WCOB\_DILLARDS\_LIVE dataset residing on the TERADATA system.
2. Prepare the dataset before working on it.
3. Create three visualizations based on an investigation of revenue generation.

**NOTE:** This tutorial assumes you have been granted access to the University of Arkansas VMWare platform. Requesting access can be done from the Enterprise Systems website at the University of Arkansas (<https://walton.uark.edu/enterprise/>). The data used should not be downloaded to your personal drives. All files should remain on the Remote Desktop S: drive provided by the University of Arkansas. This is due to our agreement with the data providers. Questions can be directed to Ron Freeze at [rfreeze@walton.uark.edu](mailto:rfreeze@walton.uark.edu).

<http://onlinehelp.tableau.com/current/guides/get-started-tutorial/en-us/get-started-tutorial-home.html>

**ESTIMATED COMPLETION TIME:** 20-40 minutes

# The Story

Suppose you are an employee for a large retail chain. You have just been hired as an area sales manager and are interested in understanding the sales amount for the area you have been assigned to as well as other areas of the company.

Initially, you are interested in visualization of the sales value of the Top 10 Zip Codes. Next, you want to take the insights gained in the first visualization and break it down to a state level. At this point, you should have a clear understanding of which zip codes and states generated the highest sales revenue. Finally, and using the previous insights, you want to narrow down your work by looking at the area you were assigned to that includes Arkansas, Kansas, Missouri, and Oklahoma.

In order to reach the desired visualizations, the following are the steps we will follow:

## 1. CONNECTING TO THE DATA ([here](#))

- Connect to **Teradata**.
- Connect to **WCOB\_DILLARDS\_LIVE** database
- Connect to the two tables: **TRANSACT** and **STORE**.

## 2. PREPARING THE DATA ([here](#))

- Use hide, create calculated field, and filter tools to reach this list of attributes:
  - Store (Table: TRANSACT)
  - Tran Date (Table: TRANSACT)
  - Tran Type (filter/Tran type = 'P') (Table: TRANSACT)
  - Tran Amt (Table: TRANSACT)
  - City (Table: STORE)
  - State (Table: STORE)
  - Zip Code (Table: STORE)

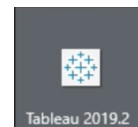
\* For further understanding of the attributes listed above, please refer to the [data dictionary](#) appended at the end of the document.

## 3. CREATE THE VISUALIZATIONS ([here](#))

1. A column chart of the top 10 Zip Codes.
2. A geographic view by state.
3. A geographic view by Zip Code of your area (Arkansas, Kansas, Missouri, and Oklahoma).


# Open Tableau Desktop and begin

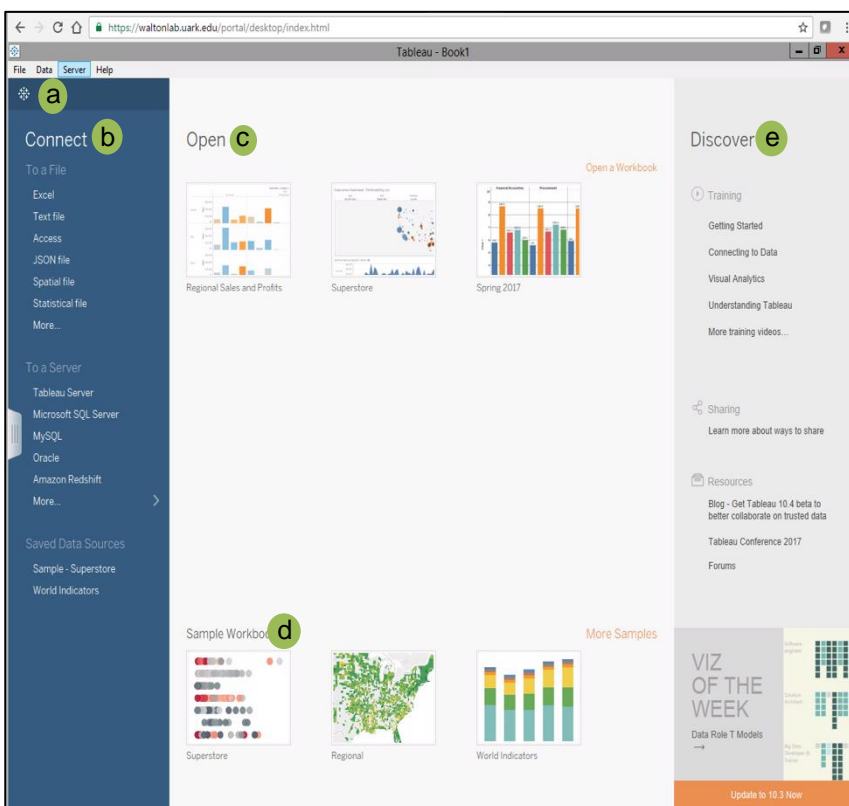
This tutorial assumes that you have logged in to University of Arkansas VMWare platform. Select the **Start Windows** icon, look for **Tableau 2019.2** and click on it. Wait 10 seconds until the application opens. If an update Tableau window pops up, exit this window.



The first thing you see after you open Tableau Desktop is the **start PAGE**. Here, you select the connection that you want to use - how you will connect to your data

The start page gives you several options to choose from:

- a. Click the Tableau icon in  the upper left-hand corner of any page to visit the start page at any time.
- b. Under **Connect**, you can:
  - o Connect to data that is stored in a file, such as Microsoft Excel or Access.
  - o Connect to data that is stored on a server, such as Tableau Server, Microsoft SQL Server, or Google Analytics.
  - o Connect to a data source that you've connected to before.



*Tableau supports the ability to connect to a wide variety of data stored in a wide variety of places. The **Connect** pane lists the most common places that you might want to connect to, or click the **More...** link to see more options. **More on connecting to data sources in the Learning Library (in the top menu).***

- c. Under **Open**, you can open workbooks that you have already created.
- d. Under **Sample Workbooks**, view sample dashboards and worksheets that come with Tableau Desktop.
- e. Under **Discover**, find additional resources like video tutorials, forums, or the “Viz of the week” to get ideas about what you can build.

# Connect to Teradata

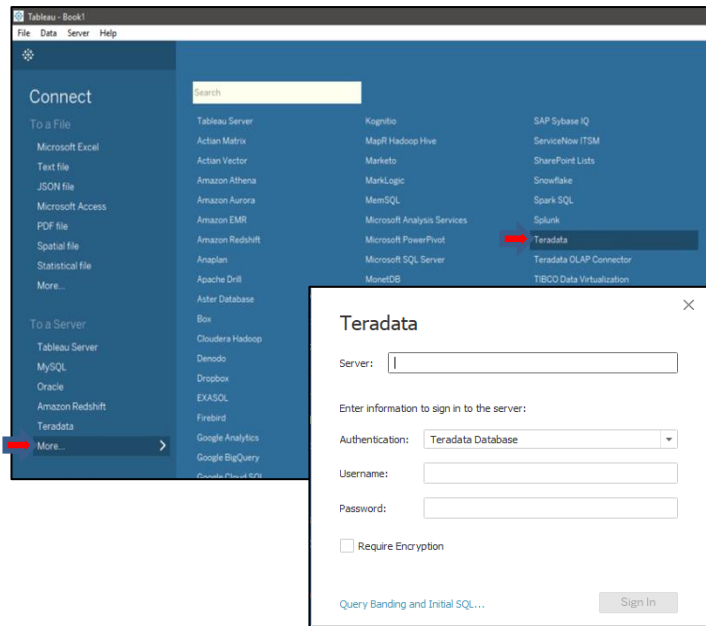
Under **Connect > To a Server**, select **More**. A new window will open, click on **Teradata** as shown in the figure alongside.

A window will ask you to enter the following information:

**Server:** 130.184.26.161

**Username:** *given to you by instructor*

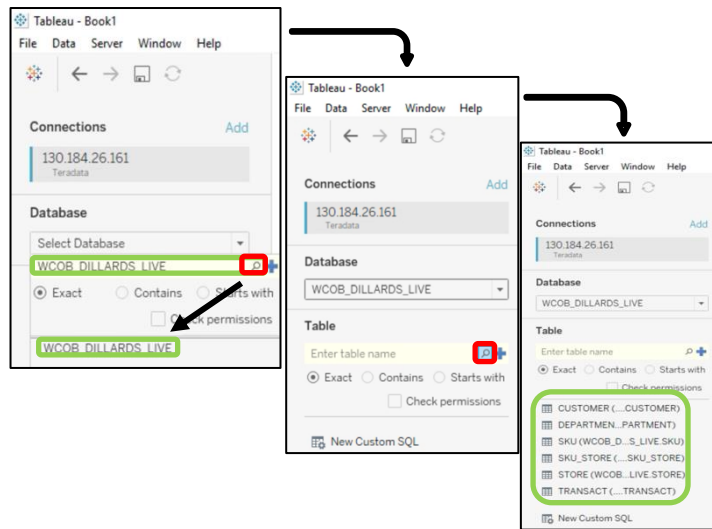
**Password:** *given to you by instructor*



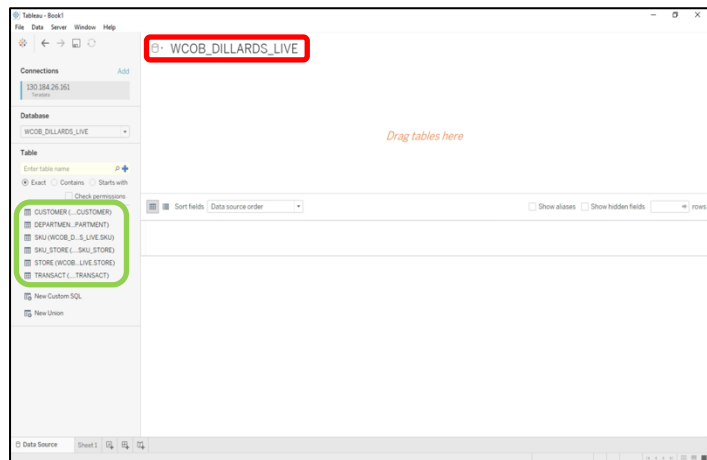
# Connect to WCOB\_DILLARDS\_LIVE

Now we have to access the **WCOB\_DILLARDS\_LIVE** database through the panel to the left:

1. Type “WCOB\_DILLARDS\_LIVE”, click on the search icon. A new “WCOB\_DILLARDS\_LIVE” will appear below, click on it.
2. Once on the database, click on the search icon.
3. Finally, tables from the database should appear on the panel to the left.



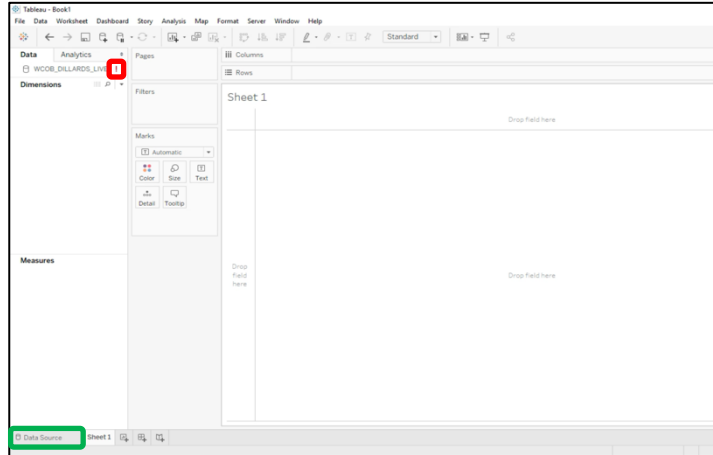
Your window should look like the picture on the right. You are connected to WCOB\_DILLARDS\_LIVE database (circled in red). The left panel shows the tables (circled in green) in the database.



At the bottom left corner, there are two tabs named **Data Source** and **Sheet 1**. Click on **Sheet 1** and a window similar to the picture below will appear.

**Sheet 1** does not currently have data. This is indicated by an exclamation point next to the database name (circled in red).

Click on the **Data Source** tab on the bottom left (circled in green). This will take you to the previous window where you connected to the Database and displayed the tables.

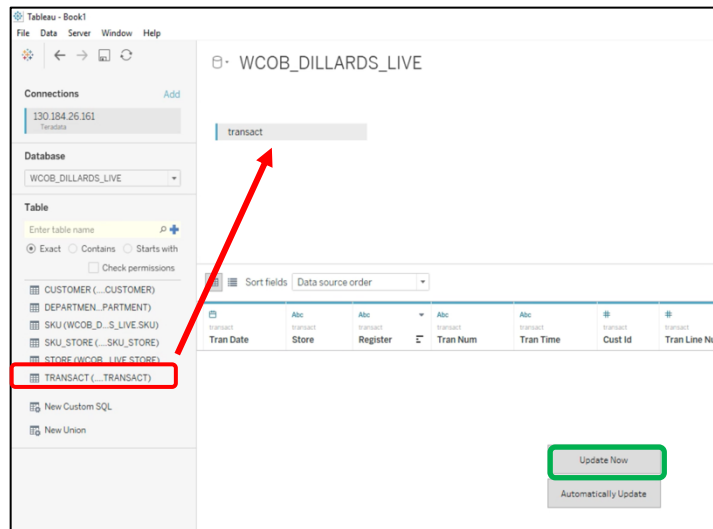


## Connect to TRANSACT table

On the Data Source window:

4. Drag and drop the **TRANSACT** table on the left panel to the blank space at the top where it says *“Drag tables here”*.
5. Click on **Update Now** (circled in green) and Tableau will fill the columns with data.
6. Click on **Sheet 1**.

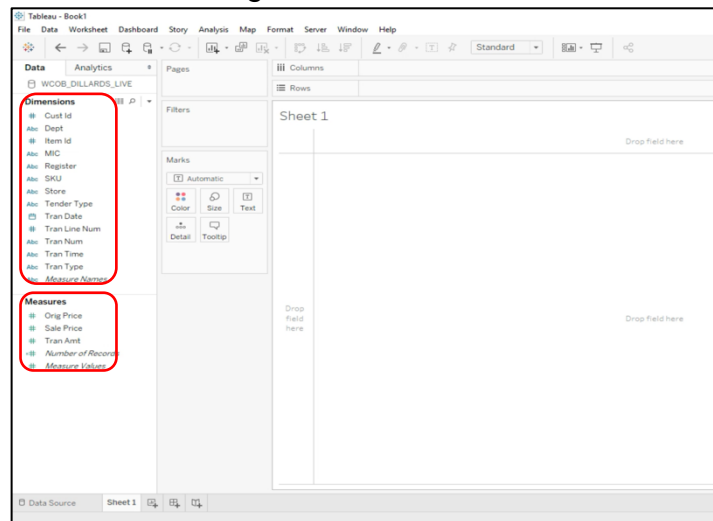
You will now note that you can see the table attributes assigned to Dimensions and **Measures**.



## As a recap:

After you connect to your data, Tableau does the following:

- Opens a new worksheet. This is a blank slate where you create your first view.
- Automatically assigns data types (such as date, number, string, etc.) and roles (dimension or measures) to your data. (We'll talk more about these terms later.)
- Adds columns from your data source to the Data pane on the left-hand side. Columns are added as fields.



If you want to review details of your data, you can select the Data Source tab in the bottom left-hand corner. Here, you can view the first 1,000 rows of your data. What happens in Tableau stays in Tableau; the data is safe.

When you connect to your own data, you might need to do some prep work before connecting to it in Tableau. This is because Tableau makes assumptions about your data so that it can display it properly. With the TRANSACT table added, we have the first three variables needed for our analysis: Store, Tran Date and Tran Type.

**Note:** If you navigated to Data Source to check out the details, just click on the tab for **Sheet 1** to get back to where you started.

## Add the STORE table

To obtain the variables for the remainder of the analysis, we need to add and connect the **STORE** table to the **TRANSACT** table. The connection is possible since both tables share a single variable attribute: **Store**.

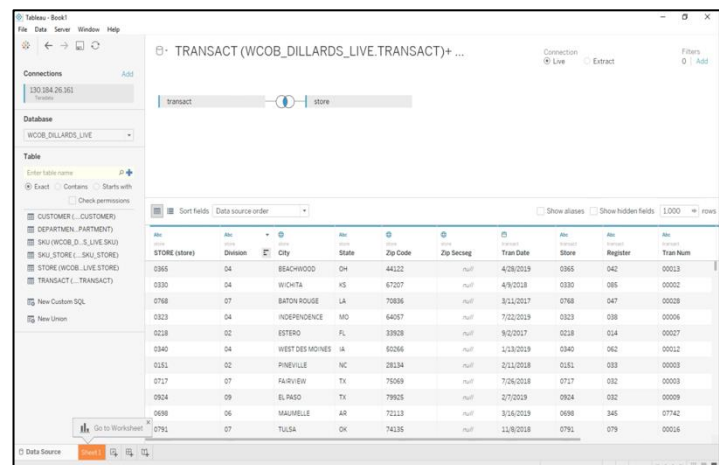
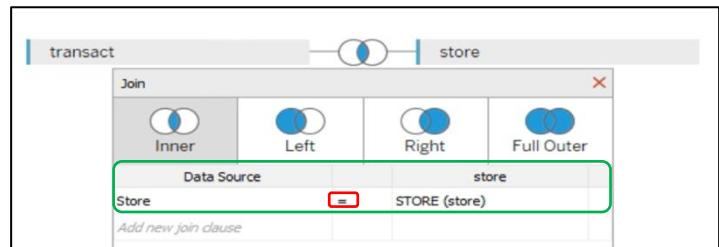
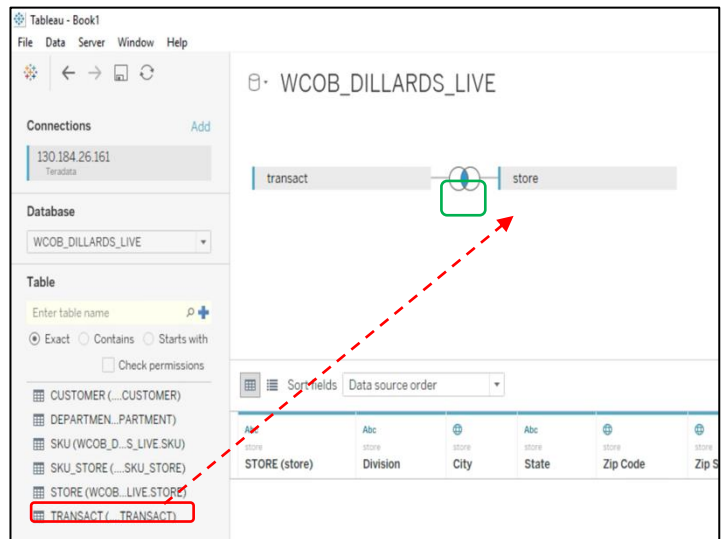
7. Return to the Data Source page.
8. Drag and drop **STORE** to the blank space like you did with **TRANSACT**.

The two intersecting circles (circled in green) mean that they were successfully connected. This is considered an inner join.

Right click on the join symbol. The popup shows all the different joins Tableau allows you to do. Here **TRANSACT** table is taken as 'Data Source' inner joining **STORE** table. Both have Store as the common attribute joined by an equal sign (circled in red).

9. Click on the **Update Now** button and you should have a window similar to the one alongside.

Now we have the attributes from both the **TRANSACT** table and the **STORE** table.

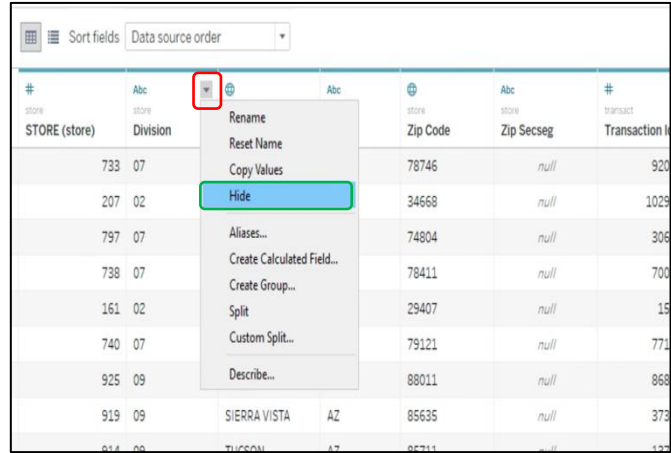


# Preparing the Data

## Hide Unneeded Fields and Create Calculated Fields

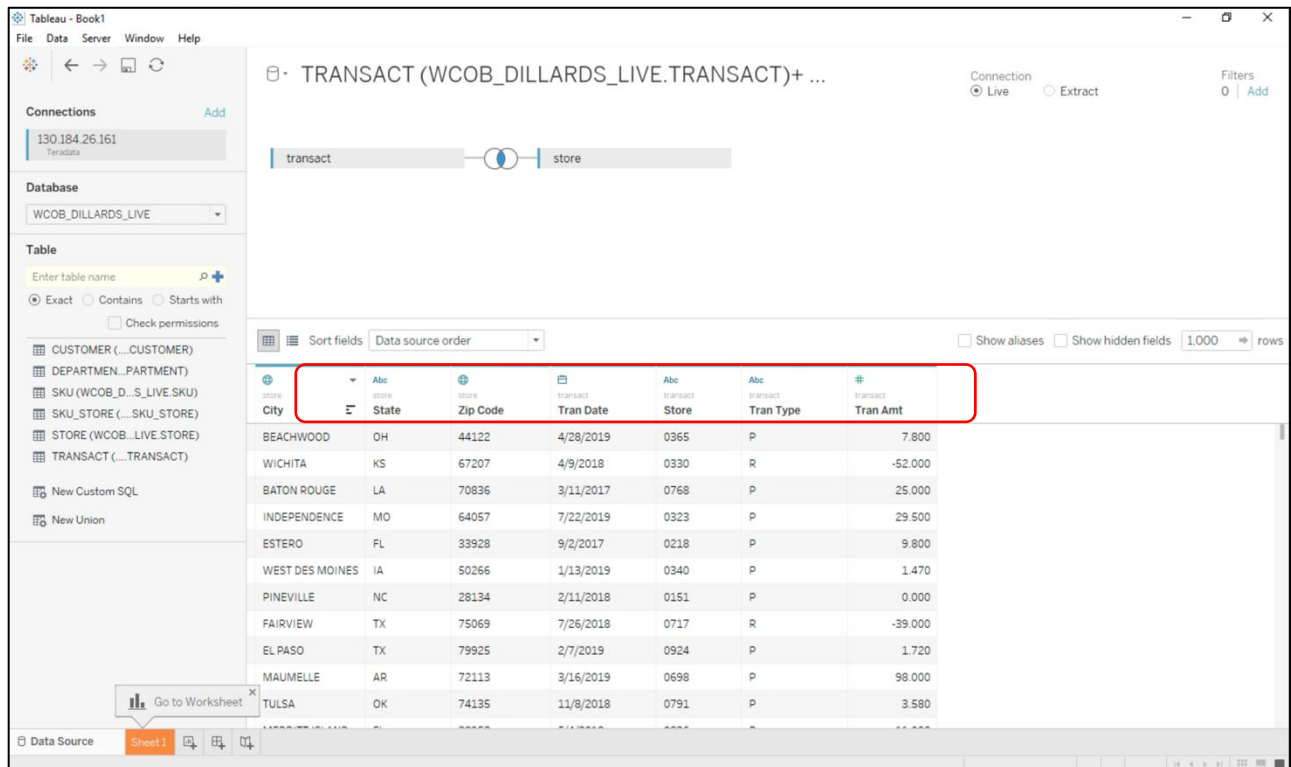
To get the list of attributes stated in [The Story](#) we need to hide the ones we don't need:

10. Click the down arrow at the top right corner of each attribute not needed in the analysis.
11. Click on Hide and the attribute will no longer appear in the data.



To undo this click the 'Show hidden fields' box at the far right of the window and manually unhide the attribute.

After hiding all the unnecessary attributes and keeping only the ones mentioned (**Tran Date, City, State, Zip Code, Store, Tran Type, and Tran Amt**) your screen should look like the following:



One more tool that can help you visualize your data before working with it is to order your data in ascending or descending order. For instance:

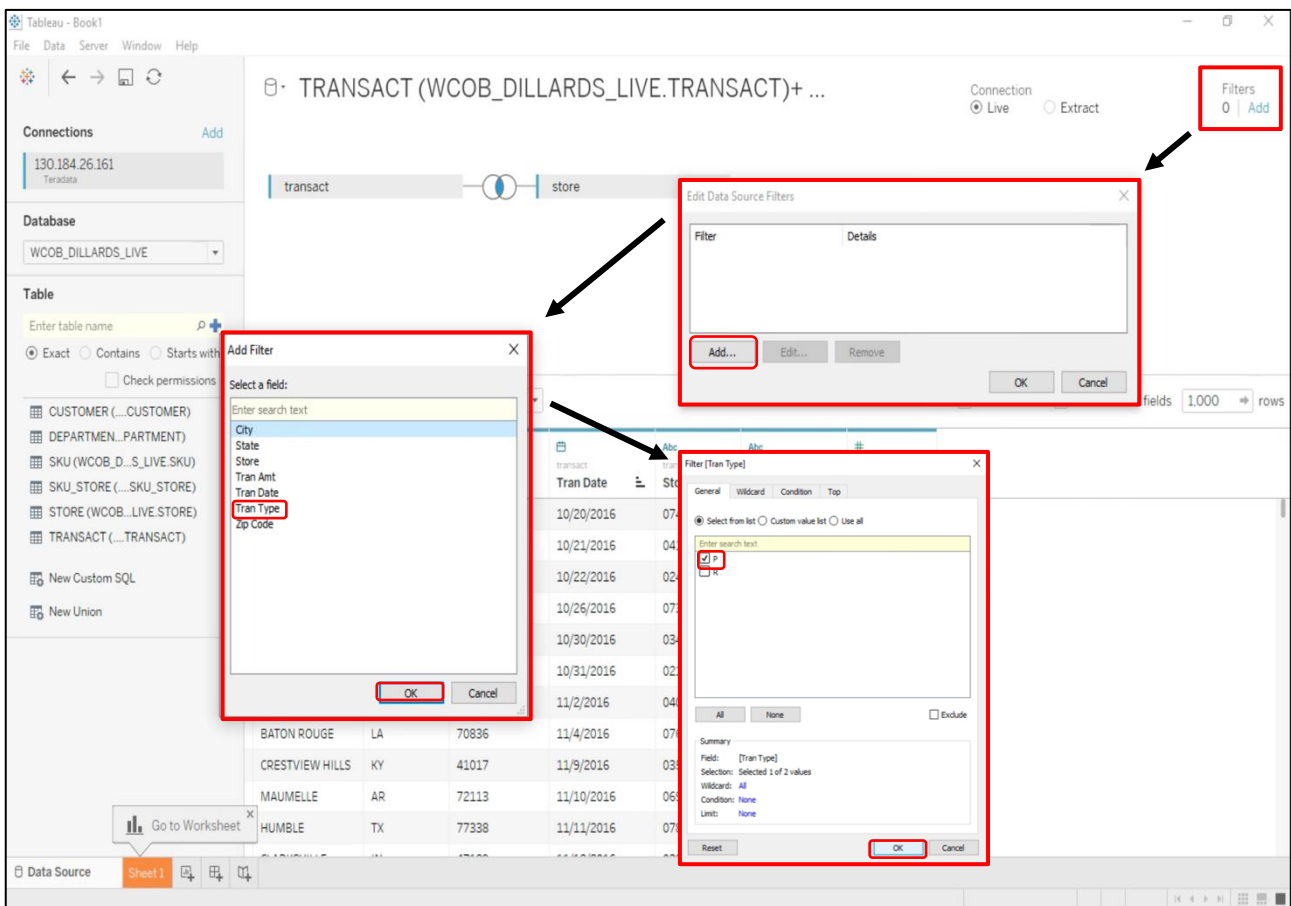
12. Click on the bar figure next to the name of the attribute (circled in red) and this will organize your column. The two figures below show how to do this.

City	State	Zip Code	Tran Date	Store	Tran Type	Tran Amt
BEACHWOOD	OH	44122	4/28/2019	0365	P	7.800
WICHITA	KS	67207	4/9/2018	0330	R	-52.000
BATON ROUGE	LA	70836	3/11/2017	0768	P	25.000
INDEPENDENCE	MO	64057	7/22/2019	0323	P	29.500
ESTERO	FL	33928	9/2/2017	0218	P	9.800

City	State	Zip Code	Tran Date	Store	Tran Type	Tran Amt
AMARILLO	TX	79121	10/20/2016	0740	P	22.500
MEMPHIS	TN	38117	10/21/2016	0410	P	76.800
TALLAHASSEE	FL	32301	10/22/2016	0243	P	48.000
TEMPLE	TX	765021803	10/26/2016	0731	R	-11.200
LINCOLN	NE	68505	10/30/2016	0344	P	24.000

## Filter Fields

Due to the size of the TRANSMACT table (425 million records), we need to reduce our focus in order to display the results more efficiently.



The screenshot shows the Tableau interface with the following elements highlighted:

- Filters Subtitle:** Located at the top right of the view, showing 'Filters 0 | Add'.
- Add Filter Dialog:** A dialog box where 'Tran Type' is selected from the list of fields.
- Edit Data Source Filters Dialog:** A dialog box with an 'Add...' button to create a new filter.
- Filter [Tran Type] Dialog:** A dialog box showing the selected values 'P' and 'R' for the Tran Type field.

On the picture above, at the top right corner, you can find a Filters subtitle:

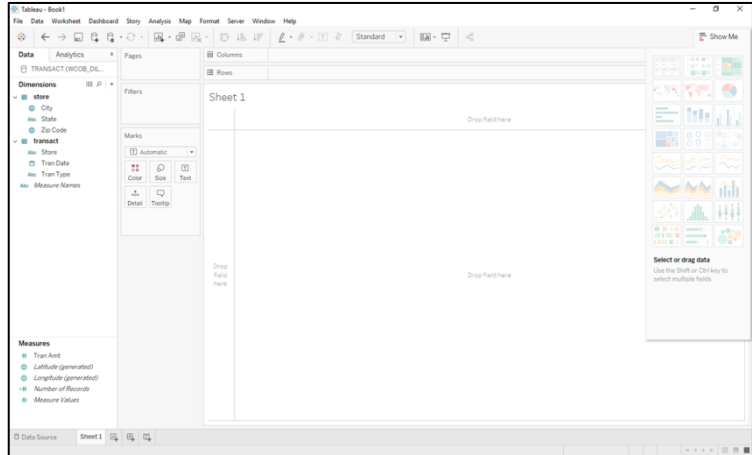
13. Click on **Add**. A new window will open, click on **Add..** again.

14. You are now in the 'Add Filter' window. Click on **Tran Type** and then **click OK**. A window will 'Filter [Tran Type]' will pop-up stating the two different values present in that field. In this case 'P' or 'R'.

15. Click on the **P** box (circled in red) to only work with **Tran Type = P** and click **OK**. The data now is updated for only **Tran Type = P**.

We have finished organizing our data:

16. Click on **Sheet 1**, your window should look like the picture alongside. →



## Create the Visualizations

### Top 10 Zip Codes

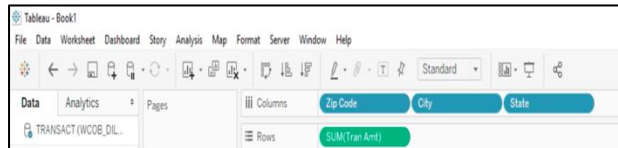
Create the column chart for the Top 10 Zip Codes in the United States.

The following are the parameters you wanted for the exploration of the sales value in your column chart.

- Top 10 Zip Codes in descending order
- For the year 2017
- Show the City and State associated with the Zip Code
- The chart should be in US Dollars

The parameters help set which dimensions and measures you are interested in using. Tableau will assume certain things for you as you create your visualization.

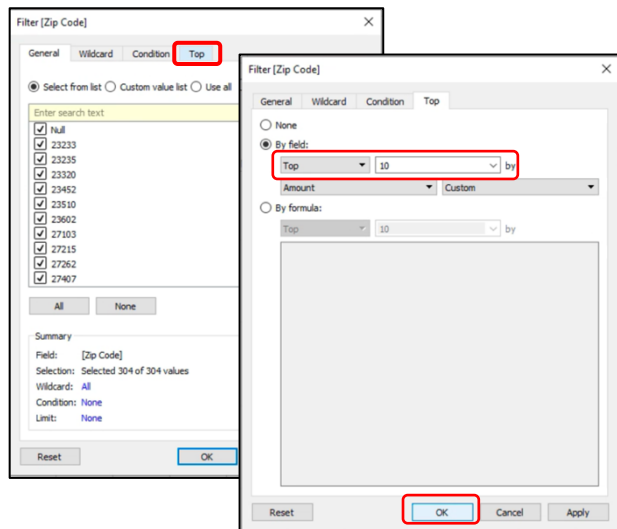
1. Start by dragging and dropping the dimensions of **Zip Code**, **City** and **State** to the Columns cell and the **Tran Amnt** measure to Rows cell.



Since we only want the top 10 Zip Codes for the year 2017, we need to drag and drop, in order, the **Zip Code** dimension and the **Tran Date** dimension to the Filters cell. A Filter [] pop-up will appear to select what is to be filtered.

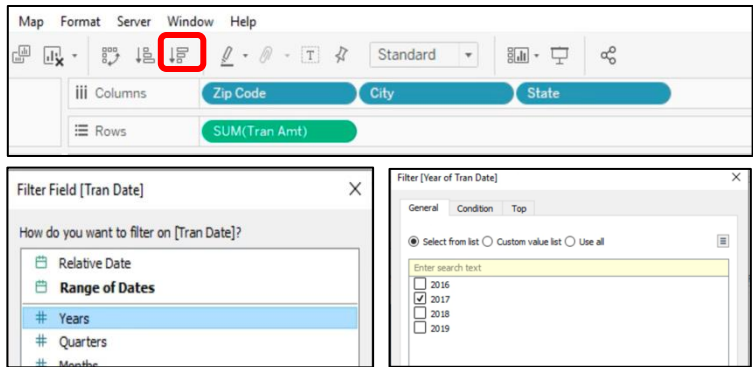
Drag **Zip Code** to *Filters*:

2. Select the *Top* tab (circled in red).
3. Select the radial button **By field**:
4. The Top 10 by **Tran Amnt** Sum should already be selected.
5. Click **OK**.



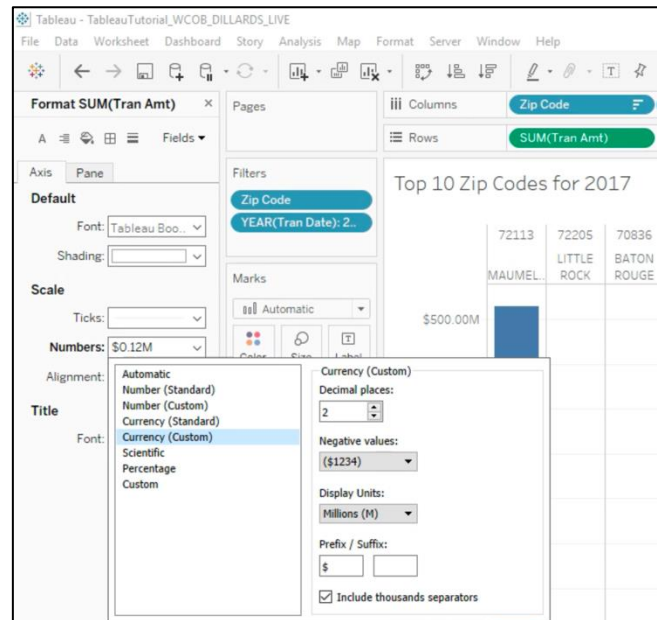
The column is not currently in descending order.

6. Select descending on the top tool bar. (circled in red)
7. Drag **Tran Date** to *Filters*
8. Select **Years** in Filter Field (highlighted in blue) and click **Next**.
9. Select 2017 in **Enter search text**. Click **OK**.




At this point, we have our visualization but need to adjust the Title, the y-axis to reflect currency and the bars in order to see the City spelling completely.

10. To adjust the Title, double click on the *Sheet 1* tab at the bottom
11. Rename *Sheet 1* to **Top 10 Zip Codes for 2017**
12. Hit *enter* to complete
13. To adjust the y-axis, right click on **Tran Amnt** and select *Format* – The left navigation pane will change to allow you to format the y-axis
14. Under *Scale*, select the down arrow on *Numbers*: and select *Currency (Custom)*
15. Change the *Decimal places*: to 0
16. Change the *Units*: to *Millions (M)*.
17. Verify the following (see screen shot):

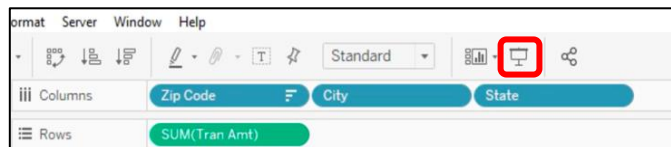


- a. Negative values: is (\$1234)
- b. Prefix / Suffix: is \$
- c. Include thousands separators is checked

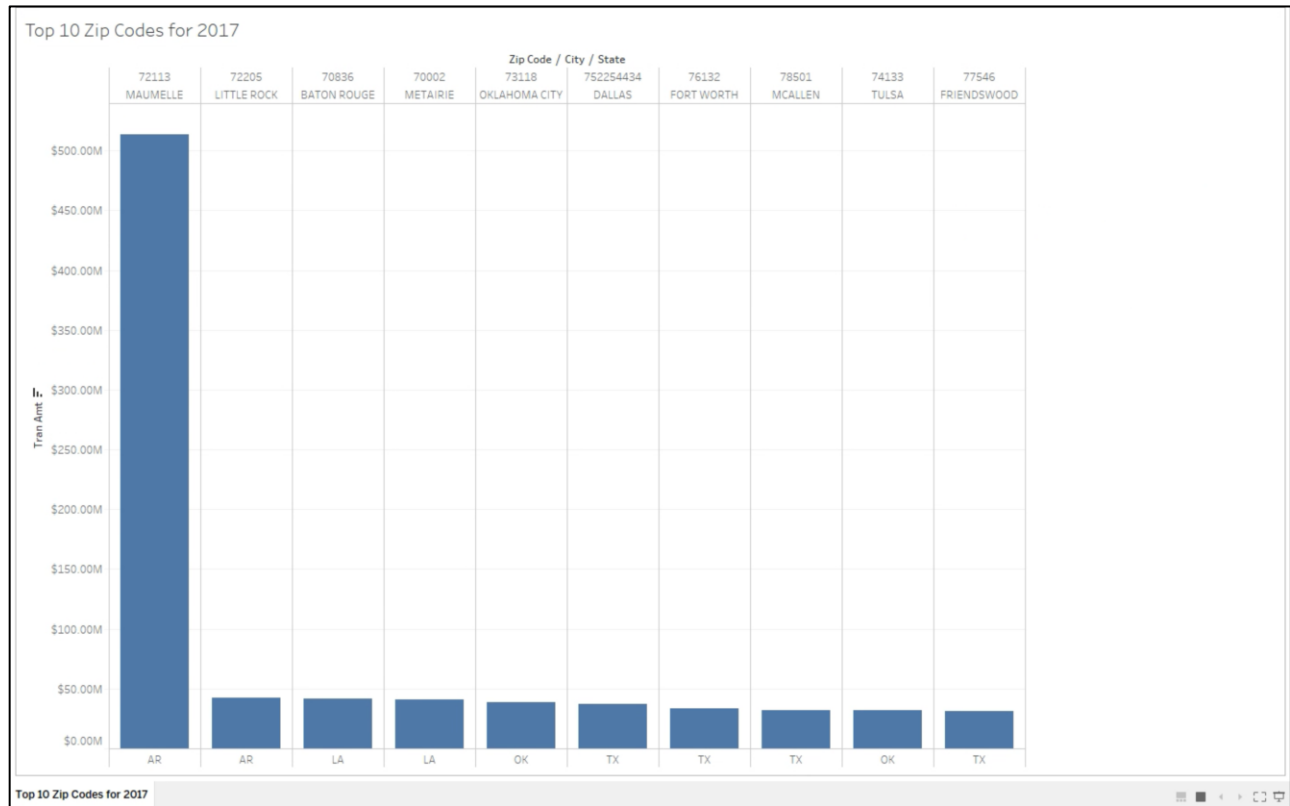
18. Click off the pop-up to save settings and return to your worksheet.
19. To adjust the column widths, mouse over one of the separators between the zip codes at the top until you get the double arrow. 

20. Right click and drag the column to the appropriate width.

21. Finally, to view your final product, click on *Presentation Mode (F7)* on the top tool bar (circled in red).



Your visualization should look similar to the following visualization:



You note that Maumelle, Arkansas with zip code 72113 leads all other zip codes with the highest revenue at almost \$80 million. Now you would like to see how this compares to revenue across all states.

## Geographic View by State

Now you would like to see how each state compares in terms of total revenue in 2017. We will need to create a new visualization.

1. Click on the tab to the right of your **Top 10 Zip Codes for 2017** tab to create a *New Worksheet*.
2. Tableau creates a new *Sheet 2*.
3. Double click on *Sheet 2* and rename it to **2017 Revenue by State**

Tableau allows multiple ways to create visualizations. For each Dimension and Measure, you can double click on your selected variables and allow Tableau to place and create your visualization.

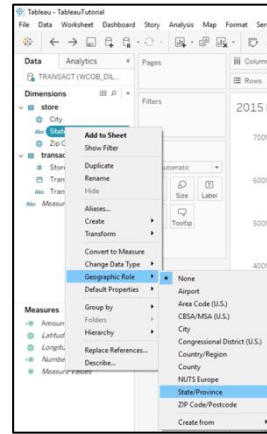
4. Double click on **Tran Amnt**.
5. Double click on **State**.
6. Filter on **Tran Date** to 2017 as was done above.

Tableau automatically created a column chart and our filter shows Texas as the highest revenue. However, we are interested in a geographic visualization to show each state. To do this, we need to change the visualization to a geographic representation.

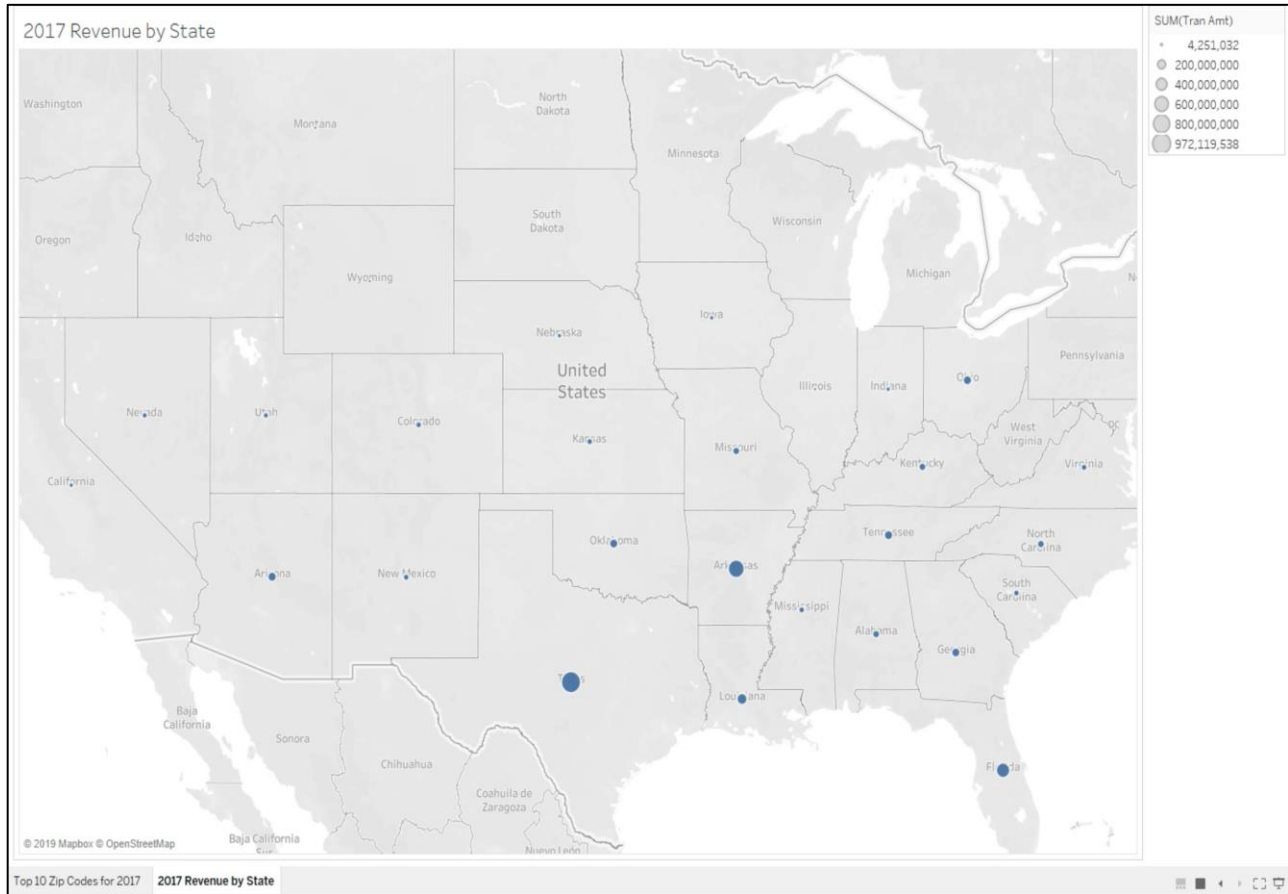
# WCOB Dillards Live Data Connection



- Right Click on **State** under Dimensions and look for **Geographic Role** and then click on **State/Province**. (Path shown in figure alongside).
- Now, locate **Show Me** at the top right of the page and select the indicated visualization.
- Go to Presentation Mode (F7).



The map confirms that Texas is the highest selling stated. A mouse over of the Texas circle is consistent with the indication in the Legend that the highest revenue is \$237 M. A mouse over of Arkansas (where the highest zip code sales resided) shows that all of Arkansas had sales revenue of \$113 M or roughly 1/2 of Texas.



## Midwest Area view by Zip Code

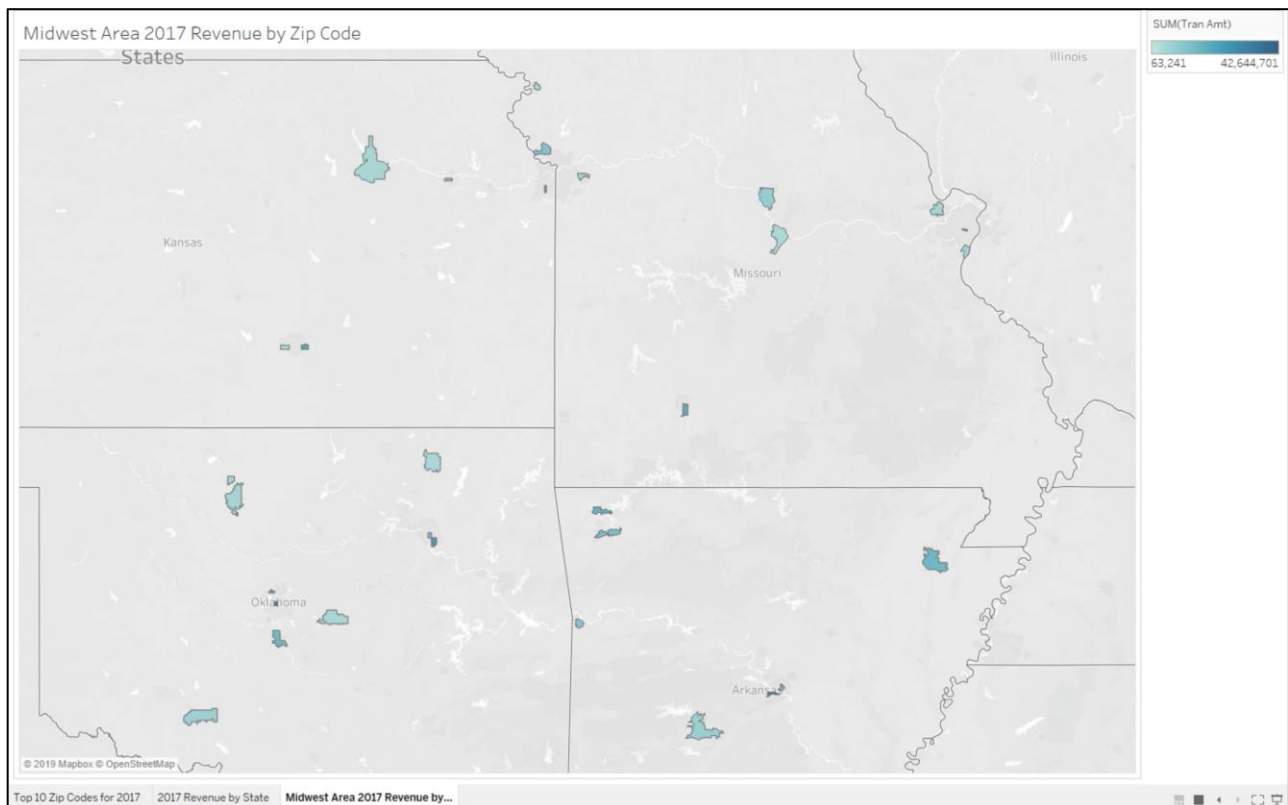
Now we are ready to drill down to the area of responsibility by zip code. Select a new worksheet and name it **Midwest Area 2017 Revenue by Zip Code**.

1. Double click on **Tran Amnt.**
2. Double Click on **Zip Code**.
3. Drag and Drop **State** to *Filters*.
4. Select your Areas (AR, KS, MO, OK) and Click *OK*.
5. Drag and Drop **Tran Date** to *Filters* and filter to Year 2017.
6. Select Show Me and select the indicated geographic visualization to the right (Circled in red).



This graph provides the outline of the zip code area and not just a circle. It is very easy to see where the highest Zip Code sales occurred in the middle of Arkansas (Maumelle, AR 72113). However, this Zip Code was so far ahead of the other zip codes that you would like to see a greater differentiation. This can be done by filtering out Zip Code 72113 from the visualization.

7. Drag **Zip Code** to *Filters*.
8. Deselect Zip Code 72113.



## On Your Own

These visualizations provide a starting point for analyzing Sales Revenue for Dillards and the Midwest Area. Here are some additional suggestions along with the steps to create additional views that can further enhance your analysis.

### I would like to change the Visualization scale for the Zip Code Areas

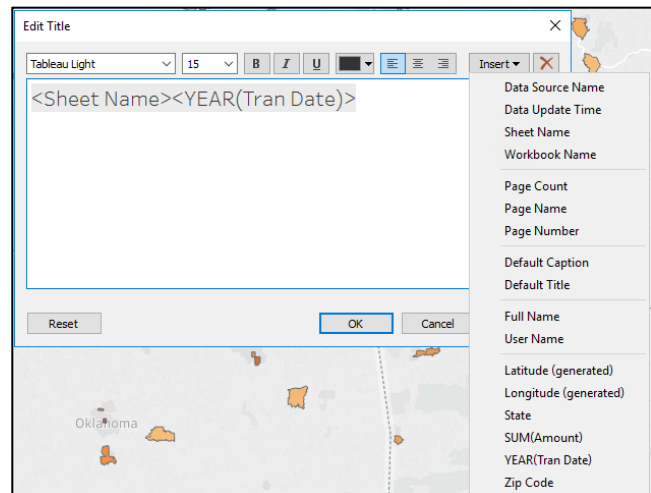
1. Select the down arrow in The SUM(**Tran Amnt**) legend on the right
2. Select Edit Colors...
3. Select the down arrow under Palette...
4. Choose your preferred palette and select OK

### I would like to change the Year of Analysis

5. Right click on YEAR(**Tran Date**): under *Filters*
6. Select Show Filter
7. The Filter selection will now appear on the Right of your chart
8. Select the single year or multiple years you would like to view
9. NOTE: The title will be incorrect with a new filter selection

### I would like to Correct my Title so the Year updates correctly

1. Rename your tab to “Midwest Area Revenue by Zip code for “
2. Right Click on the Title and select Edit Title...
3. Select YEAR(**Tran Date**) under Insert and OK
4. Verify your Title change by selecting different years



# Data Dictionary

## CUSTOMER Table

Column name	Key specifications	Datatype	Short description	Long description	Sample values
CUST_ID	Primary key	Integer (not null)	Customer Identifier	Unique identifier representing a customer instance	219948527 219930818
CITY		Varchar (28)	City	City where the customer lives.	HOUSTON COOS BAY
STATE		Char(2)	State	State where the customer lives.	FL TX
ZIP_CODE		Char(5)	Zip code	Customer's 5-digit zip code.	72701 84770
ZIP_SECSEG		Char(4)	Sector segment code	Customer's geographic segment code	5052 6474
DISTANCE_TO_NEAREST_STORE		Decimal(8,3)	Distance to nearest store for customer	Miles from the customer's house to the closest Dillard's store.	0.687 6.149
PREFERRED_STORE	Foreign Key (References STORE table)	Small Int	Customer Preferred Store Location	Dillard's store number the customer prefers to shop at regardless of distance to the customer's home address.	910 774

## DEPARTMENT Table

Column name	Key specifications	Datatype	Short description	Long description/ Additional Details	Sample values
DEPT	Primary key	Varchar (10), not null	Department Number	The Dillard's unique identifier for a collection of merchandise within a store format	0471 0029
DEPT_DESC		Varchar (25)	Department Description	The name for a department collection of merchandise within a store format.	"Christian Dior" "REBA"
DEPTDEC		Char(4)	Department Decade	The first three digits of a department code, a way to classify departments at a higher level.	047X 002X
DEPTDEC_DESC		Varchar (25)	Department Decade Description	Descriptive name representing the decade to which a department belongs.	'BASICS' 'TREATMENT'
DEPTCENT		Char(4)	Department Century	The first two digits of a department code, a way to classify departments at a higher level.	04XX 00XX
DEPTCENT_DESC		Varchar (25)	Department Century Description	The descriptive name of the century.	CHILDRENS COSMETICS

## SKU Table

Column/ Field name	Key specifications	Datatype	Short description	Long description/ Additional Details	Sample values
<b>SKU</b>	Primary key	Char(10), Not null	Stock Keeping Unit	Unique identifier for an item, identifies the item by size within a color and style for a particular vendor.	0557578 6383039
<b>DEPT</b>	Foreign key (references DEPARTMENT table)	Varchar(10), Not null	Department Number	The Dillard's unique identifier for a collection of merchandise within a store format.	0134 0343
<b>SKU_CLASS</b>		Char(10)	Stock Keeping Unit Class	Three character alpha/numeric classification code used to define the merchandise. Class requirements vary by department.	K51 220
<b>SKU_STYLE</b>		Char(15)	Stock Keeping Unit Style	The Dillard's numeric identifier for a style of merchandise.	091923690 LBF41728
<b>UPC</b>		Decimal (15,0)	Universal Product Code	A number provided by vendors to identify their product to the size level.	889448437421 44212146767
<b>COLOR</b>		Varchar(12)	Stock Keeping Unit Color	Color of an item.	BLACK PINEBARK
<b>SKU_SIZE</b>		Char(10)	Stock Keeping Unit Size	Size of an item. Product sizes are not standardized and issued by vendor	6 085M
<b>BRAND_NAME</b>		Varchar(70)	Brand Name	The item's brand.	Stride Rite UNKNOWN
<b>CLASSIFICATION</b>		Varchar(70)	Classification	Category used to sort products into logical groups.	Dress Shoe
<b>PACKSIZE</b>		Char(5)	Item Package Size	Number that describes how many of the product come in a package	001 002

## SKU\_STORE Table

Column/ Field name	Key specifications	Datatype	Short description	Long description/ Additional Details	Sample values
<b>STORE</b>	Primary key,	Small Int, Not null	Store Number	The numerical identifier for a Dillard's store.	915 701
	Foreign key (references STORE table)				
<b>SKU</b>	Primary key Foreign key (references SKU table)	Char(10)	Stock Keeping Unit	Unique identifier for an item, identifies the item by size within a color and style for a particular vendor.	4305296 6137609
<b>RETAIL</b>		Decimal(7,2) Not null	Retail Price	The price of an item.	11.90 45.15
<b>COST</b>		Decimal(7,2) Not null	Item Cost	The price charged by a vendor for an item	8.51 44.84

## STORE Table

Column/ Field name	Key specifications	Datatype	Short description	Long description/ Additional Details	Sample values
<b>STORE</b>	Primary Key	Small Int, Not null	Store	The numerical identifier for any type of Dillard's location.	767 460
<b>DIVISION</b>		Char(2)	Division	The division to which a location is assigned for operational purposes.	07 04
<b>CITY</b>		Varchar (28)	City	The city where the store is located.	IRVING MOBILE
<b>STATE</b>		Char (2)	State	The state abbreviation where the store is located.	MO AL
<b>ZIP_CODE</b>		Char (5)	Zip Code	The 5 digit zip code of a store's address.	70601 35801
<b>ZIP_SECSEG</b>		Char (4)	Zip Code Geographic Segment Code	The 4 digit code of a neighborhood within a specific zip code.	5052 6474

## TRANSACTION Table

Column/ Field name	Key specifications	Datatype	Short description	Long description/ Additional Details	Sample values
<b>TRANSACTION_ID</b>	Primary key	Integer	Transaction Identifier	Unique numerical identifier for each scan of an item at a register.	40333797 15129264
<b>TRAN_DATE</b>		Date in format MM/DD/YYYY, not null	Transaction Date	Calendar date the transaction occurred in a store.	1/1/2015 5/19/2014
<b>STORE</b>	Foreign key (references STORE table)	Small Int Not null	Store	The numerical identifier for any type of Dillard's location.	716 205
<b>REGISTER</b>		Small Int Not null	Register	The numerical identifier for the register where the item was scanned.	91 55 12
<b>TRAN_NUM</b>		Integer Not null	Transaction Number	Sequential number of transactions scanned on a register.	184 14
<b>TRAN_TIME</b>		Char(4) Not null	Transaction Time	Time of day the transaction occurred.	1839 1536
<b>CUST_ID</b>	Foreign key (references CUSTOMER table)	Integer Not null	Customer Identifier	Unique identifier representing the instance of a customer.	118458688 115935775
<b>TRAN_LINE_NUM</b>		Small Int Not null	Transaction Line Number	Sequential number of each scan or element in a transaction.	3 2
<b>MIC</b>		Char (3) Not null	Manufacturer Identification Code	Manufacturer Identification Code used to uniquely identify a vendor or brand within a department.	154 128 217

**WCOB Dillards Live Data Connection**



<b>TRAN_TYPE</b>		Char (1) Not null	Transaction Type	An identifier for a purchase or return type of transaction or line item	P R
<b>ORIG_PRICE</b>		Decimal(7,2) Not null	Original Price	The original unit price of an item before discounts.	20.00 6.00
<b>SALE_PRICE</b>		Decimal(7,2) Not null	Sale Price	The discounted unit price of an item.	15.00 2.64 6.00
<b>TRAN_AMT</b>		Decimal(7,2) Not null	Transaction Total Amount	The total pre-tax dollar amount the customer paid in a transaction.	15.00 2.64
<b>TENDER_TYPE</b>		Varchar (5) Not null	Tender Type	The type of payment a customer used to complete the transaction.	BANK DLRD DAMX
<b>SKU</b>	Index, Foreign key (references SKU table)	Char (10)	Stock Keeping Unit	Unique identifier for an item, identifies the item by size within a color and style for a particular vendor.	6107653 9999999950