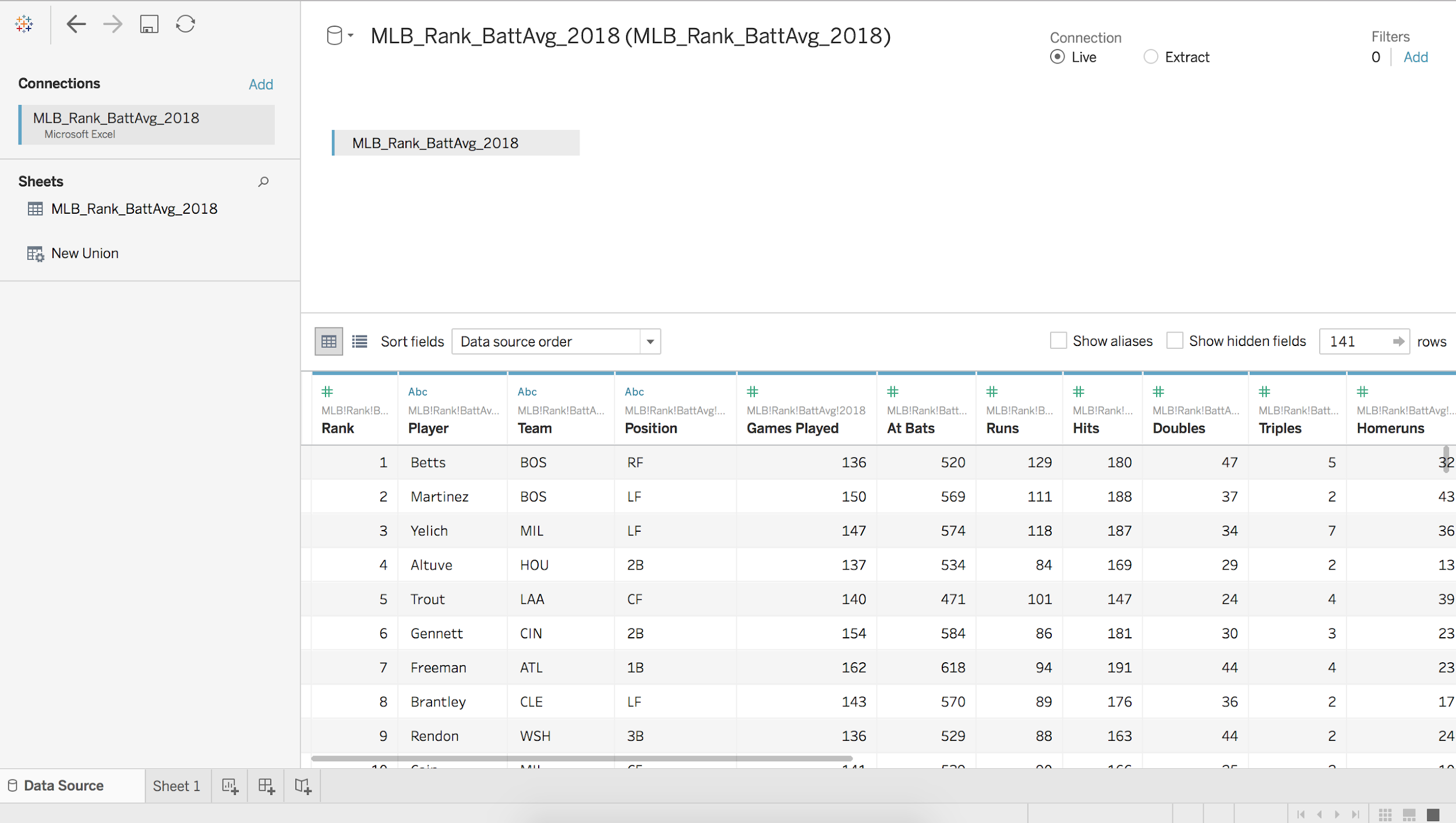
**Data Visualizations using Tableau**

**Getting Started with Tableau**

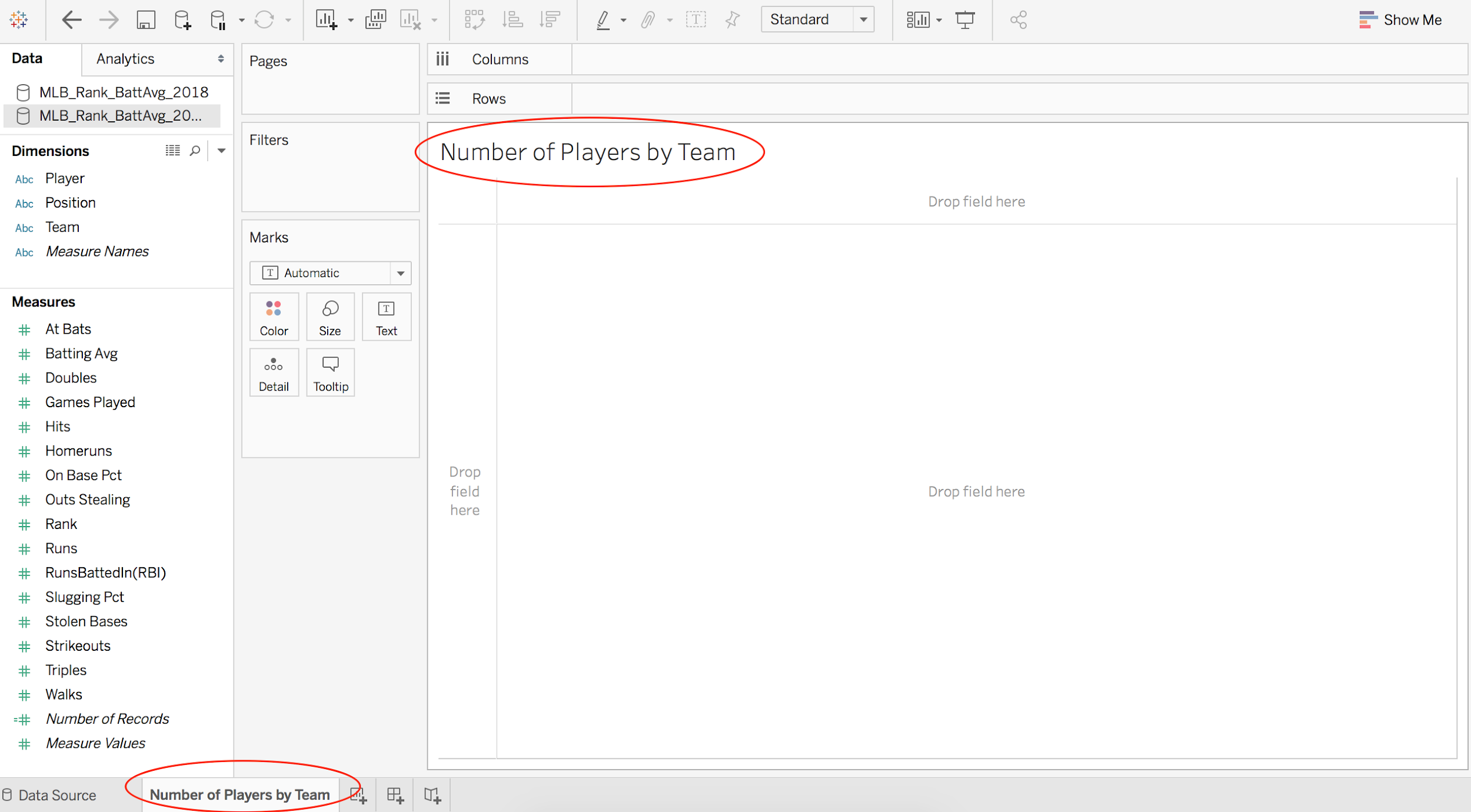
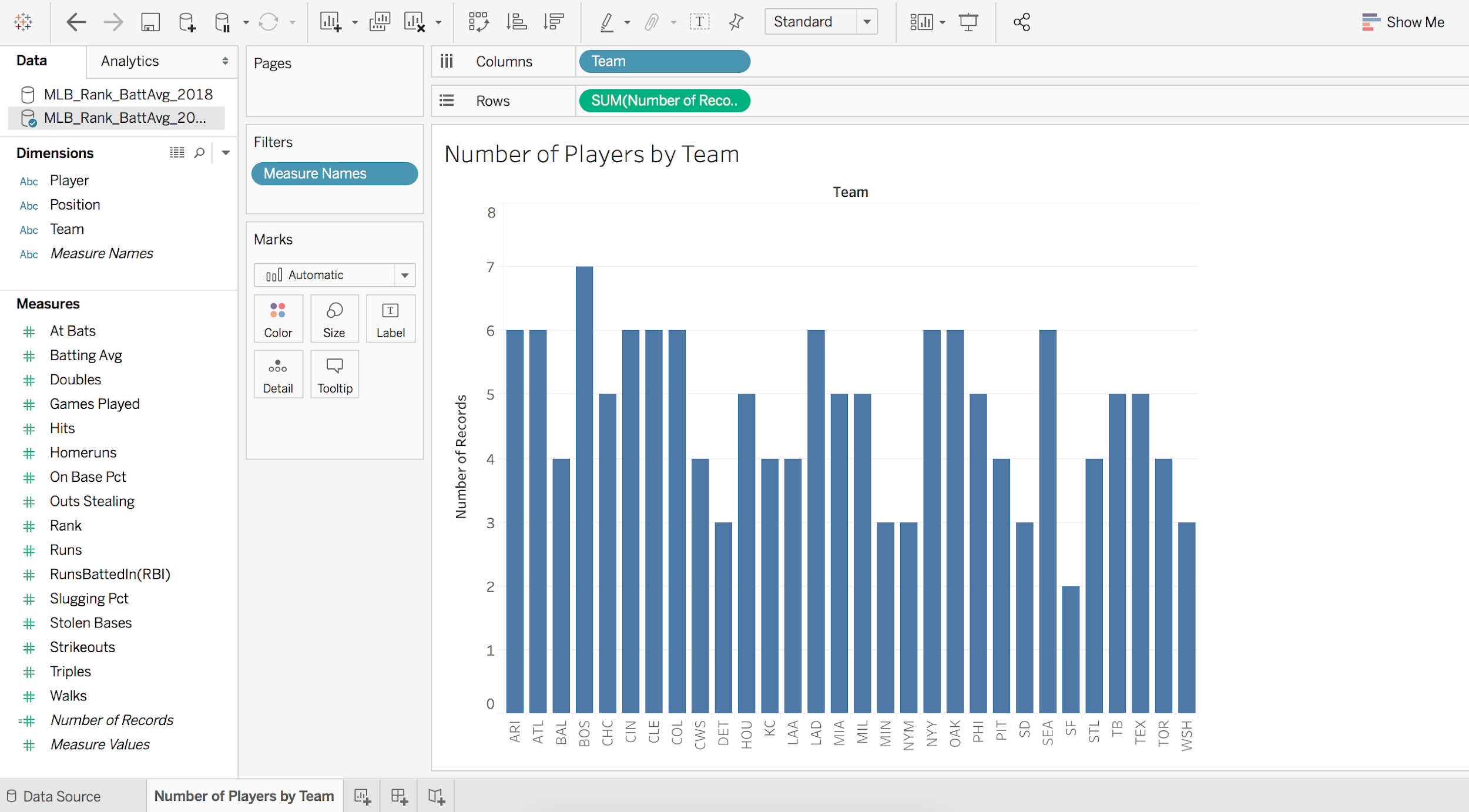
1. On the left under “Connect” you will see “To a File.” This is where you can import (or connect) your data.
2. To import an **.xlsx** file, select “Microsoft Excel”, navigate to the file on your computer, click it, and select open. You should now see data in the lower half of your window.



1. To import a **.csv** file, select “Text File”, navigate to the file on your computer, click it, and select open. You should now see data in the lower half of your window. It will look identical to the picture shown for .xlsx above.
2. If at any point you regret your import, navigate to the top left of your window where the Tableau icon is located. Clicking this will bring you back to the main screen where you can connect a file again.
3. To **Save** at any point, you can select the floppy disc icon in the top left, or select File>Save. This will save as a Tableau workbook.
4. Now that the data file is connected, we can get familiar with some options that Tableau offers for changing or sorting the data.
   1. There is a “Sort Fields” option above the data about ⅓ of the way down from the top of the window. This allows you to arrange your columns in alphabetical or reverse alphabetical order. “Per table” is applicable when you have more than one data table connected in Tableau, but we will only be working with the MLB data for this exercise. “Data Source Order” will bring you back to the original column arrangement.
   2. If at any point you make a change that you do not like, navigate to the top left where you see a forward and back arrow. Selecting the back arrow will undo the last change you made. Be careful to not click the Tableau icon, as that will undo all of your changes and you will have to start from scratch.
   3. There are “#” and “Abc” symbols above the columns in the data. These indicate whether the fields are numeric or character (string) values. Clicking on these symbols will give you the option to change the data type. This can be useful (e.g., when you have numbers that represent discrete categories), but we will not be changing our data types for these visualizations.
   4. If you scroll your mouse over the name of a field, a down arrow will appear to the right. Clicking it will give you the option to rename the field, copy values, hide, and a few other options. **You can try these out but make sure to have the data in the original form before we begin our visualizations.**
   5. Again, by scrolling over the name of a field, there will be an icon that looks like a horizontal bar chart. Clicking on this will sort the rows according to that field.
5. Now that we are familiar with changing the data fields and sorting the data, let’s start visualizing. **Be sure your data file is in the original form, as this guide will refer to fields by their original names. You can click the Tableau icon and reconnect the original dataset if you have made several changes.**

**Univariate Visualizations**

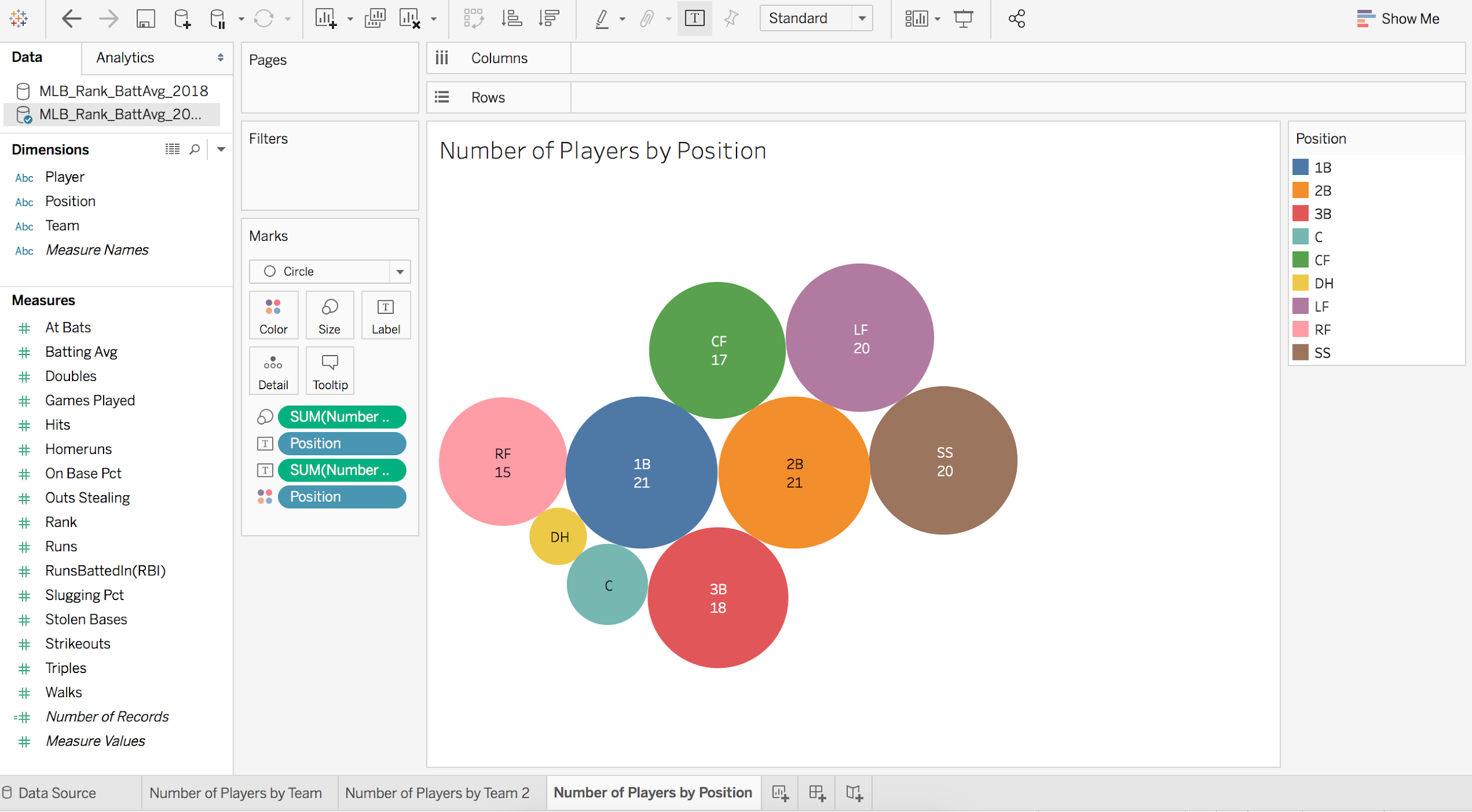
**Number of Players by Team: Bar Chart**

1. Select “Sheet 1” at the bottom left corner. This will bring you to a new display that shows dimensions and measures on the left, and an area for dropping fields in the middle of the screen.
2. Let’s rename Sheet 1 to be “Number of Players by Team.” To do this, right click on “Sheet 1” in the bottom left corner. Notice that this also changes the title displayed near the top middle of the window. You can change this title (without changing the name of the sheet) by right clicking on it and selecting Edit Title. We want to keep the title the same as name of the sheet, so we will leave it as is.
3. If you want to look at your data at any point, you can navigate between your data and the visualization sheet by selecting Data Source at the bottom left corner of the window. For now, we will be working in the Numbers of Players by Team visualization sheet.
4. To find the team variable, look to the left. Measures are numerical columns and Dimensions are character columns. Team is listed in the Dimension group because this field contains the names of the teams. If you look at the bottom of the list of Measures, you’ll notice there are two measures that were not in our dataset: *Number of Records* and *Measure Values*. These are calculated and given to us by Tableau. *Number of Records* counts the amount of times something is repeated in a field. For example, if BOS shows up 5 times in the list of Teams, *Number of Records* will indicate 5 for BOS.
5. We want to display the count of players per team. Click “Team” on the left side of your window and drag it to the right of Columns at the top of the window. Next, click on “Number of Records” and drag it to next to Rows, just under Columns. (NOTE: This field will now display as SUM(Number of Records)). This will produce as a vertical bar chart. To display as horizontal, switch Team to Rows and Number of Records to Columns. Your initial display should look like this: 
6. Direct your attention to “Marks” to the left of your visualization. Here you can change the color, the size, and labels of the bars. Under label, selecting “show mark label” will display the count above the bars (in the vertical format) or to the right of the bars (in the horizontal format). Feel free to edit the appearance of your bar chart.
7. To change the range of any numeric axis, simply double click on the axis. Click on Fixed and input the minimum and maximum values of the new range. Here, you can also change the title of the axis.
8. To change the order of the bars in the chart, locate Team at the top left of the display. You will see AZ and an arrow next to Team. Click on the arrow next to AZ, and a drop down menu will appear. If you choose SUM (Number of Records), the bars will be arranged from largest number of records at the top to fewest number of records at the bottom. You will now see a bar chart icon next to Team. If you click this, you can reverse the order of the bars.
9. To export any display from Tableau to a .png file, go to Worksheet, Export, Image. Choose the options you want, and then locate the folder where you want to save the image.

**Number of Players by Team: Circle Views**

1. Let’s create another visualization for Players by Team. To do this, look at the bottom left of your screen and click the icon to the right of Number of Players by Team - it looks like a bar chart with a plus at the bottom. If you hover your mouse over it, it should say “New Worksheet.” Now we can create another visualization. Feel free to name this whatever you wish, I named mine “Number of Players by Team 2.”
2. Go ahead and drag Team to columns and Number of Records to rows.
3. At the top right of your window, there is an icon with a horizontal bar chart with “Show Me” next to it. Click this icon. This shows the type of visualizations you can create. If you scroll your mouse over each one, at the bottom of the list you will see the requirements to make the visualization. Locate Circle Views near the middle and select it. This will automatically move Team to below the section “Marks.” Click and drag Team from Marks back up to Columns. Now you have a circle view visualization.
4. Under Marks, you can change the shape, color, size, etc.
5. At the top of the y-axis label (Number of Records is the default), you will see a bar chart icon. Clicking this will change the order of how the marks are displayed.

**Number of Players by Position: Packed Bubbles**

1. Create another sheet and name it Number of Players by Position.
2. There are many ways to create visualizations in Tableau - let’s try another way.
3. Click Position in the Dimensions list to highlight it. While holding down command (or control if you have a PC), click Number of Records so that it is also highlighted. Now click on Show Me in the top right corner. The highlighted visualizations are the ones you can create using the selected fields.
4. Locate and select packed bubbles at the bottom of the Show Me tab.
5. The size of the circles indicate the number of the records (in our case, the number of players) by Position. Tableau automatically assigned unique colors and labels for each Position.
6. Let’s add a label to indicate how many players there are per Position. To do this, click and drag Number of Records (the one in the bottom left corner, not the one already in Marks) and drop it on top of Label in the Marks section. Now each Position circle has a unique color, is sized by the number of players, a label to indicate Position, and a label to indicate the number of players in each Position. Your display should look like this: 
7. To remove a circle, click on the circle and select Exclude.
8. We can remove marks if we want. Let’s remove the label for number of positions. In the Marks area, right click SUM(Number…) (the one with the T in the box icon) and choose Remove.

**Number of Players by Position: Table**

1. Create a new sheet and name it Number of Players by Position 2. Select both Position and Number of Records. Go to Show Me and select Highlight Tables at the top right. This will give you a simple table displaying counts of players by position. Again, if you want to remove the colors, navigate to Marks, right click on Sum(Number of records) that has the four color dots next to it, and choose Remove.

**Distribution of Homeruns: Histogram**

1. Create a new sheet.
2. Highlight Homeruns under Measures.
3. Under Show Me, select Histogram.
4. You can manipulate and format this display using the tools described for other displays.

**Distribution of Batting Average: Line chart**

1. Create a new sheet.
2. Highlight Batting Avg under Measures.
3. Under Show Me, select Histogram.
4. Under Marks, select the drop down menu and select Line.
5. You can manipulate and format this display using the tools described for other displays.

**Distribution of RunsBattedIn(RBI): Area chart**

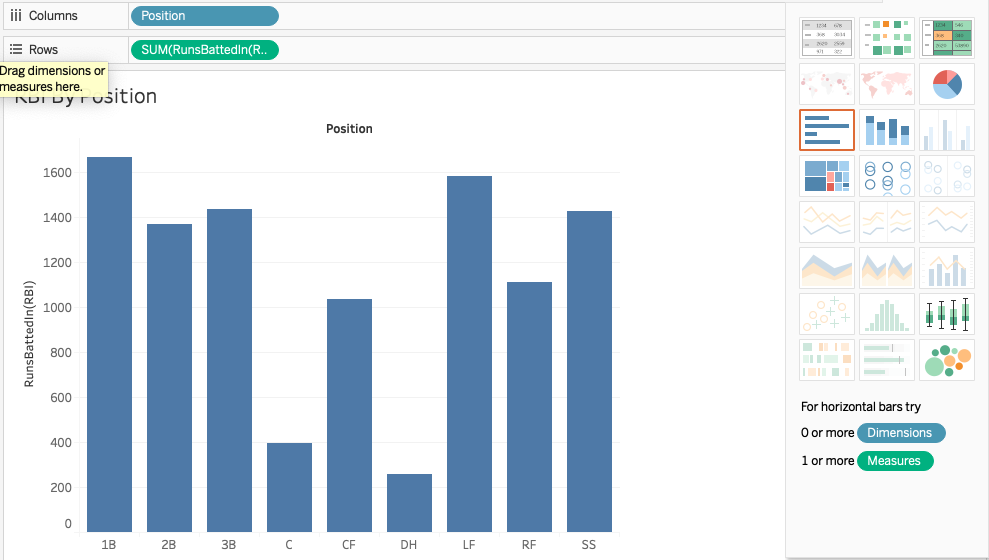
1. Create a new sheet.
2. Highlight RunsBattedIn under Measures.
3. Under Show Me, select Histogram.
4. Under Marks, select the drop down menu and select Area.
5. You can manipulate and format this display using the tools described for other displays.

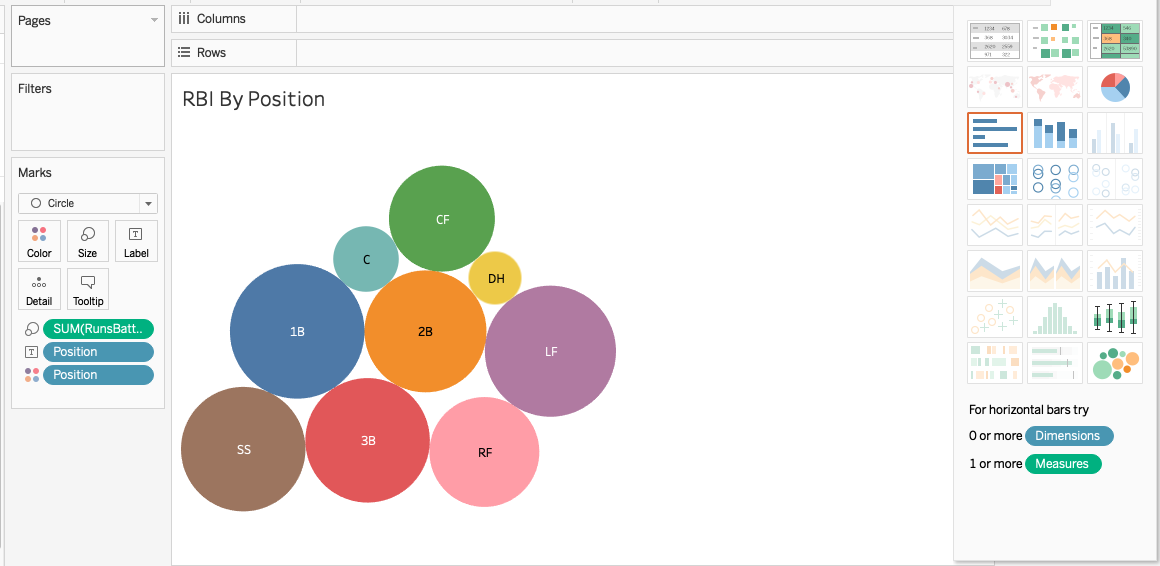
**Bivariate Visualizations**

**Number of Triples by Position: Tree Map**

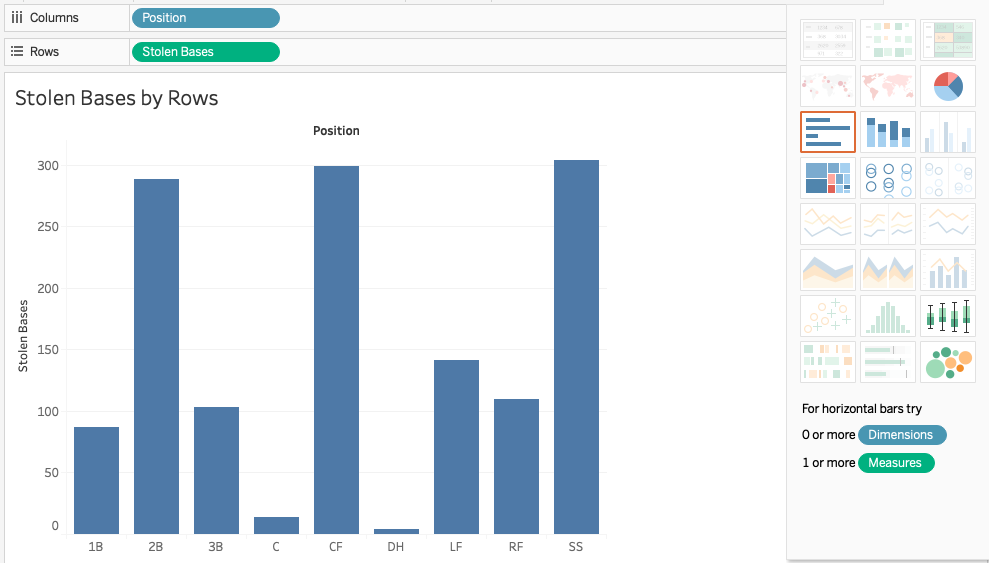
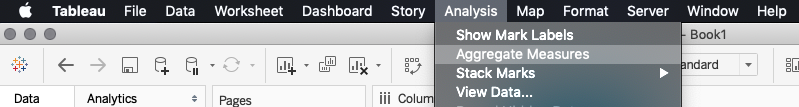
1. Create a new sheet named Triples by Position.
2. Under Dimensions, click on Position to highlight it. Hold down command or control and, under Measures, choose Triples .
3. Under Show Me, select Tree Maps. This will automatically create boxes that are sized according to the frequency of the amount of triples per position.
4. On the right, the legend shows that the darker the color, the greater the number of triples. Remember you can edit the colors under Marks.

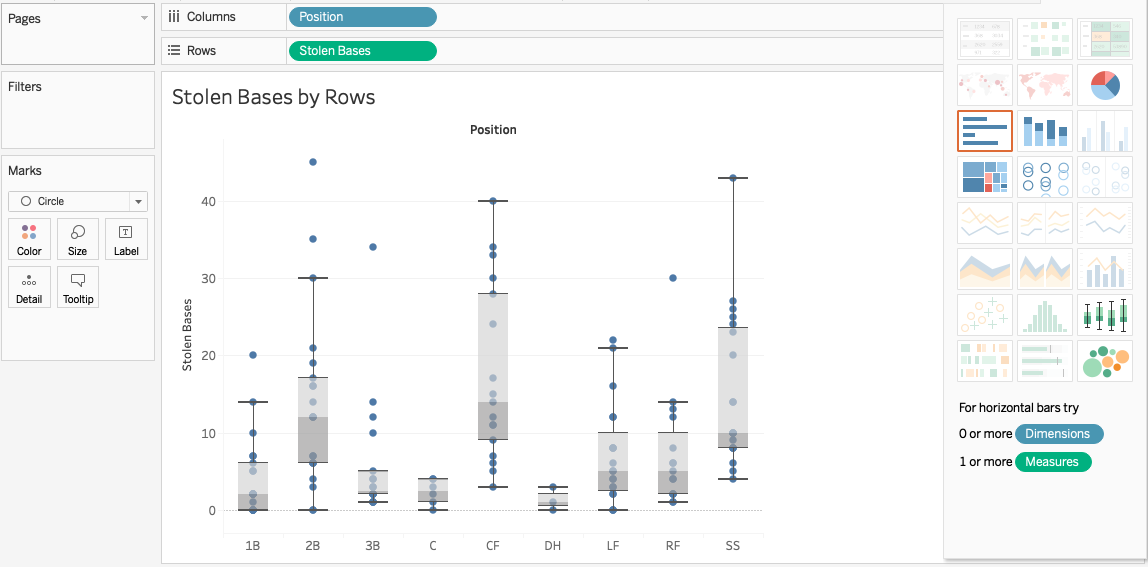
**RunsBattedIn(RBIs) by Position: Bar charts, packed bubbles**

1. We’ll first create a barplot displaying the number of RBIs by Position.
2. First, select RunsBattedIn(RBI) from the Measures panel and drop it into Rows at the top of the window.
3. Now select the Position variable from the Dimensions panel, and drop it into Columns at the top of the window.
4. You should now have a barplot and and screen that look like this: 
5. Perhaps you want to look at a different type of graph. You can create a bubble plot by clicking on Show Me in the top right of the window, and choosing the packed bubbles icon in the lower right corner. This should keep the same variables in the main panel, and display a bubble plot instead of a barplot.
6. You may notice that the variables have gone missing from Columns and Rows. Don’t mind this, Tableau automatically places the variables for us appropriately when we change the plot type.
7. Play with the different display options in Show Me until you find one that works for you.



**Stolen Bases by Position: Bar charts, box and whisker plots**

1. Now let’s create a display for stolen bases by position. Start by creating a new sheet and naming it Stolen Bases by Position.
2. First, let’s begin by creating a bar chart in the same way we have before.
3. Drag the Position to Columns, and StolenBases to Rows.
4. Now, let’s create a box and whisker plot for these variables.
5. First, click “Analysis” on the main menu bar at the top of the screen.
6. Deselect “Aggregate Measures”. This will allow us to view the distributions of stolen bases instead of just the sum.
7. Now, ensure that the Position variable is still in Columns, and the StolenBases variable is still in Rows. Select the box-and-whisker icon in the bottom right section of the Show Me tab.
8. You can manipulate and format this display using the tools described for other displays.



**Home Runs by Team: Highlight tables**

1. To start, create a new sheet called “Home Runs by Team”.
2. Per usual, drag Homeruns to Rows and Team to Columns.
3. In the Show Me tab, click the box that represents the highlight tables, located at the top right.

**Runs by At-Bats: Scatterplot**

1. To start, create a new sheet called “Runs by At-Bats”.
2. Like with the box and whisker plot above, we need to fiirst click “Analysis” on the main menu bar at the top of the screen and deselect “Aggregate Measures”. This will allow us to view the distributions of stolen bases instead of just the sum
3. Notice that we have two quantitative measures in this example. In this case, we want to place the response variable (Runs) in Rows, and the explanatory variable (At-Bats) in Columns.
4. You may want to change the scales of your axes. You can do this by double clicking the axis, choosing Fixed under Range, and changing the minimum and maximum values.
5. In the “Marks” box, you can edit the size, color, and shape of the points.