# Splitting columns/fields

Here, we will use “chapter 2 patient claims” Excel file. This dataset is similar to the one used in chapter 1 (patient-level data; each row includes information on patient healthcare cost in a particular year) with a couple of differences. There is no [Region] column/field, but the [State] column has information on both region and the state the patient resides in. We will connect the dataset to Tableau. The connected file looks like:

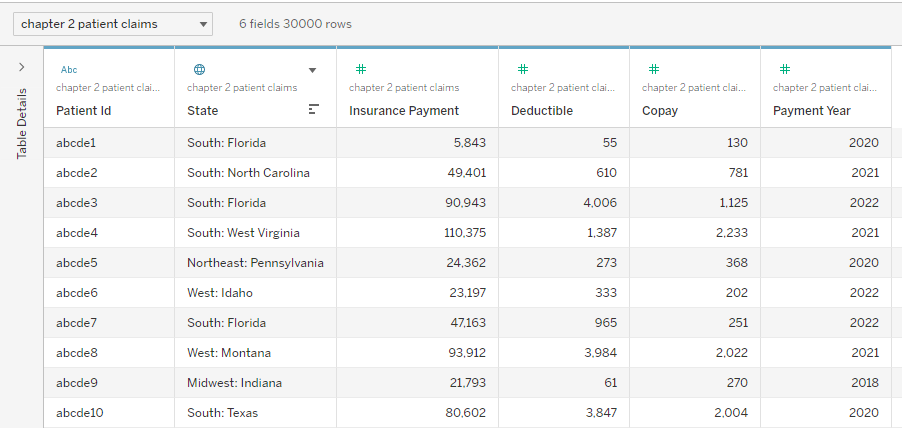


Figure : Screenshot of dataset used for chapter 02

As you can see, both state and region information is in the same column. In Tableau, we can split one column to multiple columns using different separators. In the current dataset, the value of column [State] is formatted as “Region: State”, i.e., there is a colon followed by a space that separates the name of region from the name of state. Also, Tableau recognized the column as a geographic column despite the somewhat unusual way of how the values are stored.

When we click on the little icon (▼) above [State], we get different choices. The icon pops up when we hover over any of the columns, which gives use different options for that particular column. Here, we will hover over [State] and click on the icon. We get:

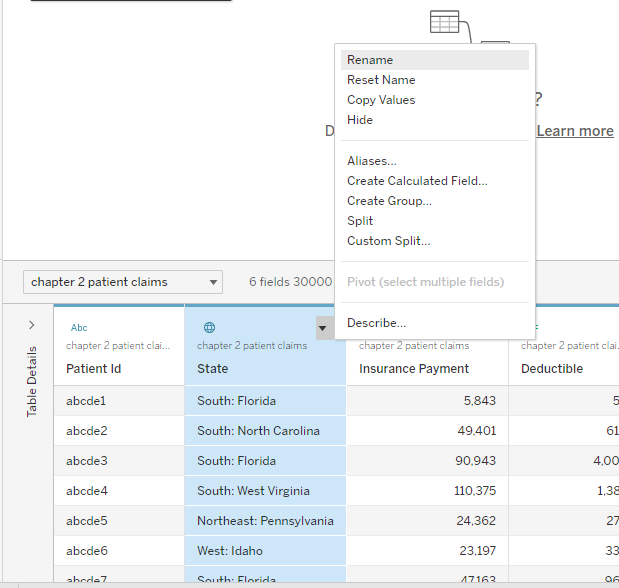


Figure : Choices we get when we click on a column

As you may notice, we have two choices pertaining to splitting the column: Split and Custom Split. If we click on Split option, Tableau will automatically split the column based on what it thinks the user wants to use as a separator. Sometimes it would be correct and sometimes not. Here, when we click on Split, Tableau will create two new columns in addition to keeping the original [State] column as such.

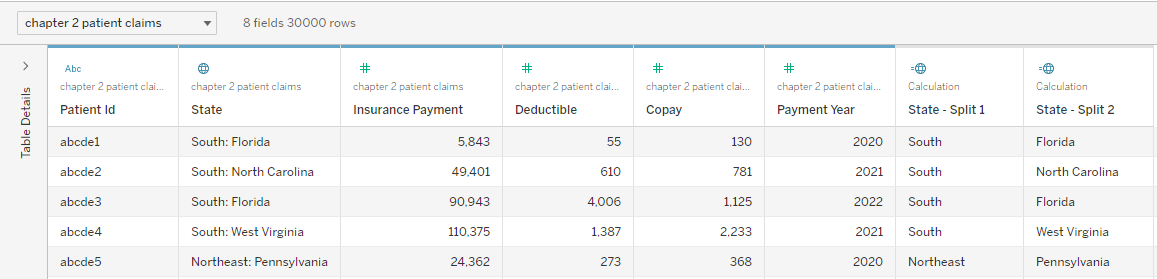


Figure : Dataset with two new columns added based on Split function

Split function is more Tableau-driven, so we may feel that we don’t have control over the splitting function. This is where Custom Split comes in; it allows us to choose the separator and also if we want to keep only one split or all splits. We will delete the two splits for now. We can delete a column by clicking on that icon that pops up over the column name and selecting delete.

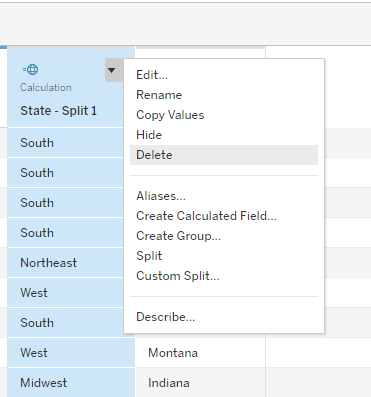


Figure : Showing how to delete a column

Now, we will use Custom Split function to split the [State] column. When we click on Custom Split function (which is one of the options that shows up when you click on the ▼ icon over the column), we get:

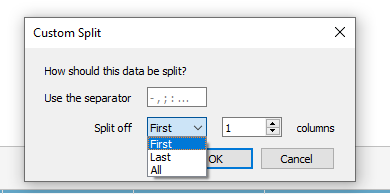


Figure : Custom split function options

Here, we can pick the separator (in our case, it’s colon). Since we want to keep both columns, we will select “All” for Split off. Our choices look like:

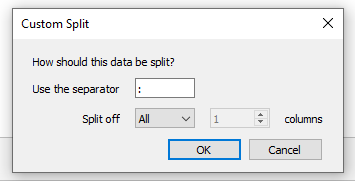


Figure : Selected choices for Custom Split function

This will give us:

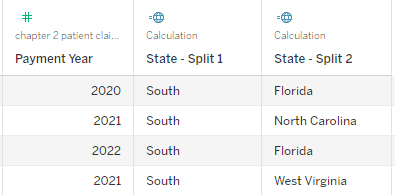


Figure : Dataset with two new columns resulting from Custom Split

We can rename the columns. There is a rename option when you click on the ▼ icon.

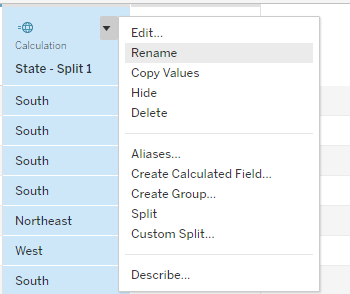


Figure : Rename option for renaming columns

We will rename the column with region information as [Region] and the column with state information as [State name].

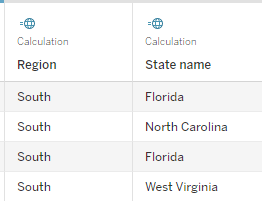


Figure : Renaming the two columns to Region and State name

The Split/Custom Split functions also work when a column could be split to more than two new columns.

# Aliases

We will now look into alias function in Tableau. Alias is used to give names to the values of text/string or geographic columns (it does not work with numeric columns) that would be displayed. In the above dataset, the [State name] column has name of the states as the values. If we want to use abbreviations of the state names, we can use Alias and assign the abbreviations to the respective state names. We can use Alias either by clicking on ▼ icon under Data Source or we can use it from the worksheet by clicking on the column of interest.

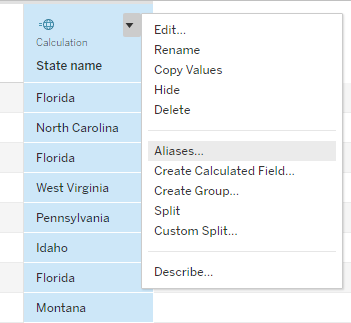


Figure : Showing Alias option from Data Source for State name column

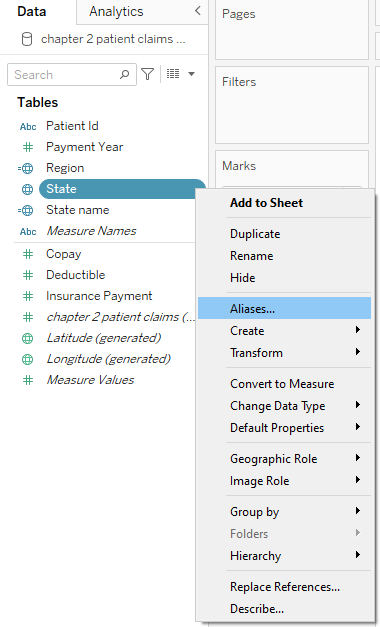


Figure : Showing Alias option from worksheet for State name column

When we click on Aliases, we get:

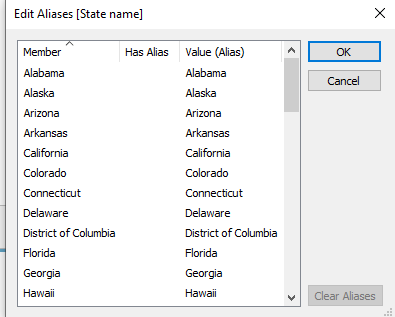


Figure : Alias box when clicked on Aliases

We can see that the actual values and Aliases are the same here. If we want to change the names, we can click on the values under “Value (Alias)” and type out what value/alias we want to assign the particular state name.

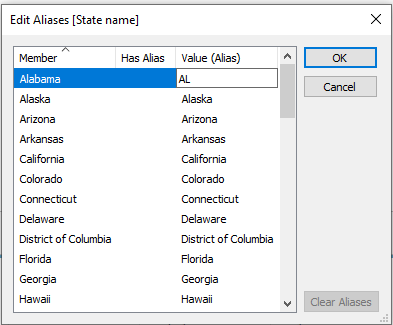


Figure : Using state abbreviations as aliases

We will add abbreviations for all states as alias. Once we are done with that, the box would look like:

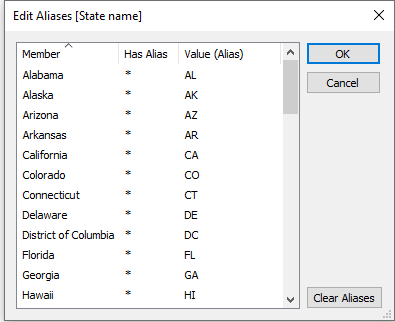


Figure : Alias box after all state abbreviations are added as alias

We can see that there are \* under “Has Alias” column in the “Edit Aliases” box. This indicates that a particular value of that column has been assigned an alias value. When we click OK, we can see that the [State name] column displays the alias values.

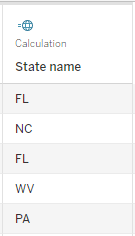


Figure : State name column with alias values

Now, if we create a table or other graphics, these alias values get displayed.

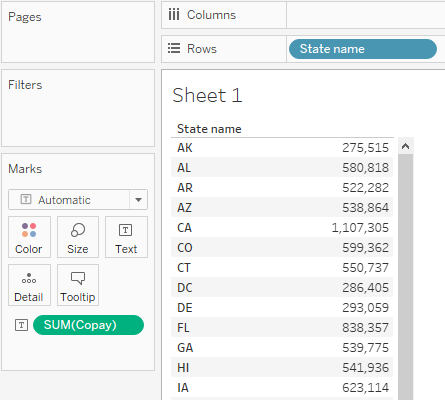


Figure : Table created using State name column (with alias values) and total copay

# Grouping

Now, we will look into grouping, which is used to group values of a column/field together. This would result in fewer values for the column.

I will first clear aliases by clicking “Clear Aliases” option that is displayed in the “Edit Aliases” box. We have four geographic regions in the [Region] column and 50 states plus Washington DC in the [State name] column. We want to create a new column by using [State name] column where we would group states into divisions (divisions are a bigger geographic unit than states but smaller than regions). For example, we will group Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island and Vermont states and name the group “New England”.

Grouping can be applied to the existing column or by creating a new column. Since we do not want to replace the [State name] column values for individual states, we will first duplicate it. To do this, we will go to an empty sheet and click on the column [State name]. We see:

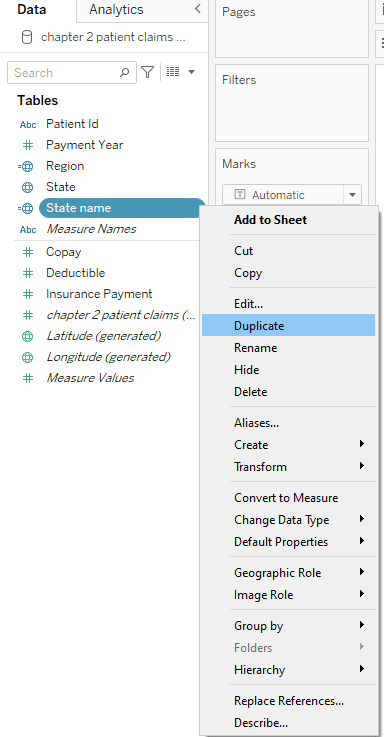


Figure : Duplicating State name column

We will use the duplicated column for state name, [State name (copy)].

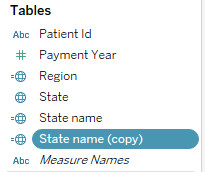


Figure : Showing newly created duplicate column for State name

We can create groups by clicking on the [State name (copy)] column and selecting “Group” under “Create”.

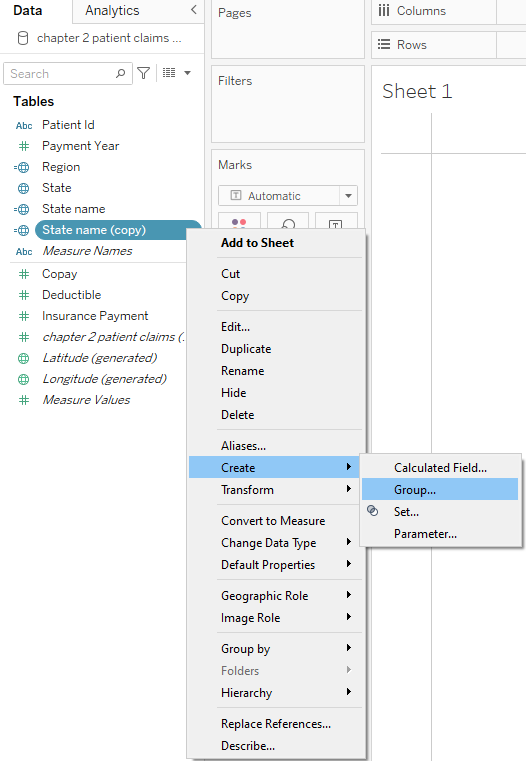


Figure : Showing Create group option

Once we click that, we get the following box:

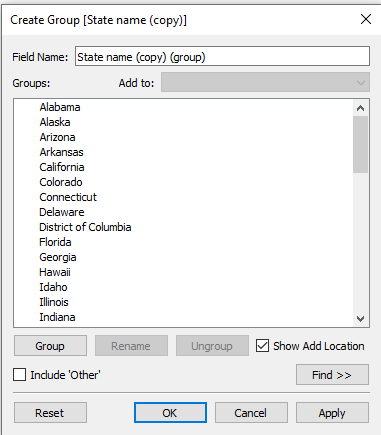


Figure : Create Group box

We can now select the states we want to group together and assign a particular value. First, I will select Alabama, Kentucky, Mississippi and Tennessee and group them. We can select each state by pressing *cntrl/cmd* and clicking on it. After we have selected the states that would go into a group, we can click “Group” option in the “Create Group” box. This will result in:

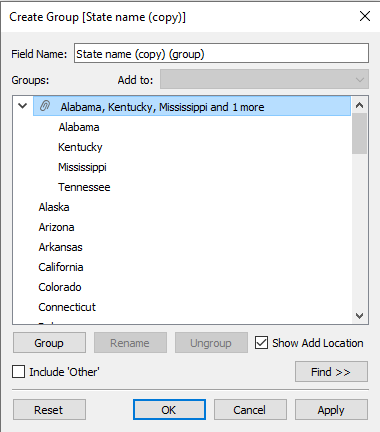


Figure : Grouping states into divisions