**Module 3 – Shape and Combine Data:**

**Shape data**

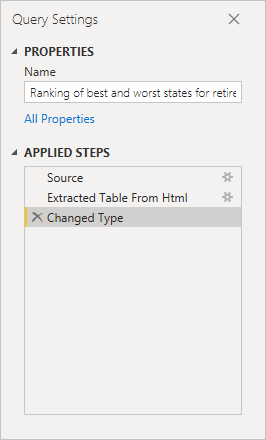
Now that you're connected to a data source, you can adjust the data to meet your needs. To *shape* data, you provide Power Query Editor with step-by-step instructions for adjusting the data while loading and presenting it. Shaping doesn't affect the original data source, only this particular view of the data.

**Note**

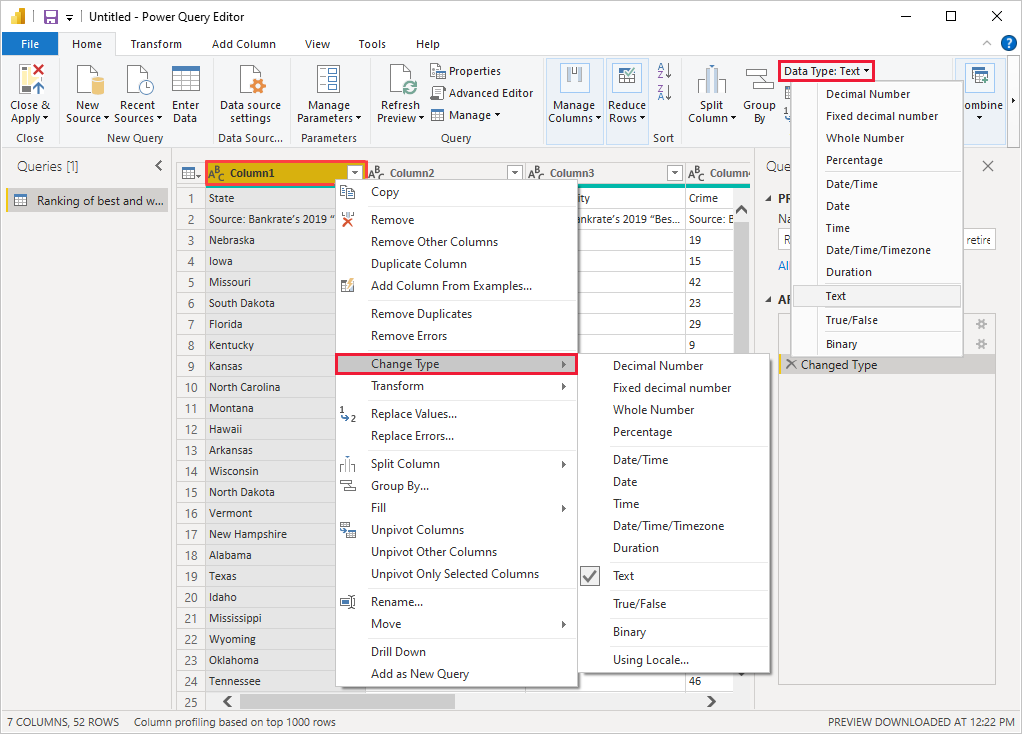
The table data used in this guide might change over time. As such, the steps you need to follow might vary, requiring you to be creative about how you adjust steps or outcomes, which is all part of the fun of learning.

Shaping can mean *transforming* the data, such as renaming columns or tables, removing rows or columns, or changing data types. Power Query Editor captures these steps sequentially under **Applied Steps** in the **Query Settings** pane. Each time this query connects to the data source, those steps are carried out, so the data is always shaped the way you specify. This process occurs when you use the query in Power BI Desktop, or when anyone uses your shared query, such as in the Power BI service.

Notice that the **Applied Steps** in **Query Settings** already contain a few steps. You can select each step to see its effect in the Power Query Editor. First, you specified a web source, and then you previewed the table in the **Navigator** window. In the third step, **Changed type**, Power BI recognized whole number data when importing it, and automatically changed the original web **Text** *data type* to **Whole numbers**.



If you need to change a data type, select the column or columns to change. Hold down the **Shift** key to select several adjacent columns, or **Ctrl** to select non-adjacent columns. Either right-click a column header, select **Change Type**, and choose a new data type from the menu, or drop down the list next to **Data Type** in the **Transform** group of the **Home** tab, and select a new data type.

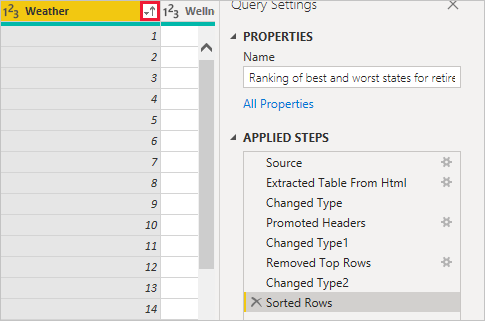


**Note**

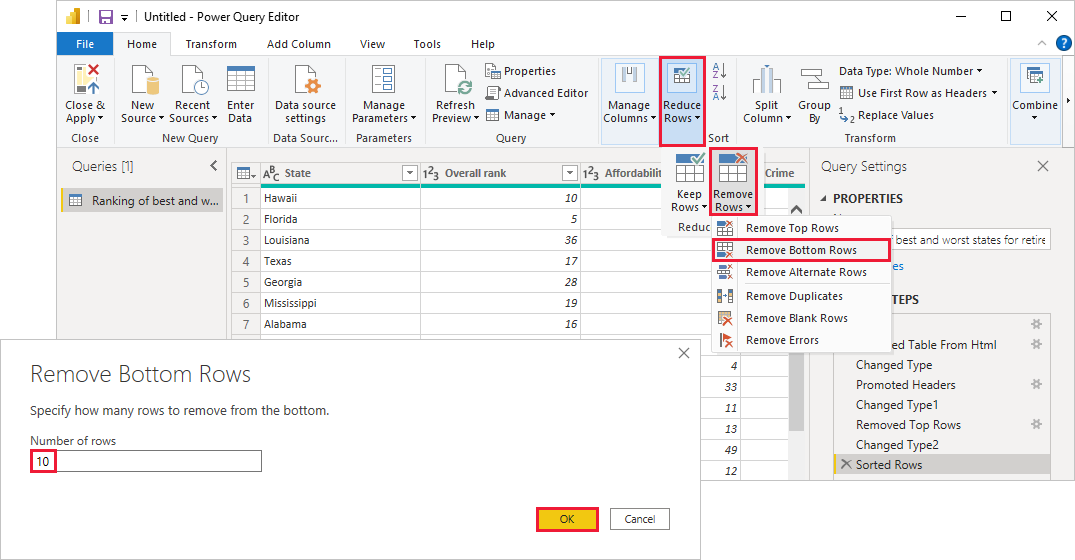
The Power Query Editor in Power BI Desktop uses the ribbon or the right-click menus for available tasks. Most of the tasks you can select on the **Home** or **Transform** tabs of the ribbon are also available by right-clicking an item and choosing from the menu that appears.

You can now apply your own changes and transformations to the data and see them in **Applied Steps**.

For example, for sunglasses sales you're most interested in the weather ranking, so you decide to sort the table by the **Weather** column instead of by **Overall rank**. Drop down the arrow next to the **Weather** header, and select **Sort ascending**. The data now appears sorted by weather ranking, and the step **Sorted Rows** appears in **Applied Steps**.



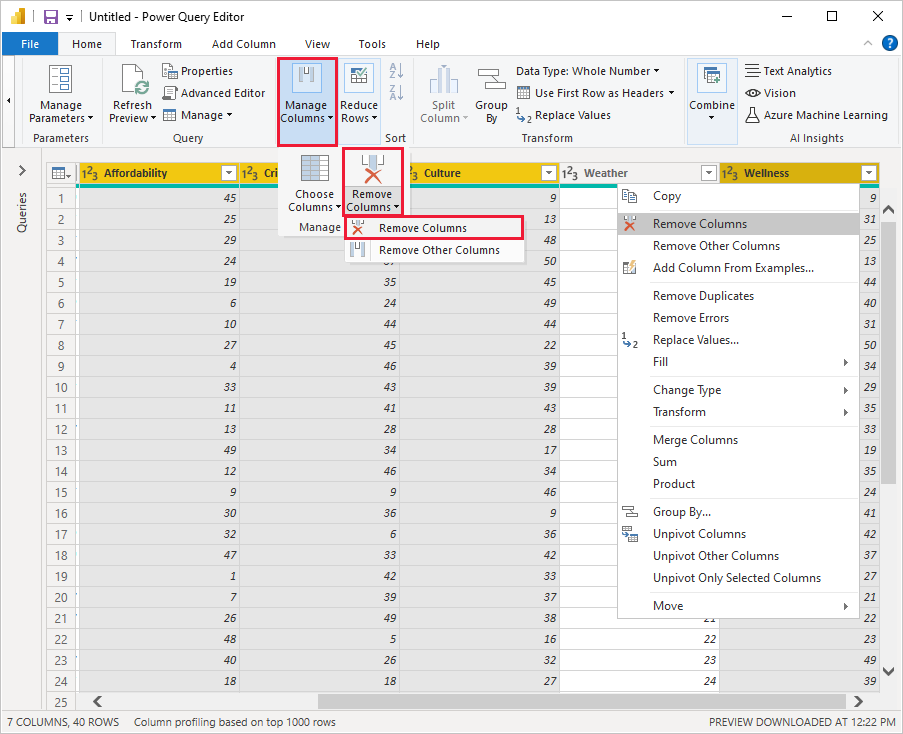
You're not very interested in selling sunglasses to the worst weather states, so you decide to remove them from the table. From the the **Home** tab, select **Reduce Rows** > **Remove Rows** > **Remove Bottom Rows**. In the **Remove Bottom Rows** dialog box, enter *10*, and then select **OK**.



The bottom 10 worst weather rows are removed from the table, and the step **Removed Bottom Rows** appears in **Applied Steps**.

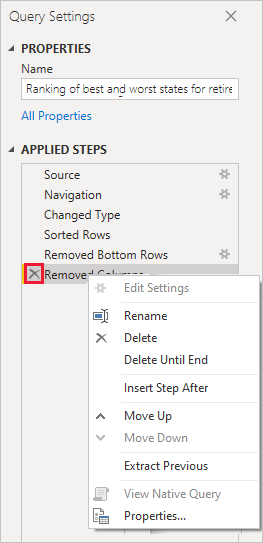
You decide the table has too much extra information for your needs, and to remove the **Affordability**, **Crime**, **Culture**, and **Wellness** columns. Select the header of each column that you want to remove. Hold down the **Shift** key to select several adjacent columns, or **Ctrl** to select non-adjacent columns.

Then, from the **Manage Columns** group of the **Home** tab, select **Remove Columns**. You can also right-click one of the selected column headers and select **Remove Columns** from the menu. The selected columns are removed, and the step **Removed Columns** appears in **Applied Steps**.

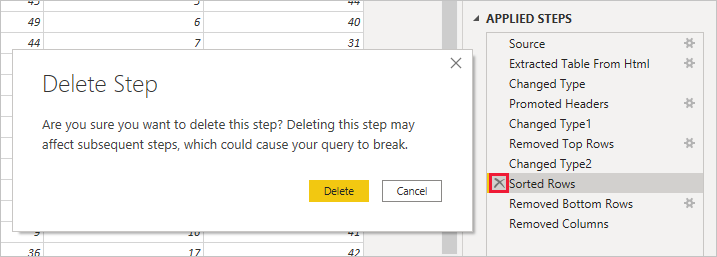


On second thought, **Affordability** might be relevant to sunglasses sales after all. You'd like to get that column back. You can easily undo the last step in the **Applied Steps** pane by selecting the **X** delete icon next to the step. Now redo the step, selecting only the columns you want to delete. For more flexibility, you could delete each column as a separate step.

You can right-click any step in the **Applied Steps** pane and choose to delete it, rename it, move it up or down in the sequence, or add or delete steps after it. For intermediate steps, Power BI Desktop will warn you if the change could affect later steps and break your query.

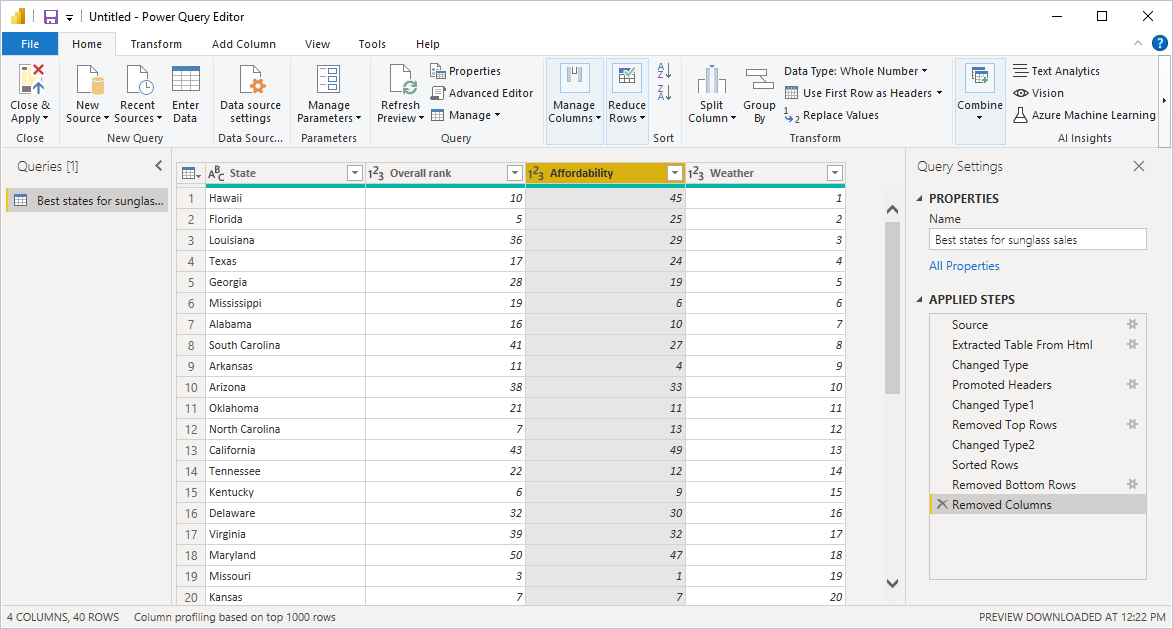


For example, if you no longer wanted to sort the table by **Weather**, you might try to delete the **Sorted Rows** step. Power BI Desktop warns you that deleting this step could cause your query to break. You removed the bottom 10 rows after you sorted by weather, so if you remove the sort, different rows will be removed. You also get a warning if you select the **Sorted Rows** step and try to add a new intermediate step at that point.



Finally, you change the table title to be about sunglass sales instead of retirement. Under **Properties** in the **Query Settings** pane, replace the old title with *Best states for sunglass sales*.

The finished query for your shaped data looks like this:



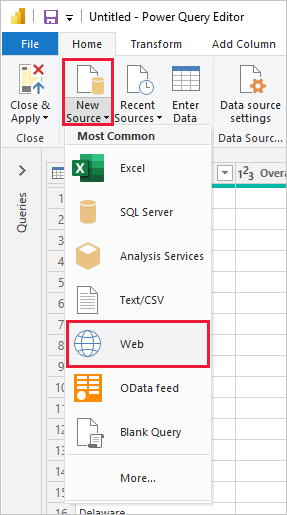
For more information about shaping data, see [Shape and combine data in Power BI Desktop](https://docs.microsoft.com/en-us/power-bi/connect-data/desktop-shape-and-combine-data).

**Combine data**

The data about various states is interesting, and will be useful for building additional analysis efforts and queries. But there's one problem: most data out there uses two-letter abbreviations for state codes, not the full names of the states. To use that data, you need some way to associate your state names with their abbreviations.

You're in luck. Another public data source does just that, but the data will need a fair amount of shaping before you can *combine* it with your sunglass table.

To import the state abbreviations data into Power Query Editor, select **New Source** > **Web** from the **New Query** group on the **Home** tab of the ribbon.

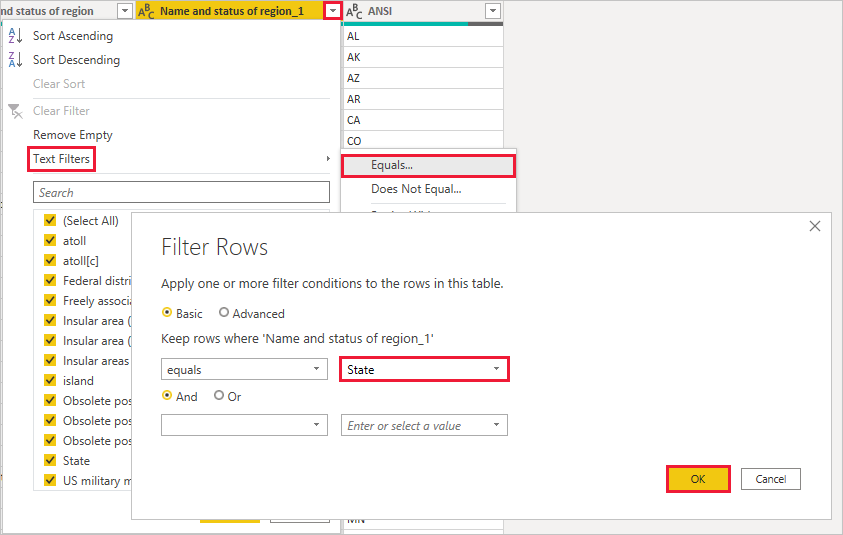


In the **From Web** dialog box, enter the URL for the state abbreviations site: *https://en.wikipedia.org/wiki/List\_of\_U.S.\_state\_abbreviations*.

In the **Navigator** window, select the table **Codes and abbreviations for U.S. states, federal district, territories, and other regions**, and then select **OK**. The table opens in Power Query Editor.

Remove all columns except for **Name and status of region**, **Name and status of region**, and **ANSI**. To keep only these columns, hold down **Ctrl** and select the columns. Then, either right-click one of the column headers and select **Remove Other Columns**, or, from the **Manage Columns** group of the **Home** tab, select **Remove Other Columns**.

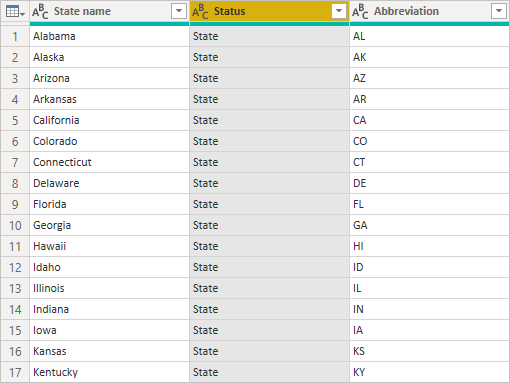
Drop down the arrow next to the **Name and status of region\_1** column header, and select **Filters** > **Equals**. In the **Filter Rows** dialog box, drop down the **Enter or select a value** field next to **equals** and select **State**. Select **OK**.



With extra values like **Federal district** and **island** removed, you now have a list of the 50 states and their official two-letter abbreviations. You can rename the columns to make more sense, for example **State name**, **Status**, and **Abbreviation**, by right-clicking the column headers and selecting **Rename**.

Note that all of these steps are recorded under **Applied Steps** in the **Query Settings** pane.

Your shaped table now looks like this:



Retitle the table to *State codes* in the **Properties** field of **Query Settings**.

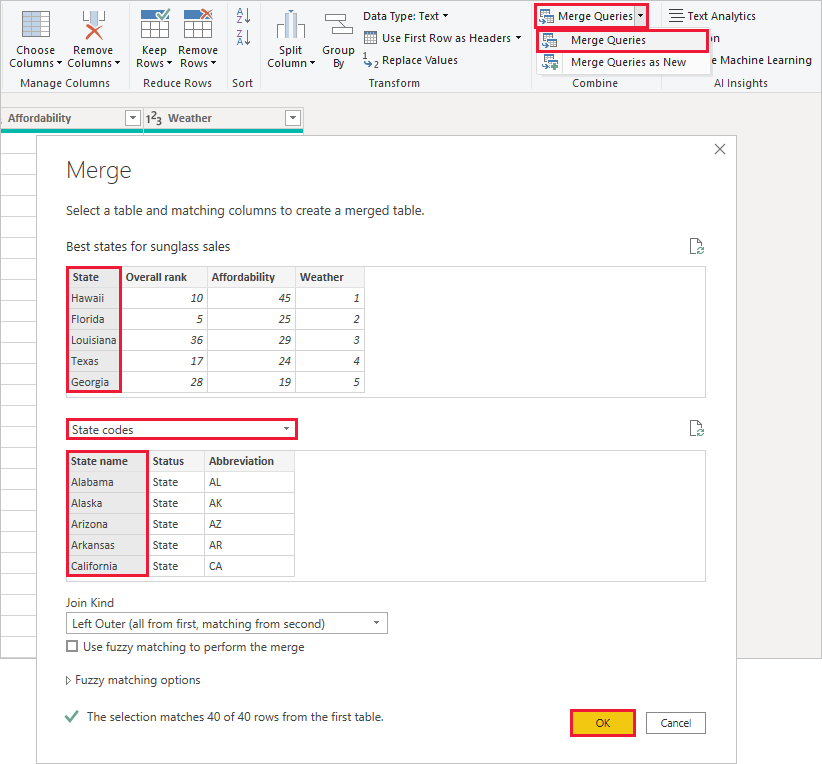
With the **State codes** table shaped, you can *combine* these two tables into one. Since the tables you now have are a result of queries you applied to the data, they're also called *queries*. There are two primary ways of combining queries: *merge* and *append*.

When you have one or more columns you'd like to add to another query, you *merge* the queries. When you have additional rows of data you'd like to add to an existing query, you *append* the query.

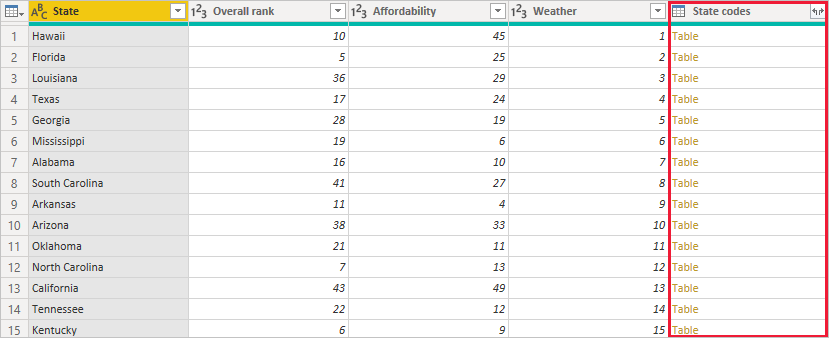
In this case, you want to *merge* the **State codes** query into the **Best states for sunglasses** query. To merge the queries, switch to the **Best states for sunglasses** query by selecting it from the **Queries** pane on the left side of Power Query Editor. Then select **Merge Queries** from the **Combine** group in the **Home** tab of the ribbon.

In the **Merge** window, drop down the field to select **State codes** from the other queries available. Select the column to match from each table, in this case **State** from the **Best states for sunglasses** query and **State name** from the **State codes** query.

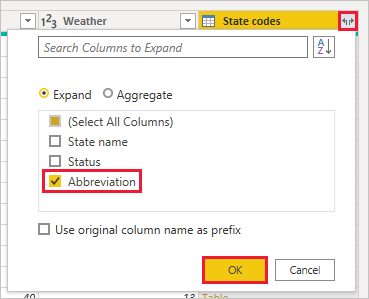
If you get a **Privacy levels** dialog, select **Ignore privacy levels checks for this file** and then select **Save**. Select **OK**.



A new column called **State codes** appears on the right of the **Best states for sunglass sales** table. It contains the state code query that you merged with the best states for sunglass sales query. All the columns from the merged table are condensed into the **State codes** column. You can *expand* the merged table and include only the columns you want.



To expand the merged table and select which columns to include, select the **Expand** icon in the column header. In the **Expand** dialog box, select only the **Abbreviation** column. Deselect **Use original column name as prefix**, and then select **OK**.

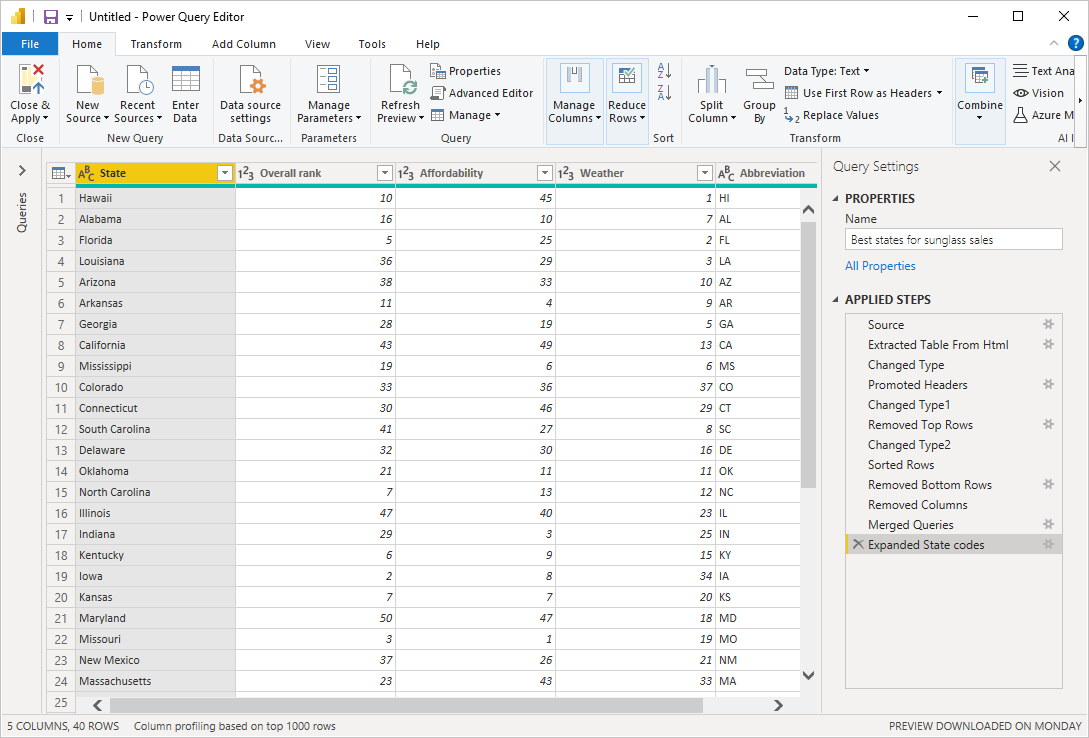


**Note**

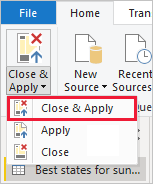
You can play around with how to bring in the **State codes** table. Experiment a bit, and if you don't like the results, just delete that step from the **Applied Steps** list in the **Query Settings** pane. It's a free do-over, which you can do as many times as you like until the expand process looks the way you want it.

For a more complete description of the shape and combine data steps, see [Shape and combine data in Power BI Desktop](https://docs.microsoft.com/en-us/power-bi/connect-data/desktop-shape-and-combine-data).

You now have a single query table that combines two data sources, each of which has been shaped to meet your needs. This query can serve as a basis for lots of additional, interesting data connections, such as demographics, wealth levels, or recreational opportunities in the states.



For now, you have enough data to create an interesting report in Power BI Desktop. Since this is a milestone, apply the changes in **Power Query Editor** and load them into Power BI Desktop by selecting **Close & Apply** from the **Home** tab of the ribbon. You can also select just **Apply** to keep the query open in Power Query Editor while you work in Power BI Desktop.



You can make more changes to a table after it is loaded into Power BI Desktop, and reload the model to apply any changes you make. To reopen **Power Query Editor** from Power BI Desktop, select **Edit Queries** on the **Home** tab of the Power BI Desktop ribbon.