..

Working with Data

Date published: 2020-10-30
Date modified: 2022-09-21

https://docs.cloudera.com/

Legal Notice

Contents

Working with datasets in CDP Data Visualization

Datasets are defined on the connections to your data, and provide access to the specific tables in the data store.

In CDP Data Visualization, visualizations are built from datasets. These datasets provide access to data annd enhance data access and usage.

Related Information

Datasets

Creating a dataset

CDP Data Visualization allows you to create datasets based on your data.

About this task

There are three options for creating datasets:

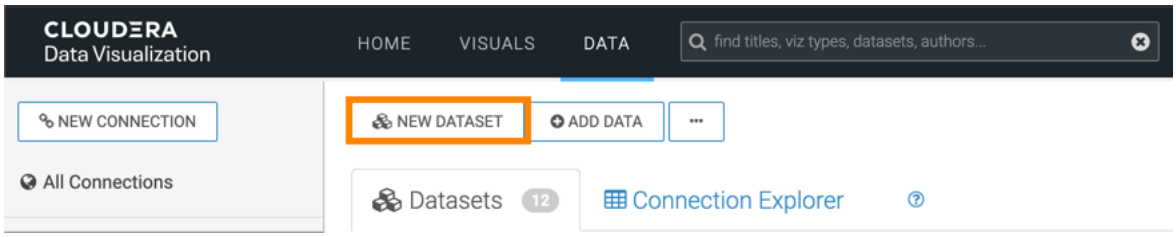
- Creating datasets in the Data interface on page 4
- Creating datasets from a query on page 5
- Creating datasets in the Connection Explorer on page 6

Creating datasets in the Data interface

Procedure

1. On the main navigation bar, click DATA.

   The Data view appears, open on the Datasets tab.
2. Click NEW DATASET near the top of the screen.



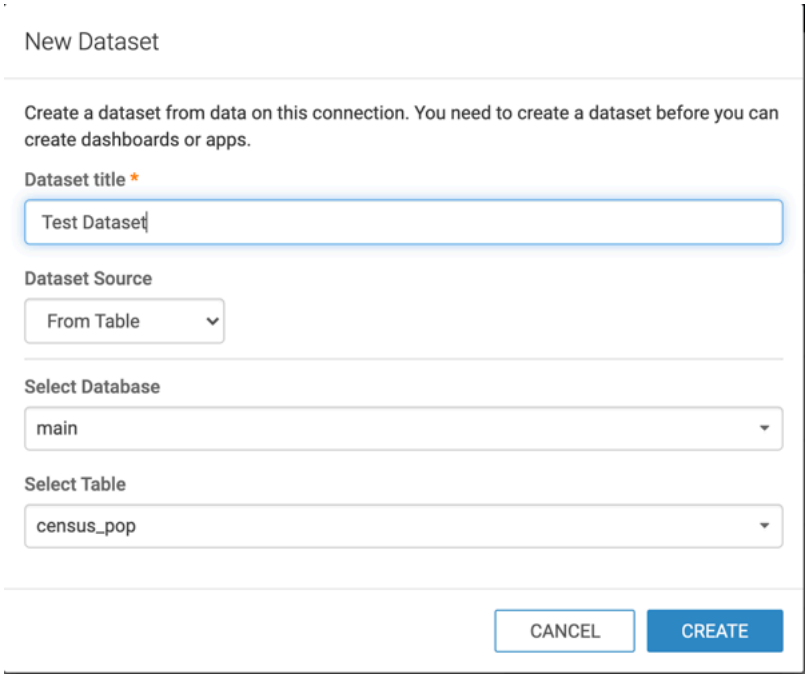
3. In the New Dataset modal window, specify the following values:
   a) Dataset title

      Specify here the name of the new dataset.
   b) Dataset Source

      In the menu, you can choose either the From Tables (default) or From SQL option. Leave the selection as From Tables.
   c) Select Database.

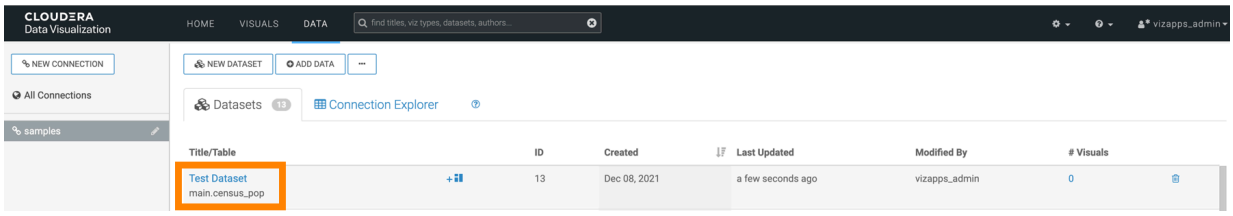
      Scroll the list of connected databases to select the correct database.
   d) Select Table

      Scroll the list of tables to select the correct table.



4. Click CREATE.

   You can now see the new dataset under the Datasets tab.



Note: To find the dataset in the future, you can scroll through the list of datasets on the connection, or use Search at the top of the page.

Creating datasets from a query

Data Visualization lets you specify a dataset from a SQL Query. This feature makes it easy to restrict access to specific table columns or rows for all users, either for security or for relevancy reasons. It also enables you to specify complex joins and analytic functions at the dataset level.

Procedure

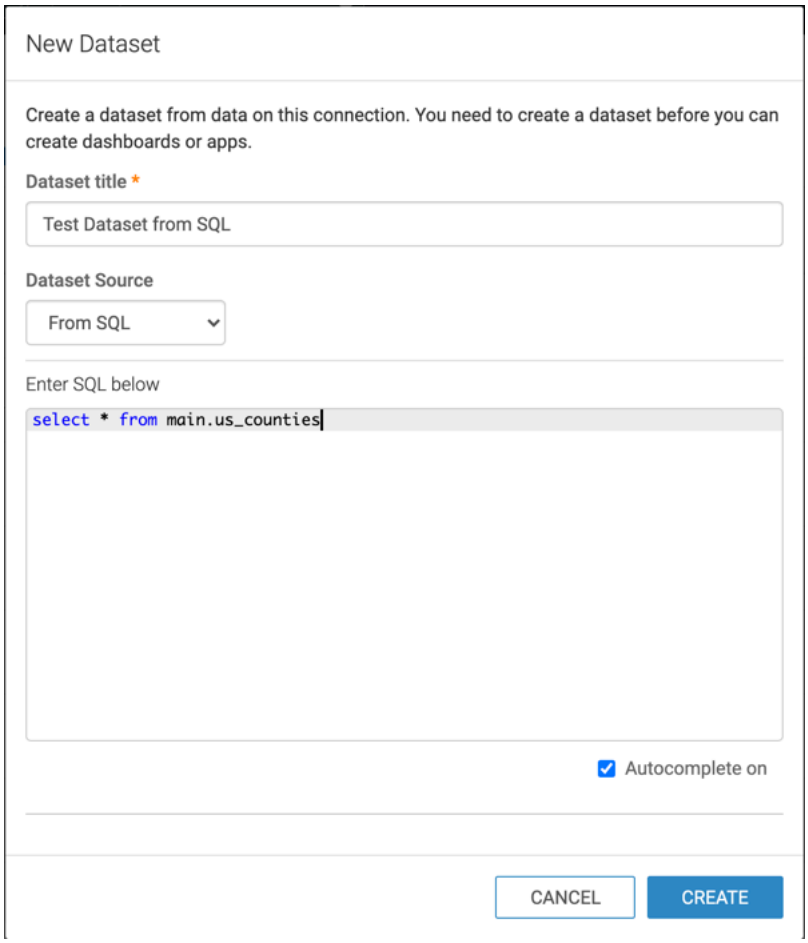
1. On the main navigation bar, click DATA.

   The Data view appears, open on the Datasets tab.
2. Click NEW DATASET near the top of the screen.
3. In the New Dataset modal window, specify the following values:
   a) Dataset title

      Specify here the name of the new dataset.
   b) Dataset Source

      In the menu, you can choose either the From Tables (default) or From SQL option. Choose From SQL.
   c) Enter SQL

      Enter the SQL query; its results from the dataset.



4. Click CREATE.

   You can now see the new dataset under the Datasets tab.

   Tip: To find the dataset in the future, you can scroll through the list of datasets on the connection, or use Search at the top of the page.

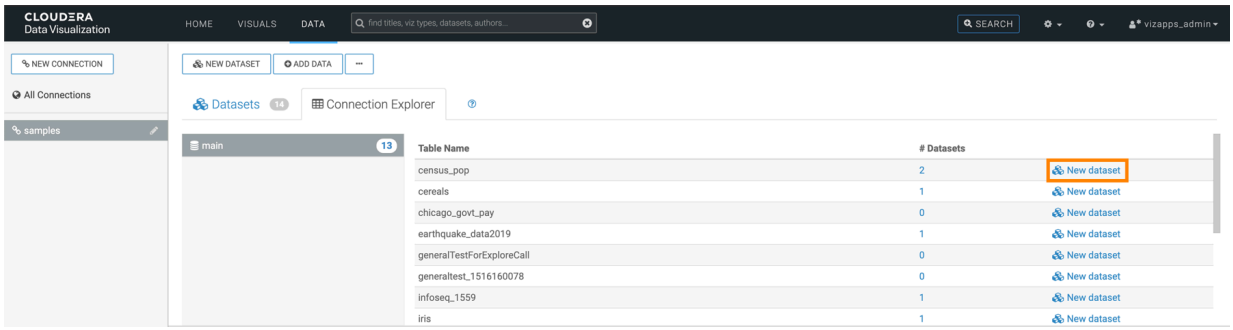
Creating datasets in the Connection Explorer

The following steps demonstrate how to create a new dataset directly on a table in the Connection Explorer.

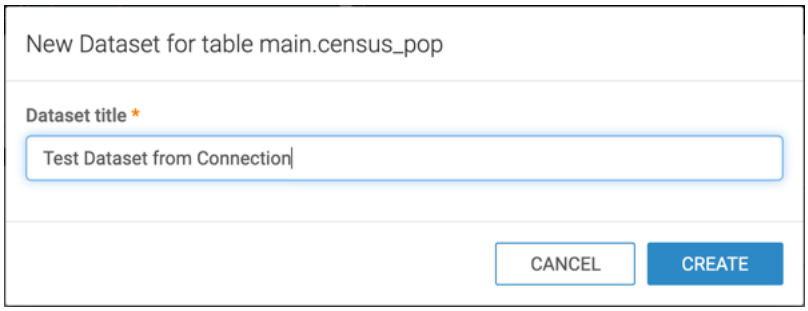
Procedure

1. On the main navigation bar, click DATA.

   The Data view appears, open on the Datasets tab.
2. Click New Dataset on the same line as the primary table.



3. In the New Dataset modal window, specify the Dataset title.



4. Click CREATE.

   You can now see the new dataset under the Datasets tab.

   Tip: To find the dataset in the future, you can scroll through the list of datasets on the connection, or use Search at the top of the page.

Finding a dataset

CDP Data Visualization makes it easy to find your datasets. You can browse the list of Datasets, check out the Connection Explorer, or use Search.

Procedure

1. On the main navigation bar, click DATA.

   The Data view appears, open on the Datasets tab.
2. On the left-side panel, select the connection on which the dataset is defined.
3. Use Search on the main navigation bar. Type the search string that matches all or part of the dataset.

   Some examples of search criteria may be data connection names, data table names, dataset names, visualization names, or application names.
4. Select the correct dataset from the abbreviated list of datasets.

Exploring dataset details

CDP Data Visualization makes it easy to check out and examine the information available about your datasets.

Procedure

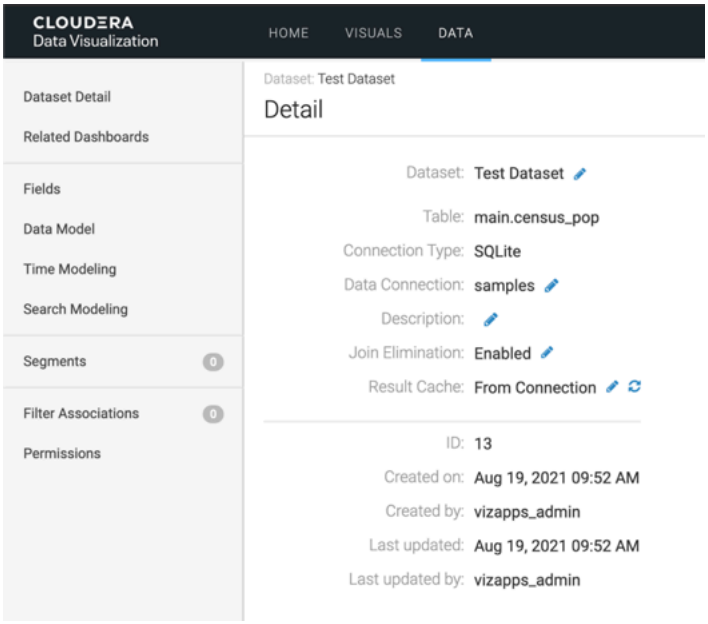
1. On the main navigation bar, click DATA.

   The Data view appears, open on the Datasets tab.
2. Find the dataset, either by browsing to a known connection and scrolling, or by using Search.

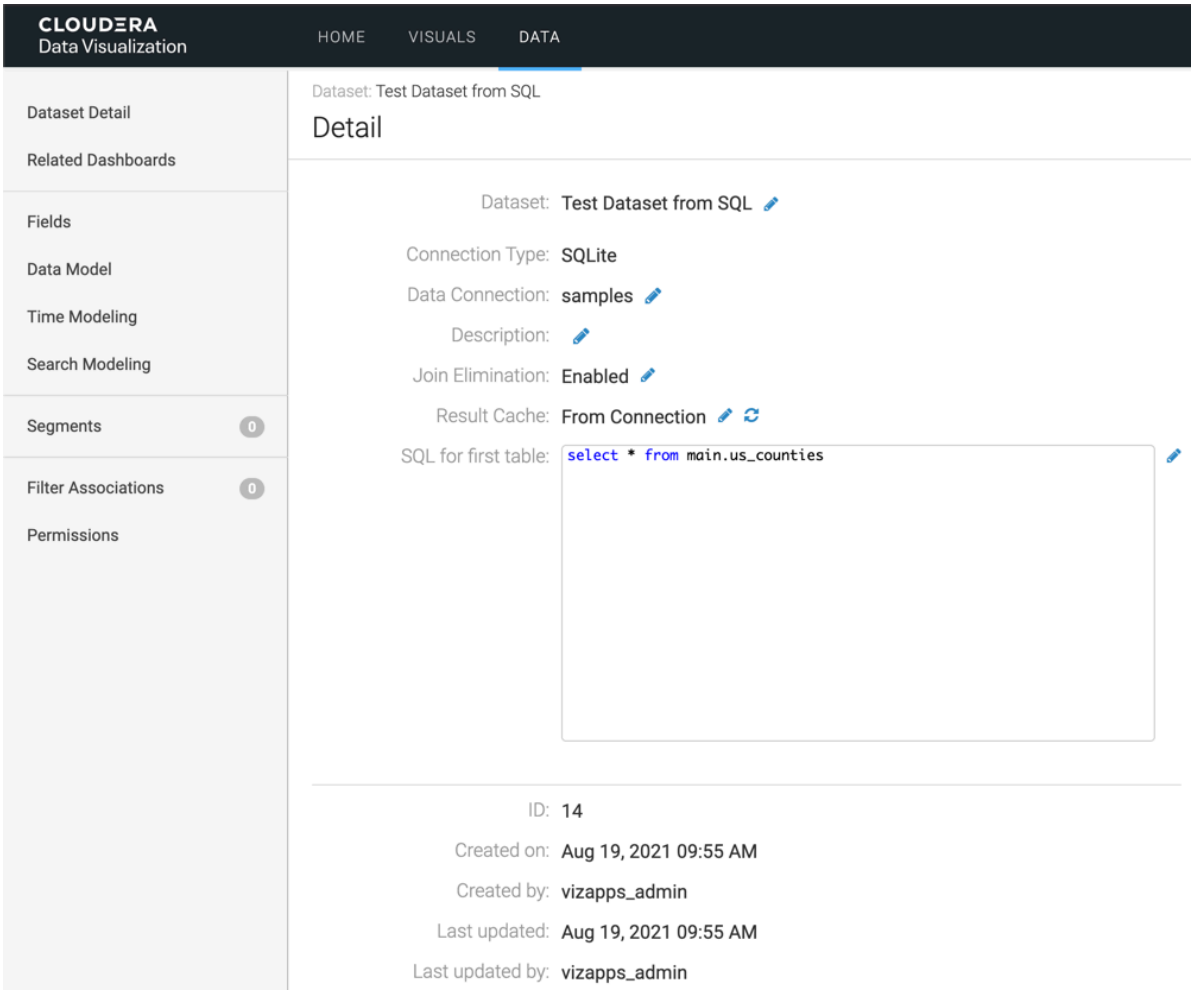
3. Click the dataset you want to examine.
   Dataset side navigation appears, open at Dataset Detail view.

   Depending on the dataset definition approach, there are two alternatives:

   - Defined on Table



   - Defined on SQL



The following information is available on Dataset Detail view, under Detail:

Tables

This is the qualified name of the data source. It appears in the form DatabaseName.DatabaseTable.

Connection type

This is the name of the database that hosts the data that appears in the form DataConnection.This feature is ideal for enterprise environments, with dashboards developed on test clusters, and then deployed to a production environment.

About this task

This assumes that the new connection has, at the minimum, the relevant base tables with metadata definitions that match those on the original connection. Users with appropriate permissions can switch the data connection of a dataset by following these steps.

Procedure

1. Click the Edit (pencil) icon.
2. Choose a different data connection from the menu.
3. Click Save.

Description

This is an optional field. You can add a description of the dataset in the available textbox.

Procedure

1. Click the Edit (pencil) icon.
2. Enter the descirption in the text box.
3. Click Save.

Join elimination

Join elimination improves query execution and visual rendering in CDP Data Visualization.

About this task

Join elimination is available both for left outer and inner joins. It is turned on by default.

When a visual uses fields and expressions that reference only a subset of the joined tables that form the dataset, this feature eliminates the unnecessary joins and access only the necessary subset of the joined tables. This improves query execution speeds, and renders the visuals faster.

Procedure

1. Click the Edit (pencil) icon.
2. You can eanble or disable join elimination:
   - Disabling: select the Disabled option.
   - Enabling: select the Enabled option.
3. Click Save.

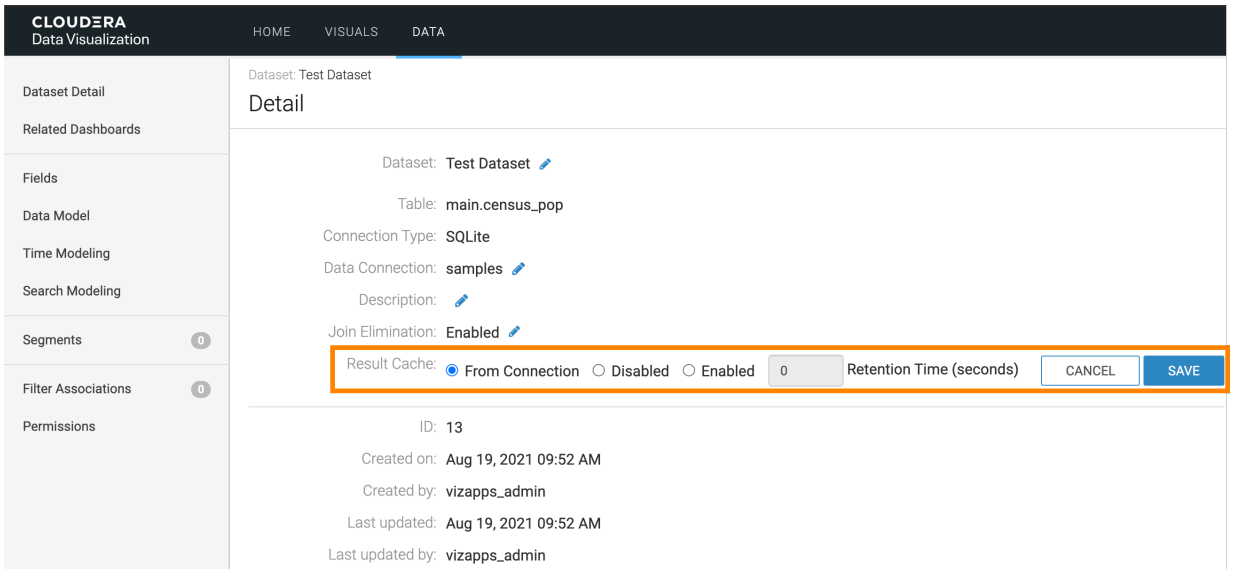
Result cache

Each dataset inherits the result caching preferences configured for its connection. The value for this field is From Connection by default, but this may be changed at the level of the dataset.

Procedure

1. Click the Edit (pencil) icon.
2. You can enable or disable the result cache:
   - Disabling: select the Disabled option.
   - Enabling: select the Enabled option, and specify the Retention Time, in seconds.
3. Click Save.

4. To clear the result cache, click the Clear icon.



SQL

In datasets initially defined on a SQL query, you can alter the query at any time to change the list of fields fetched, content of the WHERE clause, ORDER BY, and so on.

Procedure

1. Click the Edit (pencil) icon.
2. Edit the SQL statement.
3. Click Save.

Information on creation and update

The Dataset Detail interface provides information about the creation of the dataset and most recent updates to it.

Created on

This is the date, in timestamp form, when the dataset was created.

Created by

This is the username of the user who created the dataset.

Last updated

This is the date, in timestamp form, of the most recent dataset update.

Last updated by

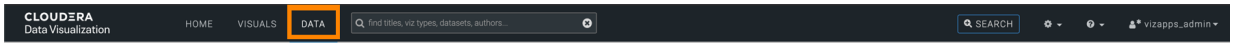
This is the username of the user who updated the dataset most recently.

Checking related dashboards

In CDP Data Visualization, you can easily determine which visuals use a particular dataset.

Procedure

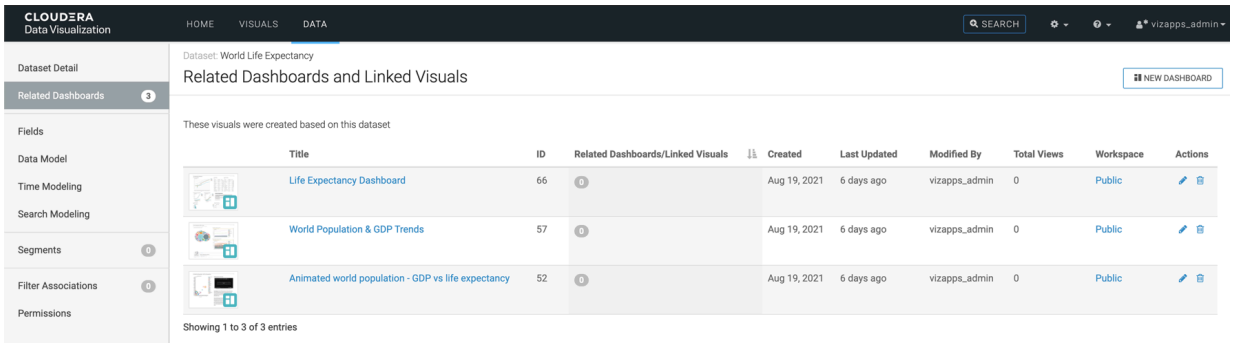
1. On the main navigation bar, click Data.



The Data view appears, open on the Datasets tab.

2. Find the dataset that you want to examine in the list of datasets, either by scrolling or by using search.
3. Click the dataset.

   Dataset side navigation appears, open at Dataset Detail view.
4. In the side navigation menu, click Related Dasboards.



   Related Dashboards view appears, which is a list of visuals that use this dataset.

   The following information is available for each app in this list, under Related Dashboards:

   - Type icon represents the style of the dashboard, or a snapshot icon of the visual (if this feature is on).
   - Title is the name of the visual.
   - ID
   - Related Dashboards/Linked Visuals
   - Created is the date when the visual was created.
   - Last Updated is the time interval after the last update of the visual. It is expressed in minutes, hours, days, or months, as appropriate.
   - Modified by is the name of the user who modified the app most recently.
   - Total Views is the number of times the app was viewed.
   - Workspace
   - Actions are the permissions available to you They may include the following:
     - Clicking Edit (pencil) icon edits the visual.
     - Clicking Delete (trash) icon deletes the visual.

Editing datasets based on a SQL query

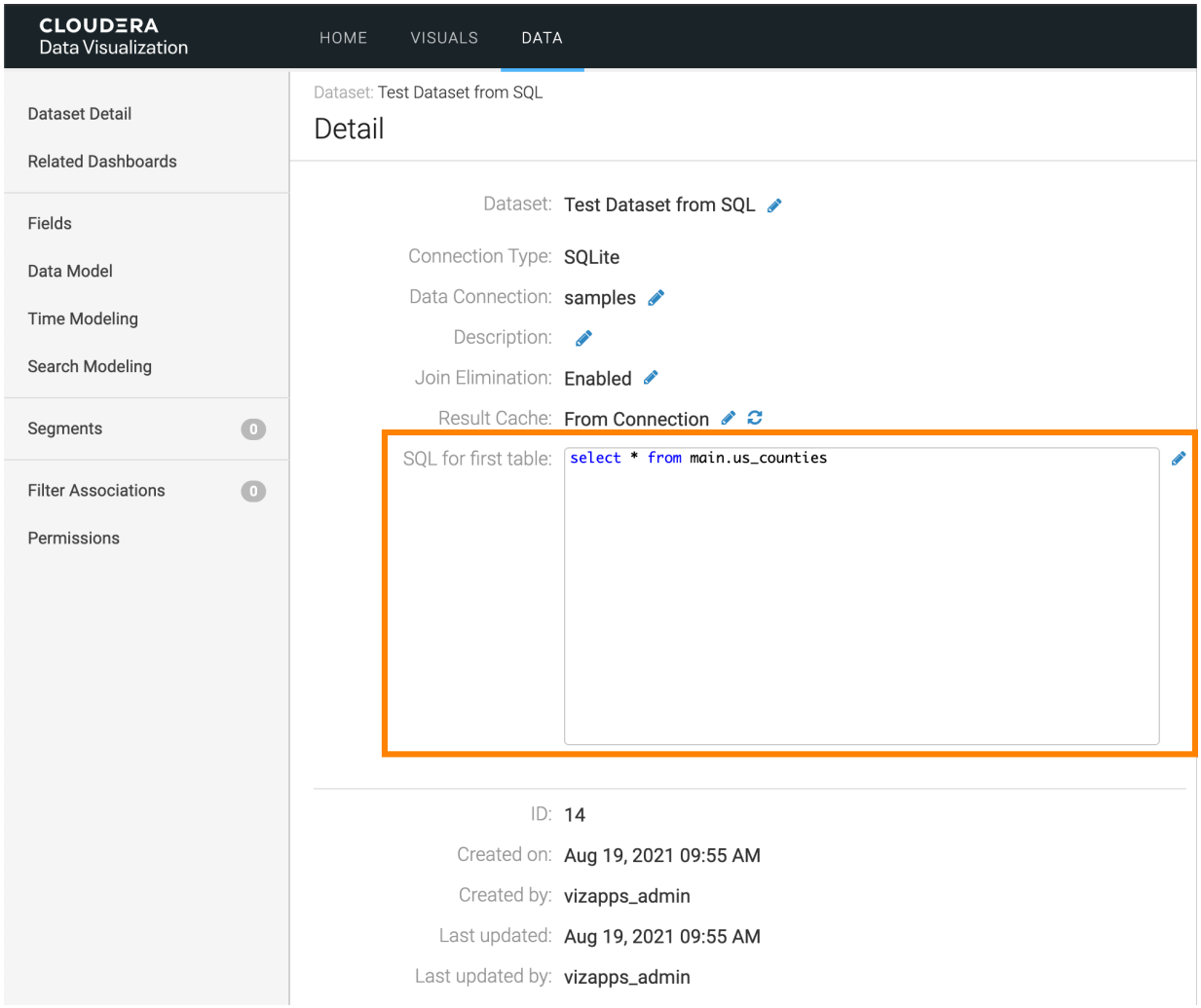
One of the major advantages CDP Data Visualization provides is the option to edit the data selection that defines the dataset.

Procedure

1. Navigate to a dataset that you created based on a SQL query.

2. Click to open the dataset, on the Dataset Detail view.

   There is a SQL text window, which you can edit.



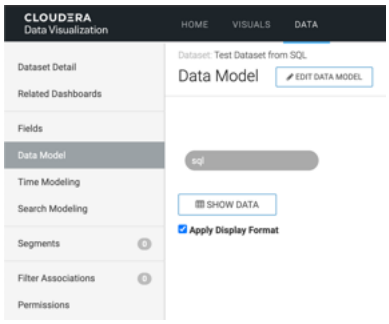
3. You can restrict rows and columns depending on what you need.

Restricting columns in datasets based on SQL query

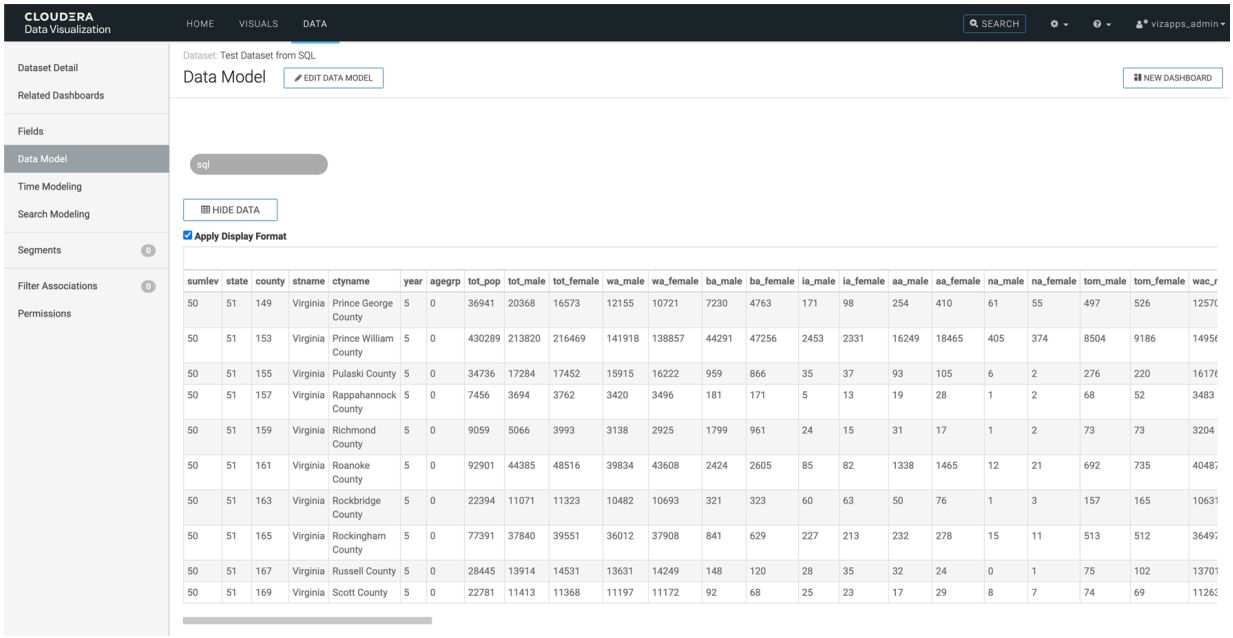
In CDP Data Visualization, you can easily restrict the table columns in the dataset by changing the SQL definition of that dataset. SQL-defined datasets make it easy to limit their content to specific columns.

Procedure

1. Switch to Data Model interface, and click Show Data.



There is a large number of columns in the query result, and many of them are not necessary when it comes to answering most common questions.

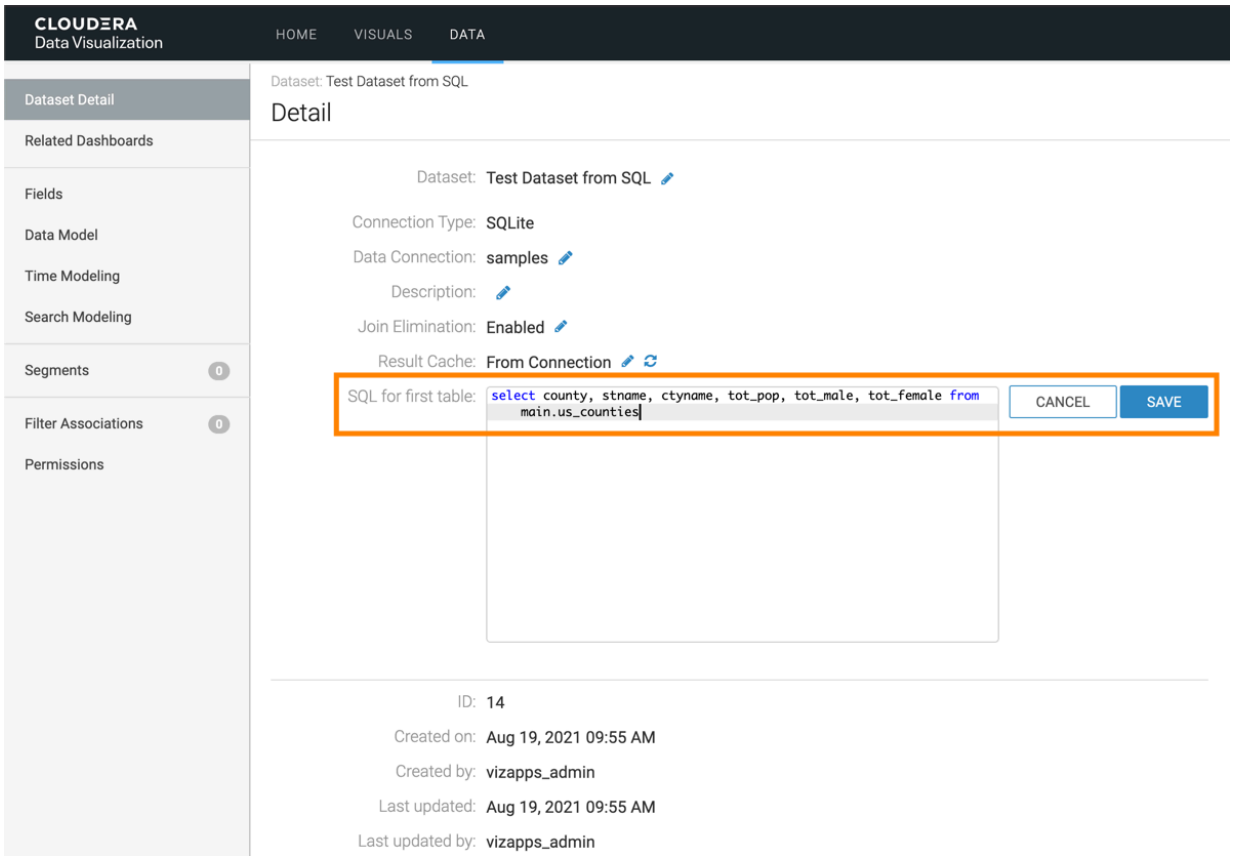


2. Find the fields that you would like to keep in the dataset definition.
3. Switch back to Dataset Detail interface, and edit SQL text window by applying the following statement:

```
select county, stname, ctyname, tot_pop, tot_male, tot_female from main.us_counties
```

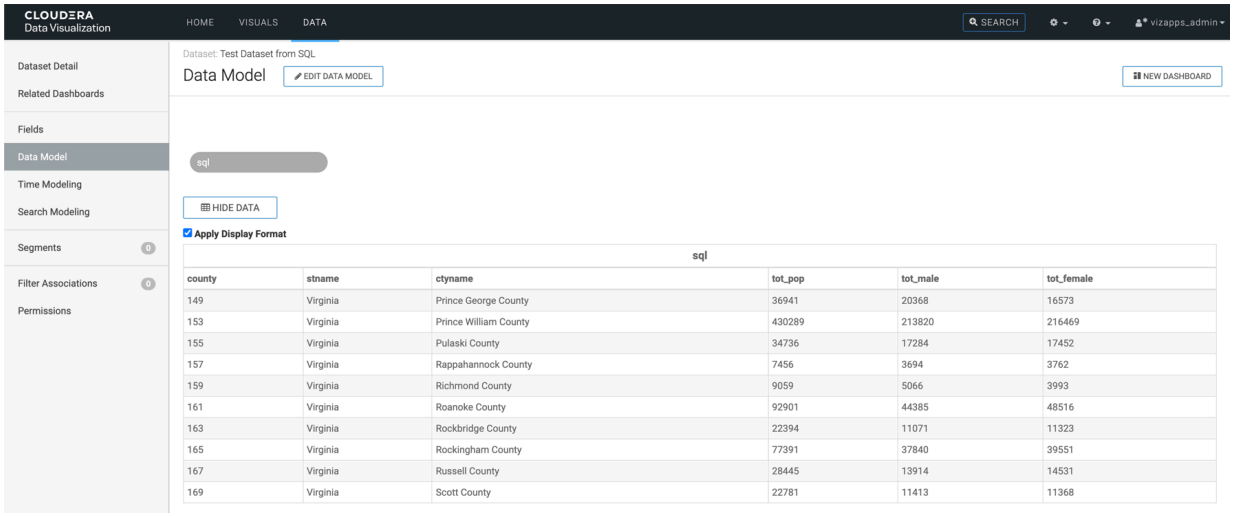
In this example we keep the columns county, stname, ctyname, tot_pop, tot_male, and tot_female.

4. Click Save.



5. In the Refresh dataset table column information modal window, click Close.

6. Switch back to the Data Model interface, click Show Data, and check that the dataset only has the explicitly specified columns:



sql

county	stname	ctyname	tot_pop	tot_male	tot_female
149	Virginia	Prince George County	36941	20368	16573
153	Virginia	Prince William County	430289	213820	216469
155	Virginia	Pulaski County	34736	17284	17452
157	Virginia	Rappahannock County	7456	3694	3762
159	Virginia	Richmond County	9059	5066	3993
161	Virginia	Roanoke County	92901	44385	48516
163	Virginia	Rockbridge County	22394	11071	11323
165	Virginia	Rockingham County	77391	37840	39551
167	Virginia	Russell County	28445	13914	14531
169	Virginia	Scott County	22781	11413	11368

In this example we have kept the columns county, stname, ctyname, tot_pop, tot_male, and tot_female.

Restricting rows in datasets based on SQL query

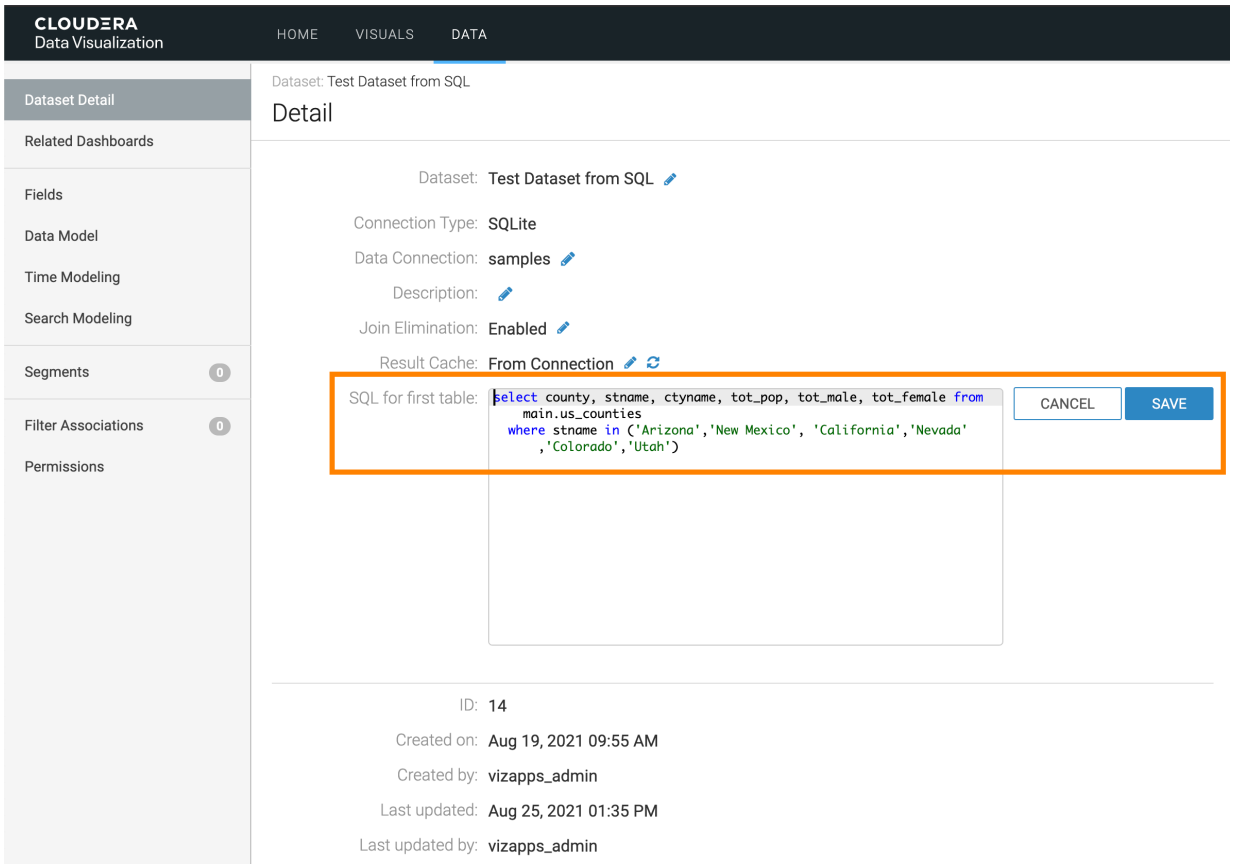
In CDP Data Visualization, you can easily restrict the table rows in the dataset by changing the SQL definition of that dataset. SQL-defined datasets make it easy to limit their content to specific rows.

Procedure

1. Switch to Dataset Detail interface, and edit SQL text window by applying the following statement:

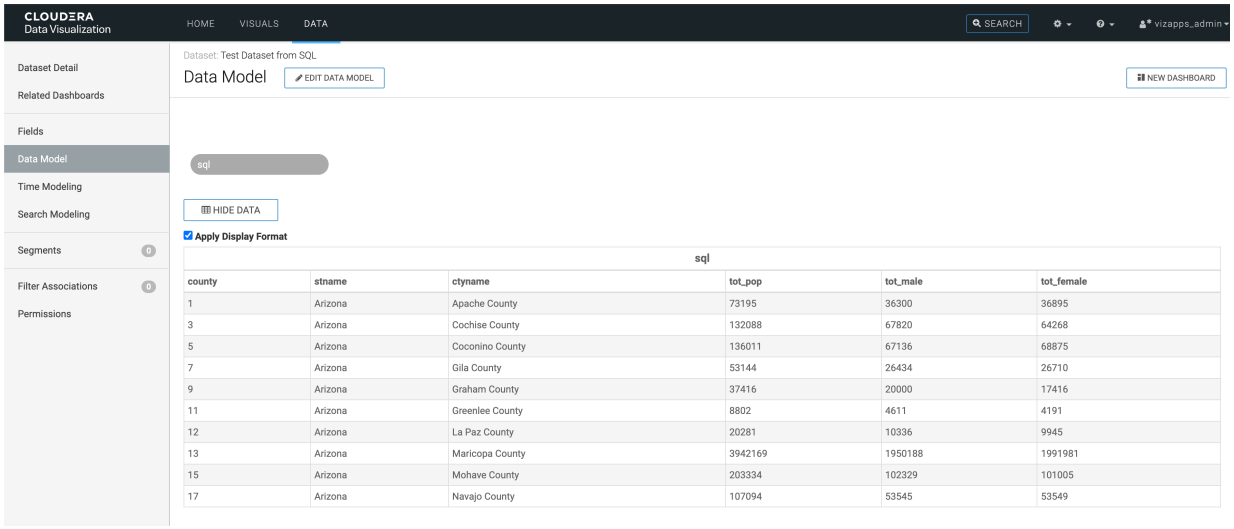
```
select county, stname, ctyname, tot_pop, tot_male, tot_female from main.
us_counties
  where stname in ('Arizona','New Mexico',
 'California','Nevada','Colorado','Utah')
```

2. Click Save.



3. In the Refresh dataset table column information modal window, click Close.

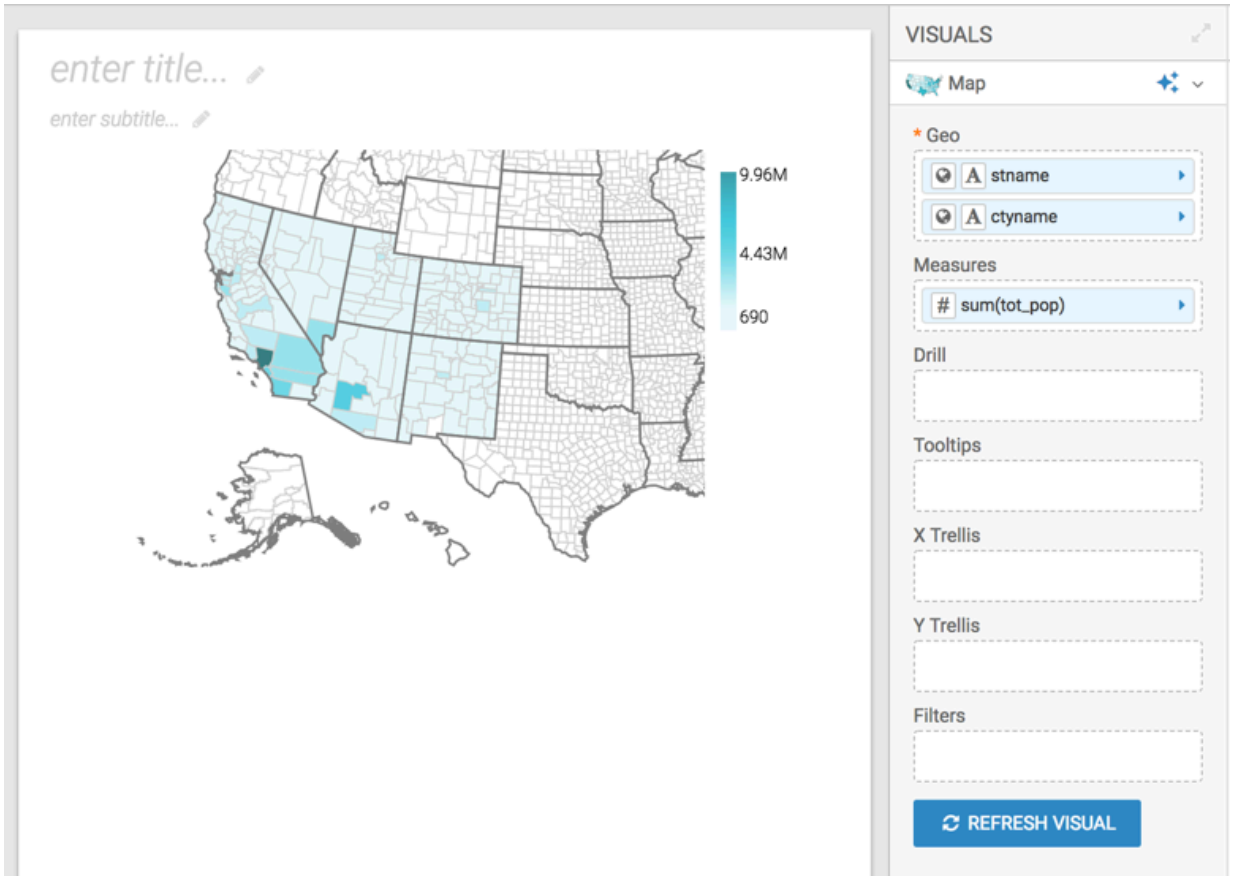
4. Switch back to the Data Model interface, click Show Data, and notice that the dataset is limited to the states specified in the SQL statement.



sql

county	stname	ctyname	tot_pop	tot_male	tot_female
1	Arizona	Apache County	73195	36300	36895
3	Arizona	Cochise County	132088	67820	64268
5	Arizona	Coconino County	136011	67136	68875
7	Arizona	Gila County	53144	26434	26710
9	Arizona	Graham County	37416	20000	17416
11	Arizona	Greenlee County	8802	4611	4191
12	Arizona	La Paz County	20281	10336	9945
13	Arizona	Maricopa County	3942169	1950188	1991981
15	Arizona	Mohave County	203334	102329	101005
17	Arizona	Navajo County	107094	53545	53549

5. If you were to test it by creating a simple map visual on the dataset, it would look something like this:



Deleting a dataset

In CDP Data Visualization, you can delete a dataset without deleting the data from the database.

About this task

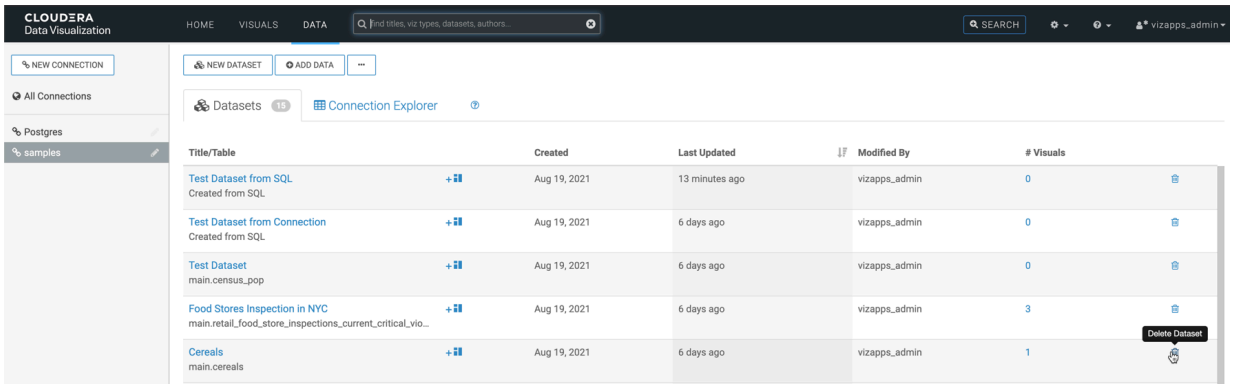
Note: Deleting a dataset deletes all dashboards and both linked and unlinked visuals that use it. However, it does not delete the data from the database. Any analytical views that were defined on the dataset remain.

Procedure

1. On the main navigation bar, click Data.

   The Data view appears, open on the Datasets tab.
2. Find the dataset in the list of datasets, either by scrolling or by using search.
3. On the row that represents the particular dataset, click the Delete (trash) icon to delete the dataset.

   In this example, the Cereals dataset has been selected, which contains 1 visual.



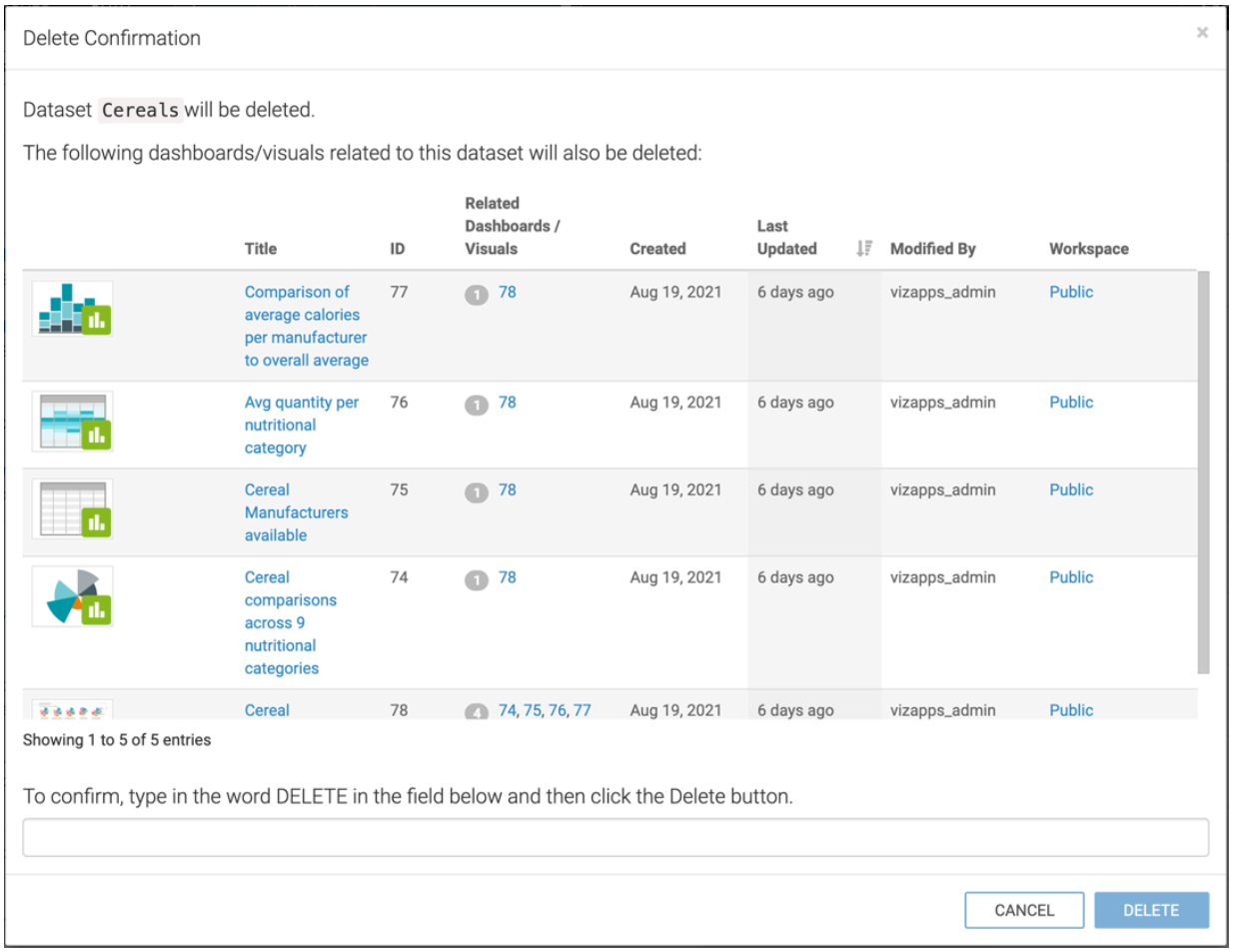
   The Delete Confirmation modal window appears.

   It contains information about all related dashboards and linked and unlinked visuals that the system deletes with the dataset:

   - *Snapshot* with a tag indicating that the artifact is a visual or a dashboard
   - *Title*
   - *ID*
   - *Related Dashboards / Visuals* lists the number of related artifacts, and their IDs.
     - For visuals, this is the number and IDs of dashboards where they appear.
     - For dashboards, it is the number and IDs of visuals that they contain.
   - *Created* date
   - Last Updated time period
   - *Modified by* username

   Note: When deleting a dataset that has associated dashboards in private worksapces, these dashboards are not listed in the Delete confirmation modal window. However, a warning will appear notifying about the dashboards.

4. In the Delete confirmation modal window's text entry field, type DELETE in uppercase, and click Delete.



In the Dataset modal window, you can see the Cereals dataset shows 1 visuals. This represents linked visuals and dashboards that reference the dataset directly. When you delete this dataset, the total number of entries display 5. This represents all dashboards, linked visuals, and all other visuals that refernce this dataset.

Changing dataset fields

It is easy to make changes to the fields of a dataset in CDP Data Visualization.

Hiding dataset fields from applications

You may find it useful to hide dataset fields that are not typically used for visualizations to prevent unintended bias in BI and analytics, or even to obscure confidential data. In CDP Data Visualization, you can do this by turning off the default visibility option of a particular dataset field.

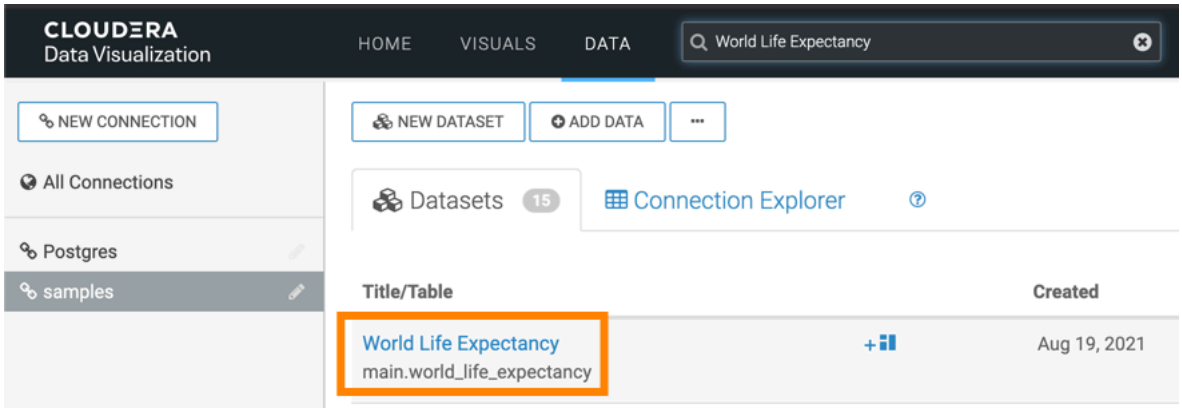
About this task

The following steps demonstrate how to prevent data fields from appearing in visualizations and applications of dataset World Life Expectancy [data source samples.world_life_expectancy]. The fields comments, lat, and lng are empty, so they are good candidates for this operation.

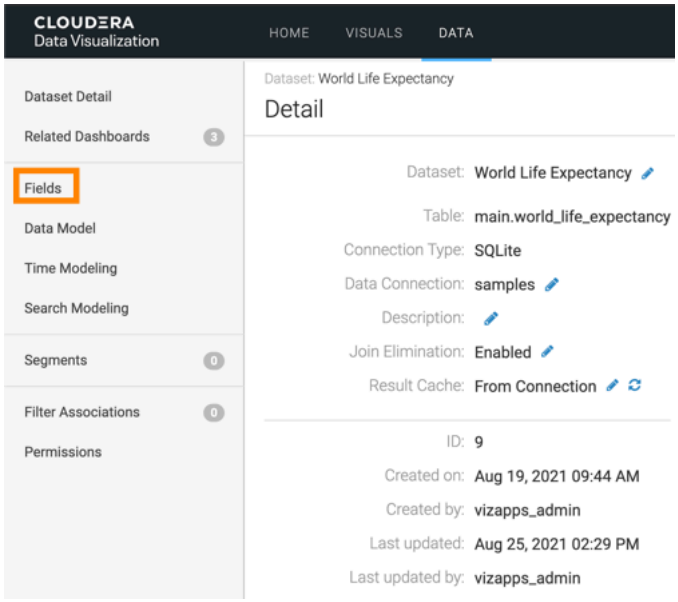
Procedure

1. On the main navigation bar, click Data.

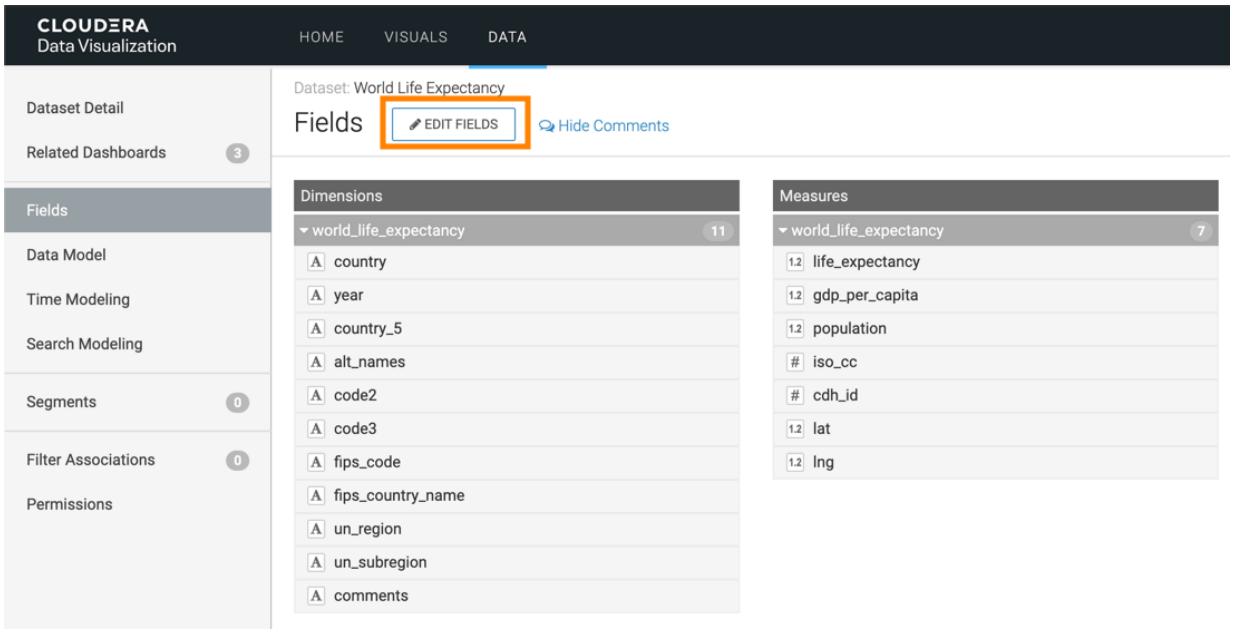
   The Data view appears, open on the Datasets tab.
2. In the left navigation menu, click Samples.
3. In the Datasets area, select World Life Expectancy (main.world_life_expectancy).



4. In the Dataset Detail menu, select Fields.

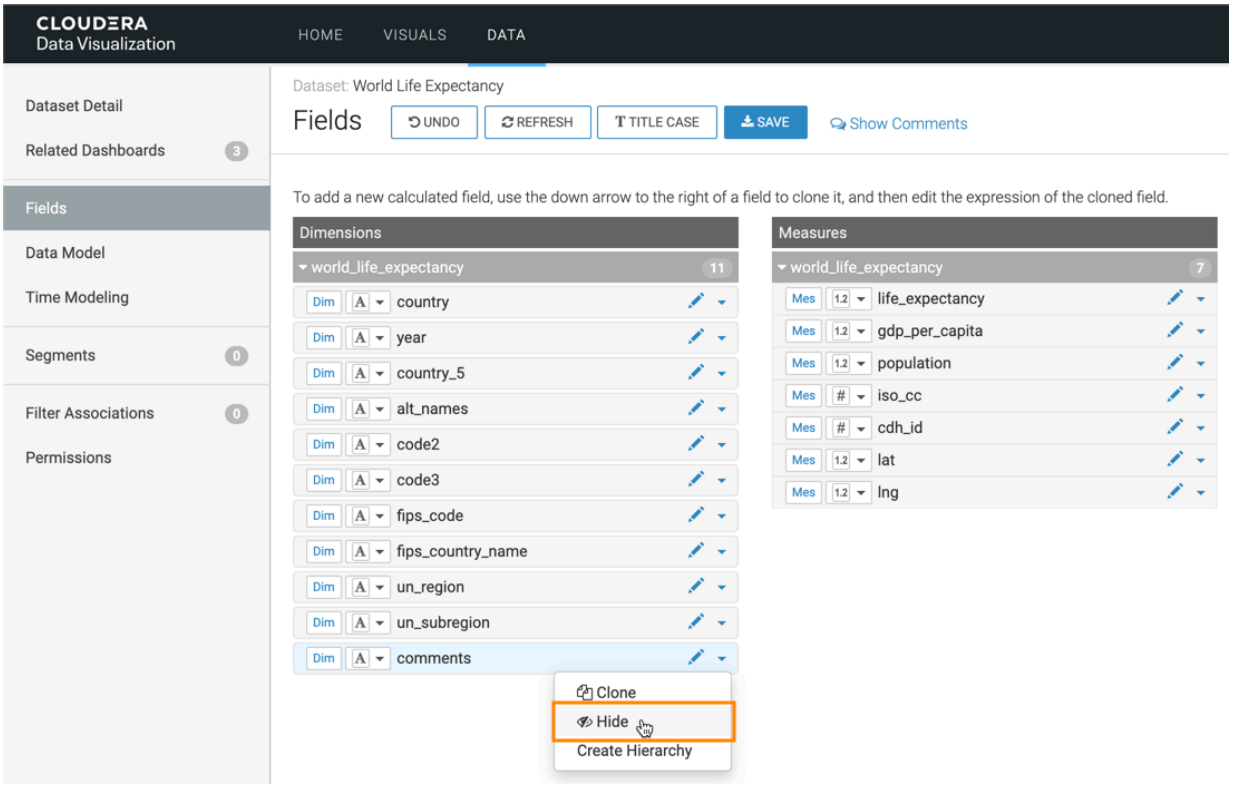


5. In the Fields interface, select Edit Field.



6. Under Dimensions, find the field comments.

7. Click (eye) icon on the comments line.



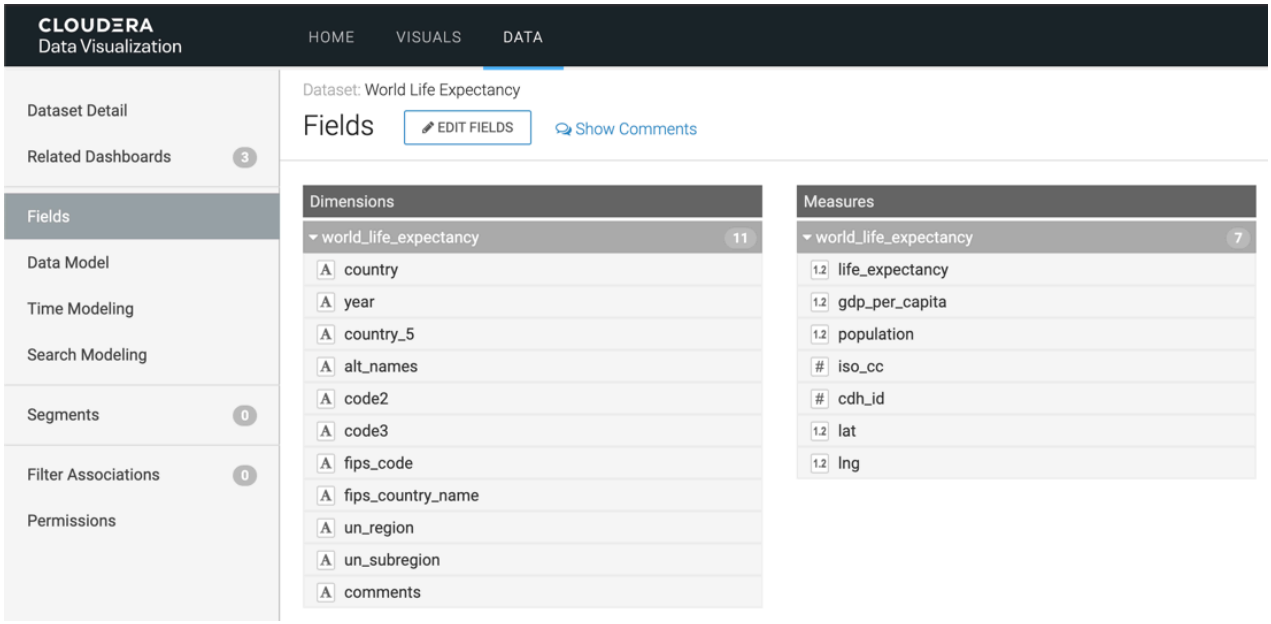
The icon next to the comment field changes to (slashed eye).

8. Under Measures, find the fields lat and lng, and hide them.

9. Click Save.

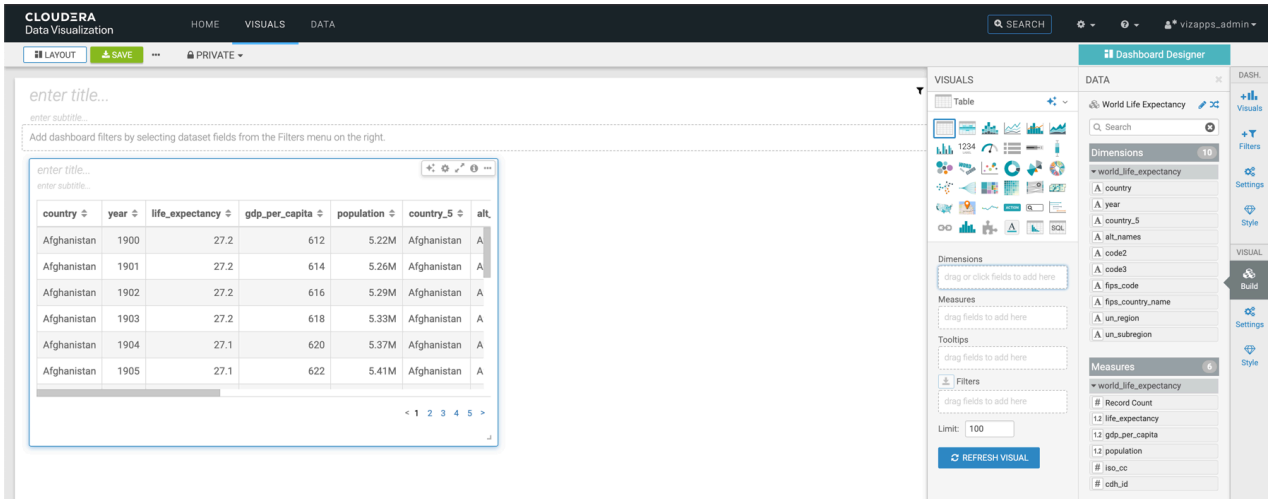
Results

In the updated Fields interface, Dimensions table has a total of 11 fields and Measures table lists a total of 7 fields, as before.



- The number of Dimensions is 11. This is calculated as All Dimensions (11) - Hidden Dimensions (1) + Segment (1).
- The number of Measures is 6. This is calculated as All Measures (7) - Hidden Measures (2) + Record Count (1).

However, when using Visual Designer, the hidden fields do not show.



Tip:

The field visibility can be changed in the Edit Field Parameters window modal.

Changing field aggregation

You may find it useful to change the basic fields defaults in your dataset.

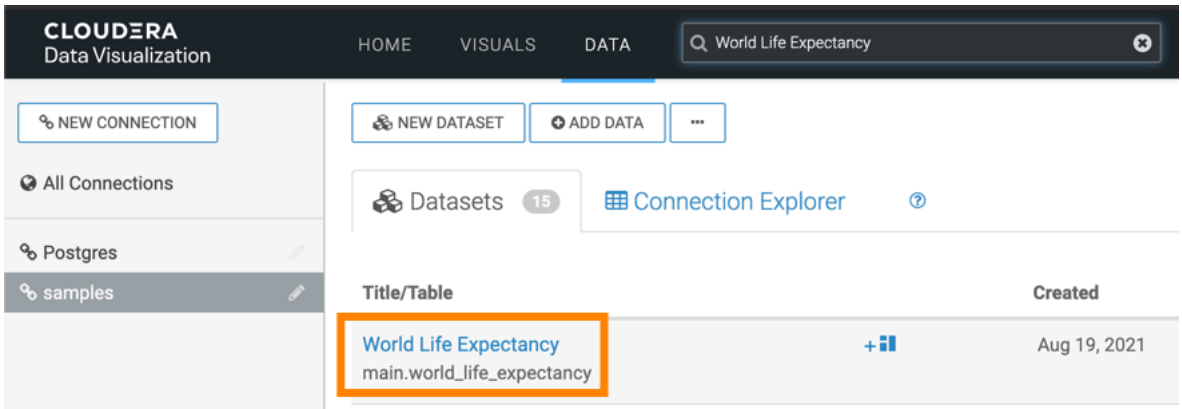
About this task

The following steps demonstrate how to change the default aggregation function from Sum to Average for the field life_expectancy in the dataset World Life Expectancy [data source: main.world_life_expectancy].

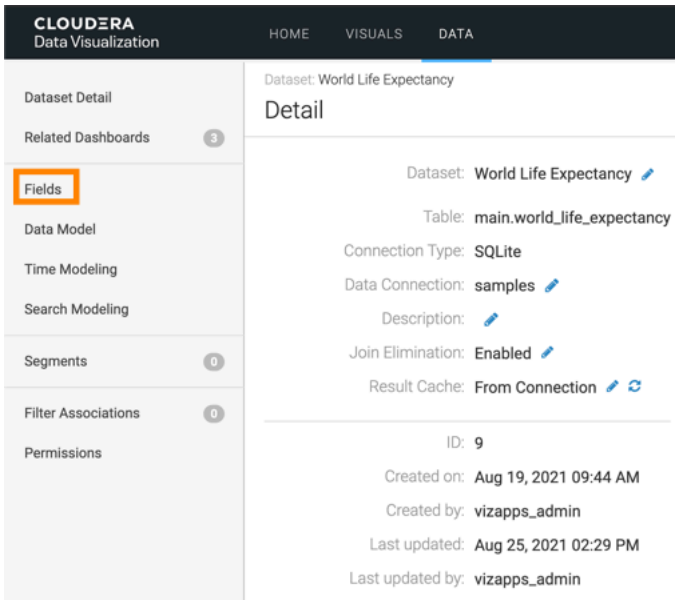
Procedure

1. On the main navigation bar, click Data.

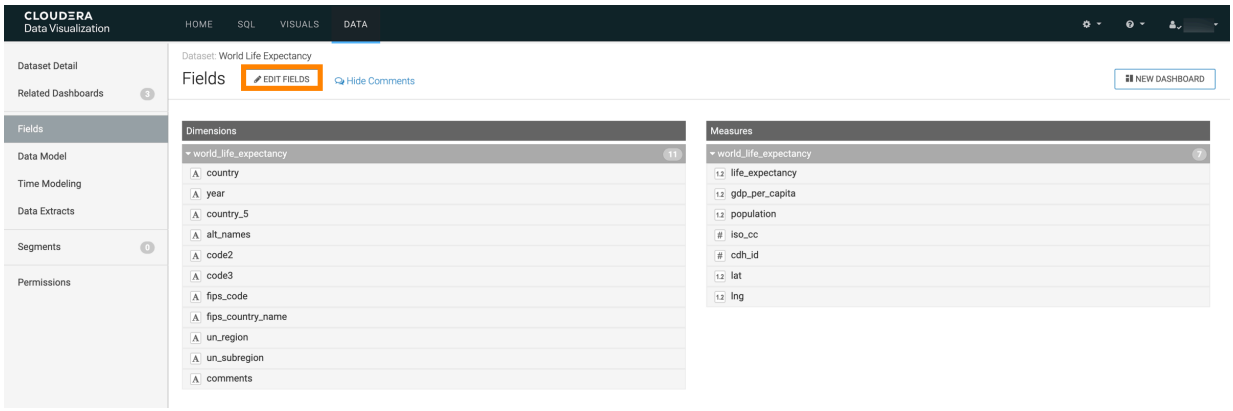
   The Data view appears, open on the Datasets tab.
2. In the left navigation menu, click samples.
3. In the Datasets area, select World Life Expectancy (samples.world_life_expectancy).



4. In the Dataset Detail menu, select Fields.



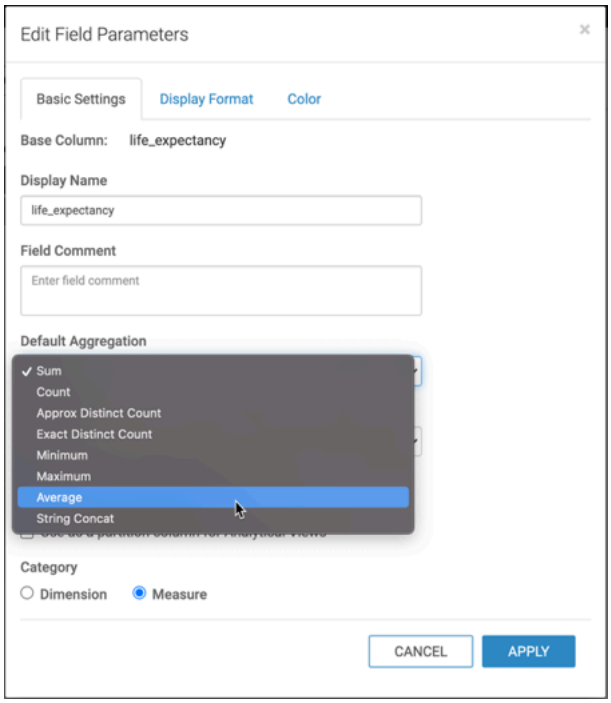
5. In the Fields interface, select Edit Fields.



6. Under Measures, find the field life_expectancy and click the (down arrow)icon on its right side.
7. From the menu, select Edit Fields.

The Edit Field Parameters window modal appears.

8. Change Default Aggregation from Sum to Average, and click Apply.

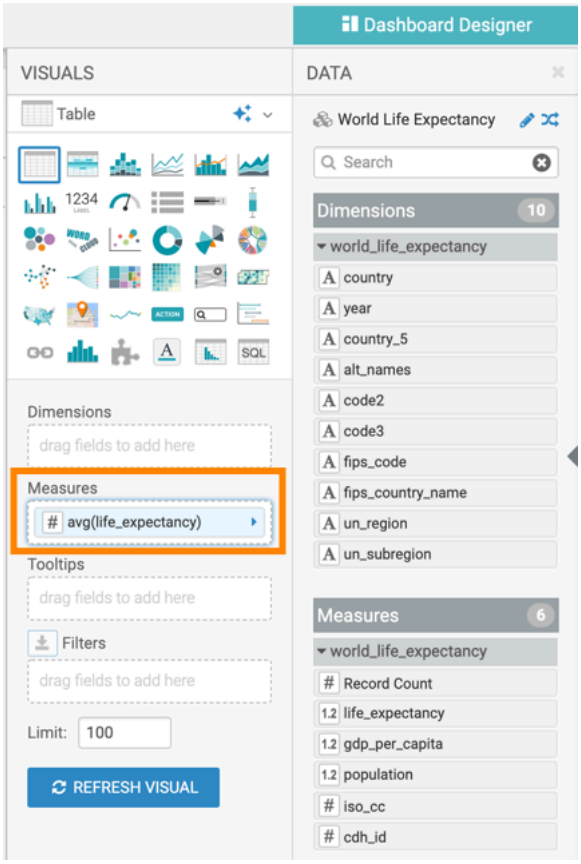


The Edit Field Parameters window modal appears.

9. Under Dataset: World Life Expectancy, click Save.

Results

As a result of this change, all new visuals created from this dataset will automatically use the new aggregation.



Creating calculated fields

Sometimes, the data in the base tables cannot be used directly, and must use an expression to change or "correct" it. For other use cases, you can create a calculation based on one or more fields. Instead of adding these expression for every visual, in CDP Data Visualization, you can easily create a new calculated field in the dataset, and subsequently use it in dashboards and visuals.

About this task

The following steps demonstrate how to create a new field gdp (gross domenstic product) in the dataset World Life Expectancy [data source main.world_life_expectancy]. We define it by the following equation:

```
gdp = gdp_per_capita x population
```

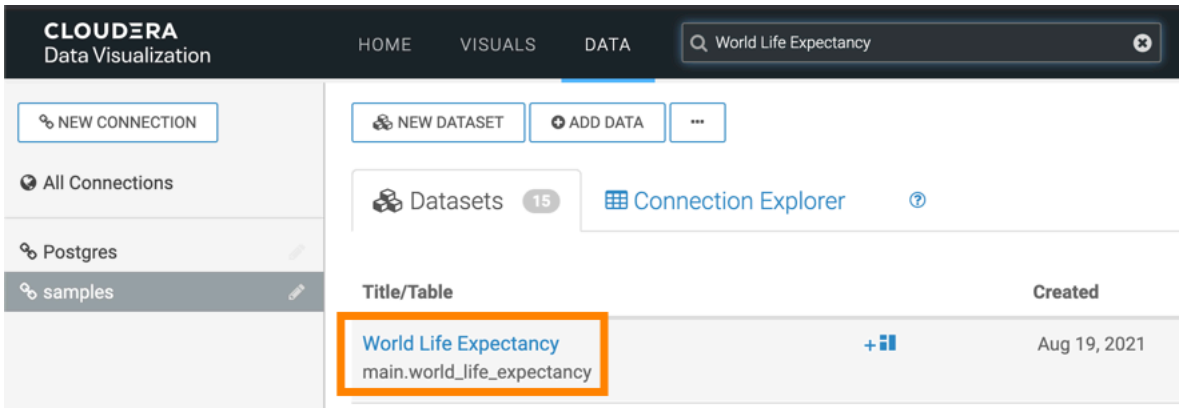
CDP Data Visualization supports three primary methods of editing fields at the dataset level: Basic, Expression, and Display Format. In this example, we make changes both on the Basic and Expression tabs. For innformationn on how to use Display Format options, see *Changing the Field Display Format.*.

Procedure

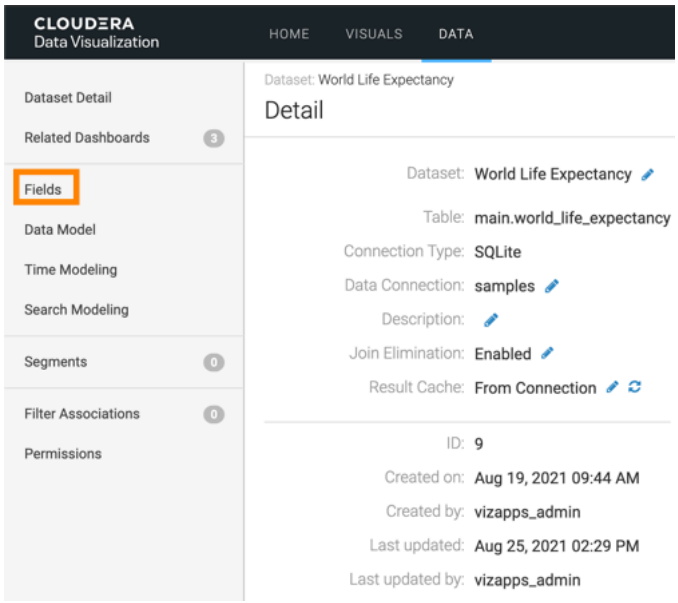
1. On the main navigation bar, click Data.

2. In the left navigation menu, click samples.

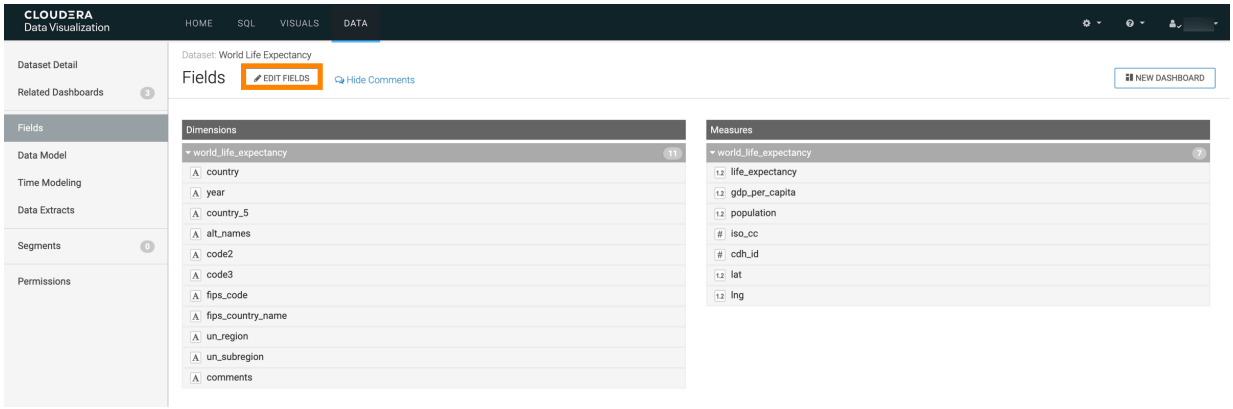
3. In the Datasets area, select World Life Expectancy (main.world_life_expectancy).



4. In the Dataset Detail menu, select Fields.

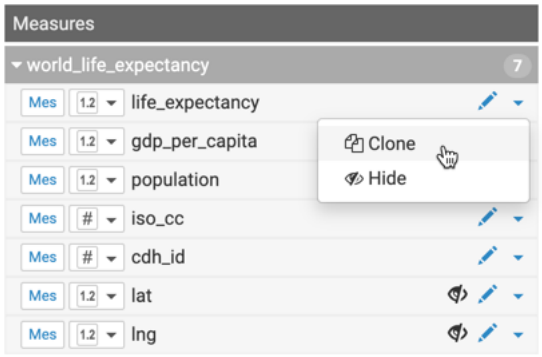


5. In the Fields interface, select Edit Field.



6. Under Measures, find the field gdp_per_capita, and click the (down arrow) icon on its right side.

7. From the menu, select Clone.



8. Under Measures, find the new cloned field Copy of gdp_per_capita, click the (down arrow) icon on its right side, and select Edit Fields.

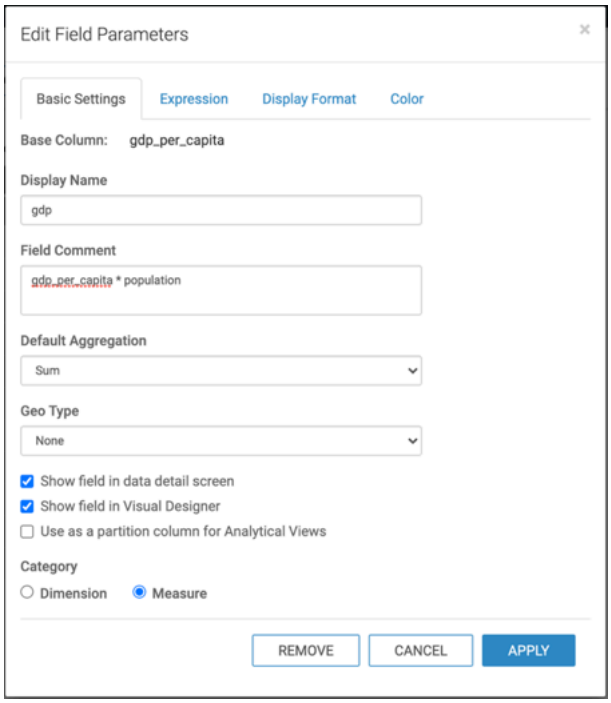
The Edit Field Parameters window modal appears, which supports three primary methods of editing fields; they match the three tabs of the modal: Basic, Expression, and Display Format.

9. In the Edit Field Parameters modal, under the Basic tab, make the following changes:

a) Change Display Name to gdp.

b) Add Field Comment gdp_per_capita * population.

c) Ensure that the Default Aggregation is Sum.

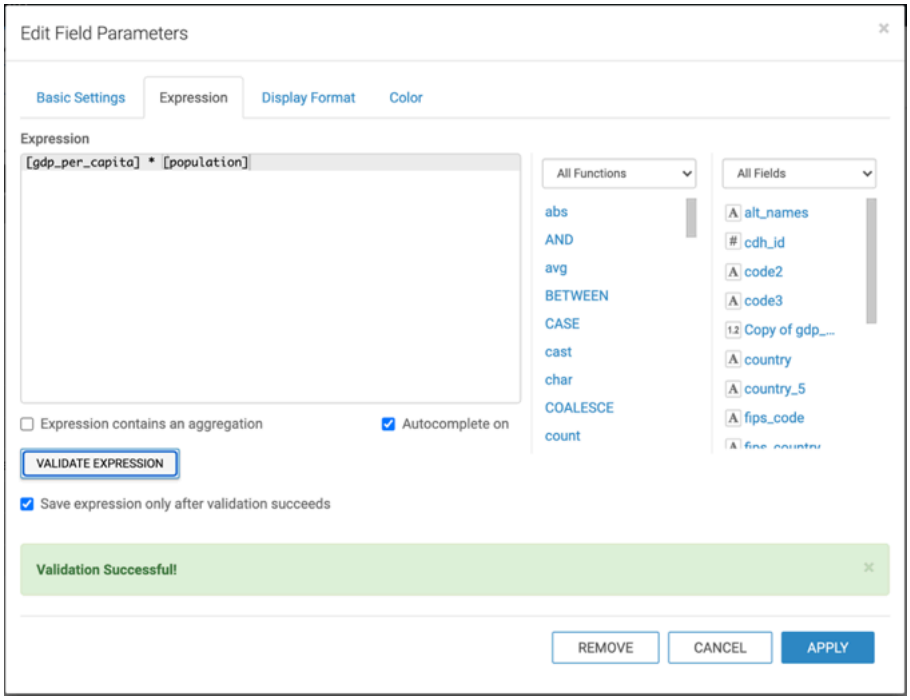


10. Click Edit Expression tab and make the following changes.

a) Change Expression to [gdp_per_capita] * [population].

b) Click Validate Expression to ensure that the calculation works.

c) When the Validation Successful message appears on the modal, click Apply.



The new calculated field has an equal sign (=) notation.

11. Under Dataset: World Life Expectancy, click Save.

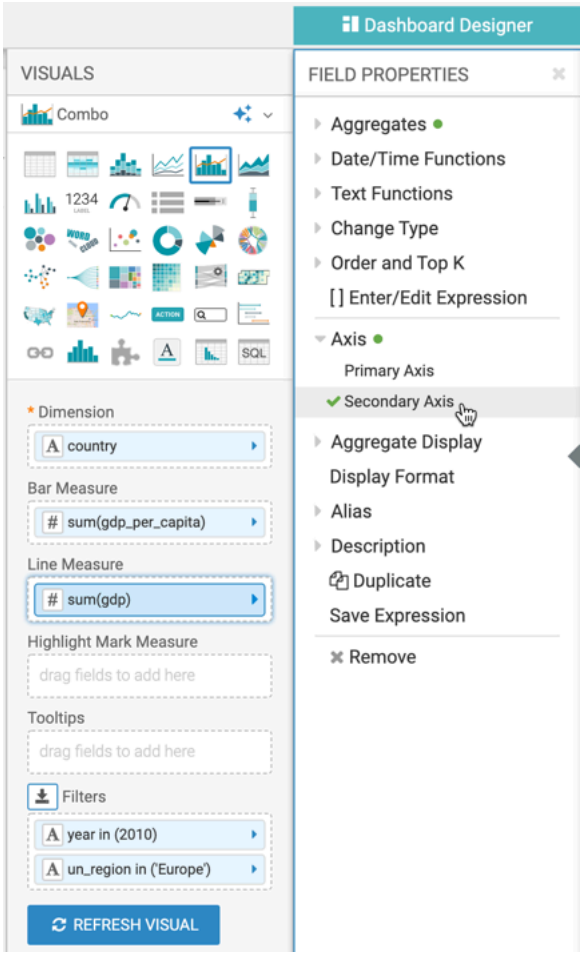
Testing calculated fields

In CDP Data Visualization, you can easily test whether a newly calculated field works correctly or not.

Procedure

1. Click New Dashboard in the top right corner of this interface.
2. Select the Combined Bar/Line visual type.

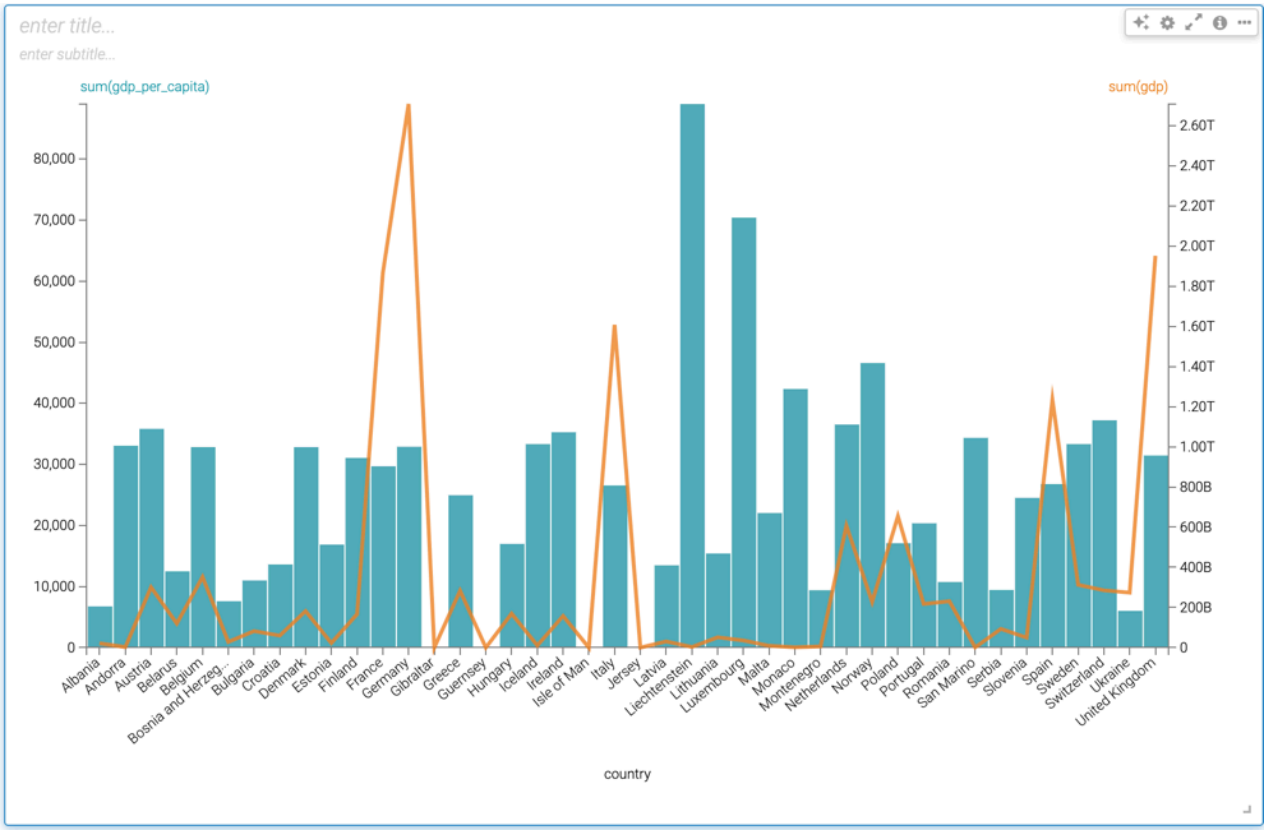
3. Populate the shelves from the available fields in the following way:
   a) Under Dimensions, select country, and place it on the Dimensions shelf.
   b) Under Dimensions, select year, and place it on the Filters shelf.
   c) In the Filter for year modal window, under the Value tab, select the year 2010.
   d) Under Dimensions, select un_region, and place it on the Filters shelf.
   e) In the Filter for un_region modal window, under the Value tab, select Europe.
   f) Under Measures, select gdp_per_capita, and place it on the Bar Measure shelf.
   g) Under Measures, select gdp, and place it on the Line Measure shelf.
   h) Click the arrow on gdp field.
   i) In the Field Properties under Axis, select Secondary Axis.



4. Click Refresh Visual.

Results

The two measurements appear on the graph, superimposed on each other: the original gdp_per_capita represented by the bars, and the calculated gdp, represented by the line.



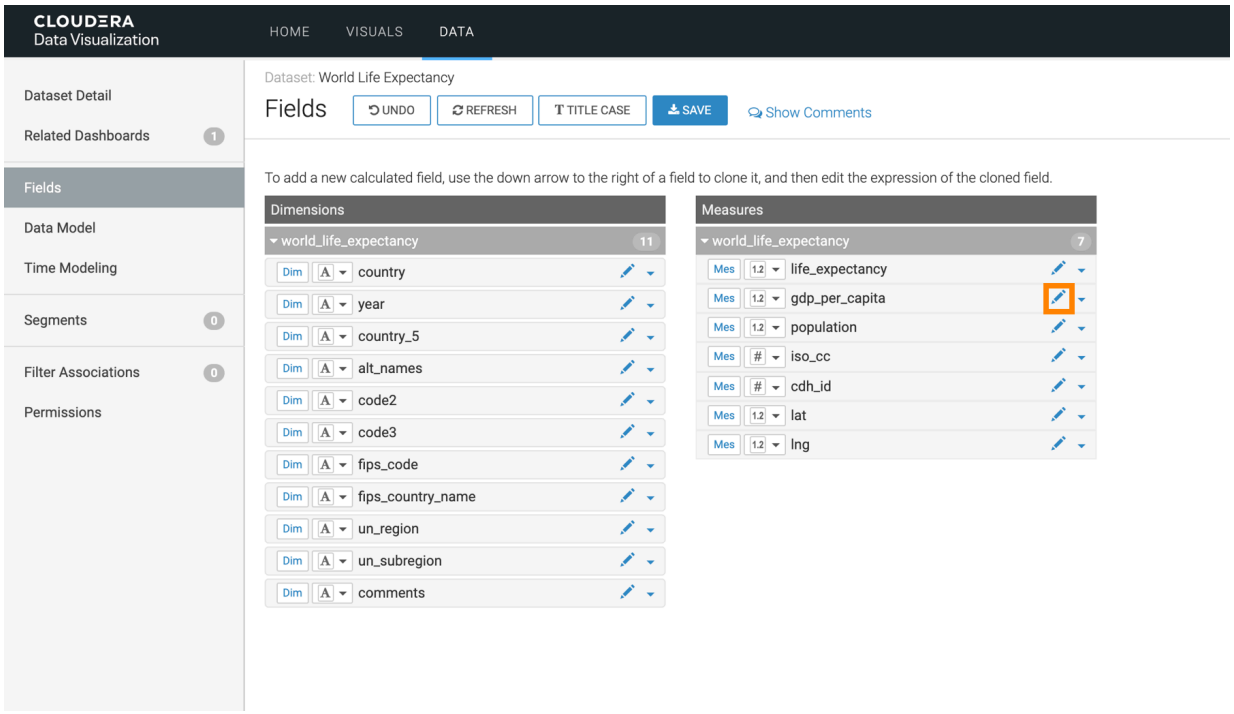
Changing the field display format

CDP Data Visualization lets you configure, at the dataset level, the display format of each field.

Procedure

1. In the World Life Expectancy dataset, click EDIT FIELDS.

2. Under Measures, find the field gdp_per_capita and click the (pencil) icon on its right side.



The Edit Field Parameters modal window appears.

3. Click the Display Format tab.

Under Display Format, you have several options in the Category menu:

- None
- Real Number
- Integer
- Percentage
- Scientific
- Currency
- Date/Time
- Custom Format
- Custom Javascript

For more information, see:

Changing currency field display format

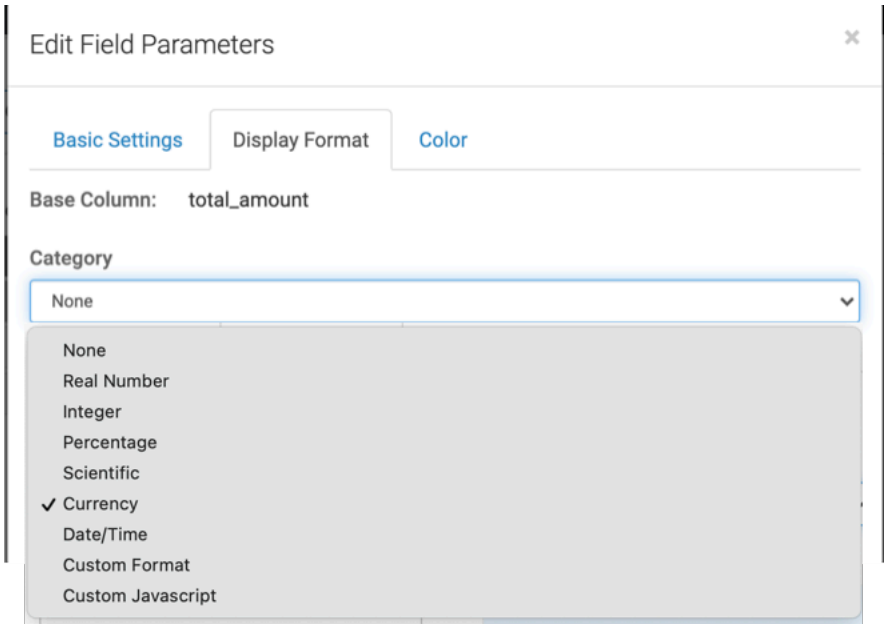
In CDP Data Visualization, you can set currency display options for numerical fields across all visuals of a dataset.

About this task

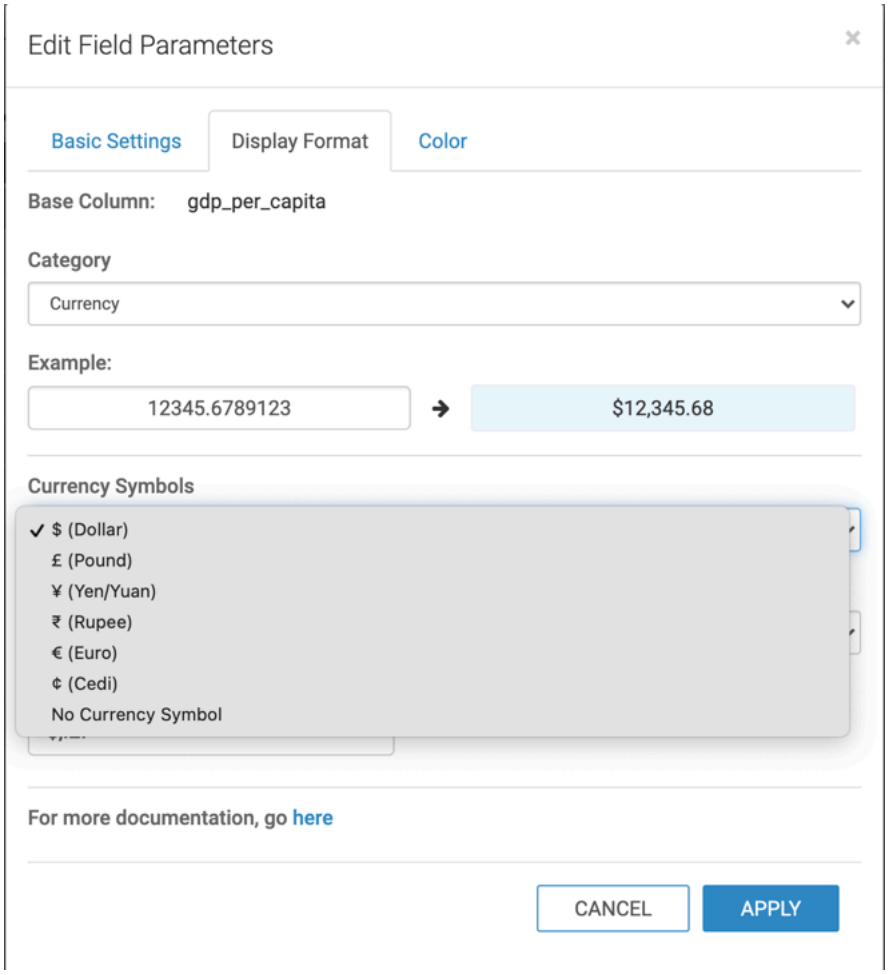
Follow these steps to continue configuring a field at the dataset level for currency format. See *Changing the field display format* for the initial navigation steps.

Procedure

1. In the Edit Field Parameters modal window, under the Display Format tab, select Currency from the Category menu.

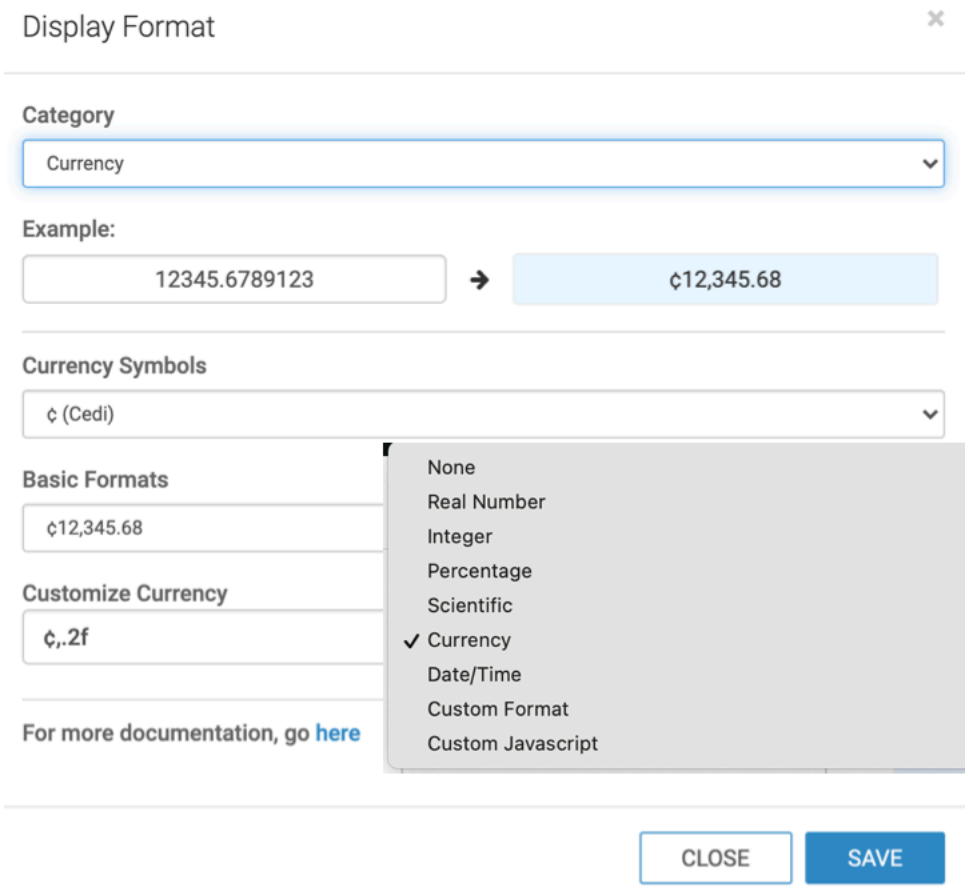


2. In the Currency Symbols menu, select the appropriate currency symbol: $ (Dollar), £ (Pound), ¥ (Yen/Yuan), # (Rupee), € (Euro), or # (Cedi).



3. Select the Basic Format for your records.

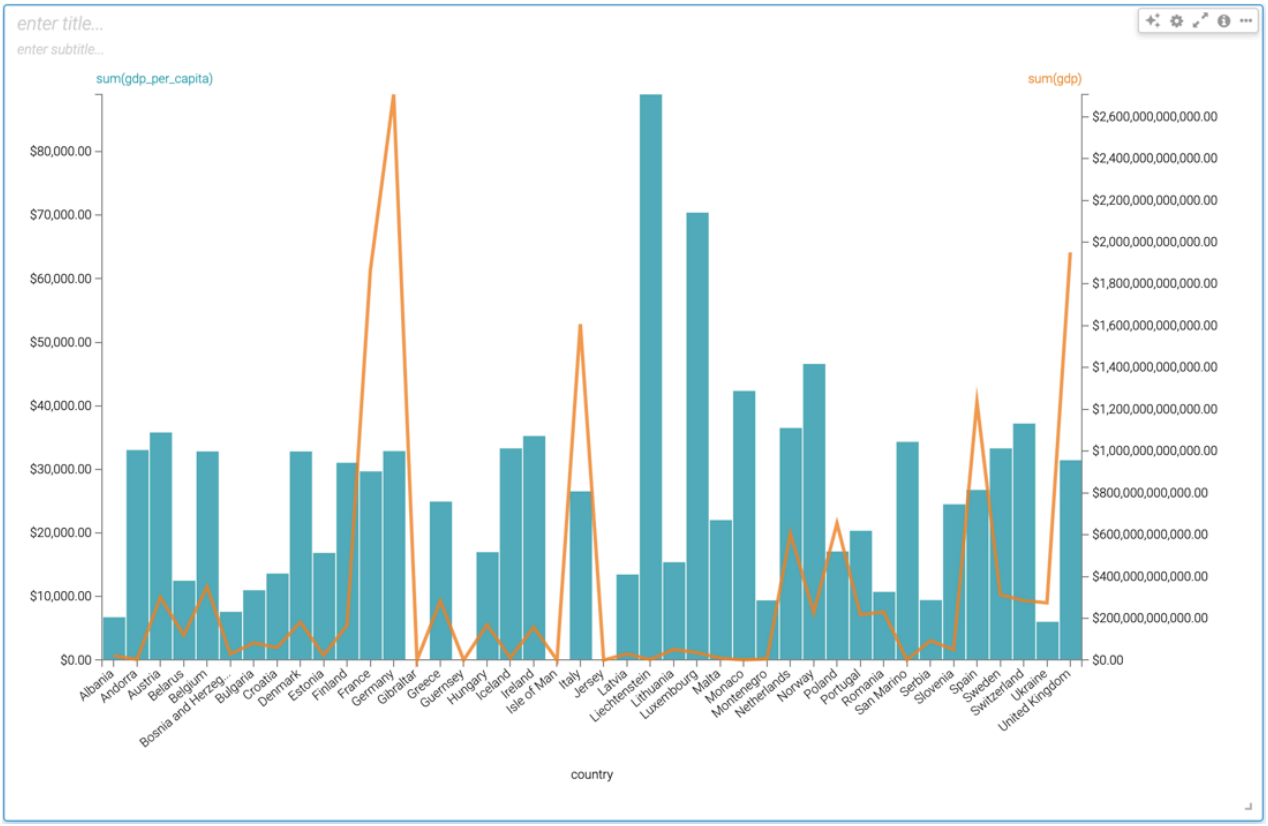
   You can also define and apply a custom format. Enter a valid format mask in the Customize Currency text box. For a full list of options, see *Display Format Reference*.



4. Click APPLY.
5. Click SAVE to save the changes to the dataset.
6. To verify that the format applies to all new visuals that use the field, create a new visual by repeating the steps in *Testing the Calculated Field*.

Results

The new visual displays the vertical axis numbers in currency format.



Related Information

Changing the field display format

Display format reference

Changing custom field display format

In CDP Data Visualization, you can Set currency display options for numerical fields across all visuals of a dataset.

About this task

Follow these steps to continue configuring a field at the dataset level for custom format. See *Changing the field display format* for the initial navigation steps.

Procedure

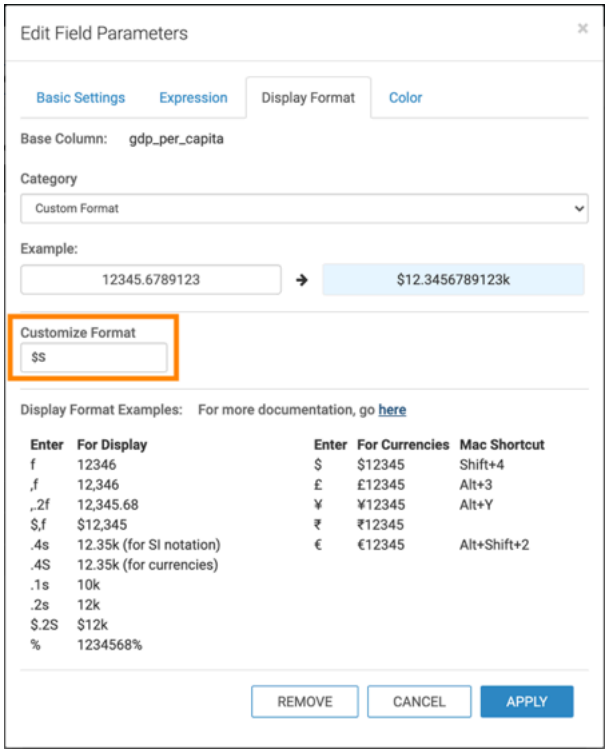
1. In the Edit Field Parameters window modal, under the Display Format tab, select Custom Format from the Category menu.

2. Under Customize Format, enter $S.

   - $ appends the dollar currency symbol to the left of the number.
   - S simplifies the number by minimizing the significant numbers (which on the axes appear with a large number of trailing zeros) and appending the appropriate non-scientific (currency) suffix to the right of the number.

   Note:

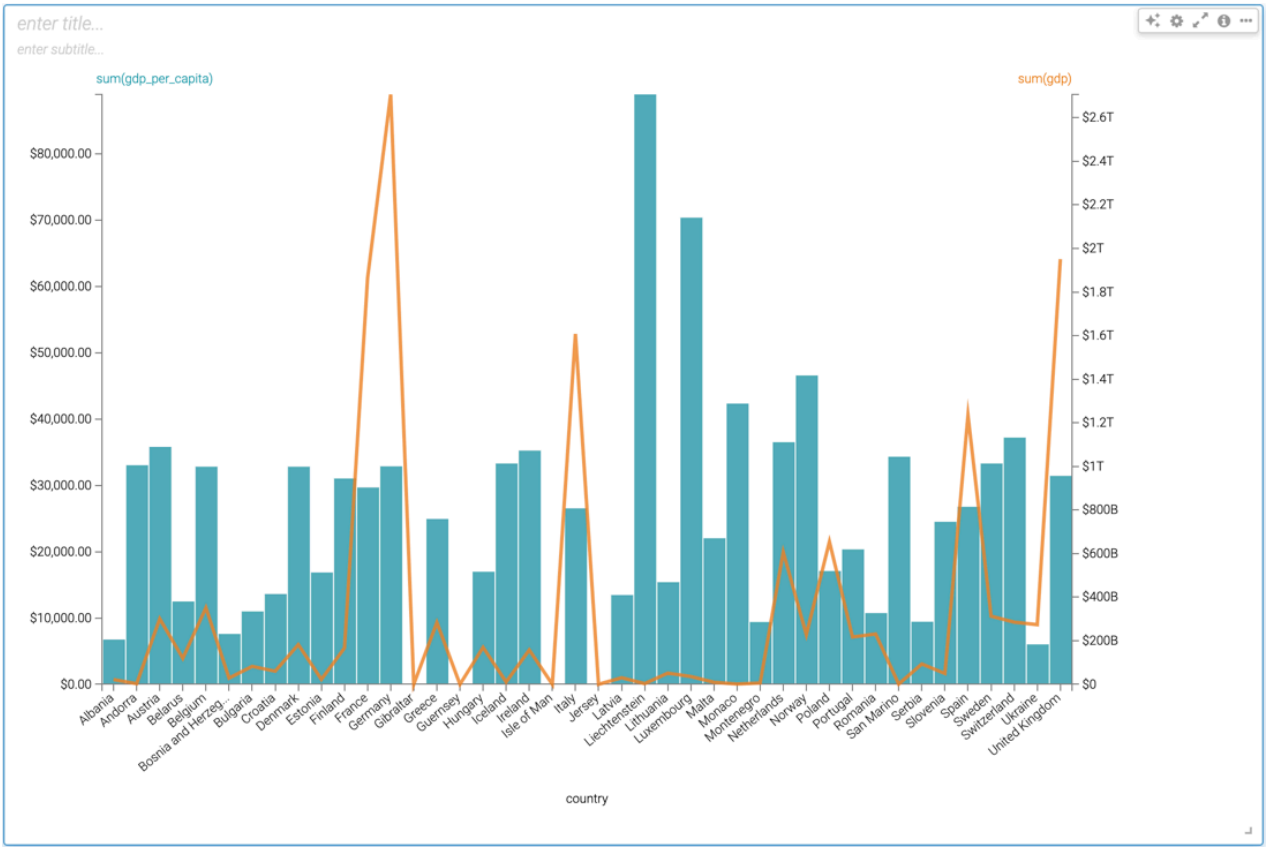
   Display Format Examples demonstrate some of the available options for defining custom format.



3. Click APPLY.
4. Save the changes to the dataset.
5. [Optional] To verify that the format applies to all new visuals that use the field, create a new visual by repeating the steps in *Testing the calculated field.*.

Results

When we use the $S custom format on the gdp field, the visual still uses the dollar currency sign, but also simplifies/abbreviates the number and shows the corresponding currency suffix. In the case of this visual, T for trillion; this custom format transforms the representation $2,500,000,000,000.00 into $2.5T.



Related Information

Changing the field display format

Changing custom Javascript field display format

About this task

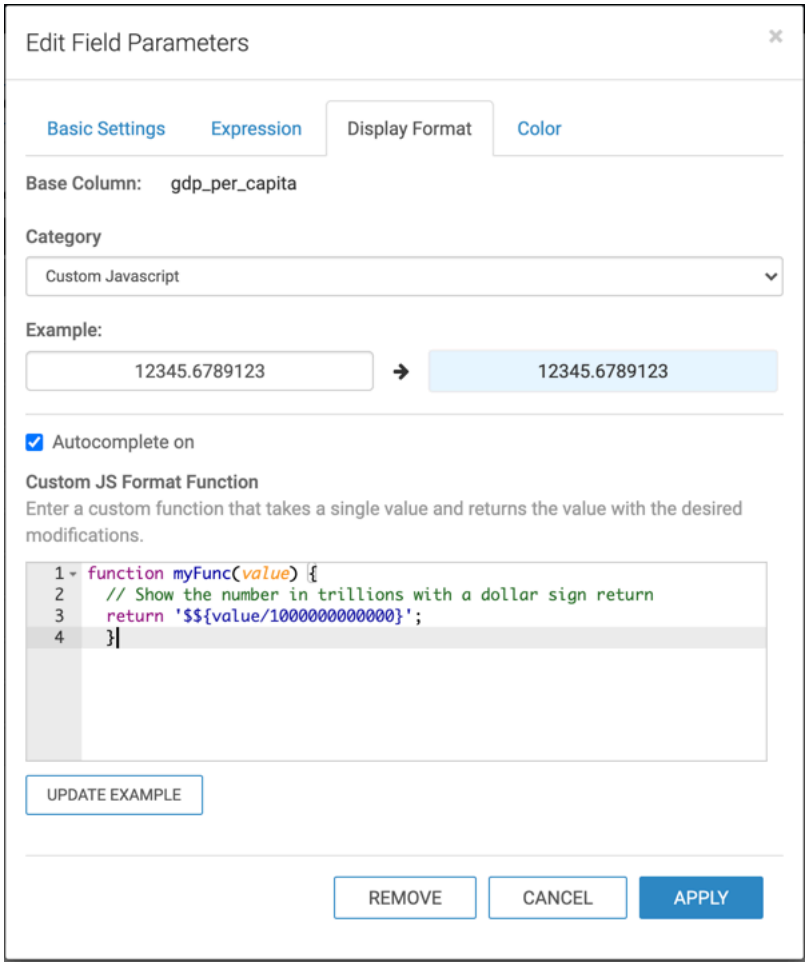
Follow these steps to continue configuring a field at the dataset level using Javascript (js). See *Changing the field display format* for initial navigation steps.

Procedure

1. In the Edit Field Parameters window modal, under the Display Format tab, select Custom Javascript from the Category menu.
2. Under Custom JS Format Function, enter the following js code:

```
function myFunc(value) {
  // Show the number in trillions with a dollar sign return
  '$${value/1000000000000}';
```

```
}
```



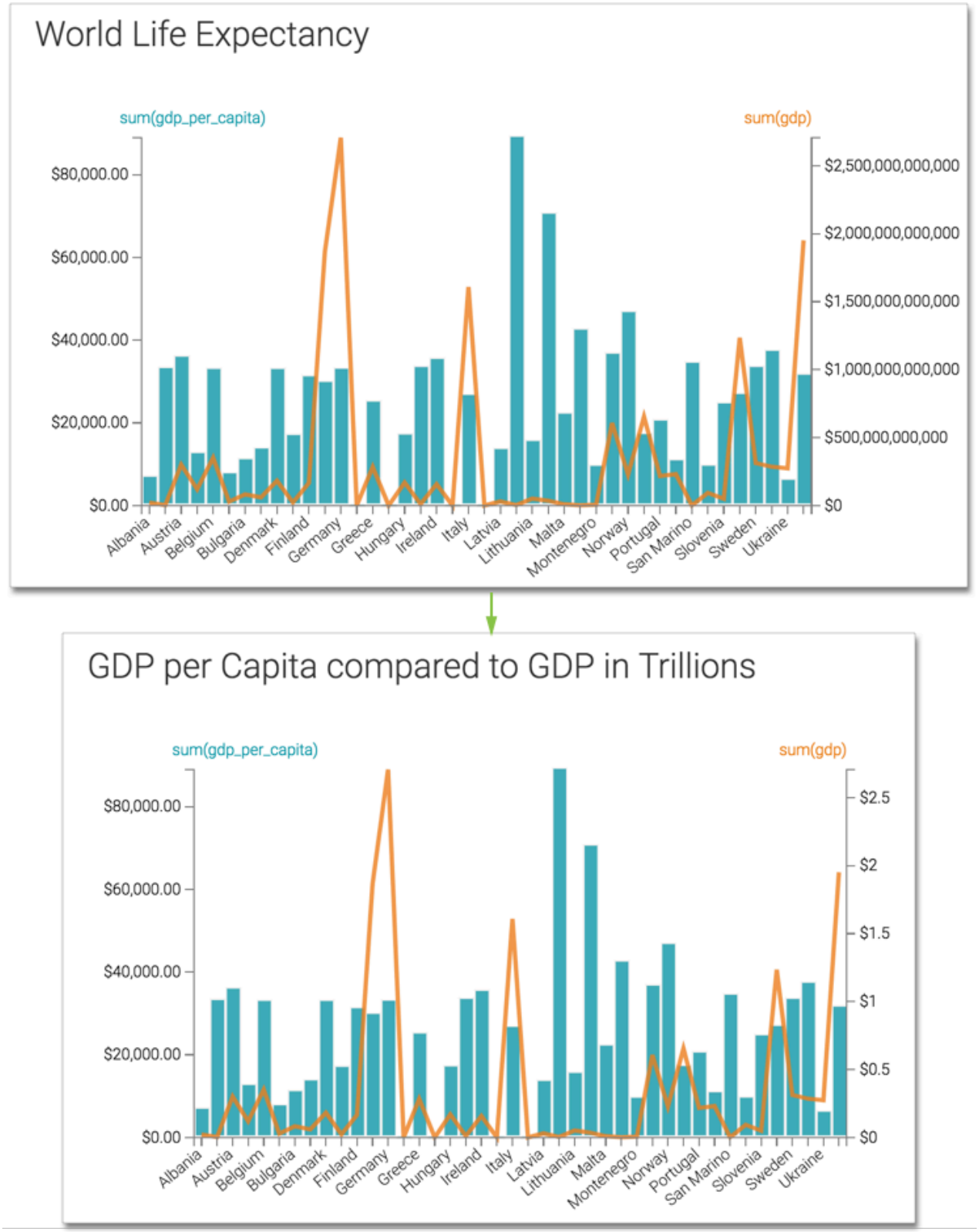
3. Click APPLY.
4. Save the changes to the dataset.

   In this example, under Dataset: World Life Expectancy, click Save.
5. To verify that the format applies to all new visuals that use the field, repeat the steps in *Testing the calculated field.*.

Results

Let's compare the visuals before and after we apply the js format on the gdp field.

- Before we apply the js format, the visual displays the currency with many trailing zeros.
- After applying the custom js format, notice that the trailing zeros no longer appear.



Related Information

Changing the field display format

Changing data type

CDP Data Visualization allows you to change the effective data type of a column in the dataset model without changing the source data. This is useful in many business environments to ensure correct processing of numerical codes, catalog numbers, event IDs, dates, and so on.

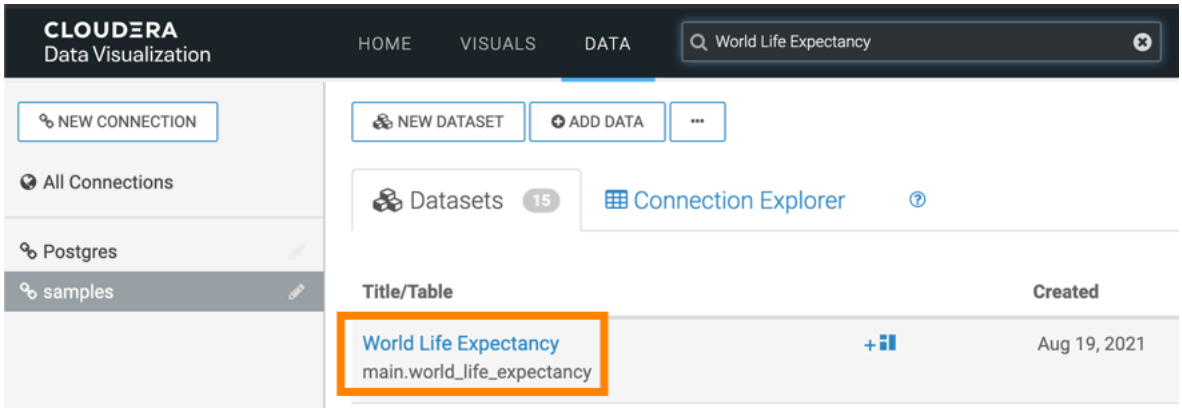
About this task

The following steps demonstrate how to change the type of a column. We are using the column iso_cc (the ISO-compliant country code) in the dataset World Life Expectancy [data source samples.world_life_expectancy].

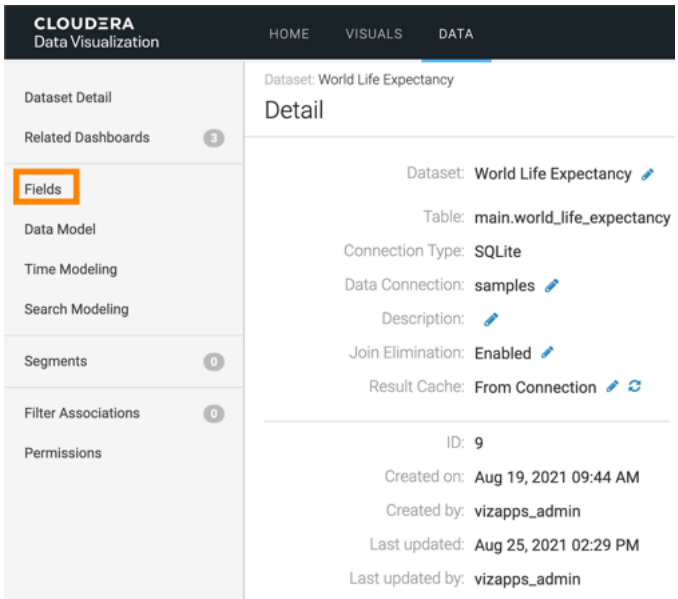
Procedure

1. On the main navigation bar, click Data.

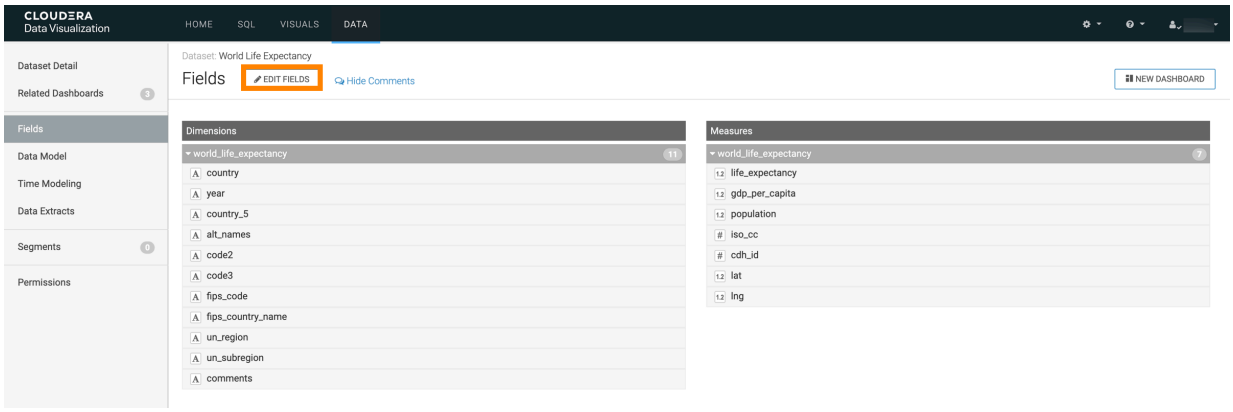
2. In the left navigation menu, click samples.
3. In the Datasets area, select World Life Expectancy (samples.world_life_expectancy).



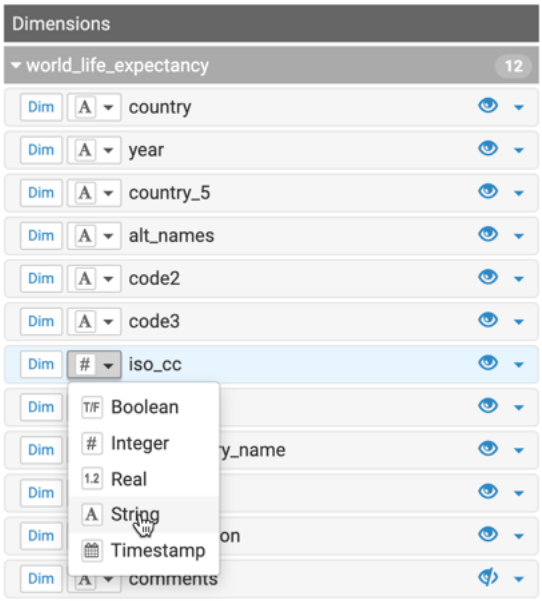
4. In the Dataset Detail menu, select Fields.



5. In the Fields interface, select Edit Field.



6. Under Dimensions, find the field fips_code, and click the down arrow icon, immediately following the # icon.
7. In the menu, select String.



8. Under Dataset: World Life Expectancy, click Save.

Results

All new visuals created from this dataset will automatically use the new type.

Note: Use this functionality carefully, as it may break the visuals that already use the column in aggregations or custom expressions.

Specifying geographic fields

In CDP Data Visualization, you can explicitly specify a dataset field as one of the many supported geographical types.

About this task

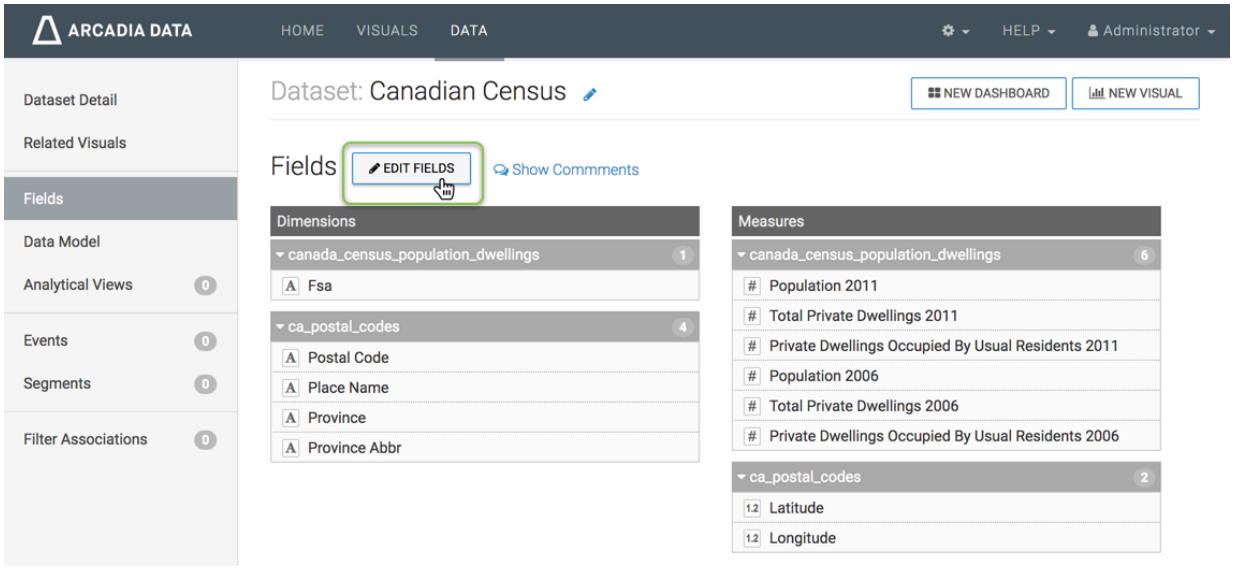
In this example we use the dataset Canadian Census, constructed from example datasets. , and joined of the fields fsa and Postal Code, respectively.

The following steps demonstrate how to assign Geo Types to a dataset field. We are using the two fields from the join of the Canadian Census dataset: fsa from canada_census_population_dwellings, and Postal Code from ca_postal_codes.

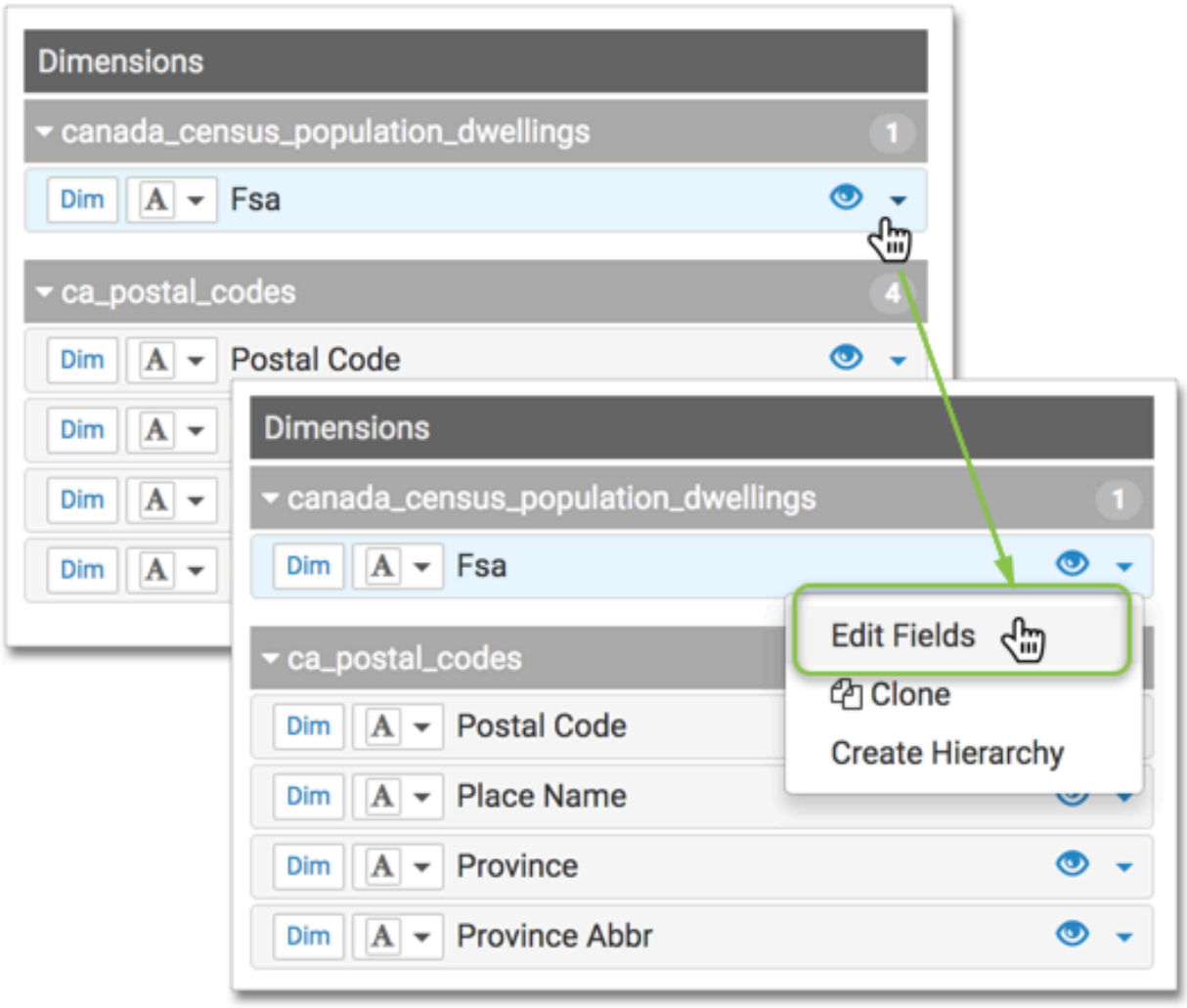
Note: Before creating visuals that use geographic data, specify the appropriate information as Geo Types.

Procedure

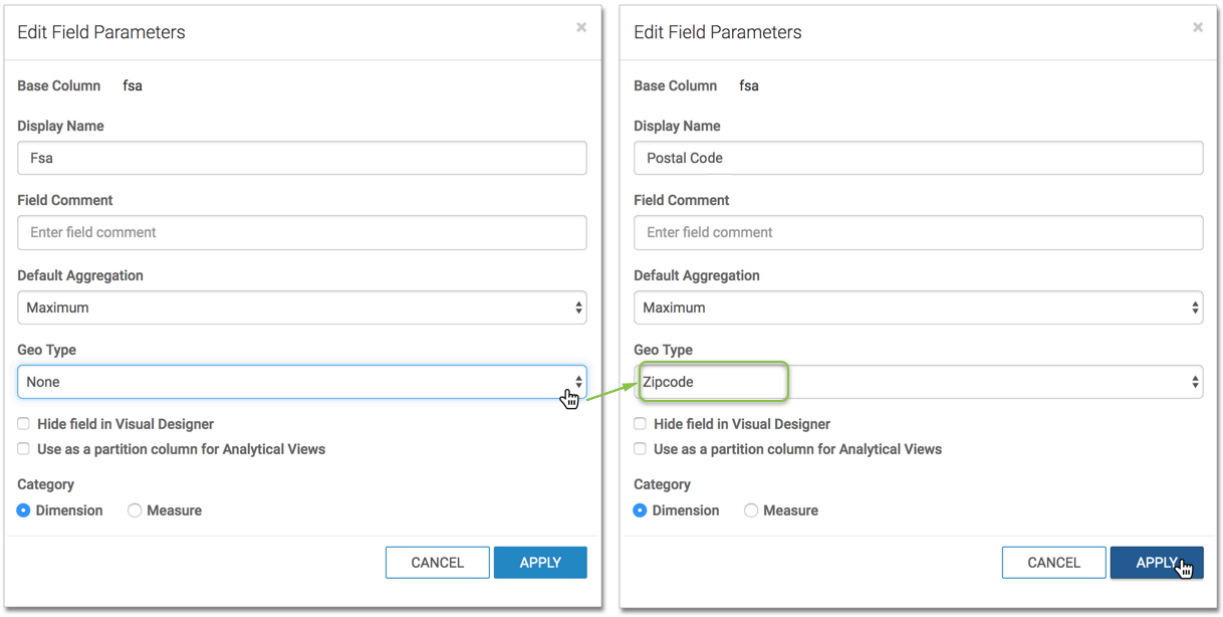
1. Navigate to the dataset that you must modify. In this example, we are using the Canadian Census dataset.
2. In the Dataset Detail menu, select Fields.
3. In the Fields interface, select Edit Fields.



4. To specify that the field is a Geo Type, In Edit mode, click the (down) icon on the right side of the field, and select the Edit Fields option.



5. In the Edit Field Parameters window modal, under Geo Type, select the appropriate option from the menu, and click Apply. For Fsa, we also changed the Display Name to Postal Code in the example.



6. Repeat for the Postal Code field of the table ca_postal_codes, and click Apply.
7. Repeat with the Province field of the table ca_postal_codes, changing the Geo Type to State/Province. Click Apply.
8. [Optional] Apply the Geo type to the fields that represent Latitude and Longitude, if they are not named appropriately.
9. Under Dataset: Canadian Census, click Save.

Results

The dataset can now be successfully used for map and interactive map visuals, without further adjustments at the level of the visual.

Adding field comments

When working with big data, it can be very helpful to have access to comprehensive field-level descriptions. In CDP Data Visualization, you can use field comments to provide the context and meaning of each dataset field.

Adding field comments in dataset

When working with large datasets data, it can be useful to have access to comprehensive field-level descriptions. In Data Visualization, you can use field comments to provide the context and meaning of each dataset field.

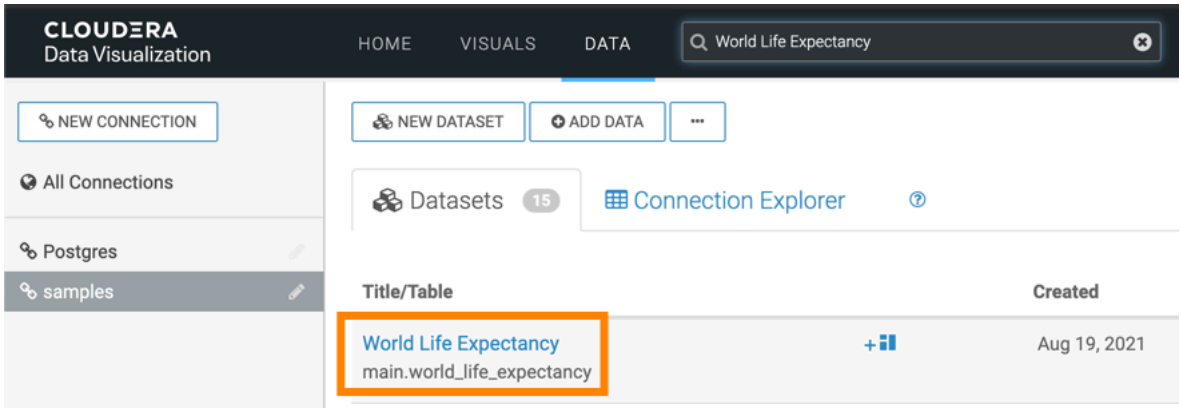
About this task

The following steps demonstrate how to add description to a column of a dataset, as a 'comment'. In this example, we are using the column iso_cc (the ISO-compliant country code) in the dataset World Life Expectancy [data source samples.world_life_expectancy].

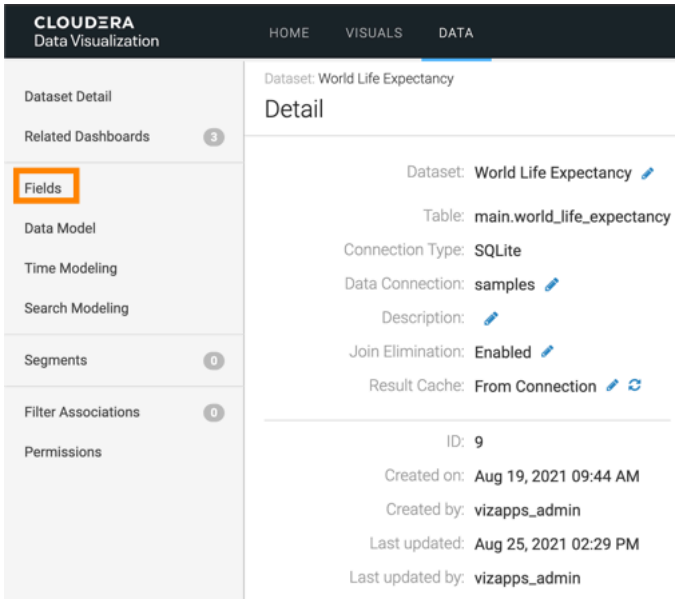
Procedure

1. On the main navigation bar, click DATA.

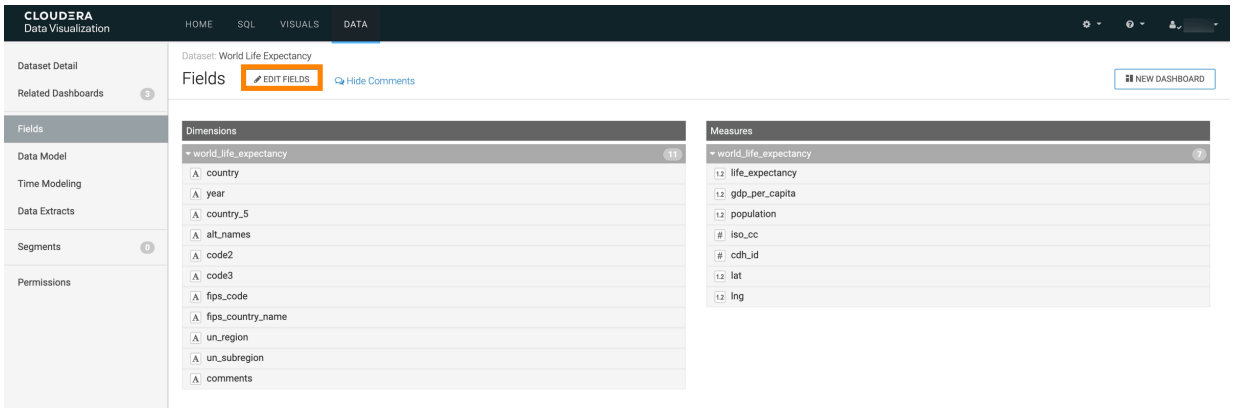
   The Data view appears, open on the Datasets tab.
2. In the left navigation menu, click samples.
3. In the Datasets area, select World Life Expectancy (samples.world_life_expectancy).



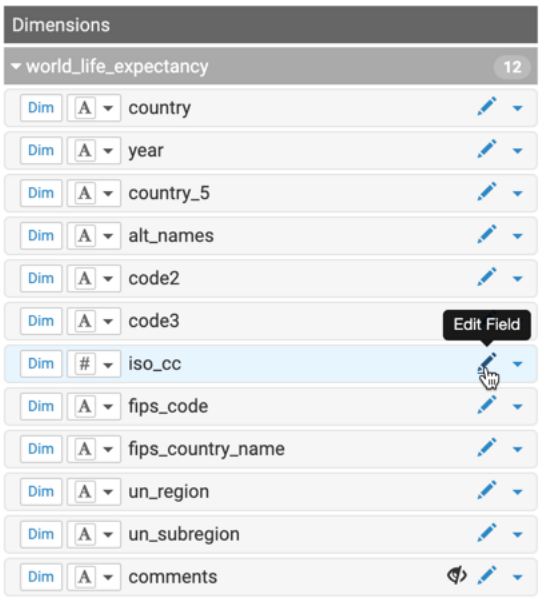
4. In the Dataset Detail menu, select Fields.



5. In the Fields interface, select EDIT FIELDS.



6. Under Dimensions, find the field iso_cc, and click the down arrow icon on the right.
7. From the menu, select Edit Field.

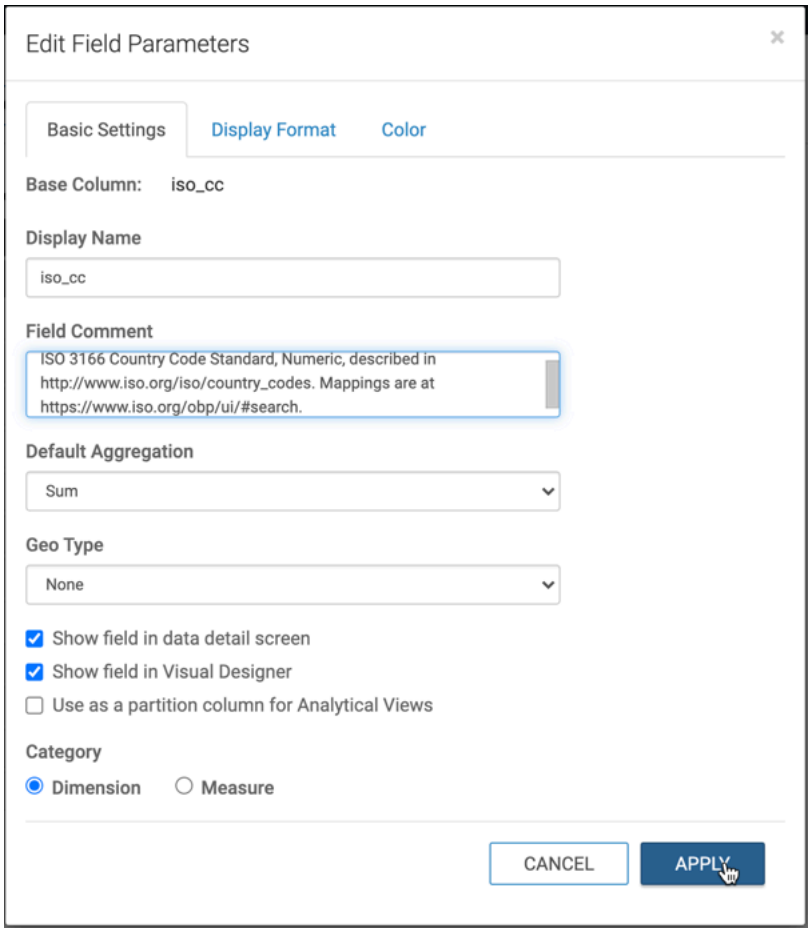


The Edit Field Parameters window modal appears.

8. Under Field Comment, add the following text:

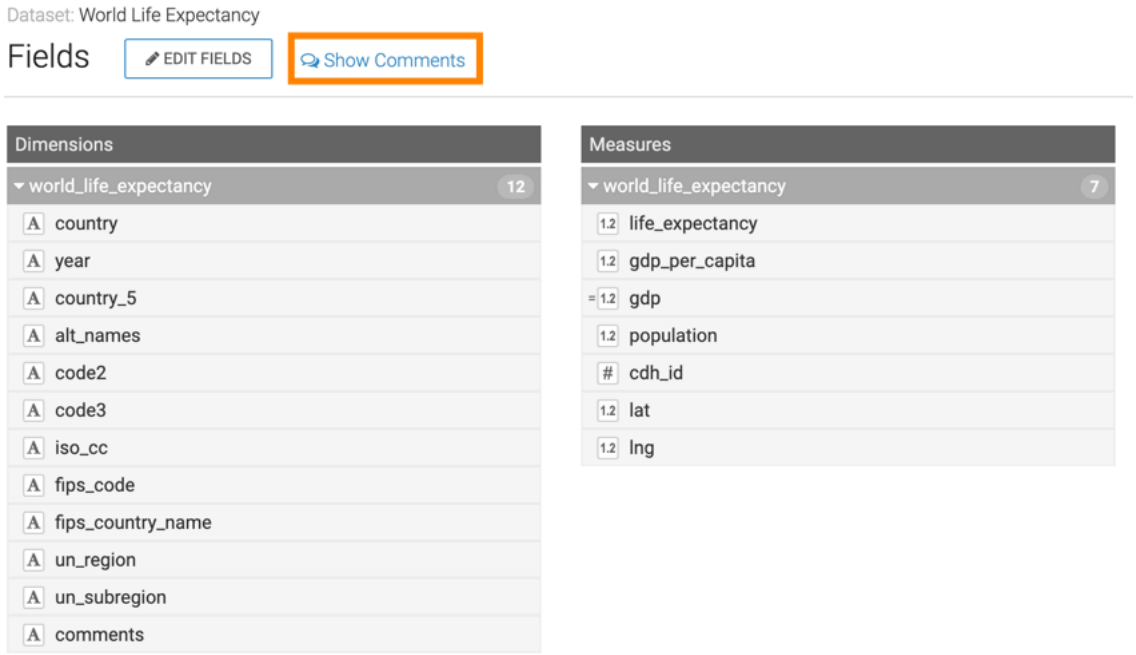
ISO 3166 Country Code Standard, Numeric, described in http://www.iso.org/iso/country_codes. Mappings are at https://www.iso.org/obp/ui/#search.

9. Click APPLY.



10. Click SAVE in the top row.

11. Click Show Comments to display the field comment.



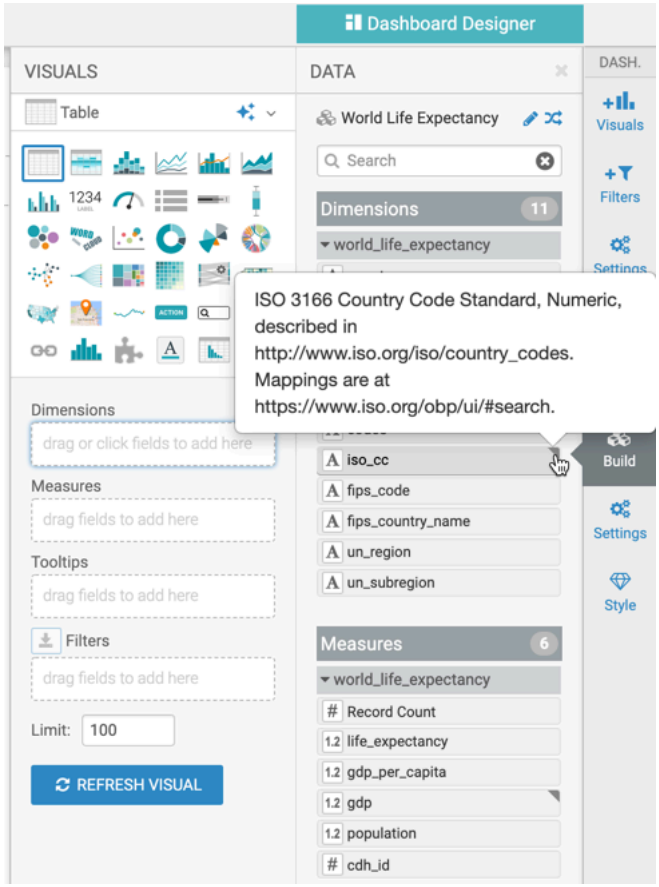
12. To hide field comments, click Hide Comments.

Viewing field comments in Visual Builder

When working with big data, it can be very helpful to have access to comprehensive field-level descriptions. In CDP Data Visualization, you can use field comments to provide the context and meaning of each dataset field.

To view field comments in the Visual Builder, create a new visual using the World Life Expectancy dataset [data source samples.world_life_expectancy]. See *Creating visuals*..

The fields iso-cc and gdp-per-capita display a gray triangle on the top right corner of the field. Hover over the triangle to view the field comments defined in the dataset.



Automatically renaming dataset fields

Quite often, the column names of the base data tables are not very human-friendly. CDP Data Visualization gives you the option to automatically adjust field names at the level of the dataset.

About this task

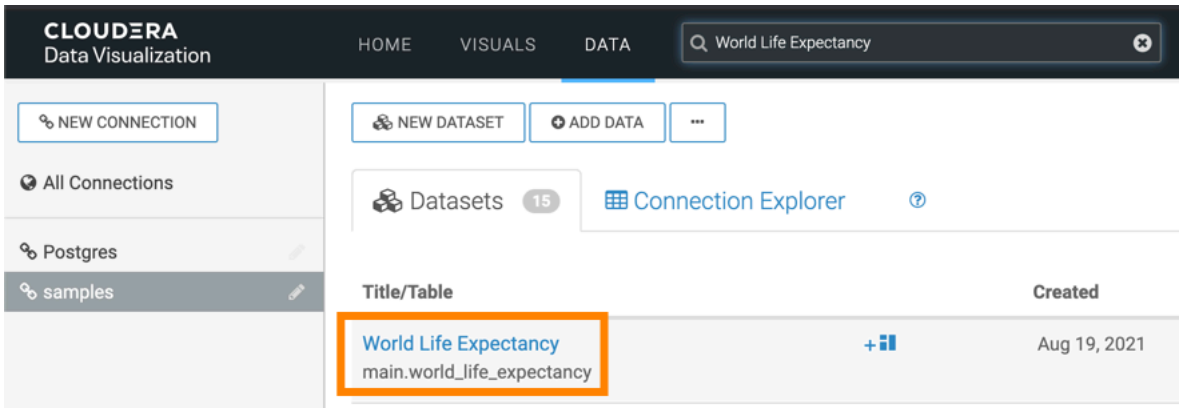
The following steps demonstrate how to prevent data fields from appearing in visualizations and applications of dataset World Life Expectancy [data source samples.world_life_expectancy]. The fields comments, lat, and lng are empty, so they are good candidates for this operation.

Procedure

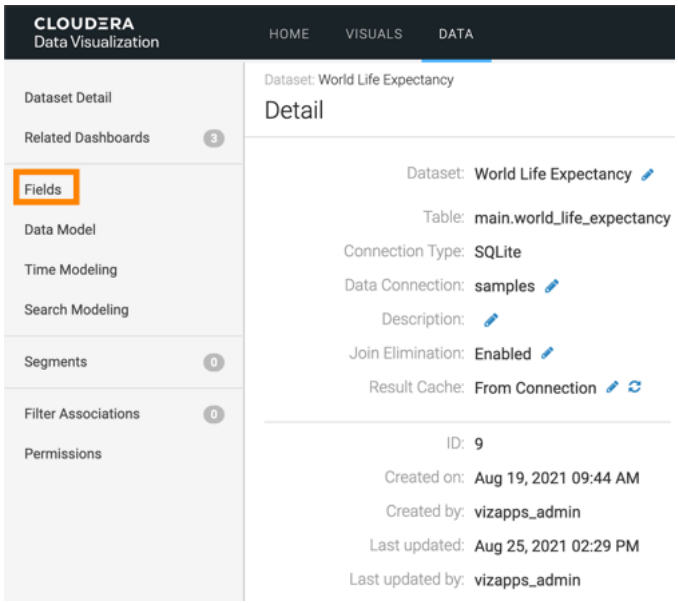
1. On the main navigation bar, click Data.

The Data view appears, open on the Datasets tab.

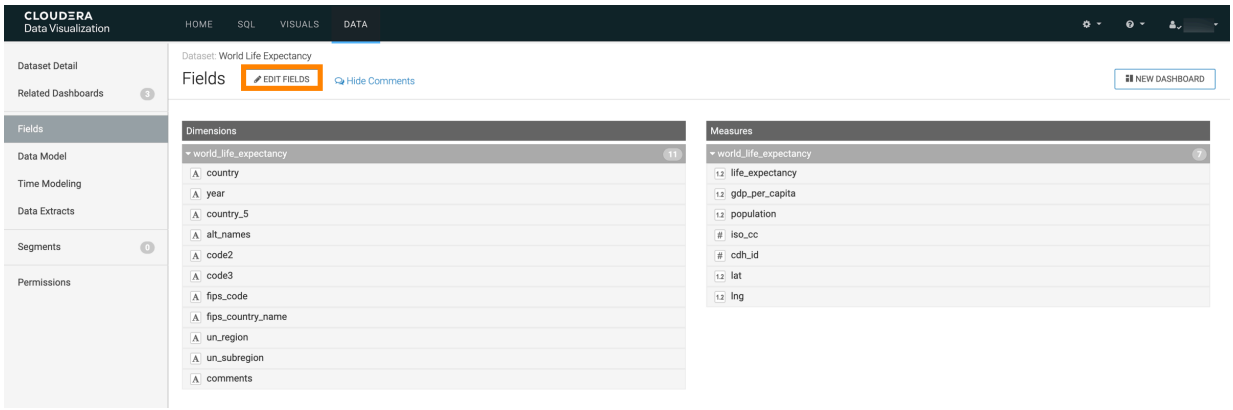
2. In the left navigation menu, click samples.
3. In the Datasets area, select World Life Expectancy (samples.world_life_expectancy).



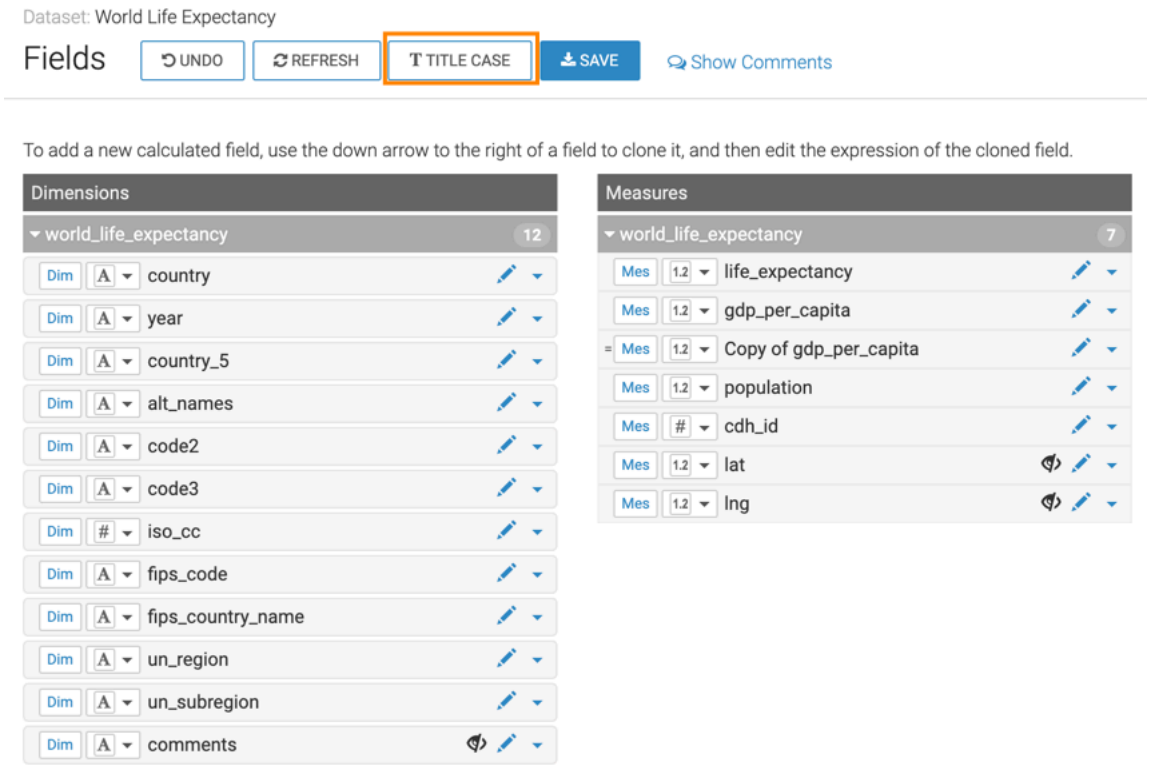
4. In the Dataset Detail menu, select Fields.



5. In the Fields interface, select Edit Field.



6. Near the top of the page, click Title Case.



The field titles change. For example, the Measure gdp_per_captia appears as GDP Per Capita.

Note: The Title Case option does not affect fields that are calculated over table fields, such as gdp.

7. Click Save.

Custom renaming dataset fields

Often, we find it useful to rename a frequently-used field directly in the dataset, instead of using an alias in each visual. CDP Visualization makes it very easy to change the display name of a dataset field.

About this task

The following steps demonstrate how to rename a field. We are using the column iso_cc (the ISO-compliant country code) in the dataset World Life Expectancy [data source: samples.world_life_expectancy].

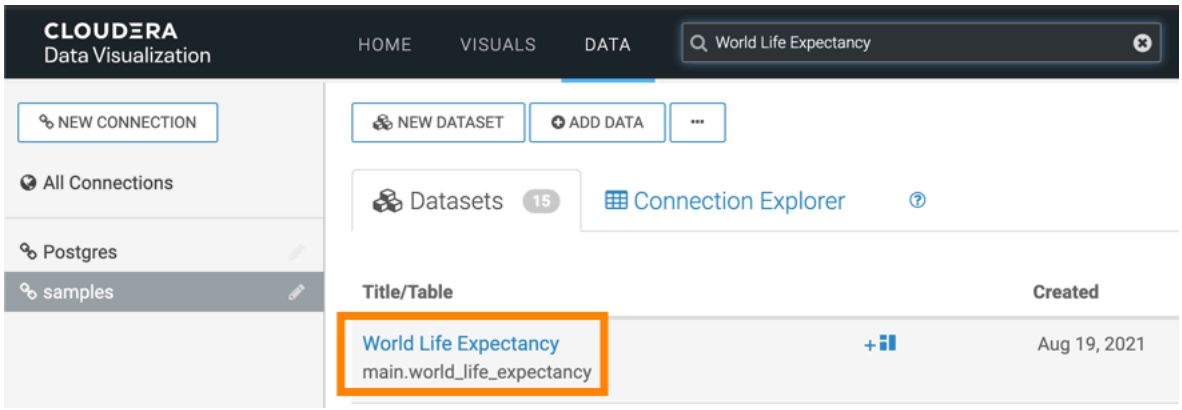
Procedure

1. On the main navigation bar, click Data.

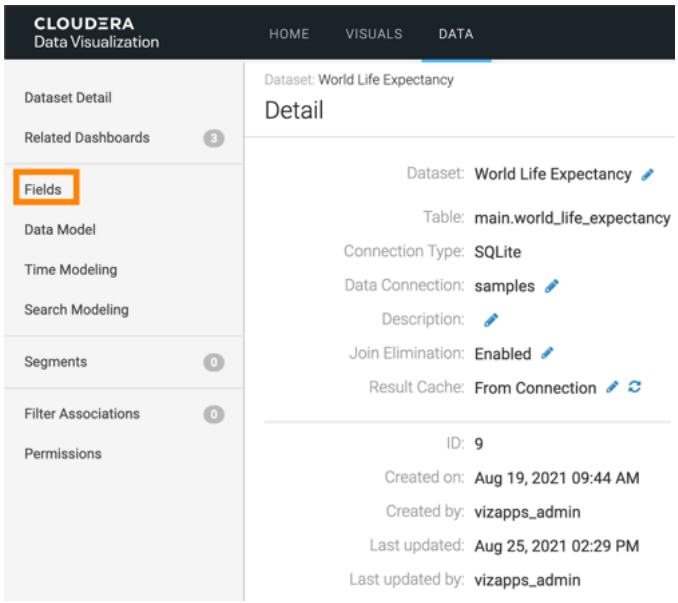
The Data view appears, open on the Datasets tab.

2. In the left navigation menu, click samples.

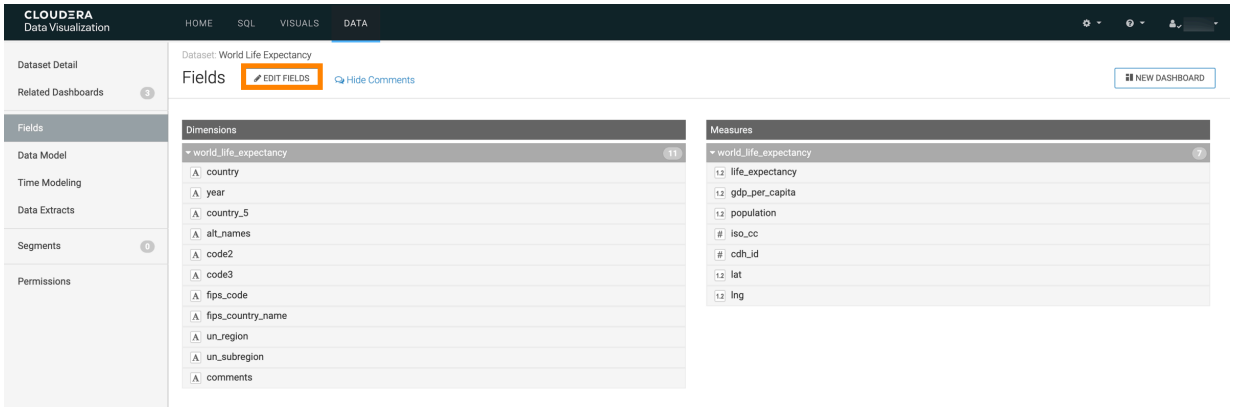
3. In the Datasets area, select World Life Expectancy (samples.world_life_expectancy).



4. In the Dataset Detail menu, select Fields.

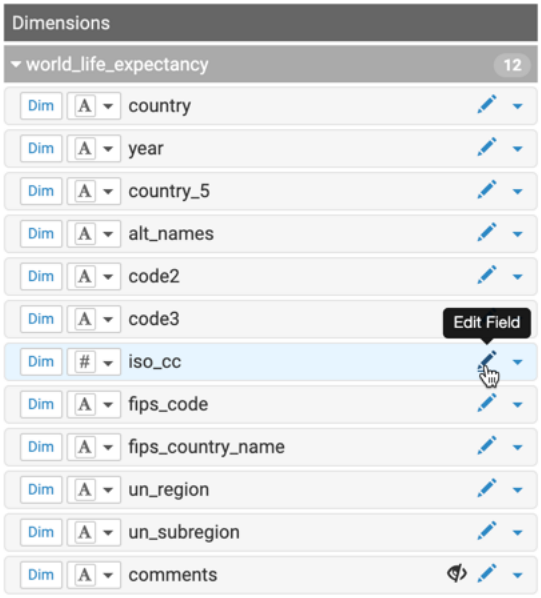


5. In the Fields interface, select Edit Field.



6. Under Dimensions, find the field iso_cc, and click the (down arrow) icon on the right.

7. From the menu, select Edit Fields.



The Edit Field Parameters window modal appears.

Note: The Base Column name cannot be edited, but you can change the Display Name of the column.

8. Change the Display Name from iso_cc to ISO Country Code.
9. Click Apply.
10. Under Dataset: World Life Expectancy, click Save.

Results

All new visuals created from this dataset use the new name automatically.

Working with data models in CDP Data Visualization

You can expand the basic one-table dataset by creating joins with other relevant tables from the same or other data stores. Combining data across multiple tables enriches the dataset considerably. It enables you to conduct much more meaningful research and produce insightful visual analytics.

There are two distinct approaches for creating data joins for the purpose of visualization:

- Defining in UI is ideal for datasets that include star-type schemas.
- Defining on back end ETL is preferable for fact-fact joins, so they may be pre-materialized.

Related Information

Data modeling

Creating a join

Learn how you can create new data joins in a dataset in CDP Data Visualization.

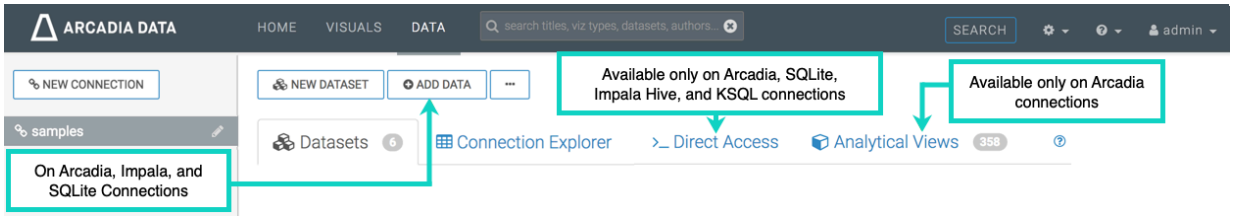
About this task

This example shows you how to create new data joins using the Flight Delays dataset.

Procedure

1. On the main navigation bar, click DATA.

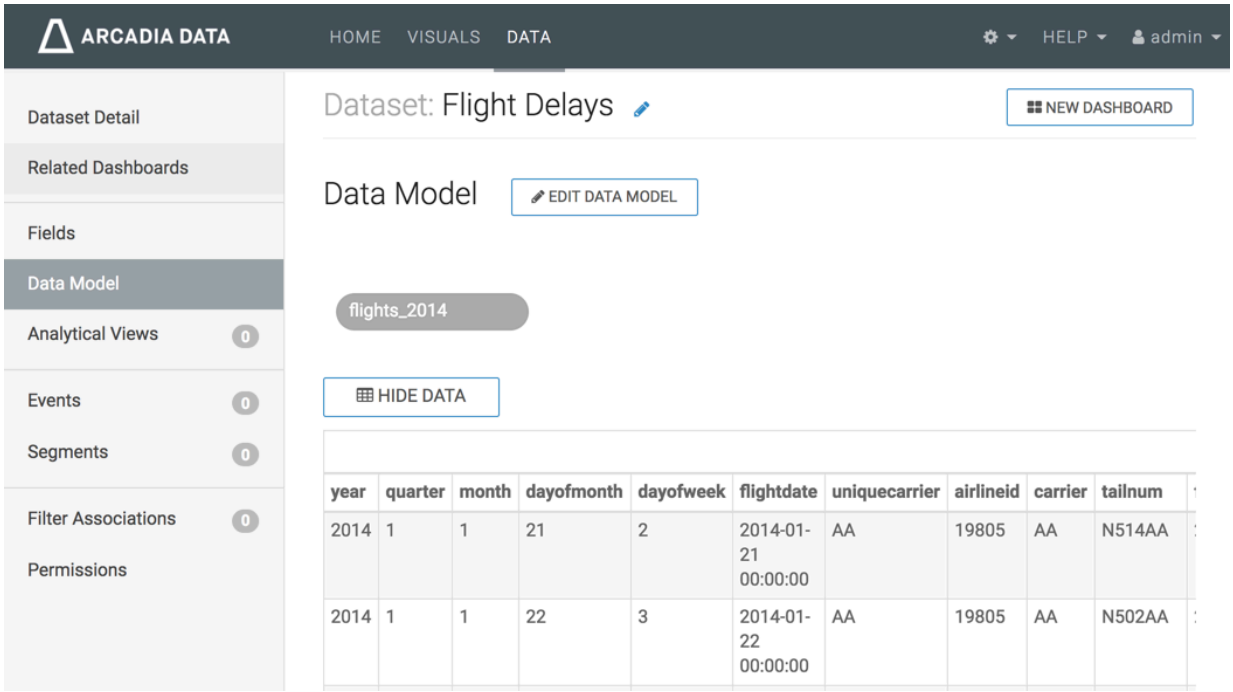
The Data view appears, open on the Datasets tab.



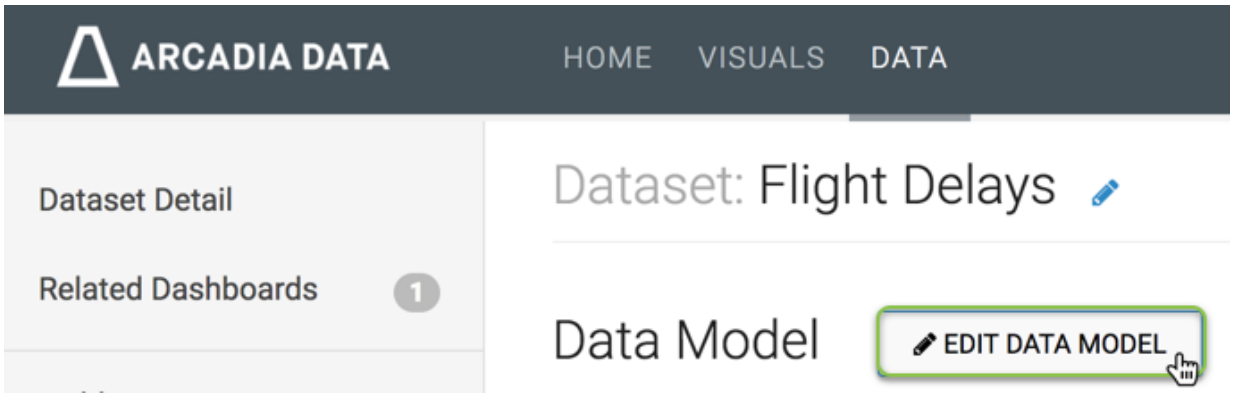
2. Create a new dataset based on the sample datafile.
3. Find the dataset in the list of datasets, either by scrolling or by using search, and click it.

   Dataset side navigation appears, open at Dataset Detail view.
4. In the side navigation menu, click Data Model.

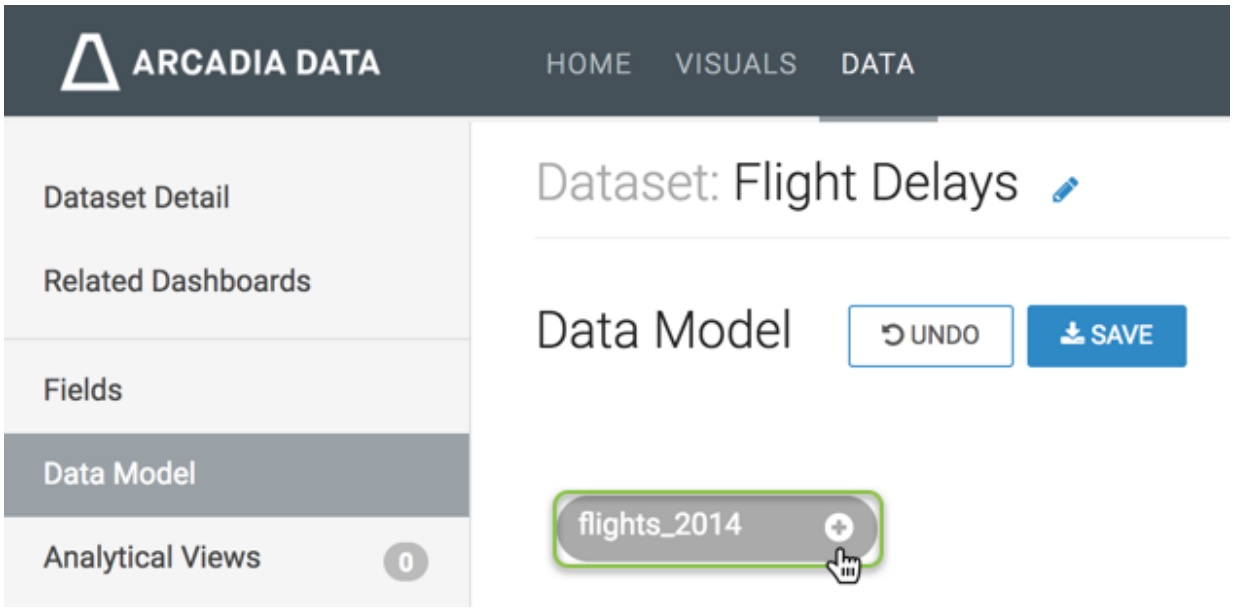
   The Data Model view appears, and shows the name of the only table in the dataset. You may click SHOW DATA to display the data of that table.



5. Click EDIT DATA MODEL.



6. Click the plus sign on the table representation.

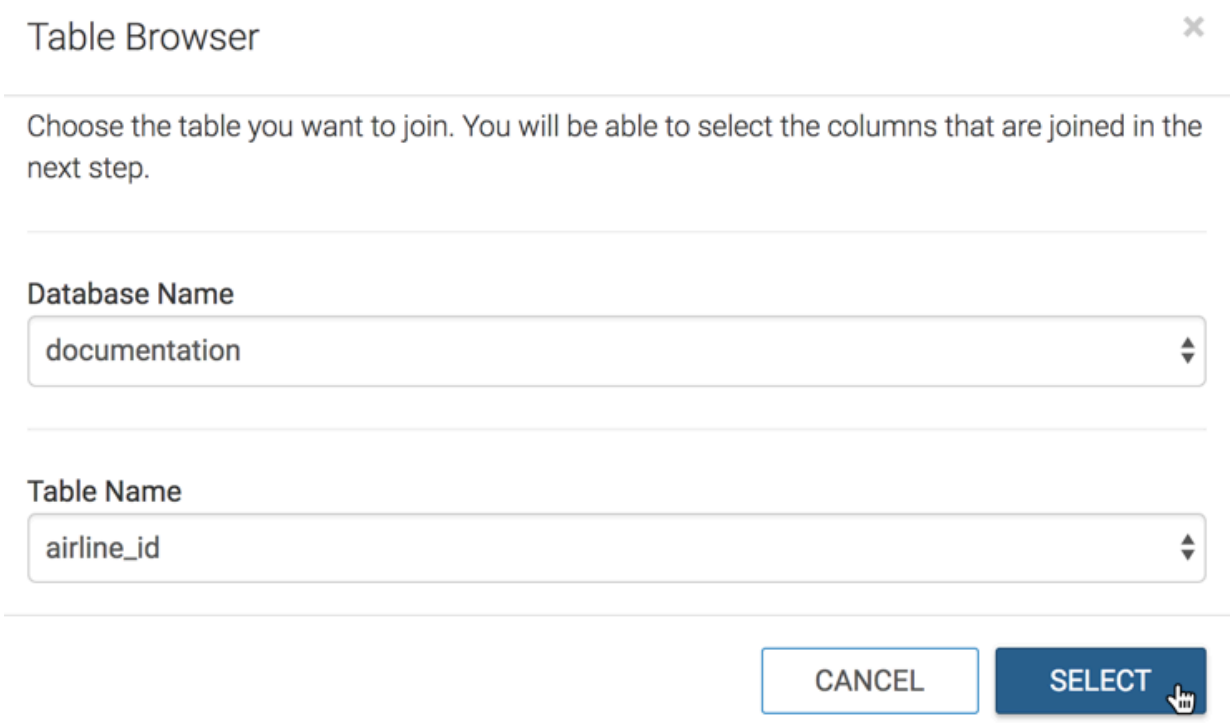


The Table Browser modal window appears.

7. In the Table Browser modal window, make the following selections:
   - In the Database Name selector, choose the database documentation.

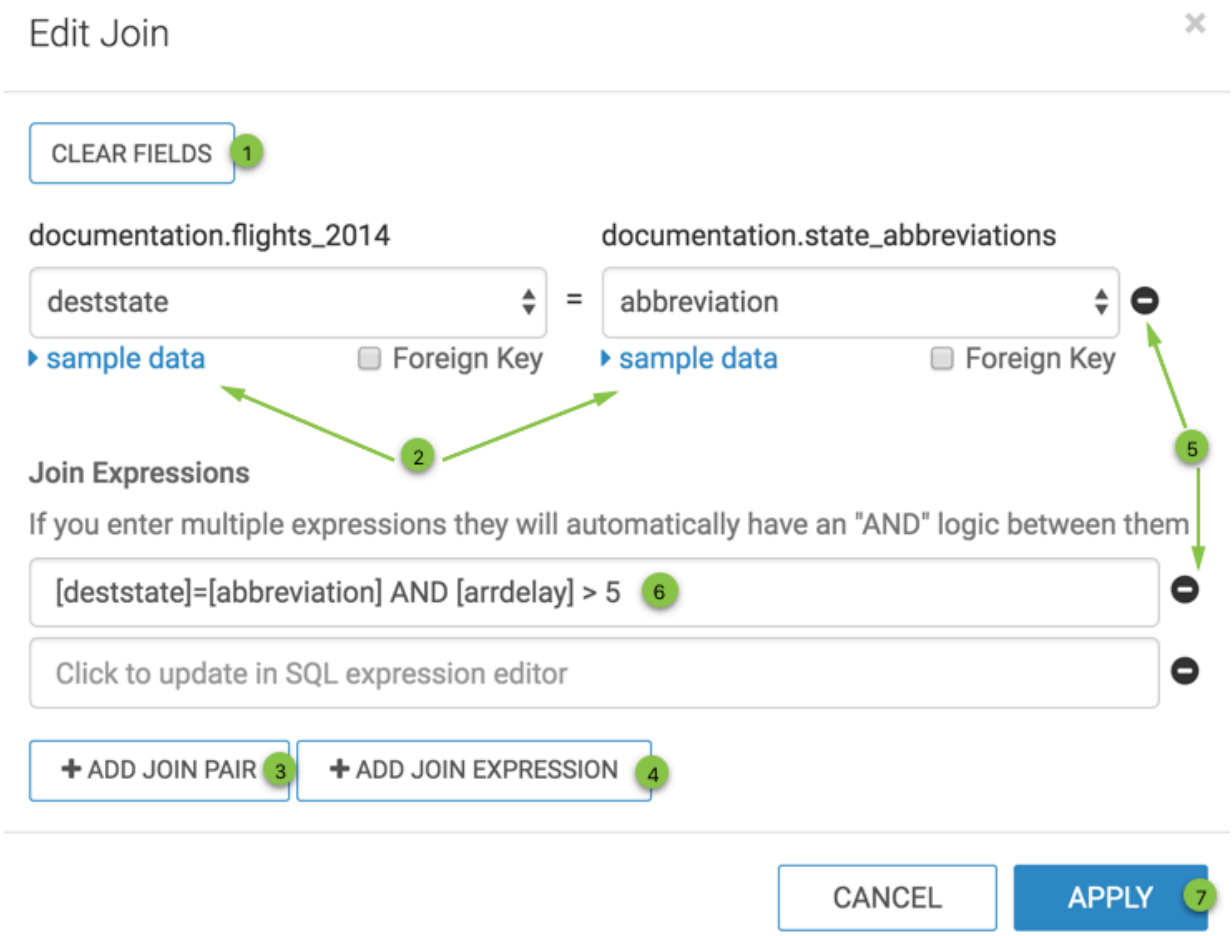
     Note: You can join tables from different databases. This value is pre-populated to match the dataset's existing table, but it may be changed.
   - In the Table Name selector, choose the table name airline_id.

     This value is pre-populated to match the existing table of the dataset, but it may be changed.
   - Click SELECT.



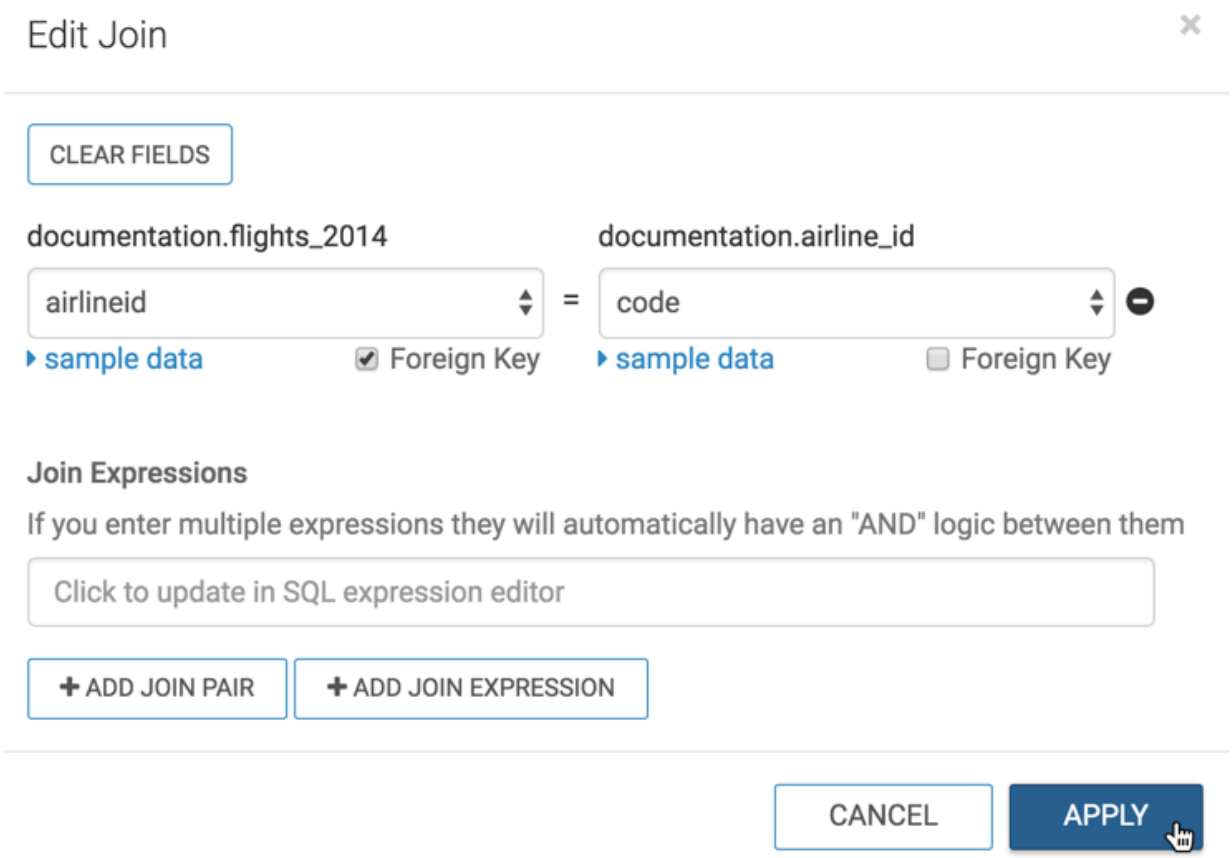
The Edit Join modal window appears.

8. In the Edit Join modal window, the following options are available:
   a. [Optional] Click Clear Fields to clear all already defined joins between the two tables.
   b. [Optional] Click sample data to preview the data. Click again to hide sample data.
   c. [Optional] Click Add Join Pair to add another column connection between the same two tables.
   d. [Optional] Click Add Join Expression to add a join between the two tables based on a custom SQL expression.
   e. [Optional] Click icon (minus) to remove an existing join pair or an existing join expression.
   f. [Optional] Under Join Expressions, click the text box to open the Join Expression interface and specify or update a custom SQL expression that defines the join conditions.
   g. Click APPLY to save the changes.



9. In the Edit Join modal window, do the following:
   - Select the matching columns for both tables. On the left side, select the field airlineid. On the right side, select the field code.
   - Click Sample Data to view some data in both columns, and verify that the join makes sense. Click again to hide sample data.

10. Click APPLY.

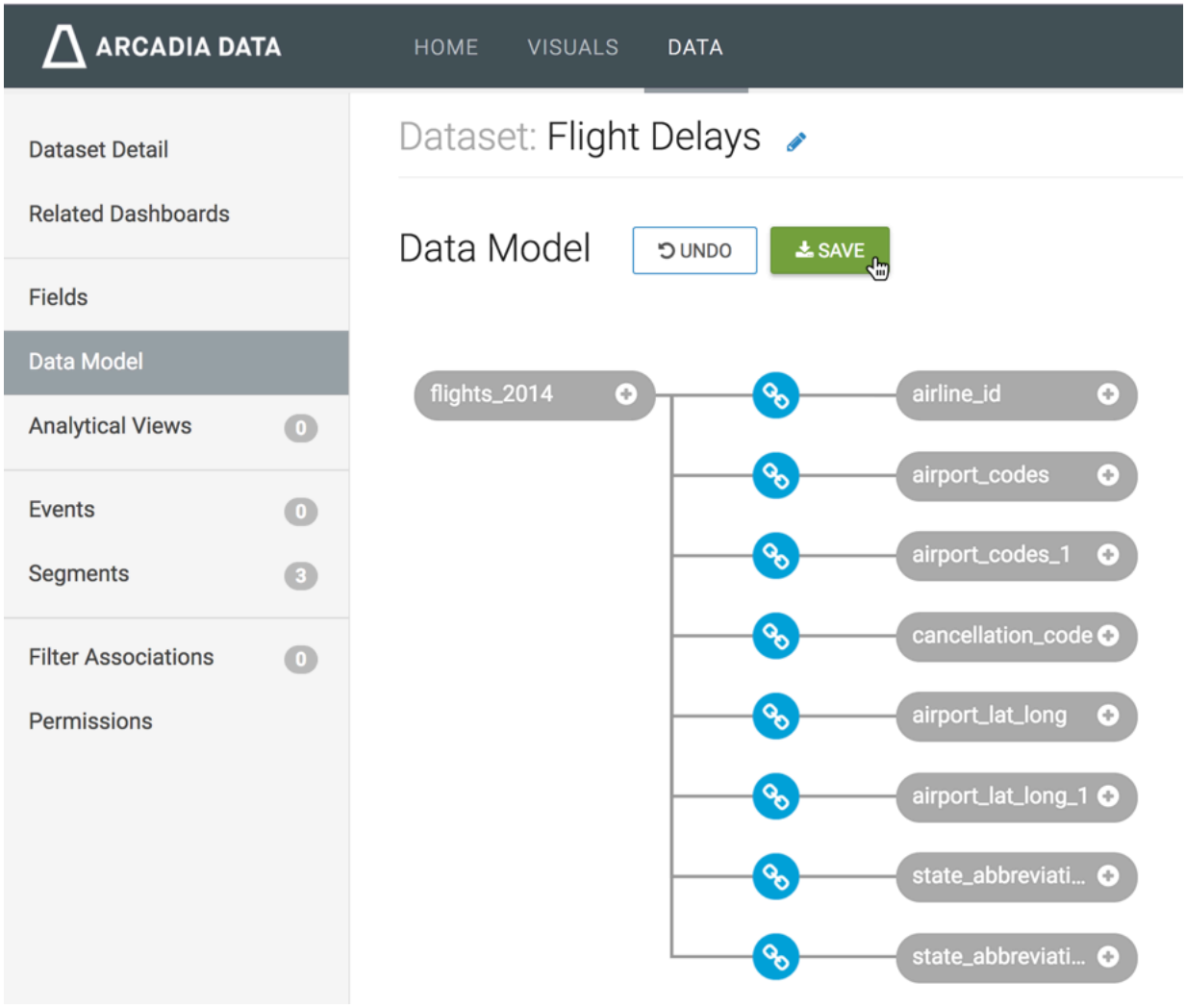


11. Repeat the previous two steps to create seven more joins as follows:

- The table airport_codes has two joins to the main table, and you must create each join separately as follows:
  - Left column origin = right column code.
  - Left column dest = right column code.
- The table cancellation_code has a join for left column cancellationcode = right column code.
- The table airport_lat_long has two joins to the main table, and you must create each join separately as follows:
  - Left column origin = right column locationid.
  - Left column dest = right column locationid.
- The table state_abbreviations has two joins to the main table, and you must create each join separately:
  - Left column deststate = right column abbreviation.
  - Left column originstate = right column abbreviation.

  This step is optional, and depends on whether your flights_* table has fully extended state names.

12. Click SAVE.



13. Click the (link) icon to edit joins or to change join type.

Changing join types

Learn how you can change the join type in a table in CDP Data Visualization.

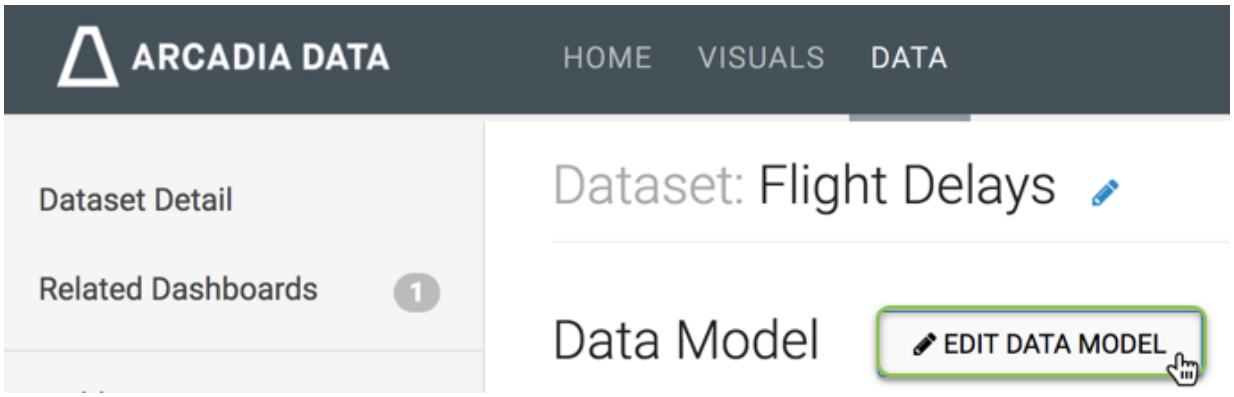
About this task

The following steps demonstrate how to change the join type in the airport_codes table from the default Left join to the Right join.

Procedure

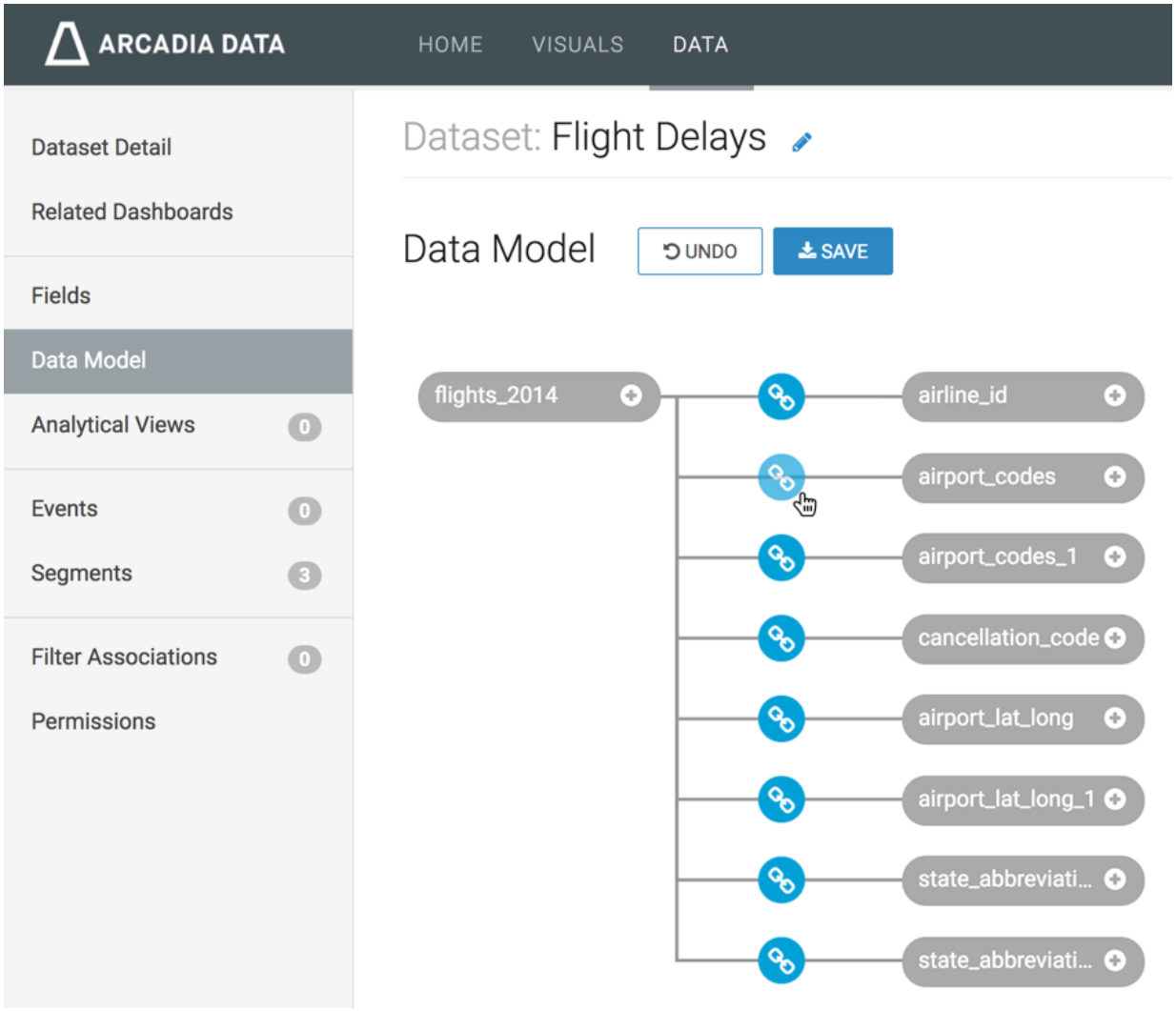
1. Navigate to the Data Model page of the dataset.

2. Click Edit Data Model to edit the data model.



3. Click the (link) icon that represents the connection that must be changed.

   In our example, we clicked the join with the table airport_codes.



   The Join Details modal window appears.

4. In the Join Details modal window, select an alternate join type.

   For example, instead of the default Left join, select Right join.

5. Click outside the Join Details modal window, or click Edit Join Details.
6. Click Save.

Editing join details

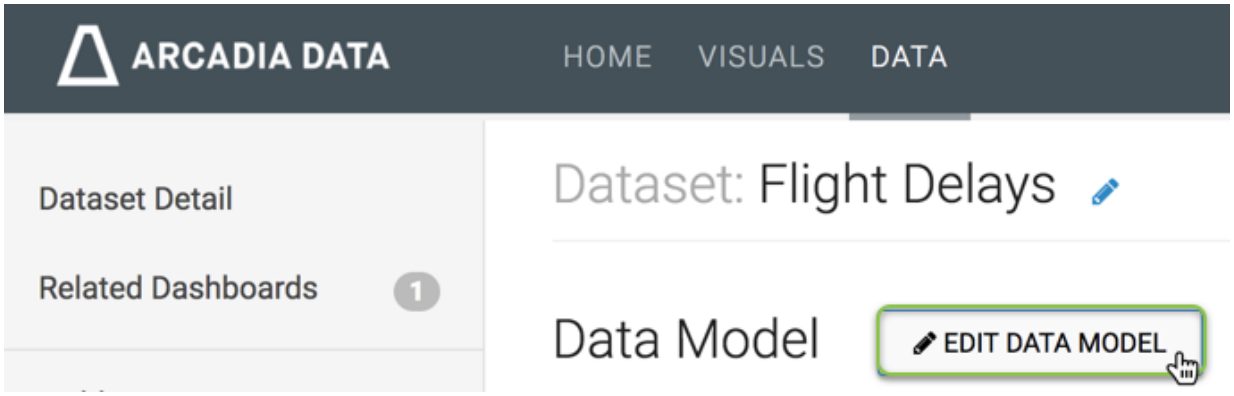
Learn how you can change the specifications of existing table joins in CDP Data Visualization.

About this task

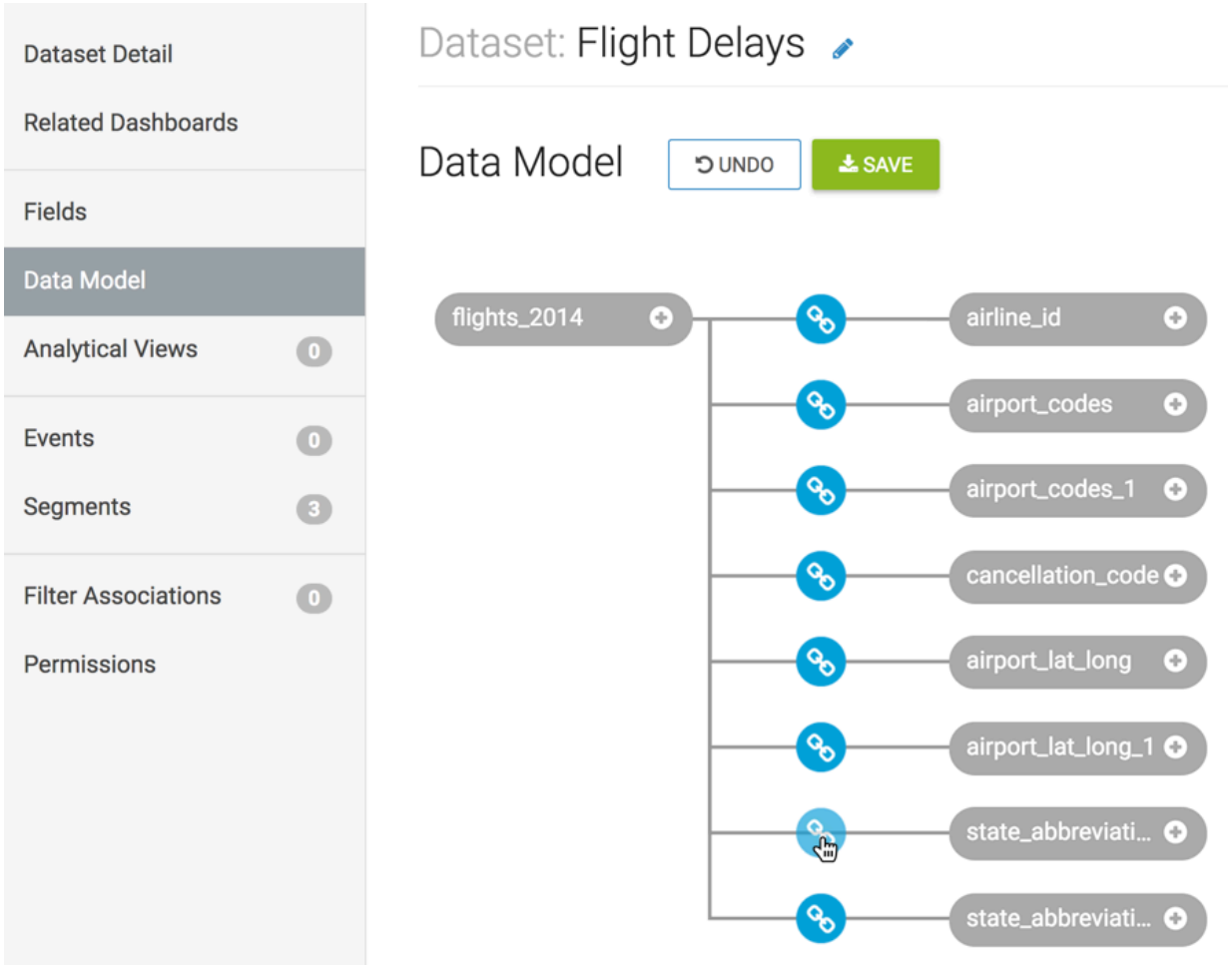
To demonstrate how to create new data joins, we used dataset Flight Delays, based on data previously imported from a sample datafile

Procedure

1. Navigate to the Data Model page of the dataset.
2. Click Edit Data Model to edit the data model.



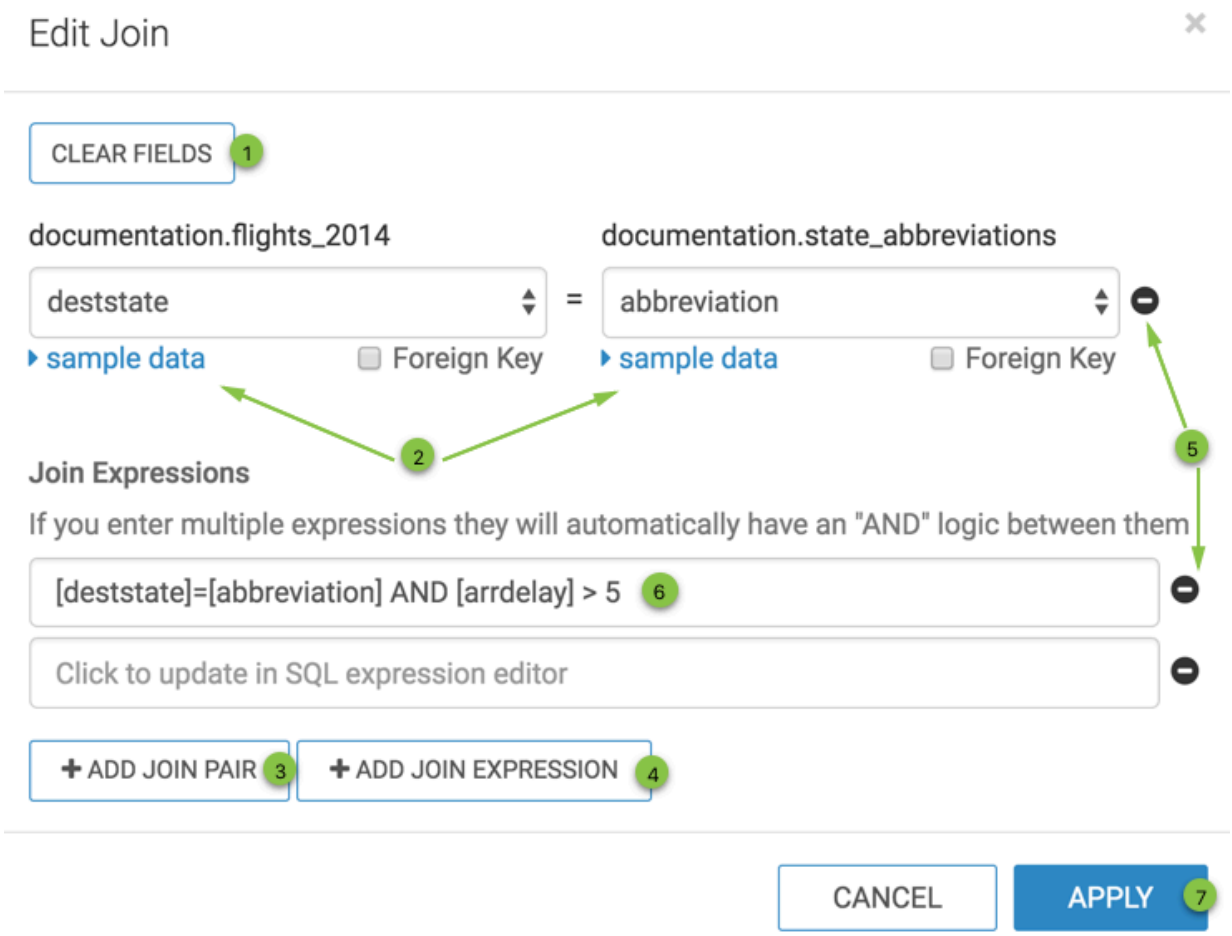
3. Click the (link) icon beside the state_abbreviations connection.



The Join Details modal window appears.

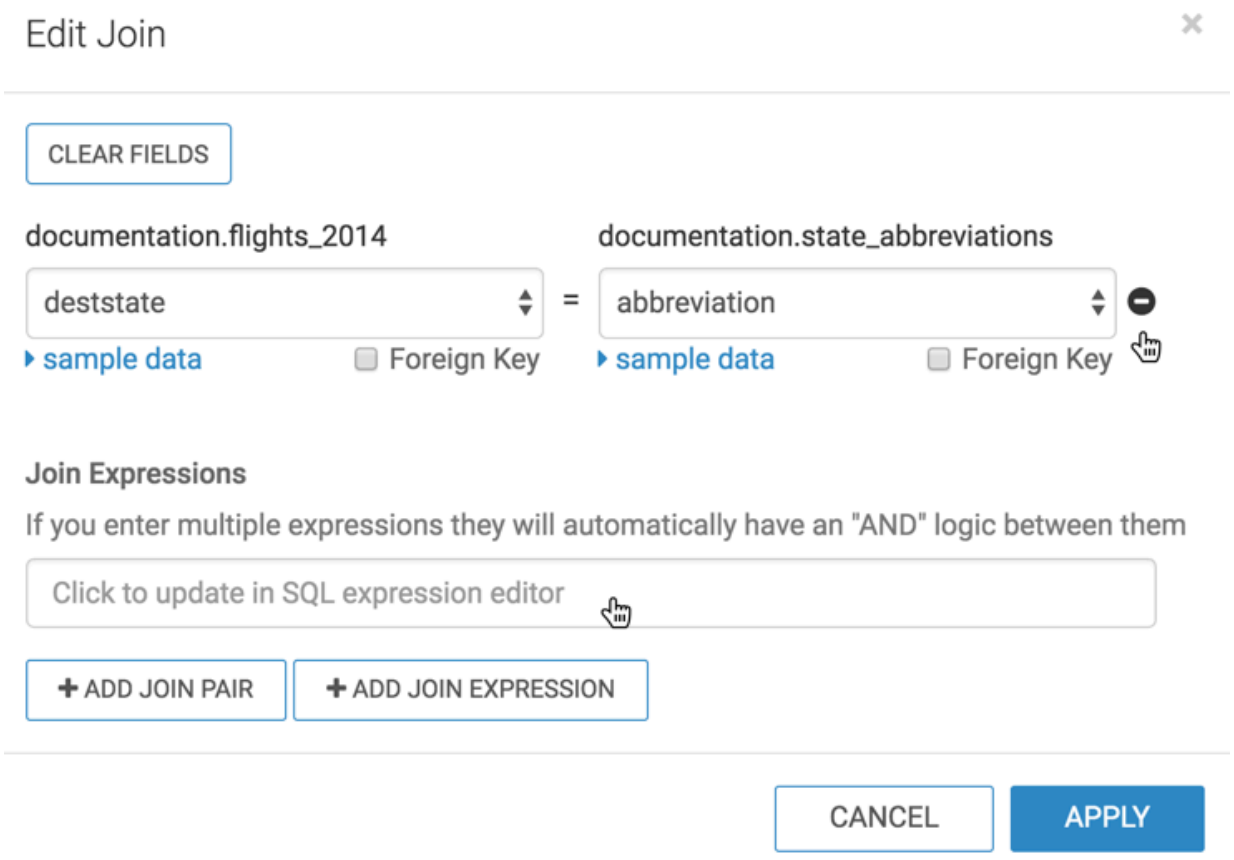
4. Click Edit Join.

5. In the Edit Join modal window, the following options are available:
   a. [Optional] Click Clear Fields to clear all already defined joins between the two tables.
   b. [Optional] Click sample data to preview the data. Click again to hide sample data.
   c. [Optional] Click Add Join Pair to add another column connection between the same two tables.
   d. [Optional] Click Add Join Expression to add a join between the two tables based on a custom SQL expression.
   e. [Optional] Click icon (minus) to remove an existing join pair or an existing join expression.
   f. [Optional] Under Join Expressions, click the text box to open the Join Expression interface. There, specify or update a custom SQL expression that defines the join conditions.
   g. Click Apply to save the changes.



6. To add a a join expression and replace the original field:field join, perform the following steps in the Edit Join modal window:
   a. Remove the initial join between the two columns by clicking the (minus) icon.
   b. Under Join Expressions, click the text box to open the Join Expression interface.

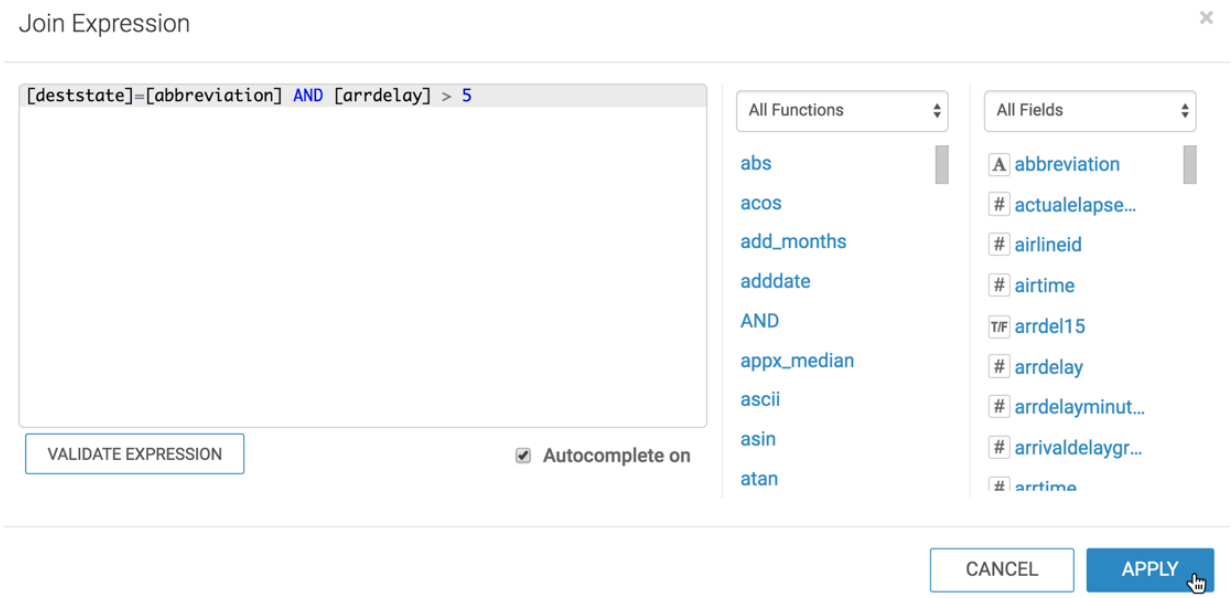
      Here you can specify or update the custom SQL expression that defines the join conditions.



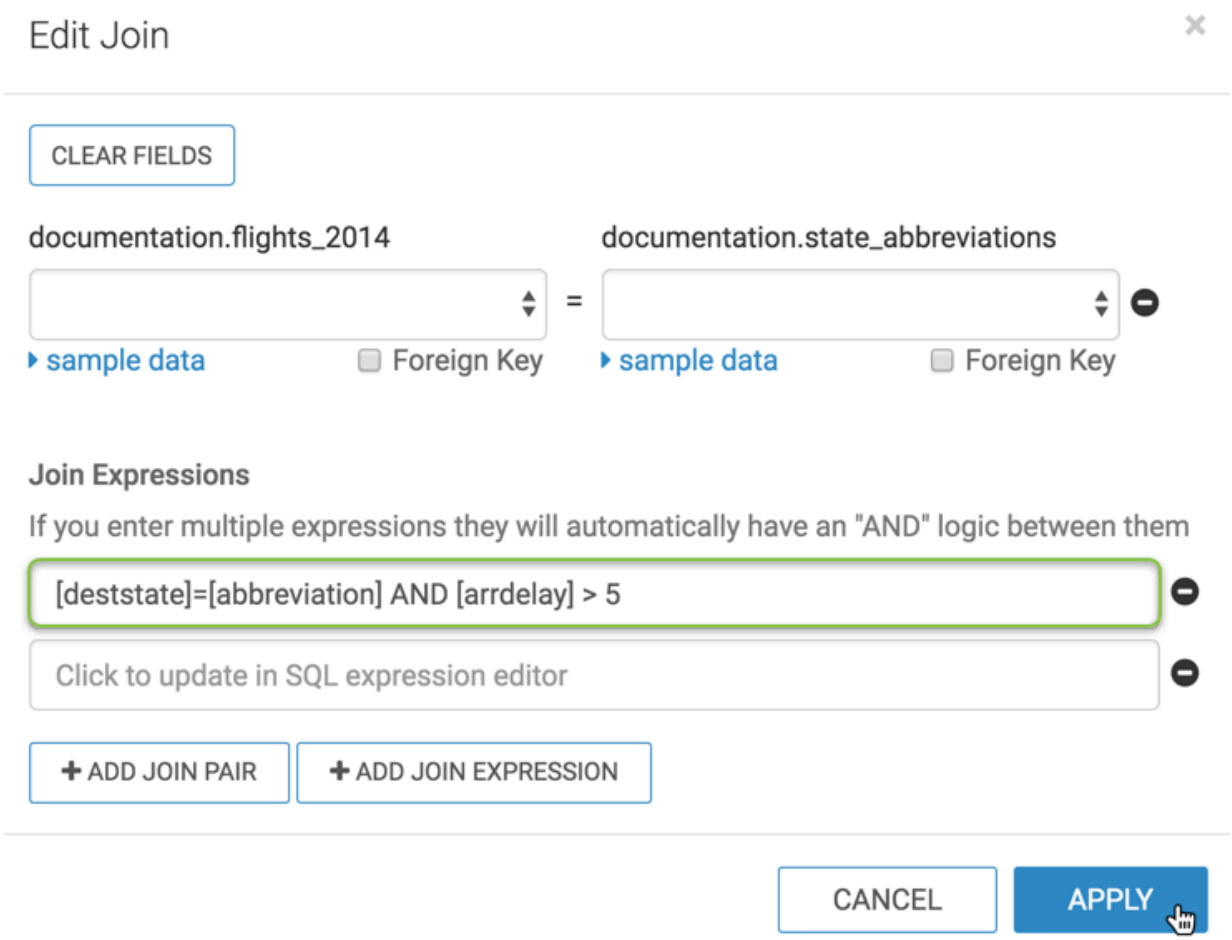
7. In the Join Expression modal window, perform the following steps:
   a. Enter the following expression to show only flights that have significant arrival delay, more than five minutes:

```
[deststate]=[abbreviation] AND [arrdelay] > 5
```

   b. Click Apply to save the expression and return to the Edit Join modal window.



8. In the Edit Join modal window, perform the following steps:
   a. Verify that the initial join between the two columns is deleted and the new join expression appears under Join Expressions.
   b. Click Apply.



   The Data Model interface appears. You can click Show Data to display the updated table.
9. [Optional] To revert this change prior to saving, click Undo.
10. Click Save.

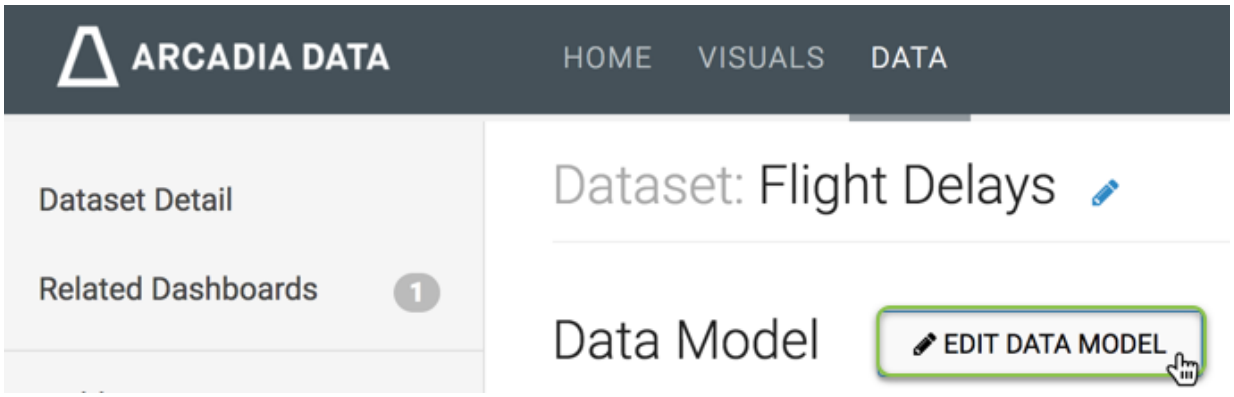
Deleting a join

Learn how you can remove an existing join between tables.

Procedure

1. Navigate to the Data Model page of the dataset.

2. Click Edit Data Model to edit the data model.

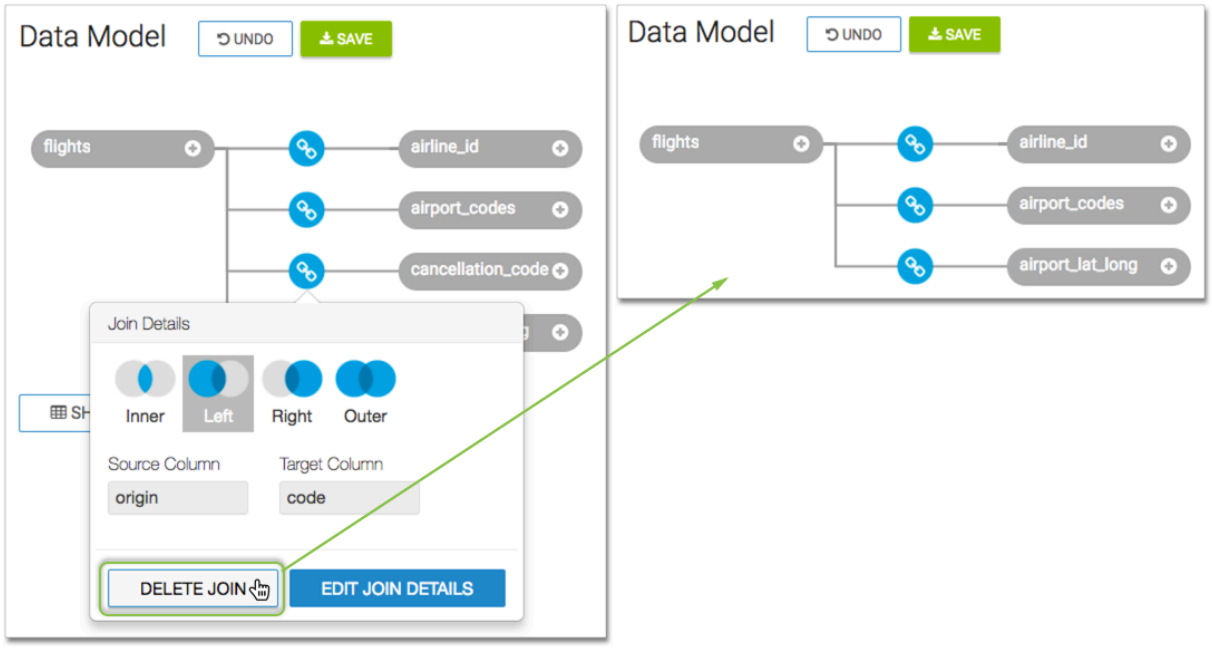


3. Click the (link) icon that represents the connection that must be changed.

   The Join Details modal window appears.

4. Click Delete Join.

   In our example, we deleted the cancellation_code connection. Notice that this table no longer appears in the Data Model.



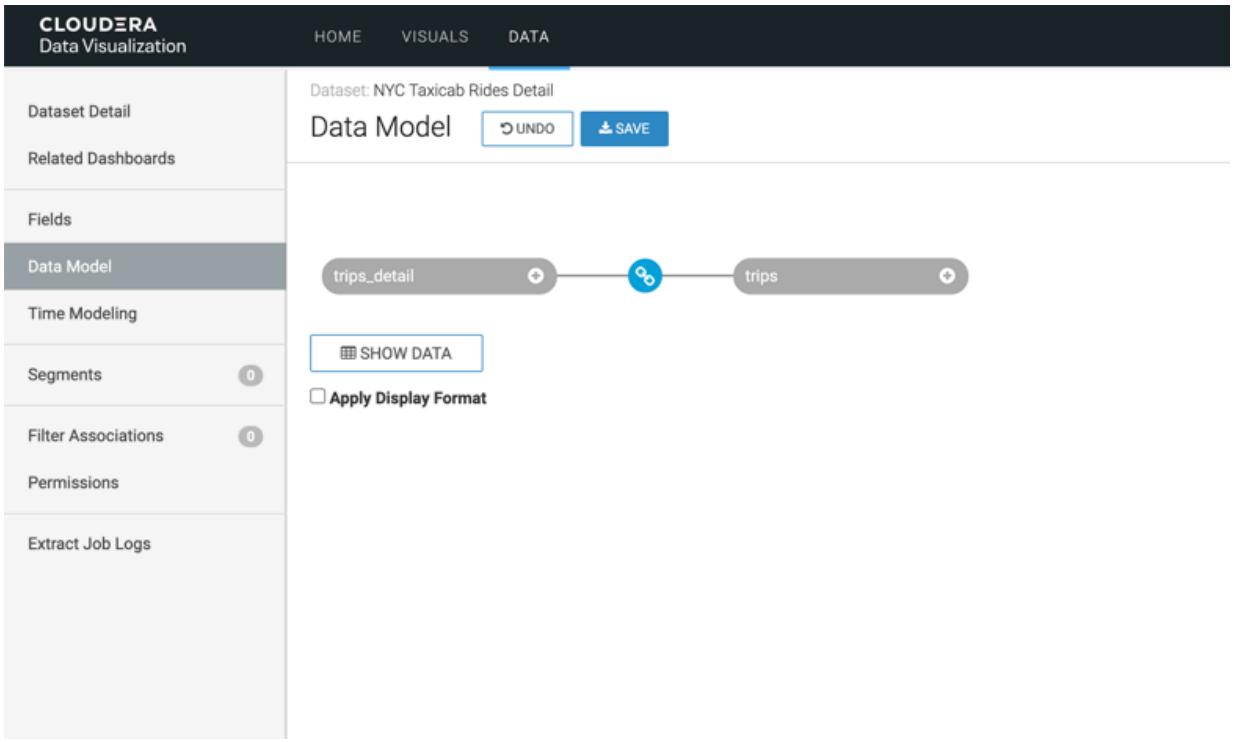
5. [Optional] To revert this change prior to saving, click Undo.
6. Click Save.

Applying field display format on sample data

Learn how you can test field display formats you have configured.

Procedure

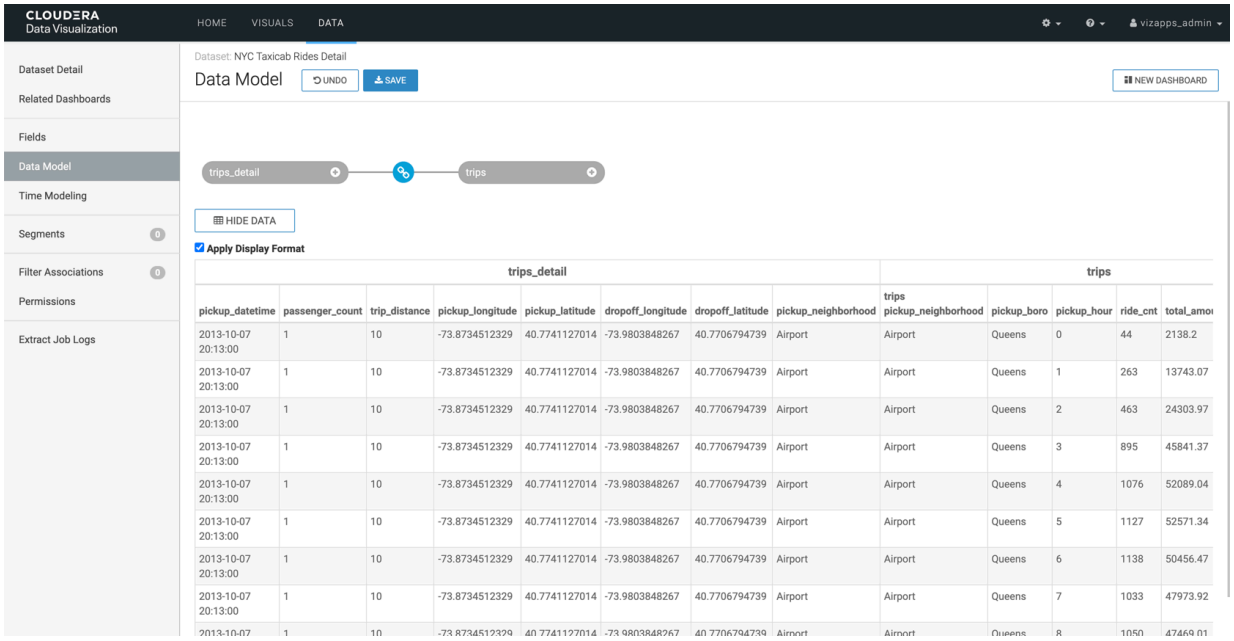
1. Navigate to the Data Model page of the dataset.



2. Select Apply Display Format.
   Selecting/deselecting the checkbox applies or removes the formatting without refetching the data.

   Note: The display format checkbox setting will be remembered.

   For information on how to configure field display at the dataset level, see *Changing the field display format*.
3. You can click SHOW DATA to view a sample of your data model.



Related Information

Changing the field display format