

# Working with Data

Date published: 2020-10-30

Date modified: 2022-09-21

# CLOUDERA

# Legal Notice

© Cloudera Inc. 2024. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

**Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.**

# Contents

<b>Working with datasets in CDP Data Visualization.....</b>	<b>4</b>
Creating a dataset.....	4
Creating datasets in the Data interface.....	4
Creating datasets from a query.....	5
Creating datasets in the Connection Explorer.....	6
Finding a dataset.....	7
Exploring dataset details.....	7
Tables.....	10
Connection type.....	10
Description.....	10
Join elimination.....	11
Result cache.....	11
SQL.....	12
Information on creation and update.....	12
Checking related dashboards.....	12
Editing datasets based on a SQL query.....	13
Restricting columns in datasets based on SQL query.....	14
Restricting rows in datasets based on SQL query.....	17
Deleting a dataset.....	19
<b>Changing dataset fields.....</b>	<b>21</b>
Hiding dataset fields from applications.....	21
Changing field aggregation.....	26
Creating calculated fields.....	29
Testing calculated fields.....	33
Changing the field display format.....	35
Changing currency field display format.....	36
Changing custom field display format.....	40
Changing custom Javascript field display format.....	42
Changing data type.....	45
Specifying geographic fields.....	46
Adding field comments.....	49
Adding field comments in dataset.....	49
Viewing field comments in Visual Builder.....	53
Automatically renaming dataset fields.....	53
Custom renaming dataset fields.....	55
<b>Working with data models in CDP Data Visualization.....</b>	<b>57</b>
Creating a join.....	57
Changing join types.....	63
Editing join details.....	65
Deleting a join.....	70
Applying field display format on sample data.....	71

# 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 and 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

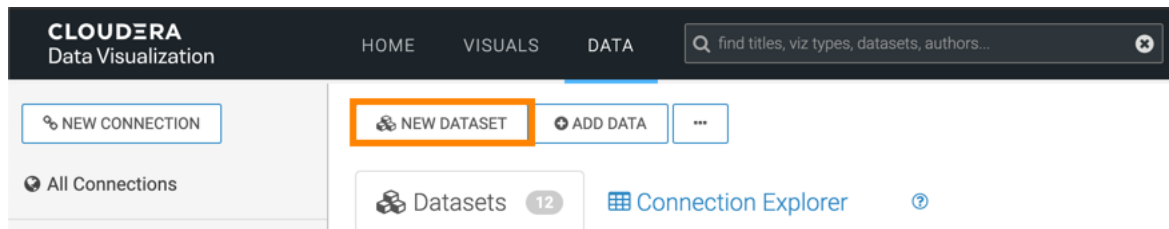
There are three options for creating datasets:

- [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.

New Dataset

Create a dataset from data on this connection. You need to create a dataset before you can create dashboards or apps.

Dataset title \*

Dataset Source

From Table ▾

Select Database

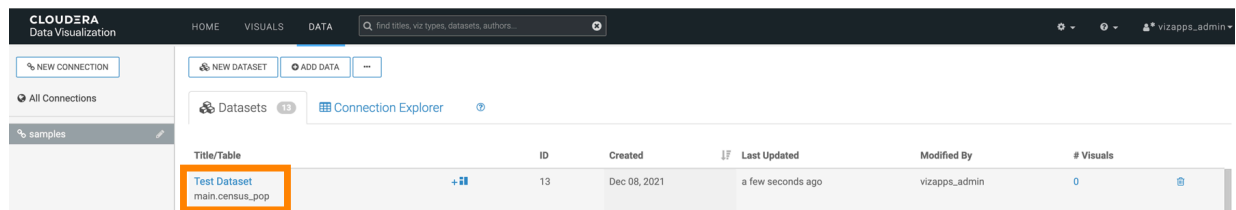
main ▾

Select Table

census\_pop ▾

4. Click CREATE.

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



Title/Table	ID	Created	Last Updated	Modified By	# Visuals
Test Dataset main.census_pop	13	Dec 08, 2021	a few seconds ago	vizapps_admin	0



**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.

New Dataset

Create a dataset from data on this connection. You need to create a dataset before you can create dashboards or apps.

Dataset title \*

Test Dataset from SQL

Dataset Source

From SQL

Enter SQL below

```
select * from main.us_counties
```

Autocomplete on

CANCEL CREATE

4. Click CREATE.

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



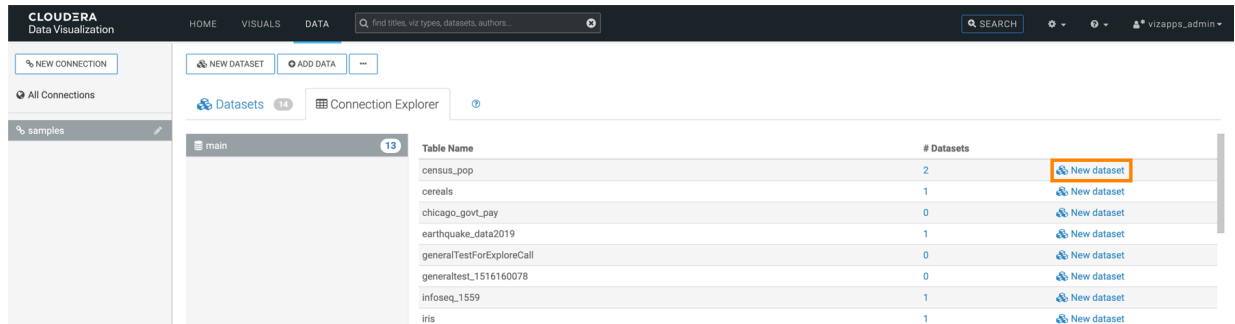
**Tip:** To find the dataset in the future, you can scroll through the list of datasets on the connection, or use Search at the top of the page.

## 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.

The screenshot shows the 'New Dataset' modal window. The title of the modal is 'New Dataset for table main.census\_pop'. Below the title, there is a label 'Dataset title \*' and a text input field containing the text 'Test Dataset from Connection'. At the bottom of the modal, there are two buttons: 'CANCEL' and 'CREATE'.

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.

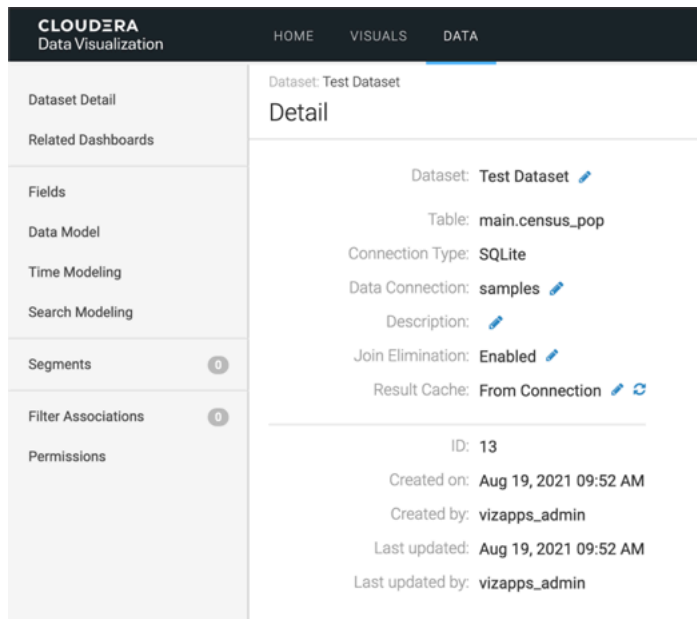


3. Click the dataset you want to examine.

Dataset side navigation appears, open at Dataset Detail view.

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

- Defined on Table



The screenshot displays the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'VISUALS', and 'DATA'. The left sidebar contains navigation options: 'Dataset Detail', 'Related Dashboards', 'Fields', 'Data Model', 'Time Modeling', 'Search Modeling', 'Segments' (with a '0' indicator), 'Filter Associations' (with a '0' indicator), and 'Permissions'. The main content area is titled 'Dataset: Test Dataset' and 'Detail'. It lists the following information: Dataset: Test Dataset, Table: main.census\_pop, Connection Type: SQLite, Data Connection: samples, Description, Join Elimination: Enabled, and Result Cache: From Connection. Below this, it shows the ID: 13, Created on: Aug 19, 2021 09:52 AM, Created by: vizapps\_admin, Last updated: Aug 19, 2021 09:52 AM, and Last updated by: vizapps\_admin.

- Defined on SQL

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'VISUALS', and 'DATA'. The left sidebar contains a menu with items: 'Dataset Detail', 'Related Dashboards', 'Fields', 'Data Model', 'Time Modeling', 'Search Modeling', 'Segments' (with a '0' badge), 'Filter Associations' (with a '0' badge), and 'Permissions'. The main content area is titled 'Dataset: Test Dataset from SQL' and 'Detail'. It displays the following information:

- Dataset: Test Dataset from SQL
- Connection Type: SQLite
- Data Connection: samples
- Description: (empty field)
- Join Elimination: Enabled
- Result Cache: From Connection
- SQL for first table: `select * from main.us_counties`
- ID: 14
- Created on: Aug 19, 2021 09:55 AM
- Created by: vizapps\_admin
- Last updated: Aug 19, 2021 09:55 AM
- Last updated by: vizapps\_admin

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 description 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 enable 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.

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

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'VISUALS', and 'DATA'. The left sidebar contains a menu with items like 'Dataset Detail', 'Related Dashboards', 'Fields', 'Data Model', 'Time Modeling', 'Search Modeling', 'Segments', 'Filter Associations', and 'Permissions'. The main content area is titled 'Dataset: Test Dataset' and 'Detail'. It displays the following information:

- Dataset: Test Dataset
- Table: main.census\_pop
- Connection Type: SQLite
- Data Connection: samples
- Description:
- Join Elimination: Enabled
- Result Cache:  From Connection  Disabled  Enabled 0 Retention Time (seconds) [CANCEL] [SAVE]
- ID: 13
- Created on: Aug 19, 2021 09:52 AM
- Created by: vizapps\_admin
- Last updated: Aug 19, 2021 09:52 AM
- Last updated by: vizapps\_admin

## 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

- Click the Edit (pencil) icon.
- Edit the SQL statement.
- 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

- On the main navigation bar, click Data.



The Data view appears, open on the Datasets tab.

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

Dataset side navigation appears, open at Dataset Detail view.

- In the side navigation menu, click Related Dashboards.

Title	ID	Related Dashboards/Linked Visuals	Created	Last Updated	Modified By	Total Views	Workspace	Actions
Life Expectancy Dashboard	66	1	Aug 19, 2021	6 days ago	vizapps_admin	0	Public	
World Population & GDP Trends	57	1	Aug 19, 2021	6 days ago	vizapps_admin	0	Public	
Animated world population - GDP vs life expectancy	52	1	Aug 19, 2021	6 days ago	vizapps_admin	0	Public	

Showing 1 to 3 of 3 entries

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

- 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.

The screenshot displays the CloudEra Data Visualization interface. The top navigation bar includes 'HOME', 'VISUALS', and 'DATA'. The left sidebar contains a menu with items: 'Dataset Detail', 'Related Dashboards', 'Fields', 'Data Model', 'Time Modeling', 'Search Modeling', 'Segments' (with a '0' indicator), 'Filter Associations' (with a '0' indicator), and 'Permissions'. The main content area is titled 'Dataset: Test Dataset from SQL' and 'Detail'. It lists the following configuration details: 'Connection Type: SQLite', 'Data Connection: samples', 'Description:', 'Join Elimination: Enabled', and 'Result Cache: From Connection'. A text box labeled 'SQL for first table:' contains the query `select * from main.us_counties`. Below this, the dataset's metadata is shown: 'ID: 14', 'Created on: Aug 19, 2021 09:55 AM', 'Created by: vizapps\_admin', 'Last updated: Aug 19, 2021 09:55 AM', and 'Last updated by: vizapps\_admin'.

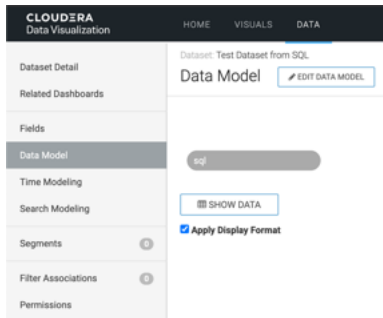
3. You can restrict rows and columns depending on what you need.

### 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.

The screenshot shows the Cloudera Data Visualization interface with the 'HIDE DATA' button highlighted. The main area displays a large table with 25 columns and 10 rows of data. The columns are: sumlev, state, county, stname, ctyname, year, agegrp, tot\_pop, tot\_male, tot\_female, wa\_male, wa\_female, ba\_male, ba\_female, ia\_male, ia\_female, aa\_male, aa\_female, na\_male, na\_female, tom\_male, tom\_female, wac\_r. The data rows represent different counties in Virginia, such as Prince George County, Prince William County, Pulaski County, Rappahannock County, Richmond County, Roanoke County, Rockbridge County, Rockingham County, Russell County, and Scott County.

sumlev	state	county	stname	ctyname	year	agegrp	tot_pop	tot_male	tot_female	wa_male	wa_female	ba_male	ba_female	ia_male	ia_female	aa_male	aa_female	na_male	na_female	tom_male	tom_female	wac_r
50	51	149	Virginia	Prince George County	5	0	36941	20368	16573	12155	10721	7230	4763	171	98	254	410	61	55	497	526	12576
50	51	153	Virginia	Prince William County	5	0	430289	213820	216469	141918	138857	44291	47256	2453	2331	16249	18465	405	374	8504	9186	14956
50	51	155	Virginia	Pulaski County	5	0	34736	17284	17452	15915	16222	959	866	35	37	93	105	6	2	276	220	16174
50	51	157	Virginia	Rappahannock County	5	0	7456	3694	3762	3420	3496	181	171	5	13	19	28	1	2	68	52	3483
50	51	159	Virginia	Richmond County	5	0	9059	5066	3993	3138	2925	1799	961	24	15	31	17	1	2	73	73	3204
50	51	161	Virginia	Roanoke County	5	0	92901	44385	48516	39834	43608	2424	2605	85	82	1338	1465	12	21	692	735	40487
50	51	163	Virginia	Rockbridge County	5	0	22394	11071	11323	10482	10693	321	323	60	63	50	76	1	3	157	165	10631
50	51	165	Virginia	Rockingham County	5	0	77391	37840	39551	36012	37908	841	629	227	213	232	278	15	11	513	512	36497
50	51	167	Virginia	Russell County	5	0	28445	13914	14531	13631	14249	148	120	28	35	32	24	0	1	75	102	13701
50	51	169	Virginia	Scott County	5	0	22781	11413	11368	11197	11172	92	68	25	23	17	29	8	7	74	69	11262

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.

CLUDERA  
Data Visualization

HOME VISUALS DATA

Dataset: Test Dataset from SQL

Detail

Dataset: [Test Dataset from SQL](#)

Connection Type: SQLite

Data Connection: [samples](#)

Description: [Description](#)

Join Elimination: [Enabled](#)

Result Cache: [From Connection](#)

SQL for first table: `select county, stname, ctyname, tot_pop, tot_male, tot_female from main.us_counties` [CANCEL](#) [SAVE](#)

ID: 14

Created on: Aug 19, 2021 09:55 AM

Created by: vizapps\_admin

Last updated: Aug 19, 2021 09:55 AM

Last updated by: vizapps\_admin

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

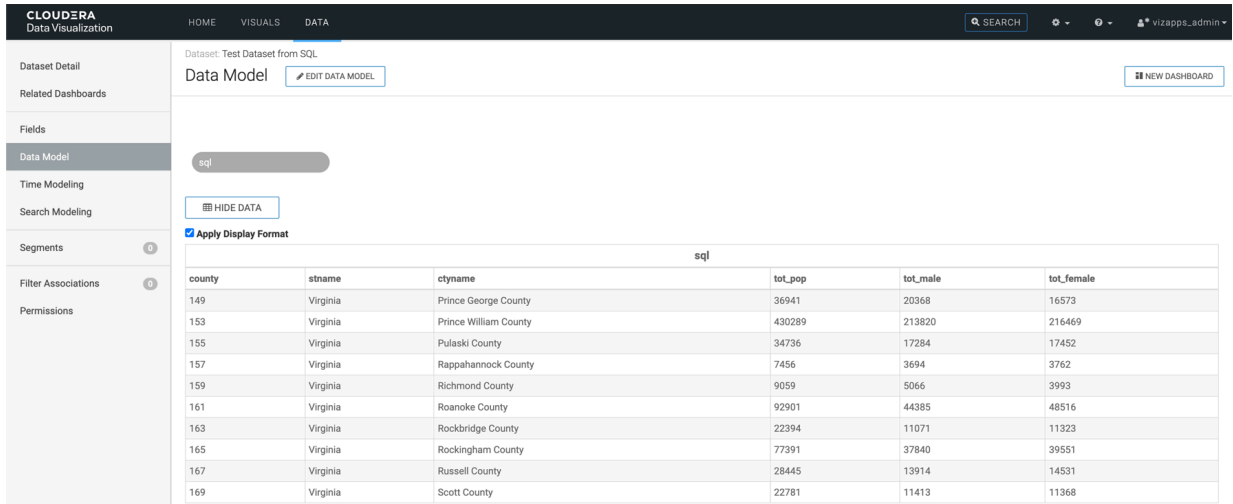
Refresh dataset table column information

Table columns updated

[CLOSE](#)



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



The screenshot shows the Cloudera Data Visualization interface. The left sidebar contains navigation options: Dataset Detail, Related Dashboards, Fields, Data Model (selected), Time Modeling, Search Modeling, Segments, Filter Associations, and Permissions. The main area is titled 'Data Model' and shows a dataset named 'Test Dataset from SQL'. Below the title, there is a search bar containing 'sql', a 'HIDE DATA' button, and a checked 'Apply Display Format' checkbox. The data is presented in a table with the following columns: county, stname, ctyname, tot\_pop, tot\_male, and tot\_female. The table contains 11 rows of data for various counties in Virginia.

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

- 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.

CLUDERA  
Data Visualization

HOME VISUALS DATA

Dataset: Test Dataset from SQL

Detail

Dataset: Test Dataset from SQL

Connection Type: SQLite

Data Connection: samples

Description:

Join Elimination: Enabled

Result Cache: From Connection

SQL for first table: `select county, stname, ctynome, tot_pop, tot_male, tot_female from main.us_counties where stname in ('Arizona', 'New Mexico', 'California', 'Nevada', 'Colorado', 'Utah')`

CANCEL SAVE

ID: 14

Created on: Aug 19, 2021 09:55 AM

Created by: vizapps\_admin

Last updated: Aug 25, 2021 01:35 PM

Last updated by: vizapps\_admin

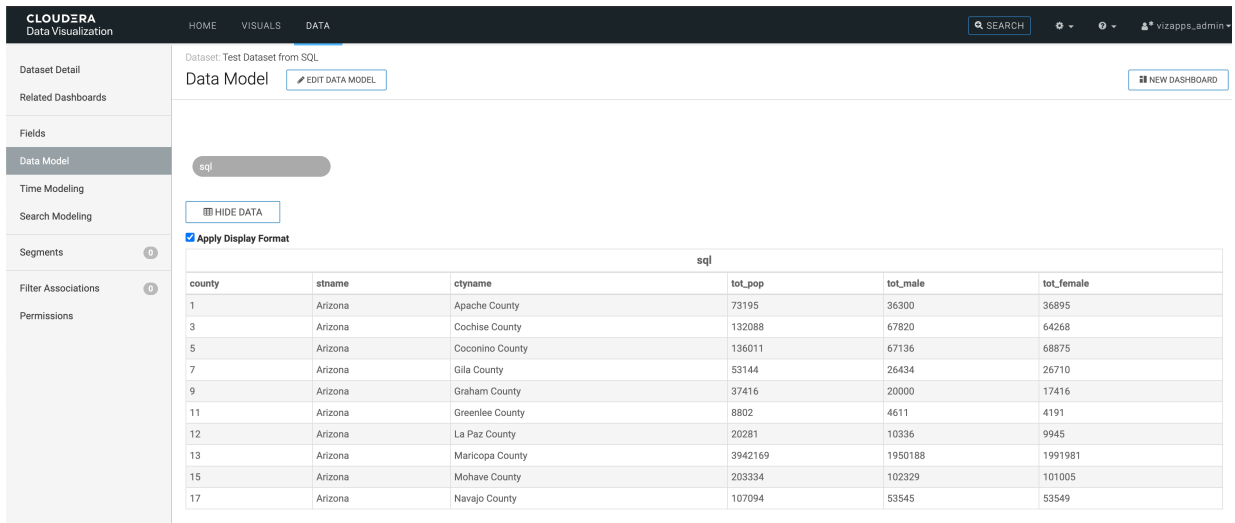
## 3. In the Refresh dataset table column information modal window, click Close.

Refresh dataset table column information

Table columns updated

CLOSE

- 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.



Dataset: Test Dataset from SQL

Data Model [EDIT DATA MODEL](#) [NEW DASHBOARD](#)

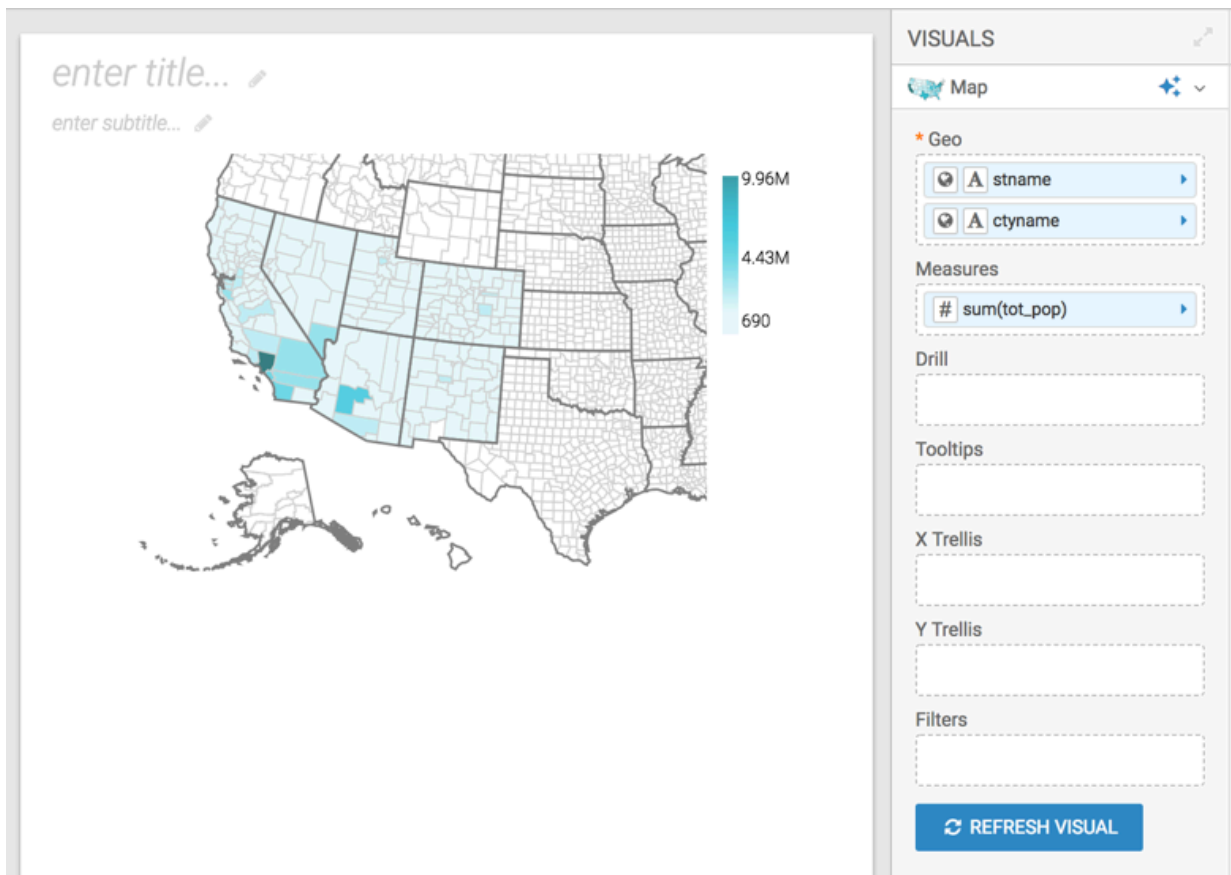
sql

[HIDE DATA](#)

Apply Display Format

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	Cocconino 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

- 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

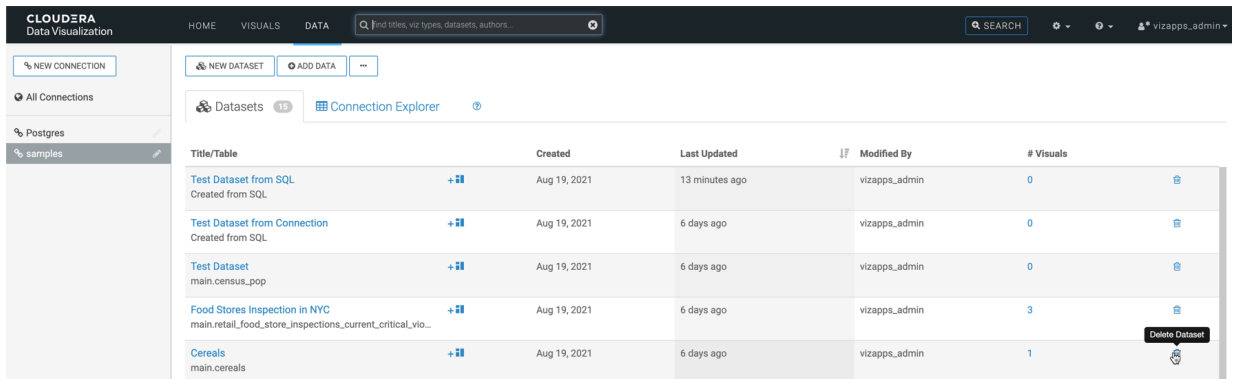
1. On the main navigation bar, click Data.



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 workspaces, 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.

Delete Confirmation ✕

Dataset **Cereals** will be deleted.

The following dashboards/visuals related to this dataset will also be deleted:

	Title	ID	Related Dashboards / Visuals	Created	Last Updated	Modified By	Workspace
	Comparison of average calories per manufacturer to overall average	77	1 78	Aug 19, 2021	6 days ago	vizapps_admin	Public
	Avg quantity per nutritional category	76	1 78	Aug 19, 2021	6 days ago	vizapps_admin	Public
	Cereal Manufacturers available	75	1 78	Aug 19, 2021	6 days ago	vizapps_admin	Public
	Cereal comparisons across 9 nutritional categories	74	1 78	Aug 19, 2021	6 days ago	vizapps_admin	Public
	Cereal	78	4 74, 75, 76, 77	Aug 19, 2021	6 days ago	vizapps_admin	Public

Showing 1 to 5 of 5 entries

To confirm, type in the word DELETE in the field below and then click the Delete button.

CANCEL
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 reference 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.

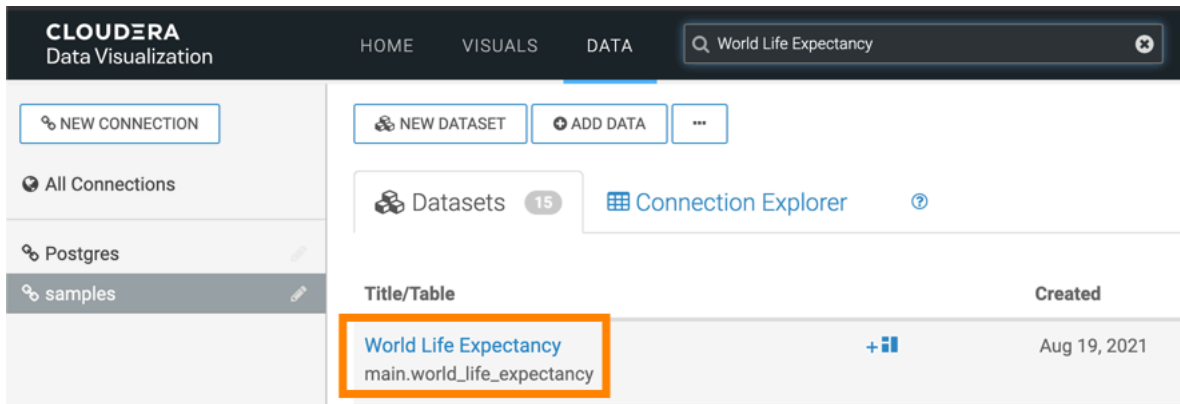
## Procedure

1. On the main navigation bar, click Data.

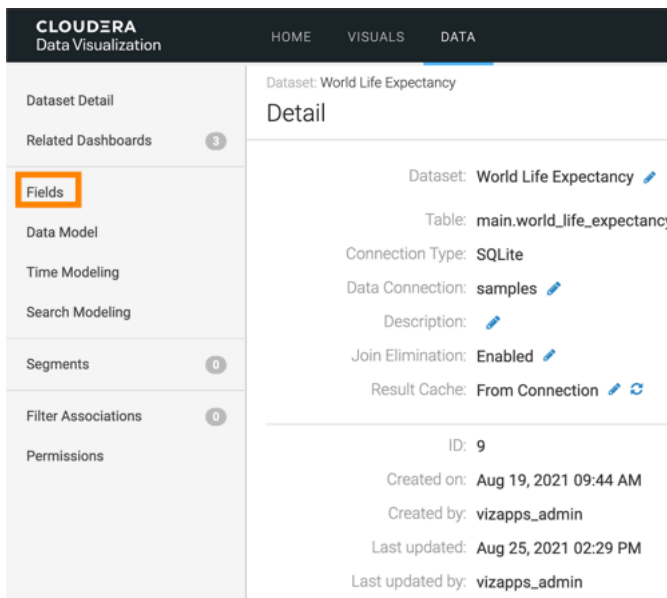


The Data view appears, open on the Datasets tab.

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.

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'VISUALS', and 'DATA'. The main content area is titled 'Dataset: World Life Expectancy' and 'Fields'. A button labeled 'EDIT FIELDS' is highlighted with an orange box. Below the 'Fields' header, there are two columns: 'Dimensions' and 'Measures'. The 'Dimensions' column lists 11 fields, including 'country', 'year', 'country\_5', 'alt\_names', 'code2', 'code3', 'fips\_code', 'fips\_country\_name', 'un\_region', 'un\_subregion', and 'comments'. The 'Measures' column lists 7 fields, including 'life\_expectancy', 'gdp\_per\_capita', 'population', 'iso\_cc', 'cdh\_id', 'lat', and 'lng'.

Dimensions	Measures
world_life_expectancy (11)	world_life_expectancy (7)
country	life_expectancy
year	gdp_per_capita
country_5	population
alt_names	iso_cc
code2	cdh_id
code3	lat
fips_code	lng
fips_country_name	
un_region	
un_subregion	
comments	

6. Under Dimensions, find the field comments.

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

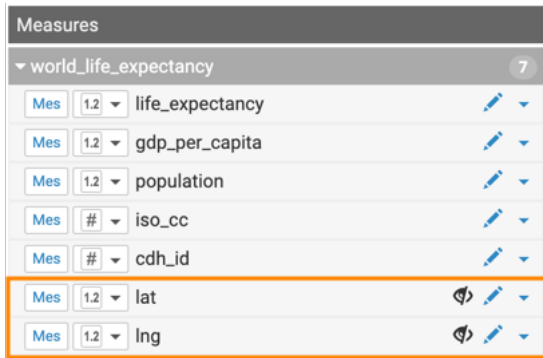
The screenshot shows the Cloudera Data Visualization interface. On the left is a sidebar with navigation options: Dataset Detail, Related Dashboards (3), Fields, Data Model, Time Modeling, Segments (0), Filter Associations (0), and Permissions. The main area is titled 'Dataset: World Life Expectancy' and 'Fields'. It includes buttons for UNDO, REFRESH, TITLE CASE, SAVE, and Show Comments. Below this is a tip: 'To add a new calculated field, use the down arrow to the right of a field to clone it, and then edit the expression of the cloned field.' There are two panels: 'Dimensions' (11 fields) and 'Measures' (7 fields). The 'comments' field in the Dimensions list has a context menu open with options: Clone, Hide (highlighted with an orange box and a mouse cursor), and Create Hierarchy.

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

This screenshot shows the 'Dimensions' list from the previous image. The 'comments' field at the bottom is highlighted with an orange box. The icon next to it is now a slashed eye, indicating that the field has been hidden. The other dimensions remain unchanged.



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.

The screenshot shows the Cloudera Data Visualization interface. The main area displays a table with the following data:

country	year	life_expectancy	gdp_per_capita	population	country_5	alt
Afghanistan	1900	27.2	612	5.22M	Afghanistan	A
Afghanistan	1901	27.2	614	5.26M	Afghanistan	A
Afghanistan	1902	27.2	616	5.29M	Afghanistan	A
Afghanistan	1903	27.2	618	5.33M	Afghanistan	A
Afghanistan	1904	27.1	620	5.37M	Afghanistan	A
Afghanistan	1905	27.1	622	5.41M	Afghanistan	A

The right-hand side of the interface shows the 'Dashboard Designer' panel, which includes a 'DATA' section for 'World Life Expectancy' and a 'Measures' section with a list of fields like 'Record Count', 'life\_expectancy', 'gdp\_per\_capita', 'population', 'iso\_cc', and 'csh\_id'.

**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`).

The screenshot shows the Cloudera Data Visualization interface with the 'World Life Expectancy' dataset selected in the Datasets area. The dataset name 'World Life Expectancy' and its source 'main.world\_life\_expectancy' are highlighted with an orange box. The 'Created' date is 'Aug 19, 2021'.

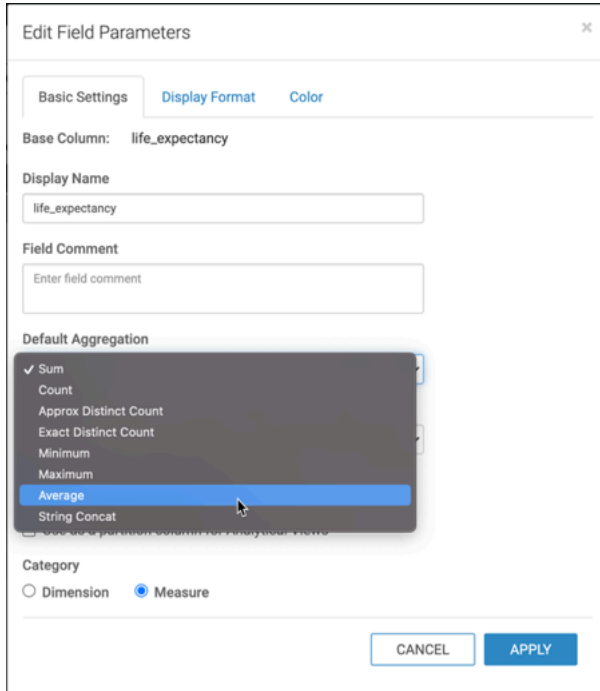
- In the Dataset Detail menu, select Fields.

- In the Fields interface, select Edit Fields.

- Under Measures, find the field life\_expectancy and click the (down arrow) icon on its right side.
- 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 screenshot shows the 'Edit Field Parameters' dialog box for the 'life\_expectancy' field. The 'Basic Settings' tab is active. The 'Base Column' is 'life\_expectancy'. The 'Display Name' is 'life\_expectancy'. The 'Field Comment' is empty. The 'Default Aggregation' dropdown menu is open, showing a list of aggregation options: Sum (checked), Count, Approx Distinct Count, Exact Distinct Count, Minimum, Maximum, Average (highlighted), and String Concat. The 'Category' is set to 'Measure' (radio button selected). The 'CANCEL' and 'APPLY' buttons are at the bottom right.

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.

The screenshot shows the 'Dashboard Designer' interface. On the left, the 'VISUALS' panel is active, showing a 'Table' visual. Below it, there are sections for 'Dimensions', 'Measures', 'Tooltips', and 'Filters'. The 'Measures' section is highlighted with an orange box and contains the calculated field '# avg(life\_expectancy)'. Below the 'Measures' section is a 'Limit' field set to '100' and a 'REFRESH VISUAL' button. On the right, the 'DATA' panel is active, showing the dataset 'World Life Expectancy'. It has a search bar and two sections: 'Dimensions' (10 items) and 'Measures' (6 items). The dimensions listed are: country, year, country\_5, alt\_names, code2, code3, fips\_code, fips\_country\_name, un\_region, and un\_subregion. The measures listed are: Record Count, life\_expectancy, gdp\_per\_capita, population, iso\_cc, and cdh\_id.

## 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 domenic product) in the dataset World Life Expectancy [data source main.world\_life\_expectancy]. We define it by the following equation:

$$\text{gdp} = \text{gdp\_per\_capita} \times \text{population}$$

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 information 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.

- In the Datasets area, select World Life Expectancy (main.world\_life\_expectancy).

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'VISUALS', and 'DATA'. A search bar contains 'World Life Expectancy'. Below the navigation bar, there are buttons for 'NEW CONNECTION', 'NEW DATASET', 'ADD DATA', and a menu icon. A 'Datasets' section shows 15 datasets, with 'World Life Expectancy' (main.world\_life\_expectancy) highlighted in an orange box. The 'Created' column shows 'Aug 19, 2021'.

- In the Dataset Detail menu, select Fields.

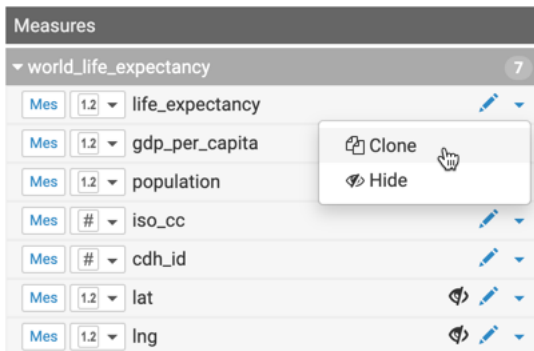
The screenshot shows the Cloudera Data Visualization interface in the Dataset Detail view for 'World Life Expectancy'. The left sidebar contains a menu with options: 'Dataset Detail', 'Related Dashboards', 'Fields' (highlighted with an orange box), 'Data Model', 'Time Modeling', 'Search Modeling', 'Segments', 'Filter Associations', and 'Permissions'. The main content area shows details for the dataset: 'World Life Expectancy', 'Table: main.world\_life\_expectancy', 'Connection Type: SQLite', 'Data Connection: samples', 'Description:', 'Join Elimination: Enabled', and 'Result Cache: From Connection'. Below this, it shows 'ID: 9', 'Created on: Aug 19, 2021 09:44 AM', 'Created by: vizapps\_admin', 'Last updated: Aug 25, 2021 02:29 PM', and 'Last updated by: vizapps\_admin'.

- In the Fields interface, select Edit Field.

The screenshot shows the Cloudera Data Visualization interface in the Fields view for 'World Life Expectancy'. The left sidebar contains a menu with options: 'Dataset Detail', 'Related Dashboards', 'Fields' (selected), 'Data Model', 'Time Modeling', 'Data Extracts', 'Segments', and 'Permissions'. The main content area shows the 'Fields' interface with an 'EDIT FIELDS' button highlighted in an orange box. The interface is divided into two columns: 'Dimensions' and 'Measures'. The 'Dimensions' column lists fields like 'country', 'year', 'country\_5', 'alt\_names', 'code2', 'code3', 'fips\_code', 'fips\_country\_name', 'un\_region', 'un\_subregion', and 'comments'. The 'Measures' column lists fields like 'life\_expectancy', 'gdp\_per\_capita', 'population', 'iso\_cc', 'cdh\_id', 'lat', and 'lng'.

- Under Measures, find the field gdp\_per\_capita, and click the (down arrow) icon on its right side.

- From the menu, select Clone.



- 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.

Edit Field Parameters ×  
 Basic Settings **Expression** Display Format Color  
 Base Column: gdp\_per\_capita  
 Display Name: gdp  
 Field Comment: gdp\_per\_capita \* population  
 Default Aggregation: Sum  
 Geo Type: None  
 Show field in data detail screen  
 Show field in Visual Designer  
 Use as a partition column for Analytical Views  
 Category:  
 Dimension  Measure  
 REMOVE CANCEL APPLY

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.

Edit Field Parameters ×  
 Basic Settings **Expression** Display Format Color  
 Expression: [gdp\_per\_capita] \* [population]  
 All Functions All Fields  
 Expression contains an aggregation  Autocomplete on  
 VALIDATE EXPRESSION  
 Save expression only after validation succeeds  
 Validation Successful! ×  
 REMOVE CANCEL APPLY

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



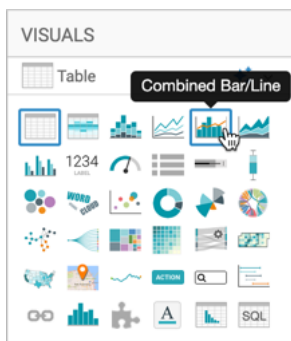
11. Under Dataset: World Life Expectancy, click Save.

## 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.



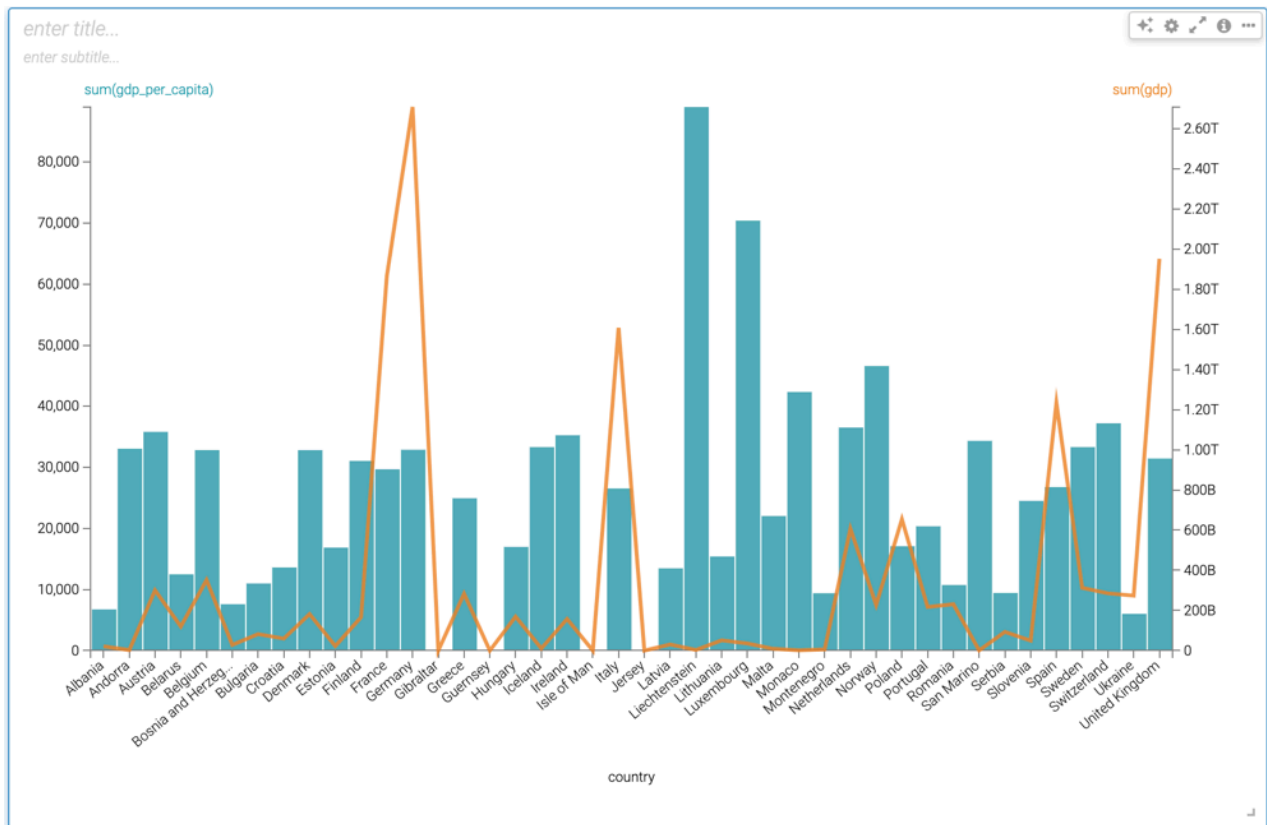
3. Populate the shelves from the available fields in the following way:
  - a) Under Dimensions, select country, and place it on the Dimensions shelf.
  - b) Under Dimensions, select year, and place it on the Filters shelf.
  - c) In the Filter for year modal window, under the Value tab, select the year 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.

The screenshot shows the Tableau Dashboard Designer interface. On the left, the 'VISUALS' pane displays a 'Combo' chart. The 'Dimension' shelf contains 'country'. The 'Bar Measure' shelf contains 'sum(gdp\_per\_capita)'. The 'Line Measure' shelf contains 'sum(gdp)'. The 'Filters' shelf contains 'year in (2010)' and 'un\_region in (Europe)'. A 'REFRESH VISUAL' button is at the bottom. On the right, the 'FIELD PROPERTIES' pane is open for the 'sum(gdp)' field. Under the 'Axis' section, 'Secondary Axis' is selected, indicated by a green checkmark and a mouse cursor. Other options include 'Primary Axis', 'Aggregate Display', 'Display Format', 'Alias', 'Description', 'Duplicate', 'Save Expression', and 'Remove'.

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.

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

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'VISUALS', and 'DATA'. The main content area is titled 'Dataset: World Life Expectancy' and 'Fields'. Below this, there are buttons for 'UNDO', 'REFRESH', 'TITLE CASE', 'SAVE', and 'Show Comments'. A text box explains: 'To add a new calculated field, use the down arrow to the right of a field to clone it, and then edit the expression of the cloned field.' The interface is divided into two main sections: 'Dimensions' and 'Measures'. The 'Dimensions' section lists 11 fields: country, year, country\_5, alt\_names, code2, code3, fips\_code, fips\_country\_name, un\_region, un\_subregion, and comments. The 'Measures' section lists 7 fields: life\_expectancy, gdp\_per\_capita, population, iso\_cc, cdh\_id, lat, and lng. The 'gdp\_per\_capita' field in the Measures section is highlighted with an orange box, and a pencil icon is visible next to it, indicating it is selected for editing.

The Edit Field Parameters modal window appears.

- 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

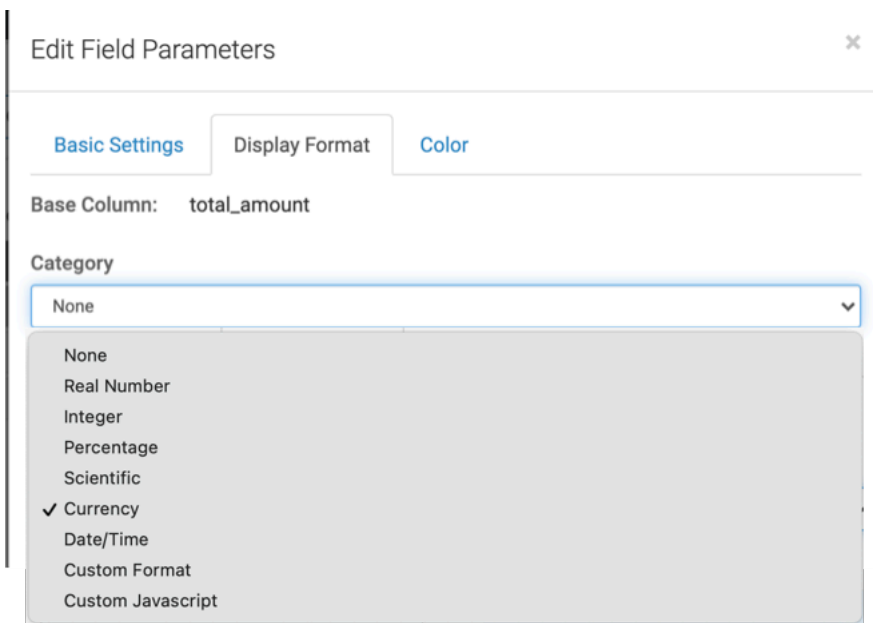
In CDP Data Visualization, you can set currency display options for numerical fields across all visuals of a dataset.

### 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).

### Edit Field Parameters ×

**Basic Settings** | Display Format | Color

Base Column: `gdp_per_capita`

Category: Currency

Example:  →

Currency Symbols

- \$ (Dollar)
- £ (Pound)
- ¥ (Yen/Yuan)
- ₹ (Rupee)
- € (Euro)
- ₵ (Cedi)
- No Currency Symbol

For more documentation, go [here](#)

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*.

Display Format
✕

---

**Category**

Currency
▼

**Example:**

12345.6789123

→

¢12,345.68

---

**Currency Symbols**

¢ (Cedi)
▼

**Basic Formats**

¢12,345.68
▼

**Customize Currency**

¢,.2f
▼

For more documentation, go [here](#)

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

CLOSE

SAVE

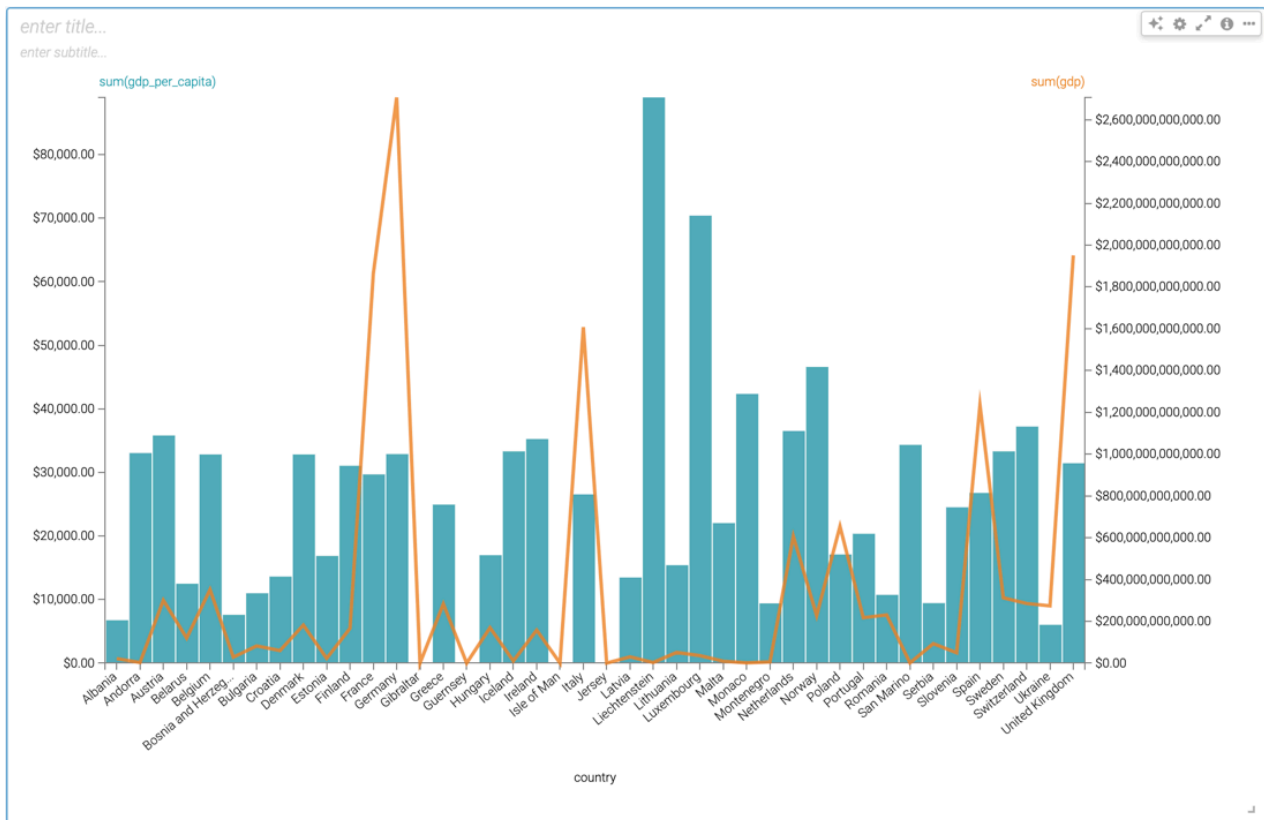
4. Click APPLY.

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 \$\$.

- \$ 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.

Edit Field Parameters

Basic Settings Expression Display Format Color

Base Column: gdp\_per\_capita

Category: Custom Format

Example: 12345.6789123 → \$12.3456789123k

Customize Format: \$\$

Display Format Examples: For more documentation, go [here](#)

Enter	For Display	Enter	For Currencies	Mac Shortcut
f	12346	\$	\$12345	Shift+4
,f	12,346	£	£12345	Alt+3
.,2f	12,345.68	¥	¥12345	Alt+Y
\$.f	\$12,345	₹	₹12345	
.4s	12.35k (for SI notation)	€	€12345	Alt+Shift+2
.4S	12.35k (for currencies)			
.1s	10k			
.2s	12k			
\$.2S	\$12k			
%	1234568%			

REMOVE CANCEL APPLY

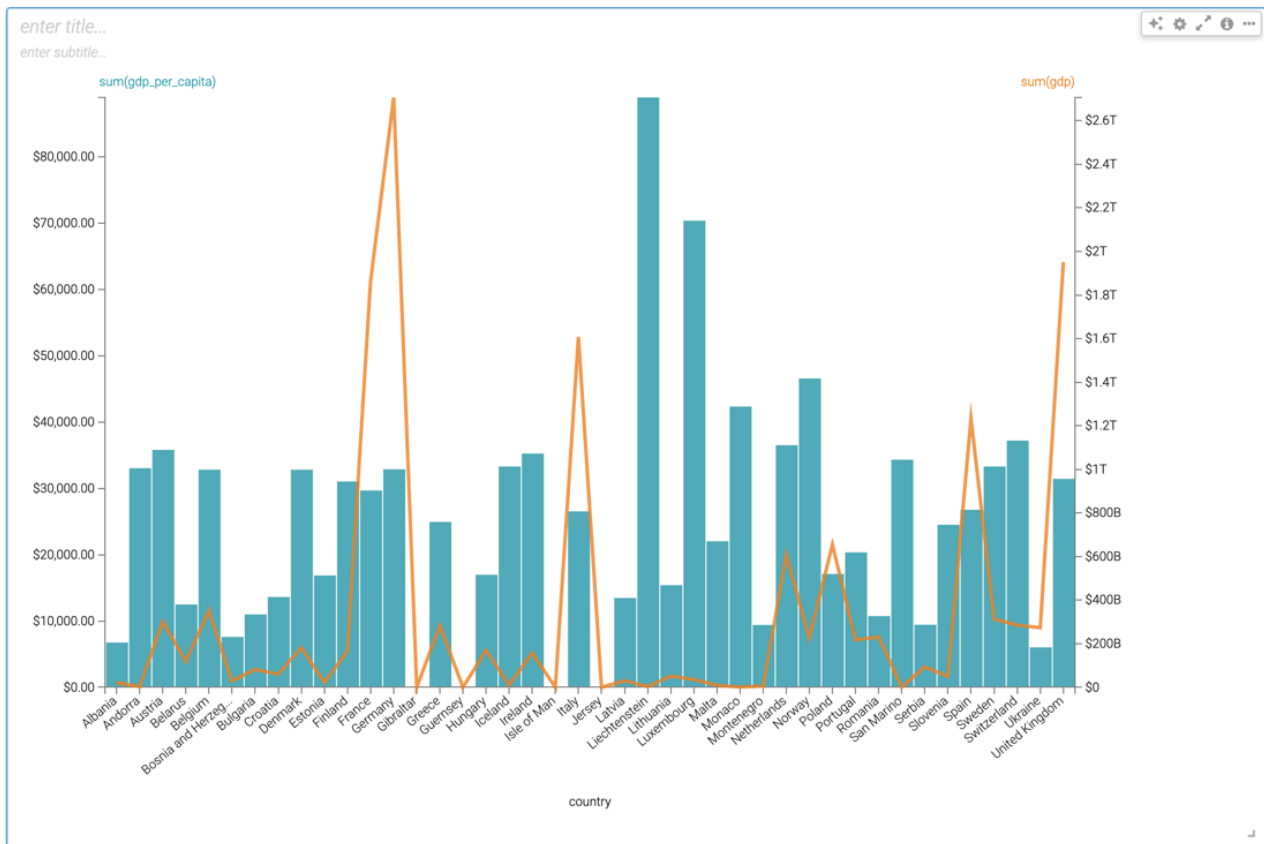
3. Click APPLY.

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 \$\$ 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}';
}
```

}

Edit Field Parameters
✕

Basic Settings
Expression
Display Format
Color

Base Column: gdp\_per\_capita

Category: Custom Javascript

Example:

12345.6789123

→

12345.6789123

Autocomplete on

Custom JS Format Function

Enter a custom function that takes a single value and returns the value with the desired modifications.

```

1 function myFunc(value) {
2   // Show the number in trillions with a dollar sign return
3   return '$${value/1000000000000}';
4 }

```

UPDATE EXAMPLE

REMOVE
CANCEL
APPLY

3. Click APPLY.

4. Save the changes to the dataset.

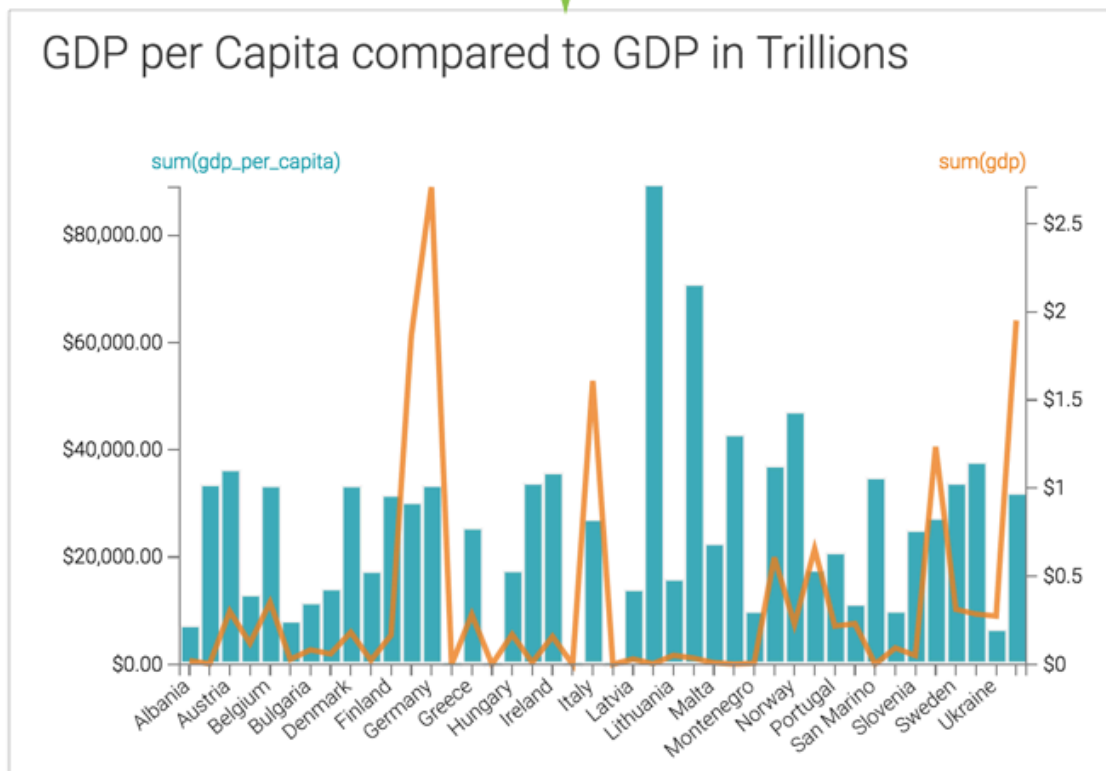
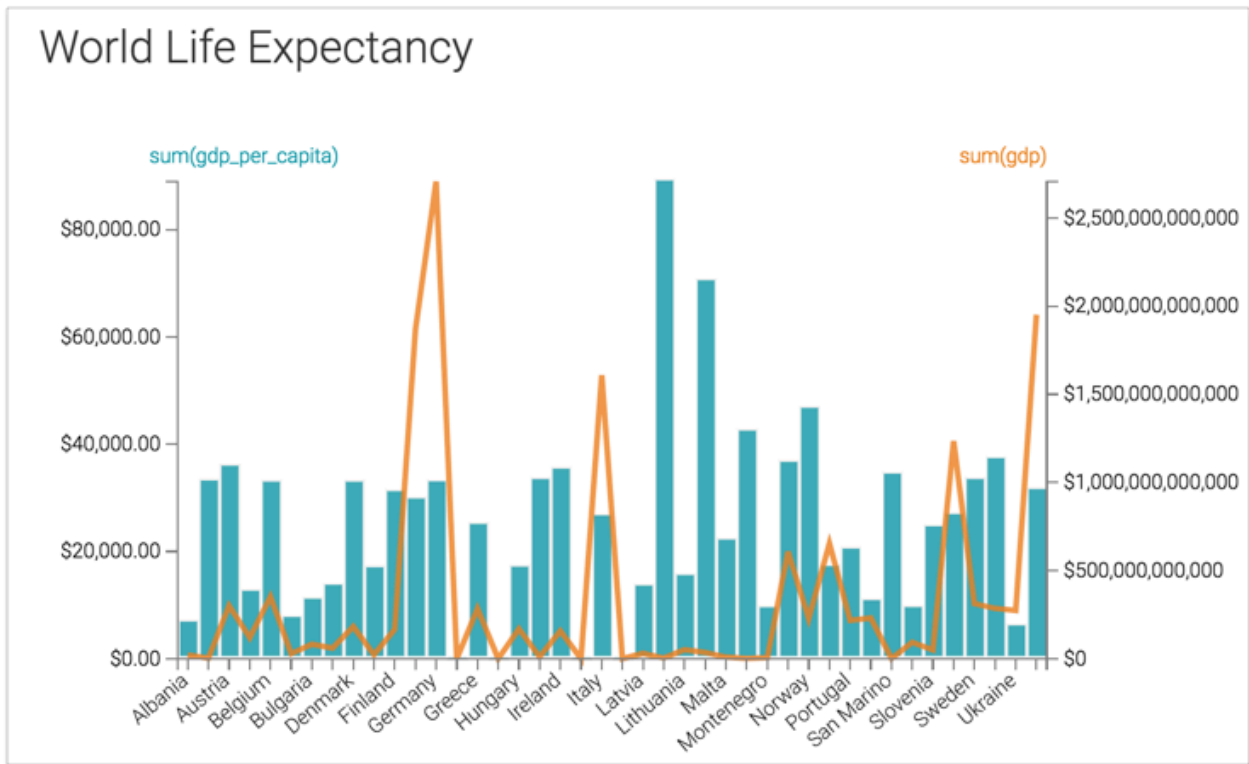
In this example, under Dataset: World Life Expectancy, click Save.

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`].

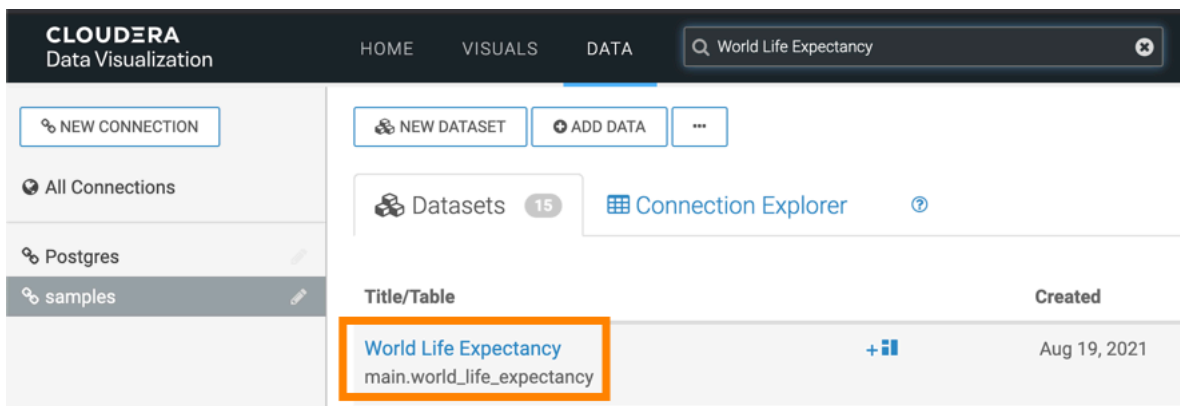
### Procedure

1. On the main navigation bar, click Data.

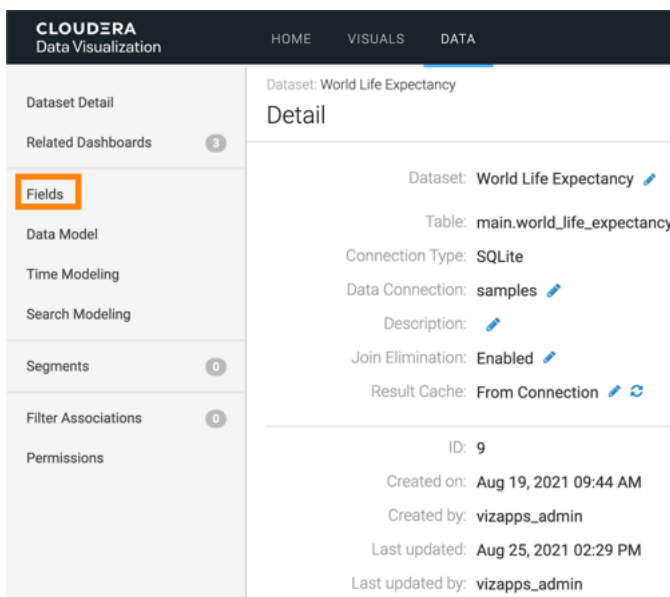


2. In the left navigation menu, click samples.

3. In the Datasets area, select World Life Expectancy (samples.world\_life\_expectancy).



4. In the Dataset Detail menu, select Fields.



- In the Fields interface, select Edit Field.

- Under Dimensions, find the field fips\_code, and click the down arrow icon, immediately following the # icon.
- In the menu, select String.

- 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.



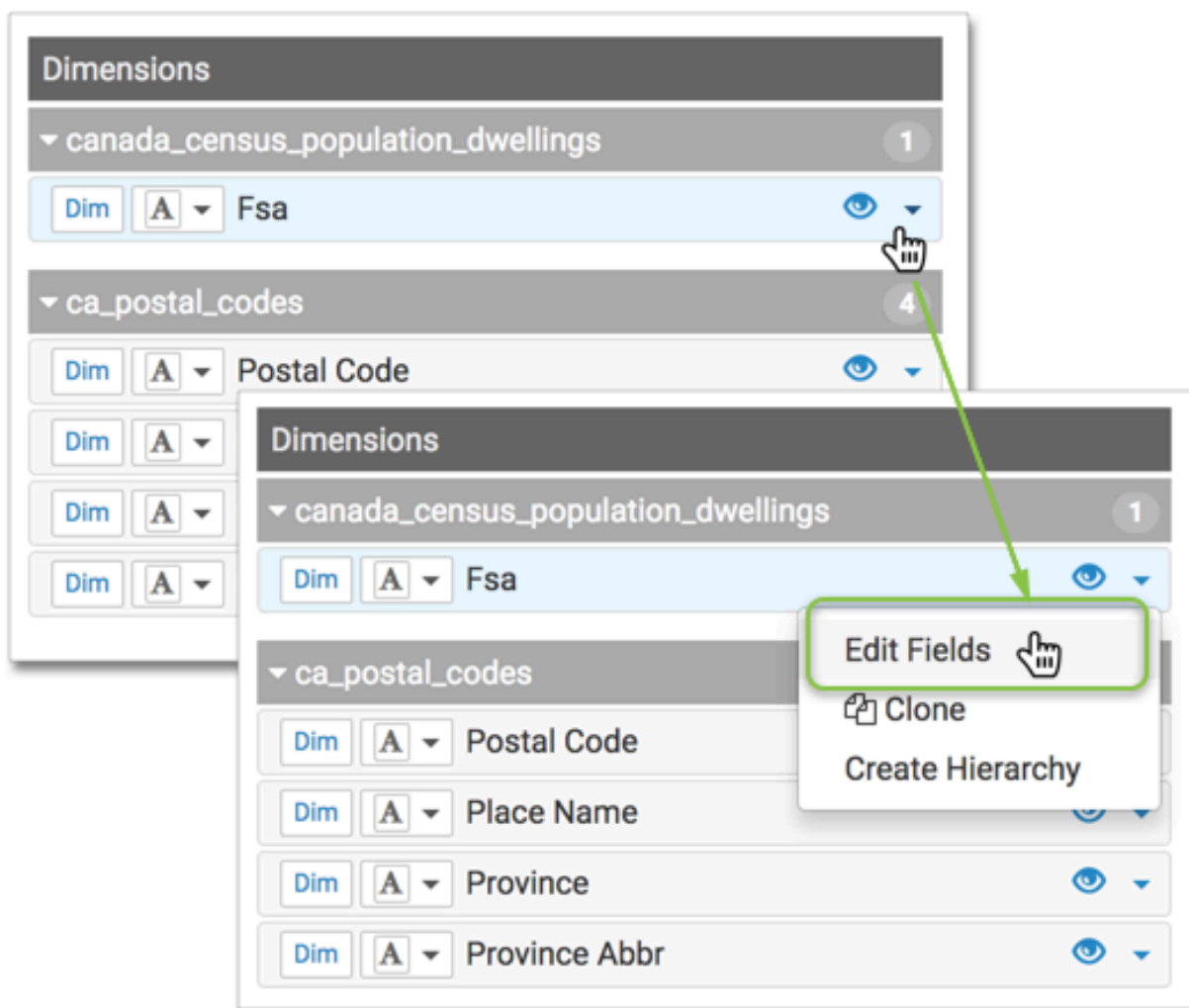
**Note:** Before creating visuals that use geographic data, specify the appropriate information as Geo Types.

### 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.

The screenshot shows the Arcadia Data interface for the 'Canadian Census' dataset. The left sidebar contains navigation options: Dataset Detail, Related Visuals, Fields (selected), Data Model, Analytical Views (0), Events (0), Segments (0), and Filter Associations (0). The main content area is titled 'Dataset: Canadian Census' and includes buttons for 'NEW DASHBOARD' and 'NEW VISUAL'. Below the title, there is a 'Fields' section with an 'EDIT FIELDS' button (highlighted with a green box and a mouse cursor) and a 'Show Comments' link. The 'Fields' section is divided into 'Dimensions' and 'Measures'. The 'Dimensions' section shows two datasets: 'canada\_census\_population\_dwellings' (1 field) with 'Fsa' and 'ca\_postal\_codes' (4 fields) with 'Postal Code', 'Place Name', 'Province', and 'Province Abbr'. The 'Measures' section shows two datasets: 'canada\_census\_population\_dwellings' (6 measures) with 'Population 2011', 'Total Private Dwellings 2011', 'Private Dwellings Occupied By Usual Residents 2011', 'Population 2006', 'Total Private Dwellings 2006', and 'Private Dwellings Occupied By Usual Residents 2006'; and 'ca\_postal\_codes' (2 measures) with 'Latitude' and 'Longitude'.

- 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.





- 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.

The image shows two side-by-side screenshots of the 'Edit Field Parameters' modal. Both modals have the following fields: 'Base Column' (fsa), 'Display Name' (Fsa/Postal Code), 'Field Comment' (Enter field comment), 'Default Aggregation' (Maximum), 'Geo Type' (None/Zipcode), 'Hide field in Visual Designer' (checkbox), 'Use as a partition column for Analytical Views' (checkbox), 'Category' (Dimension/Measure), 'CANCEL' button, and 'APPLY' button. A green box highlights the 'Zipcode' option in the 'Geo Type' dropdown of the right modal, with a green arrow pointing from the 'None' option in the left modal to it.

- Repeat for the Postal Code field of the table `ca_postal_codes`, and click Apply.
- Repeat with the Province field of the table `ca_postal_codes`, changing the Geo Type to State/Province. Click Apply.
- [Optional] Apply the Geo type to the fields that represent Latitude and Longitude, if they are not named appropriately.
- 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`].

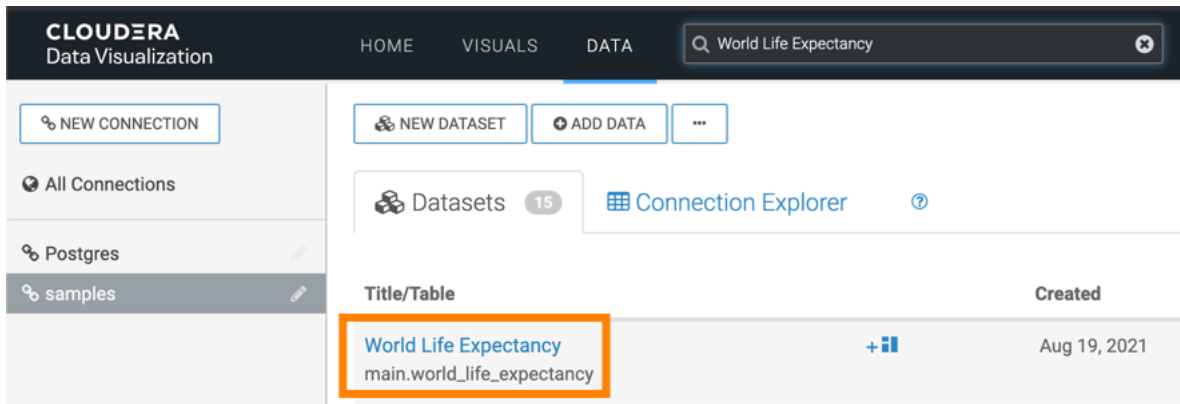
## Procedure

1. On the main navigation bar, click DATA.

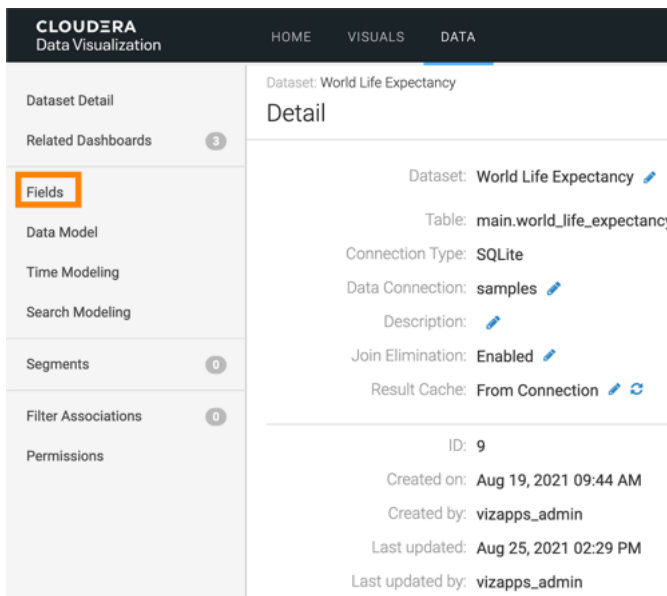


The Data view appears, open on the Datasets tab.

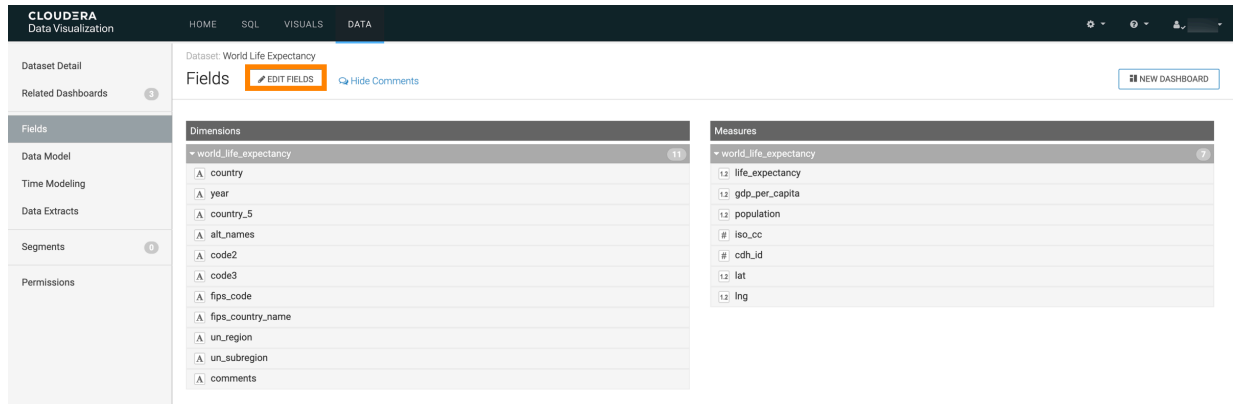
2. In the left navigation menu, click samples.
3. In the Datasets area, select World Life Expectancy (samples.world\_life\_expectancy).



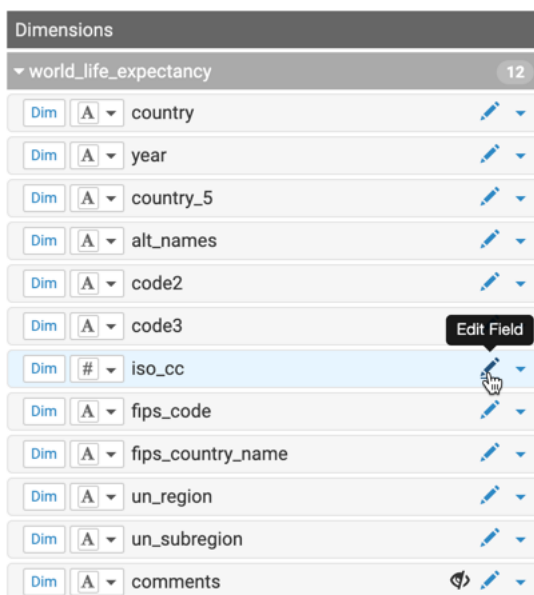
4. In the Dataset Detail menu, select Fields.



- In the Fields interface, select EDIT FIELDS.



- Under Dimensions, find the field iso\_cc, and click the down arrow icon on the right.
- From the menu, select Edit Field.



The Edit Field Parameters window modal appears.

- Under Field Comment, add the following text:

*ISO 3166 Country Code Standard, Numeric, described in [http://www.iso.org/iso/country\\_codes](http://www.iso.org/iso/country_codes). Mappings are at <https://www.iso.org/obp/ui/#search>.*

9. Click APPLY.

Edit Field Parameters
✕

---

Basic Settings
Display Format
Color

Base Column: iso\_cc

Display Name

Field Comment

Default Aggregation

Geo Type

Show field in data detail screen  
 Show field in Visual Designer  
 Use as a partition column for Analytical Views

Category  
 Dimension    Measure

CANCEL
APPLY

10. Click SAVE in the top row.

11. Click Show Comments to display the field comment.

Dataset: World Life Expectancy

Fields

EDIT FIELDS
Show Comments

Dimensions	Measures
<div style="display: flex; justify-content: space-between; align-items: center;"> <span>▼ world_life_expectancy</span> <span style="border: 1px solid #ccc; border-radius: 50%; padding: 2px 5px;">12</span> </div> <ul style="list-style-type: none"> <li style="border-bottom: 1px solid #ccc; padding: 2px 5px;"><input type="checkbox"/> country</li> <li style="border-bottom: 1px solid #ccc; padding: 2px 5px;"><input type="checkbox"/> year</li> <li style="border-bottom: 1px solid #ccc; padding: 2px 5px;"><input type="checkbox"/> country_5</li> <li style="border-bottom: 1px solid #ccc; padding: 2px 5px;"><input type="checkbox"/> alt_names</li> <li style="border-bottom: 1px solid #ccc; padding: 2px 5px;"><input type="checkbox"/> code2</li> <li style="border-bottom: 1px solid #ccc; padding: 2px 5px;"><input type="checkbox"/> code3</li> <li style="border-bottom: 1px solid #ccc; padding: 2px 5px;"><input type="checkbox"/> iso_cc</li> <li style="border-bottom: 1px solid #ccc; padding: 2px 5px;"><input type="checkbox"/> fips_code</li> <li style="border-bottom: 1px solid #ccc; padding: 2px 5px;"><input type="checkbox"/> fips_country_name</li> <li style="border-bottom: 1px solid #ccc; padding: 2px 5px;"><input type="checkbox"/> un_region</li> <li style="border-bottom: 1px solid #ccc; padding: 2px 5px;"><input type="checkbox"/> un_subregion</li> <li style="border-bottom: 1px solid #ccc; padding: 2px 5px;"><input type="checkbox"/> comments</li> </ul>	<div style="display: flex; justify-content: space-between; align-items: center;"> <span>▼ world_life_expectancy</span> <span style="border: 1px solid #ccc; border-radius: 50%; padding: 2px 5px;">7</span> </div> <ul style="list-style-type: none"> <li style="border-bottom: 1px solid #ccc; padding: 2px 5px;"><input type="checkbox"/> life_expectancy</li> <li style="border-bottom: 1px solid #ccc; padding: 2px 5px;"><input type="checkbox"/> gdp_per_capita</li> <li style="border-bottom: 1px solid #ccc; padding: 2px 5px;">= <input type="checkbox"/> gdp</li> <li style="border-bottom: 1px solid #ccc; padding: 2px 5px;"><input type="checkbox"/> population</li> <li style="border-bottom: 1px solid #ccc; padding: 2px 5px;"># <input type="checkbox"/> cdh_id</li> <li style="border-bottom: 1px solid #ccc; padding: 2px 5px;"><input type="checkbox"/> lat</li> <li style="border-bottom: 1px solid #ccc; padding: 2px 5px;"><input type="checkbox"/> lng</li> </ul>

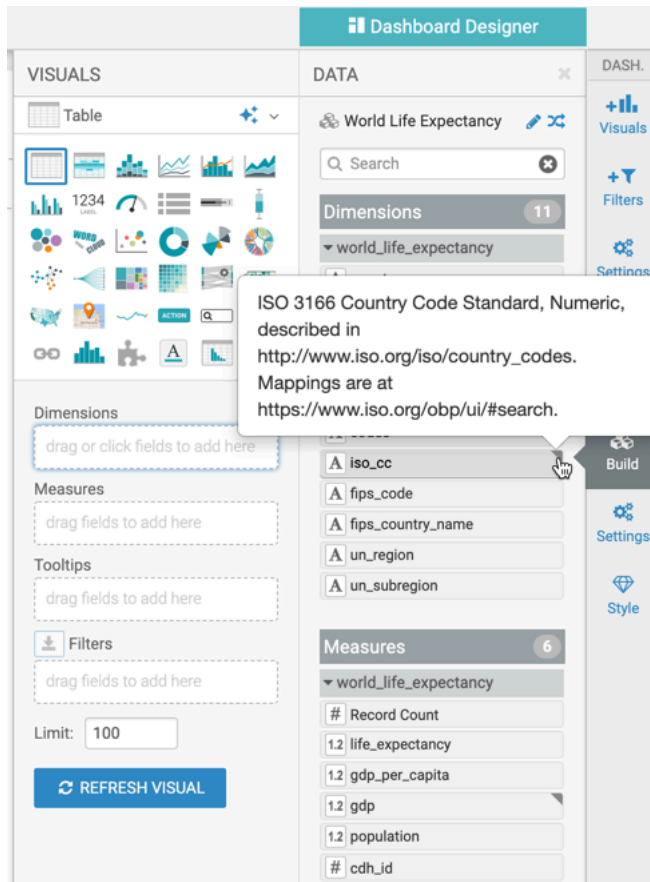
12. To hide field comments, click Hide Comments.

## 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.

- In the left navigation menu, click samples.
- In the Datasets area, select World Life Expectancy (samples.world\_life\_expectancy).

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'VISUALS', and 'DATA'. A search bar contains 'World Life Expectancy'. The left sidebar has 'NEW CONNECTION', 'All Connections', 'Postgres', and 'samples'. The main area shows 'NEW DATASET', 'ADD DATA', and a 'Datasets' list with 15 items. The 'World Life Expectancy' dataset is highlighted with an orange box. Below it, a table lists the dataset details:

Title/Table	Created
World Life Expectancy main.world_life_expectancy	Aug 19, 2021

- In the Dataset Detail menu, select Fields.

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'VISUALS', and 'DATA'. The left sidebar has 'Dataset Detail', 'Related Dashboards', 'Fields', 'Data Model', 'Time Modeling', 'Search Modeling', 'Segments', 'Filter Associations', and 'Permissions'. The 'Fields' option is highlighted with an orange box. The main area shows the 'Dataset: World Life Expectancy' detail page. The 'Detail' section includes:

- Dataset: World Life Expectancy
- Table: main.world\_life\_expectancy
- Connection Type: SQLite
- Data Connection: samples
- Description:
- Join Elimination: Enabled
- Result Cache: From Connection

Below the detail section, the following information is displayed:

- ID: 9
- Created on: Aug 19, 2021 09:44 AM
- Created by: vizapps\_admin
- Last updated: Aug 25, 2021 02:29 PM
- Last updated by: vizapps\_admin

- In the Fields interface, select Edit Field.

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'SQL', 'VISUALS', and 'DATA'. The left sidebar has 'Dataset Detail', 'Related Dashboards', 'Fields', 'Data Model', 'Time Modeling', 'Data Extracts', 'Segments', and 'Permissions'. The 'Fields' option is highlighted with an orange box. The main area shows the 'Dataset: World Life Expectancy' Fields interface. The 'Fields' section includes:

- Fields: EDIT FIELDS, Hide Comments
- NEW DASHBOARD

The 'Dimensions' section lists the following fields:

- world\_life\_expectancy
- country
- year
- country\_5
- alt\_names
- code2
- code3
- fips\_code
- fips\_country\_name
- un\_region
- un\_subregion
- comments

The 'Measures' section lists the following fields:

- world\_life\_expectancy
- life\_expectancy
- gdp\_per\_capita
- population
- iso\_cc
- cdh\_id
- lat
- lng

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

Dataset: World Life Expectancy

Fields ↶ UNDO ↻ REFRESH T TITLE CASE SAVE Show Comments

To add a new calculated field, use the down arrow to the right of a field to clone it, and then edit the expression of the cloned field.

Dimensions	Measures
world_life_expectancy 12 <ul style="list-style-type: none"> <li>Dim A country</li> <li>Dim A year</li> <li>Dim A country_5</li> <li>Dim A alt_names</li> <li>Dim A code2</li> <li>Dim A code3</li> <li>Dim # iso_cc</li> <li>Dim A fips_code</li> <li>Dim A fips_country_name</li> <li>Dim A un_region</li> <li>Dim A un_subregion</li> <li>Dim A comments</li> </ul>	world_life_expectancy 7 <ul style="list-style-type: none"> <li>Mes 1.2 life_expectancy</li> <li>Mes 1.2 gdp_per_capita</li> <li>= Mes 1.2 Copy of gdp_per_capita</li> <li>Mes 1.2 population</li> <li>Mes # cdh_id</li> <li>Mes 1.2 lat</li> <li>Mes 1.2 lng</li> </ul>

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



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

- 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

- On the main navigation bar, click Data.



The Data view appears, open on the Datasets tab.

- In the left navigation menu, click samples.

- In the Datasets area, select World Life Expectancy (samples.world\_life\_expectancy).

The screenshot shows the Cloudera Data Visualization interface. At the top, there are navigation tabs for HOME, VISUALS, and DATA. A search bar contains 'World Life Expectancy'. Below the navigation, there are buttons for 'NEW CONNECTION', 'NEW DATASET', and 'ADD DATA'. A 'Datasets' section shows a list of datasets. The 'World Life Expectancy' dataset is highlighted with an orange box. The table below shows the following data:

Title/Table	Created
World Life Expectancy main.world_life_expectancy	Aug 19, 2021

- In the Dataset Detail menu, select Fields.

The screenshot shows the Cloudera Data Visualization interface. The left sidebar has a 'Fields' option highlighted with an orange box. The main content area shows the 'Detail' page for the 'World Life Expectancy' dataset. The table below shows the following data:

Field	Value
Dataset:	World Life Expectancy
Table:	main.world_life_expectancy
Connection Type:	SQLite
Data Connection:	samples
Description:	
Join Elimination:	Enabled
Result Cache:	From Connection
ID:	9
Created on:	Aug 19, 2021 09:44 AM
Created by:	vizapps_admin
Last updated:	Aug 25, 2021 02:29 PM
Last updated by:	vizapps_admin

- In the Fields interface, select Edit Field.

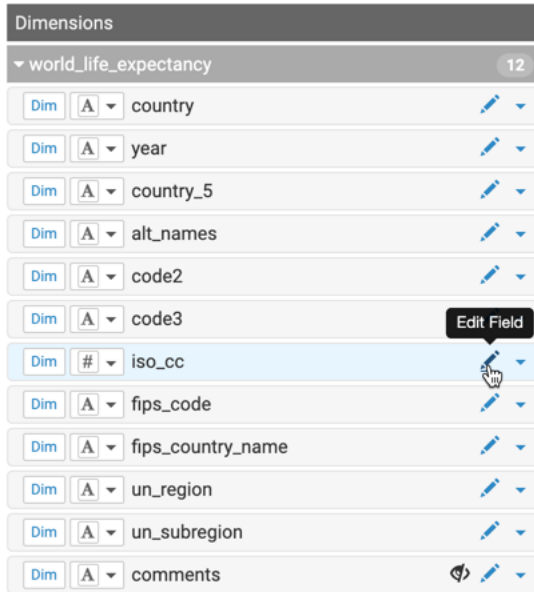
The screenshot shows the Cloudera Data Visualization interface. The left sidebar has a 'Fields' option highlighted with an orange box. The main content area shows the 'Fields' page for the 'World Life Expectancy' dataset. The 'EDIT FIELDS' button is highlighted with an orange box. The table below shows the following data:

Field	Type
world_life_expectancy	Dimension
country	Dimension
year	Dimension
country_5	Dimension
alt_names	Dimension
code2	Dimension
code3	Dimension
fips_code	Dimension
fips_country_name	Dimension
un_region	Dimension
un_subregion	Dimension
comments	Dimension
world_life_expectancy	Measure
life_expectancy	Measure
gdp_per_capita	Measure
population	Measure
iso_cc	Measure
cdh_id	Measure
lat	Measure
lng	Measure

- 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

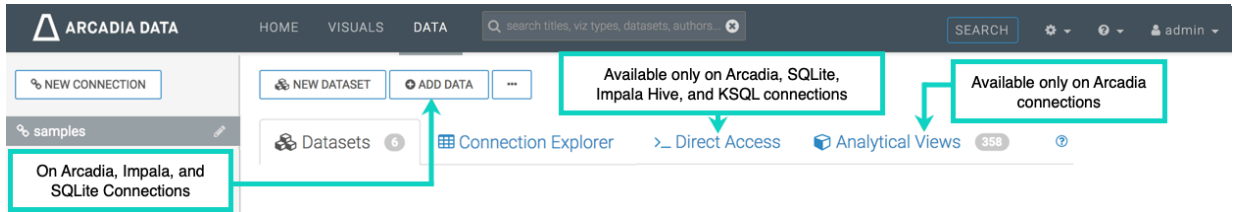
This example shows you how to create new data joins using the Flight Delays dataset.

## 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.

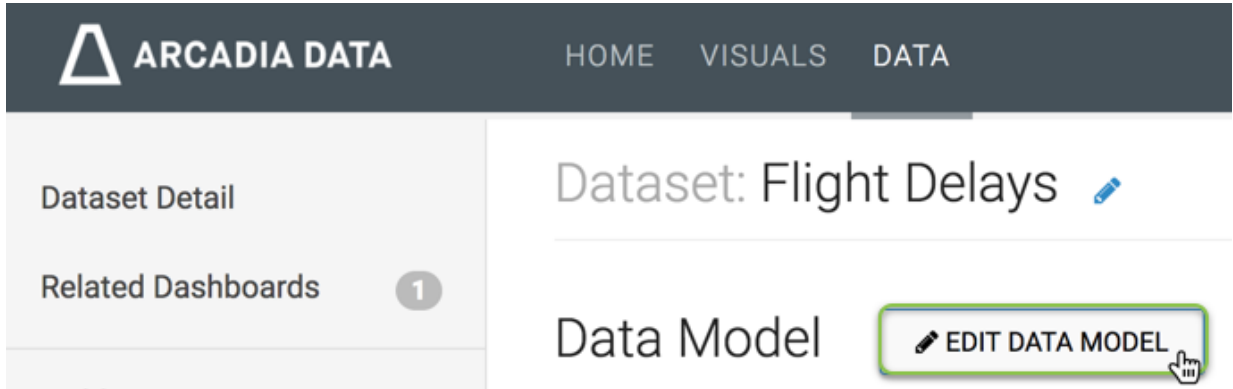
4. In the side navigation menu, click Data Model.

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.

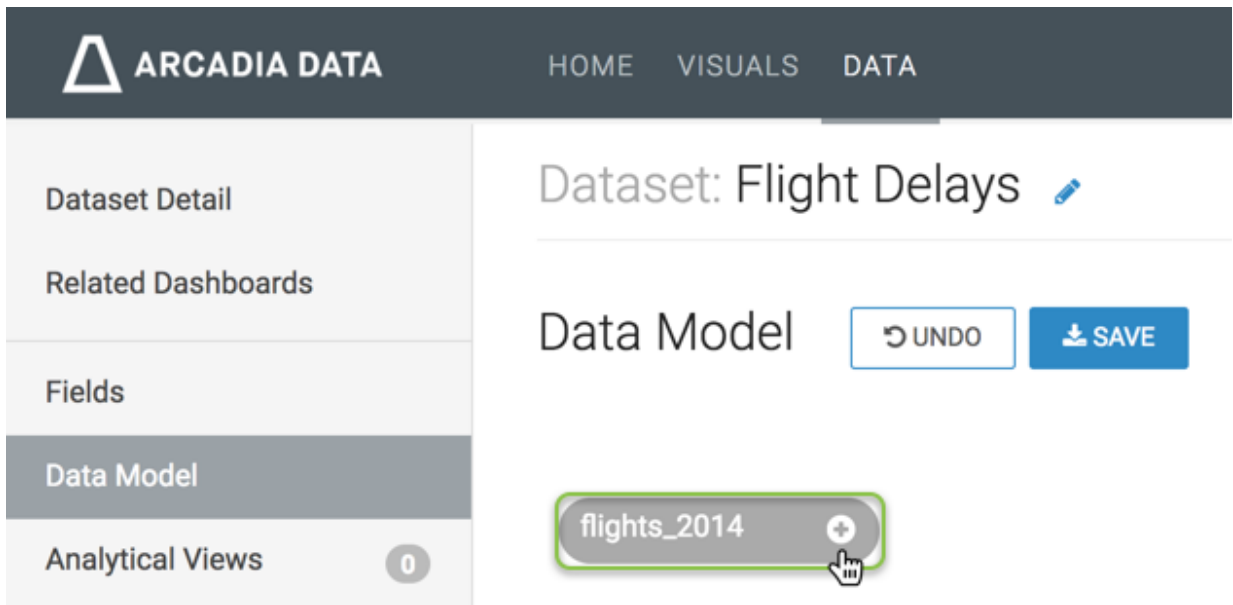
 A screenshot of the Arcadia Data interface showing the 'Dataset: Flight Delays' detail view. The left sidebar contains a navigation menu with items: 'Dataset Detail', 'Related Dashboards', 'Fields', 'Data Model' (which is selected), 'Analytical Views', 'Events', 'Segments', 'Filter Associations', and 'Permissions'. The main content area shows the 'Data Model' section with a button 'EDIT DATA MODEL'. Below this, the table name 'flights\_2014' is displayed. A 'HIDE DATA' button is visible. At the bottom, a table displays the data for the 'flights\_2014' dataset.
 

year	quarter	month	dayofmonth	dayofweek	flightdate	uniquecarrier	airlineid	carrier	tailnum
2014	1	1	21	2	2014-01-21 00:00:00	AA	19805	AA	N514AA
2014	1	1	22	3	2014-01-22 00:00:00	AA	19805	AA	N502AA

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.

### Table Browser ×

Choose the table you want to join. You will be able to select the columns that are joined in the next step.

---

**Database Name**

---

**Table Name**

---

The Edit Join modal window appears.

8. In the Edit Join modal window, the following options are available:
- [Optional] Click Clear Fields to clear all already defined joins between the two tables.
  - [Optional] Click sample data to preview the data. Click again to hide sample data.
  - [Optional] Click Add Join Pair to add another column connection between the same two tables.
  - [Optional] Click Add Join Expression to add a join between the two tables based on a custom SQL expression.
  - [Optional] Click icon (minus) to remove an existing join pair or an existing join expression.
  - [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.
  - Click APPLY to save the changes.

Edit Join
✕

---

CLEAR FIELDS 1

documentation.flights\_2014

deststate
▾

▶ sample data  Foreign Key

=

documentation.state\_abbreviations

abbreviation
▾

▶ sample data  Foreign Key

➔ ➔

**Join Expressions**

If you enter multiple expressions they will automatically have an "AND" logic between them

[deststate]=[abbreviation] AND [arrdelay] > 5 6

Click to update in SQL expression editor

+ ADD JOIN PAIR 3

+ ADD JOIN EXPRESSION 4

CANCEL

APPLY 7

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.

Edit Join
✕

---

CLEAR FIELDS

documentation.flights\_2014

airlineid

▶ sample data  Foreign Key

=

documentation.airline\_id

code

▶ sample data  Foreign Key

**Join Expressions**

If you enter multiple expressions they will automatically have an "AND" logic between them

Click to update in SQL expression editor

+ ADD JOIN PAIR

+ ADD JOIN EXPRESSION

CANCEL

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.

- Click Edit Data Model to edit the data model.

The screenshot shows the Arcadia Data interface. At the top, there is a navigation bar with 'HOME', 'VISUALS', and 'DATA'. Below this, the main content area is titled 'Dataset: Flight Delays'. On the left, there is a sidebar with 'Dataset Detail' and 'Related Dashboards' (with a '1' badge). In the main content area, the 'Data Model' section is visible, and the 'EDIT DATA MODEL' button is highlighted with a green border and a mouse cursor pointing to it.

- Click the (link) icon that represents the connection that must be changed. In our example, we clicked the join with the table `airport_codes`.

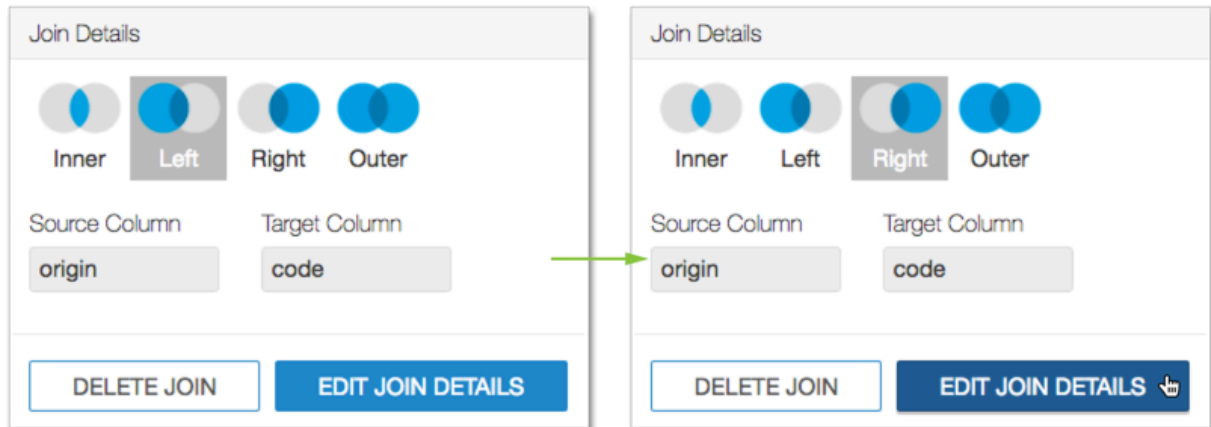
The screenshot shows the Arcadia Data interface with the 'Data Model' section expanded. The 'airline\_id' join is highlighted, and the 'airline\_id' field is selected. The interface includes a navigation bar with 'HOME', 'VISUALS', and 'DATA'. The main content area is titled 'Dataset: Flight Delays'. Below this, the 'Data Model' section is visible, with 'UNDO' and 'SAVE' buttons. The 'Data Model' section shows a list of fields: 'flights\_2014', 'airline\_id', 'airport\_codes', 'airport\_codes\_1', 'cancellation\_code', 'airport\_lat\_long', 'airport\_lat\_long\_1', 'state\_abbreviati...', and 'state\_abbreviati...'. The 'airline\_id' field is selected, and the 'airline\_id' join is highlighted.

The Join Details modal window appears.



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

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



- Click outside the Join Details modal window, or click Edit Join Details.

- Click Save.

## Editing join details

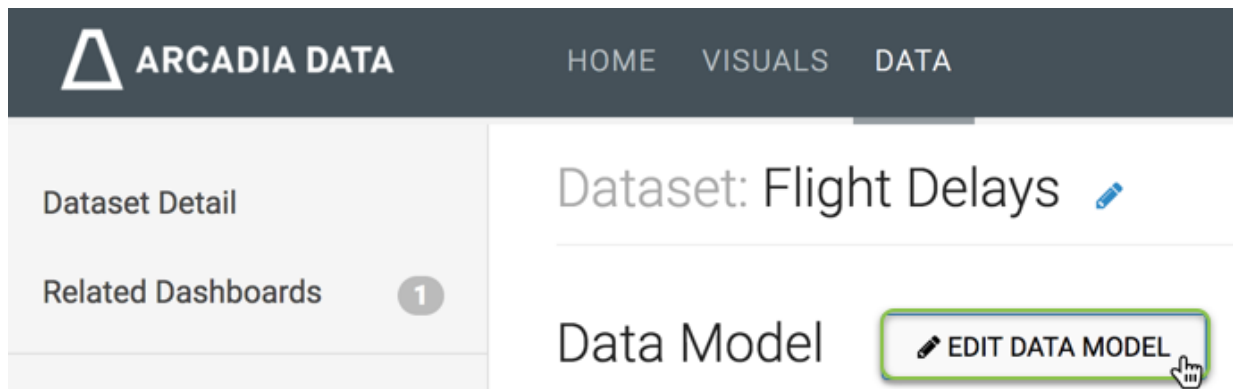
Learn how you can change the specifications of existing table joins in CDP Data Visualization.

### 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

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



- Click the (link) icon beside the state\_abbreviations connection.

Dataset: Flight Delays

Data Model UNDO SAVE

flights\_2014

- airline\_id
- airport\_codes
- airport\_codes\_1
- cancellation\_code
- airport\_lat\_long
- airport\_lat\_long\_1
- state\_abbreviations
- state\_abbreviations

The Join Details modal window appears.

- Click Edit Join.

Join Details

Inner **Left** Right Outer

Source Column = Target Column

deststate = abbreviation

DELETE JOIN **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.

Edit Join
✕

---

CLEAR FIELDS 1

documentation.flights\_2014

deststate
⌵

▶ sample data
 Foreign Key

=

documentation.state\_abbreviations

abbreviation
⌵

▶ sample data
 Foreign Key

**Join Expressions**

If you enter multiple expressions they will automatically have an "AND" logic between them

[deststate]=[abbreviation] AND [arrdelay] > 5 6

Click to update in SQL expression editor

+ ADD JOIN PAIR 3

+ ADD JOIN EXPRESSION 4

---

CANCEL

APPLY

7

6. To add 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.

### Edit Join ✕

---

CLEAR FIELDS

documentation.flights\_2014

deststate

=

documentation.state\_abbreviations

abbreviation

-

[▶ sample data](#)  Foreign Key

[▶ sample data](#)  Foreign Key

#### Join Expressions

If you enter multiple expressions they will automatically have an "AND" logic between them

Click to update in SQL expression editor

+ ADD JOIN PAIR

+ ADD JOIN EXPRESSION

CANCEL

APPLY

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

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

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

Join Expression ✕

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

Autocomplete on

All Functions

- abs
- acos
- add\_months
- adddate
- AND
- appx\_median
- ascii
- asin
- atan

All Fields

- A abbreviation
- # actualelapse...
- # airlineid
- # airtime
- T/F arrdel15
- # arrdelay
- # arrdelayminut...
- # arrivaldelaygr...
- # artime

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.

Edit Join
✕

---

CLEAR FIELDS

documentation.flights\_2014

▶ sample data  Foreign Key

=

documentation.state\_abbreviations

▶ sample data  Foreign Key

**Join Expressions**

If you enter multiple expressions they will automatically have an "AND" logic between them

+ ADD JOIN PAIR

+ ADD JOIN EXPRESSION

CANCEL

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.

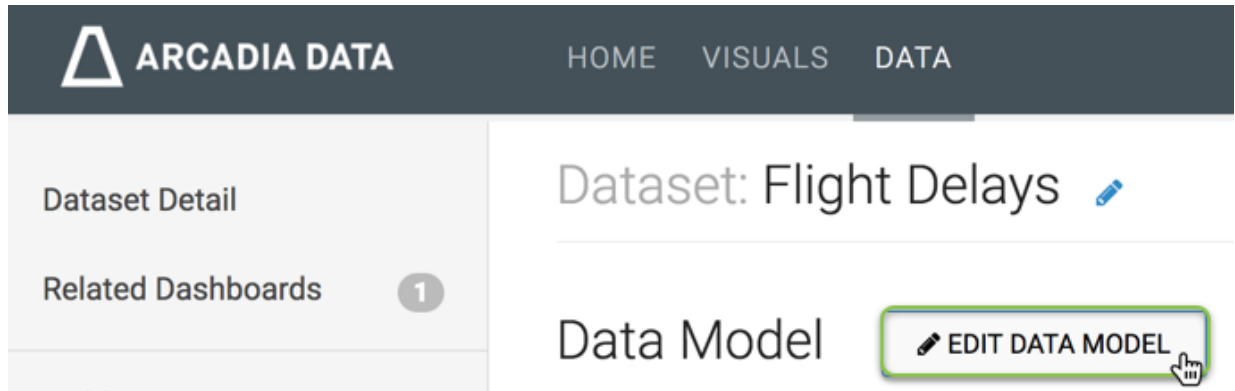
## Deleting a join

Learn how you can remove an existing join between tables.

### Procedure

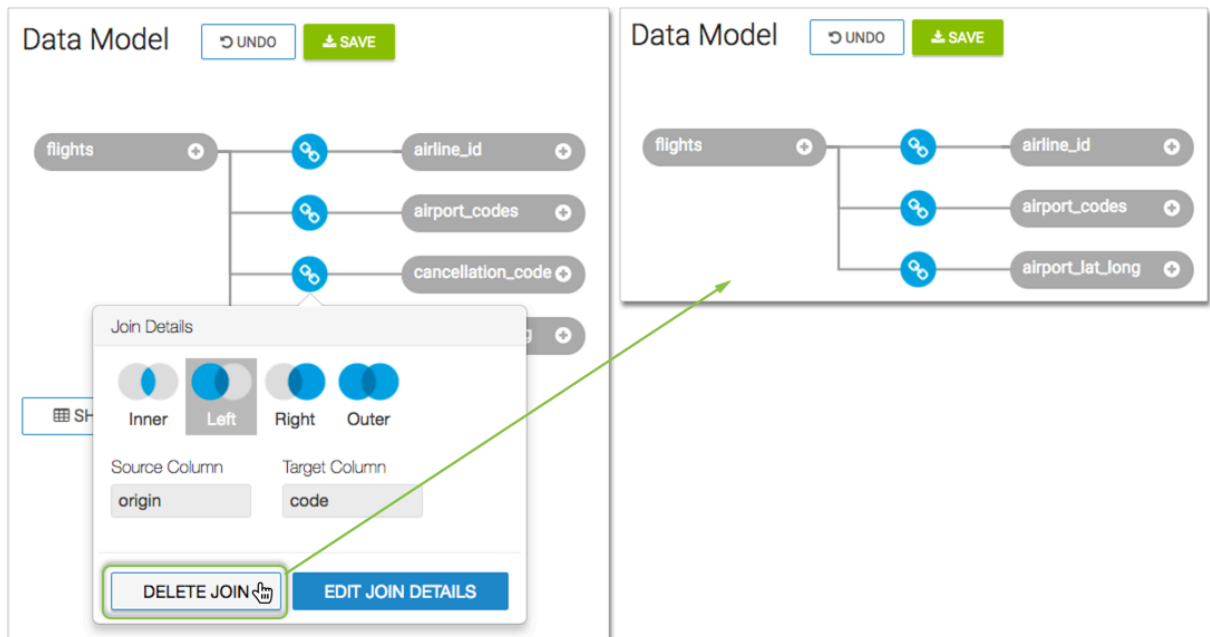
1. Navigate to the Data Model page of the dataset.

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.

trips_detail								trips				
pickup_datetime	passenger_count	trip_distance	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_latitude	pickup_neighborhood	trips_pickup_neighborhood	pickup_boro	pickup_hour	ride_cnt	total_amo
2013-10-07 20:13:00	1	10	-73.8734512329	40.7741127014	-73.9803848267	40.7706794739	Airport	Airport	Queens	0	44	2138.2
2013-10-07 20:13:00	1	10	-73.8734512329	40.7741127014	-73.9803848267	40.7706794739	Airport	Airport	Queens	1	263	13743.07
2013-10-07 20:13:00	1	10	-73.8734512329	40.7741127014	-73.9803848267	40.7706794739	Airport	Airport	Queens	2	463	24303.97
2013-10-07 20:13:00	1	10	-73.8734512329	40.7741127014	-73.9803848267	40.7706794739	Airport	Airport	Queens	3	895	45841.37
2013-10-07 20:13:00	1	10	-73.8734512329	40.7741127014	-73.9803848267	40.7706794739	Airport	Airport	Queens	4	1076	52089.04
2013-10-07 20:13:00	1	10	-73.8734512329	40.7741127014	-73.9803848267	40.7706794739	Airport	Airport	Queens	5	1127	52571.34
2013-10-07 20:13:00	1	10	-73.8734512329	40.7741127014	-73.9803848267	40.7706794739	Airport	Airport	Queens	6	1138	50456.47
2013-10-07 20:13:00	1	10	-73.8734512329	40.7741127014	-73.9803848267	40.7706794739	Airport	Airport	Queens	7	1033	47973.92
2013-10-07 20:13:00	1	10	-73.8734512329	40.7741127014	-73.9803848267	40.7706794739	Airport	Airport	Queens	8	1050	47469.01



**Related Information**

[Changing the field display format](#)