

Using Analytic Functions

Date published: 2020-10-30

Date modified: 2024-05-30

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

Supported connections.....	4
Using analytic functions - basic steps.....	4
Setting up a basic visual for aggregates.....	7
Setting up a basic visual for single values.....	8
Previous value analytic function.....	9
Difference from previous value analytic function.....	12
Percentage difference from previous value analytic function.....	15
Next value analytic function.....	17
Difference from next value analytic function.....	20
Percentage difference from next value analytic function.....	23
Row number function.....	26
Rank analytic function.....	29
Dense rank analytic function.....	31
Moving average analytic function.....	33
Percent of group analytic function.....	36
Running total analytic function.....	38

Supported connections

CDP Data Visualization supports several analytic functions that examine overlapping groupings of data.

Analytic functions are similar to aggregate functions because both use the contents of multiple data input rows to calculate the result. Analytic functions use flexible conditions that are specified by the OVER(...) clause to order and group input so that specific rows may be part of the calculation for several output values.



Note: Analytic functions do not validate in the Enter/Edit Expression editor. Refresh the visual to see if the function generates any errors.

Syntax for analytic functions is slightly different depending on the type of data connection used.

Supported Data Connections include the following:

- Impala
- Hive
- PostgreSQL
- SparkSQL
- RedShift

The Analytic Functions field properties are not available on MariaDB, MySQL, non-Oracle SQLite, Apache Drill, and Apache Solr connections.



Note:

- Syntax for analytic functions is slightly different depending on the type of data connection used. Analytic functions are not available for connections to MySQL, SQLite, Drill, MS Sql Server, Teradata, Solr, KSql, and MariaDB.
- In the query execution order, analytic functions follow the WHERE and GROUP BY clauses. Therefore, the function excludes the rows that are filtered out by these mechanisms, and they never become part of the analytic function data subset.
- When using both analytic functions and ordering, the available ordering options include all fields that are on the shelves, less the fields that are on the Filters shelf. To sort a visual on the results of an analytic functions, place the field used in the analytic function onto the Tooltips shelf. For more information, see *Customizing order and top K*.
- Use the Customizing enter/edit expressions option for running analytic functions that are not automated within CDP Data Visualization.

Related Information

[Customizing order and top K](#)

[Customizing enter/edit expressions](#)

Using analytic functions - basic steps

Procedure

1. Open the visual where you want to specify an analytic function, in Edit mode.
2. On a measurement shelf of a visual, click the field you plan to modify to open the Field Properties menu.

In this examples, the population field is used on the X Axis shelf.

3. In the FIELD PROPERTIES menu, click to expand the Analytic Functions menu.

The image shows the Cloudera Data Visualization interface. On the left is the 'VISUALS' panel, and on the right is the 'FIELD PROPERTIES' panel. The 'FIELD PROPERTIES' panel is open, showing a list of function categories. The 'Analytic Functions' category is highlighted with an orange box and a mouse cursor. The 'VISUALS' panel shows a 'Lines' chart with the following configurations:

- X Axis:** # year
- Y Axis:** # Population
- Colors:** A country
- Tooltips:** drag fields to add here
- Drill:** drag fields to add here
- Labels:** drag fields to add here
- Filters:**
 - ((year >= 1950) AND (year <=...
 - un_subregion in ('Northern Af...

A 'REFRESH VISUAL' button is located at the bottom of the 'VISUALS' panel.

The 'FIELD PROPERTIES' panel lists the following categories and functions:

- Aggregates ●
- Date/Time Functions
- Text Functions
- Analytic Functions** (highlighted)
- Previous Value
- Difference from Previous
- % Difference from Previous
- Next Value
- Difference from Next
- % Difference from Next
- Row Number
- Rank
- Dense Rank
- Moving Average
- % of Group
- Running Total
- Change Type
- Order and Top K
- [] Enter/Edit Expression
- Aggregate Display
- Display Format
- Alias ●
- Description
- Duplicate
- Save Expression
- Remove

On the far right, a vertical sidebar contains icons for Data, Settings, Colors, Style, Custom Style, and Segments.

4. Select one of the following analytic functions, directly supported by CDP Data Visualization.
5. In addition to these, you may use the expression builder to specify other analytic functions.

For more information, see *Customizing enter/edit expressions*.

Related Information

[Customizing enter/edit expressions](#)

Setting up a basic visual for aggregates

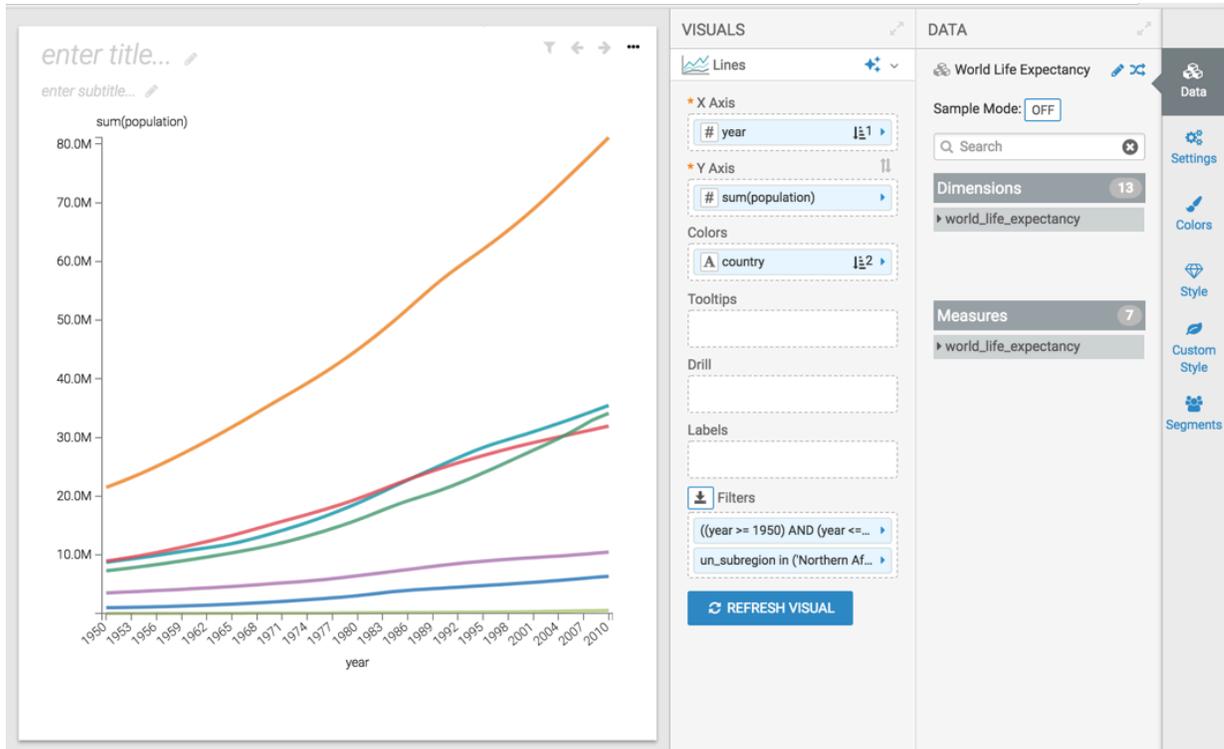
About this task

In this example, a basic line visual is created on the World Life Expectancy dataset. In subsequent topics, this visual is used to demonstrate how the various analytic functions for aggregates work.

Procedure

1. Open a new visual.
2. In the VISUALS menu, select the Lines visual type.
3. Populate the shelves of the visual from the fields listed in the Data menu:
 - *X Axis*
Add the field year. Order it in ascending order.
 - *Y Axis*
Add the field population.
 - *Colors*
Add the field country. Order it in ascending order.
 - *Filters*
Add the field year, and set it to the interval of 1950 through 2010.
Add the field un_subregion, and set it to Northern Africa.

- Click REFRESH VISUAL to see the new line visual.



- Name the visual Basic Lines.
- Click SAVE.

Setting up a basic visual for single values

Procedure

- Open a new visual.
- In the Visuals menu, select the Table visual type.
- Populate the shelves of the visual from the fields listed in the Data menu:
 - Dimensions*
Add the fields country, life_expectancy, gdp_per_capita, and population.
 - Filters*
Add the field year, and select the value 2010.
Add the field un_subregion, and set it to Africa.
- Specify descending order on life_expectancy.
- In the Enter/Edit Expression editor, change the gdp_per_capita calculation on the shelf, and rename it: $[gdp_per_capita] * [population]$ as 'GDP'
- Change the Display Format options for the fields on the shelves to remove extra decimals, specify currency, and so on.

- Change the Alias for the fields on the shelves. Click Refresh Visual to update the table visual.

The screenshot displays the Cloudera Data Visualization interface. On the left, a table visual is shown with the following data:

country	life expectancy	GDP	population
Mayotte	78.5	\$0	204,114
Tunisia	75.1	\$80,334,260,876	10,480,934
Libya	74.7	\$75,259,780,831	6,355,112
Cape Verde	74.1	\$1,813,273,168	495,999
Mauritius	73.1	\$16,002,161,612	1,299,172
Seychelles	72.7	\$1,413,868,538	86,518
Algeria	70.6	\$223,470,994,789	35,468,208
Egypt	70.5	\$491,593,744,800	81,121,080

On the right, the configuration panels are visible:

- VISUALS:** Shows the visual type as 'Table'. Dimensions include 'country', 'life expectancy', 'GDP', and 'population'. Measures are empty. Filters include 'year in (2010)' and 'un_region in (Africa)'. A 'Limit: 100' is set. A 'REFRESH VISUAL' button is present.
- DATA:** Shows the data source as 'World Life Expectancy'. Sample Mode is 'OFF'. Dimensions include 'world_life_expectancy'. Measures include 'world_life_expectancy'.

- Name the visual Basic Table.
- Click Save.

Previous value analytic function

This article describes how to use the Previous Value analytic function in CDP Data Visualization.

Procedure

- Click the field on the measurement shelf.
In this example, the sum(population) field on the Y Axis shelf.

2. In the FIELD PROPERTIES menu, expand Analytic Functions, and select Previous Value.

The screenshot displays the Cloudera Data Visualization interface. On the left, a line chart titled "enter title..." shows "sum(population)" on the y-axis (ranging from 0 to 80.0M) and "year" on the x-axis (ranging from 1950 to 2010). The chart contains several lines representing different data series. On the right, the "FIELD PROPERTIES" panel is open, showing a list of functions. Under the "Analytic Functions" category, the "Previous Value" option is highlighted with an orange box. Other options in the list include "Difference from Previous", "% Difference from Previous", "Next Value", "Difference from Next", "% Difference from Next", "Row Number", "Rank", "Dense Rank", "Moving Average", "% of Group", "Running Total", "Change Type", "Order and Top K", "Aggregate Display", "Display Format", "Alias", "Description", and "Duplicate". The "VISUALS" panel on the left of the Field Properties menu shows the current visualization settings: X Axis is "# year", Y Axis is "# sum(population)", Colors is "country", and Filters includes "((year >= 1950) AND (year <= ...))" and "un_subregion in ('Northern Af...')". A "REFRESH VISUAL" button is located at the bottom of the VISUALS panel.

3. In the Analytic Function: Previous Value modal window:

- In Select aggregate for this measure, leave the default Sum. The other aggregation options are Count, Minimum, Maximum, and Average. In addition, some connection types also support Approximate Distinct Count and Exact Distinct Count.
- Under Select entities over which to take previous value, select state. This is the grouping that applies when calculating previous values.
- Under Select the sort order, click the Plus icon, select year, and then set it to ascending sort order. This order specifies what determines what the previous row is, and the row where the value is taken.



Note: You can add additional sort parameters, by clicking the (Plus) icon.

- Under Offset, specify the number of previous rows over which the function executes.

Analytic Function : Previous Value ✕

1 Select aggregate for this measure

✓ Sum

Count

Approx Distinct Count

Exact Distinct Count

Minimum

Maximum

Average

2 Select entities over which to take previous value (optional)

year

country

3 * Select the sort order

1 year ⌵ -

2 country ⌵ -

(If you do not select a sort order, the aggregate of the current column will be applied)

+

4 Offset: rows

CANCEL

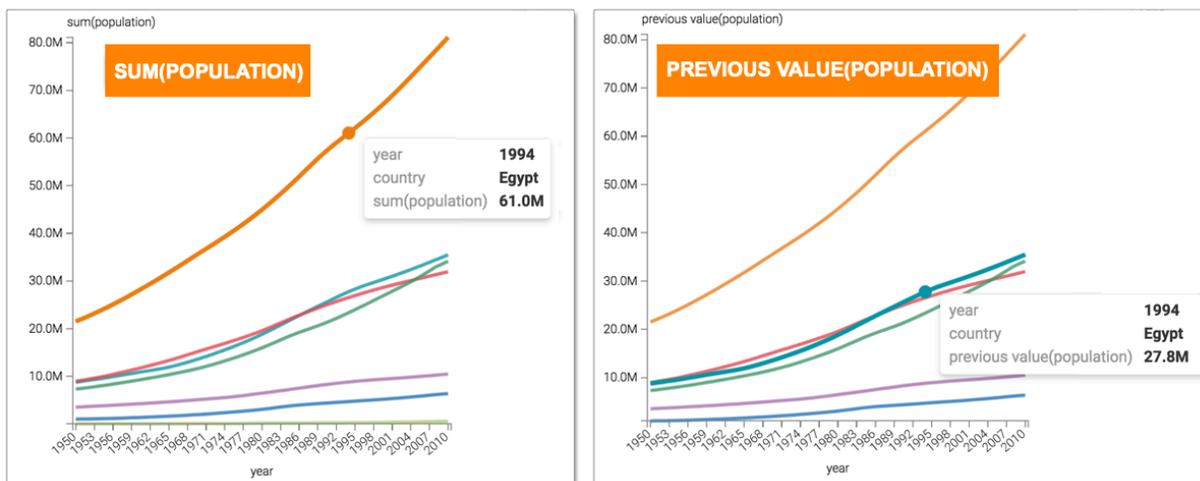
APPLY

4. Click APPLY.

- If you examine the expression in the Enter/Edit Expression interface, it shows that the function has the following form: `LAG(sum([population]), 1) over (partition by [year] order by [year], [country]) as 'previous value(population)'`

6. Click REFRESH VISUAL.

You can see the change to the appearance of the visual, and the change in values that appear on the tooltip.



Difference from previous value analytic function

This article describes how to use the Difference from Previous Value analytic function in CDP Data Visualization.

Procedure

1. Click the field on the measurement shelf.

In this example, the sum(population) field on the Y Axis shelf.

2. In the FIELD PROPERTIES menu, expand Analytic Functions, and select Difference from Previous.

The screenshot displays the Cloudera Data Visualization interface. On the left, a line chart titled "sum(population)" shows population trends from 1950 to 2010. The Y-axis ranges from 0.0M to 80.0M. The X-axis is labeled "year" and shows years from 1950 to 2010 in 2-year increments. The chart contains several lines representing different countries, with a prominent green line showing exponential growth.

On the right, the "FIELD PROPERTIES" panel is open, showing a tree view of analytic functions. The "Analytic Functions" category is expanded, and "Difference from Previous" is highlighted with an orange box. Other functions listed include "Previous Value", "% Difference from Previous", "Next Value", "Difference from Next", "% Difference from Next", "Row Number", "Rank", "Dense Rank", "Moving Average", "% of Group", "Running Total", "Change Type", "Order and Top K", "Aggregate Display", "Display Format", "Alias", "Description", and "Duplicate".

The "VISUALS" panel on the left of the Field Properties menu shows the following configuration:

- Visual: Lines
- X Axis: # year
- Y Axis: # sum(population)
- Colors: A country
- Filters: ((year >= 1950) AND (year <= ...)) and un_subregion in ('Northern Af...')

A "REFRESH VISUAL" button is located at the bottom of the VISUALS panel.

3. In the Analytic Function: Difference from Previous Value modal window:

- In Select aggregate for this measure, leave the default Sum. The other aggregation options are Count, Minimum, Maximum, and Average. In addition, some connection types also support Approximate Distinct Count and Exact Distinct Count.
- Under Select entities over which to take difference from previous value, select year. This is the grouping that applies when calculating the difference from previous values.
- Under Select the sort order, you can specify the order of sorting. This order specifies what the previous row is, and the row where the value is taken. If your visual already specifies sorting order, these values appear in the modal.



Note: You can add additional sort parameters, by clicking the (Plus) icon.

In this example, leave the default order: year ascending, followed by country ascending.

- Under Offset, specify the number of previous rows over which the function executes.

Analytic Function : Difference from Previous Value ×

1 Select aggregate for this measure

- ✓ Sum
- Count
- Approx Distinct Count
- Exact Distinct Count
- Minimum
- Maximum
- Average

2 Select entities over which to take difference from previous value

year

3 * Select the sort order

1 year ⌵ -

2 country ⌵ -

(If you do not select a sort order, the aggregate of the current column will be applied)

+

4 Offset: rows

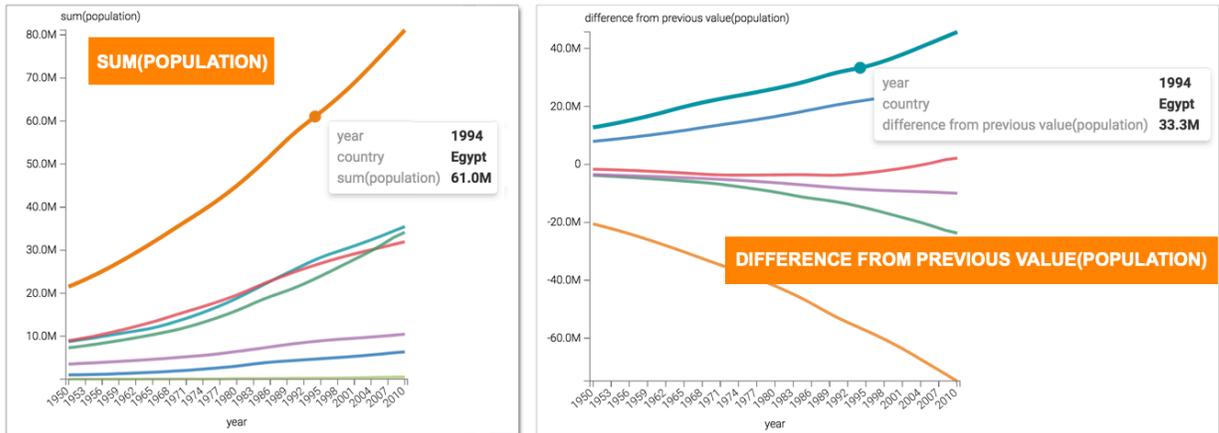
CANCEL
APPLY

4. Click APPLY.

- If you examine the expression in the Enter/Edit Expression interface, it shows that the function has the following form: `sum([population]) - LAG(sum([population]), 1) over (partition by [year] order by [year], [country]) as 'difference from previous value(population)'`

6. Click REFRESH VISUAL.

You can see the change to the appearance of the visual, and the change in values that appear on the tooltip.



Percentage difference from previous value analytic function

This article describes how to use the Percent Difference from Previous Value analytic function in CDP Data Visualization.

Procedure

1. Click the field on the measurement shelf.
In this example, the sum(population) field on the Y Axis shelf.
2. In the FIELD PROPERTIES menu, expand Analytic Functions, and select % Difference from Previous.

The screenshot shows the Cloudera Data Visualization interface. The main chart is titled 'US Population by Year and State' and displays population data from 1950 to 2010. The 'FIELD PROPERTIES' panel is open on the right, showing the 'Analytic Functions' section with '% Difference from Previous' selected. The 'VISUALS' panel on the left shows the chart configuration, including the X Axis (year), Y Axis (sum(population)), and Filters.

3. In the Analytic Function: Percentage Difference from Previous Value modal window:

- In Select aggregate for this measure, leave the default Sum. The other aggregation options are Count, Minimum, Maximum, and Average. In addition, some connection types also support Approximate Distinct Count and Exact Distinct Count.
- Under Select entities over which to take % difference from previous value, select country. This is the grouping that applies when calculating the % difference from previous values.
- Under Select the sort order, you can specify the order of sorting. This order specifies what the previous row is, and the row where the value is taken. If your visual already specifies sorting order, these values appear in the modal.



Note: You can add additional sort parameters, by clicking the (Plus) icon.

In this example, edit the sort order to be: country ascending, followed by year ascending.

- Under Offset, specify the number of previous rows over which the function executes.

Analytic Function : Percentage Difference from Previous Value ✕

1 Select aggregate for this measure Sum ▼

2 Select entities over which to take % difference from previous value (optional)

year

country

3 * Select the sort order

1 country ▼ -

2 year ▼ -

(If you do not select a sort order, the aggregate of the current column will be applied)

+

4 Offset: 1 rows

CANCEL APPLY

4. Click APPLY.

- If you examine the expression in the Enter/Edit Expression interface, it shows that the function has the following form:

```
(sum([population]) - LAG(sum([population]), 1)
```

```

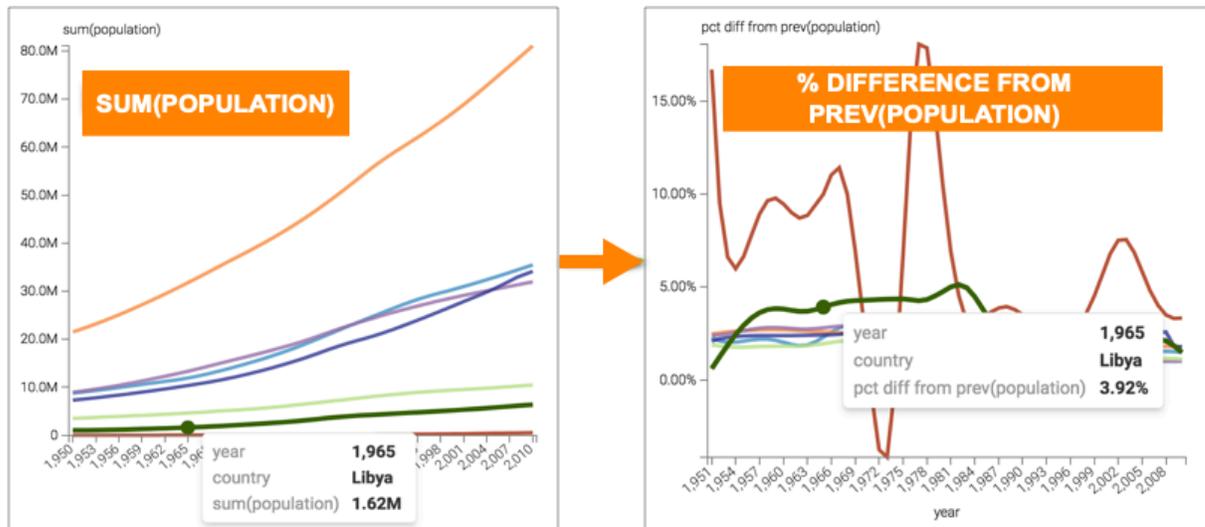
over (partition by [country] order by [country], [year])) / LAG(sum([pop
ulation]), 1)
over (partition by [country] order by [country], [year])
as 'pct diff from prev(population)'

```

6. Click REFRESH VISUAL.

You can see the change in the appearance of the Y axis and the tooltip, from displaying the sum(population) in millions to population as a percentage difference from previous value.

Hover over any country to view the percentage difference of population from the previous year. In the following example, the tooltip shows an increase in the population of Libya from previous year 1964 to the current year 1965 by 3.92%.



Next value analytic function

This article describes how to use the Next Value analytic function in CDP Data Visualization.

Procedure

1. Click the field on the measurement shelf.
In this example, the sum(population) field on the Y Axis shelf.

2. In the FIELD PROPERTIES menu, expand Analytic Functions, and select Next Value.

The screenshot displays the Cloudera Data Visualization interface. On the left, a line chart titled 'sum(population)' shows population growth from 1950 to 2010, categorized by country. The Y-axis ranges from 0 to 80.0M. The X-axis shows years from 1950 to 2010. The chart features several lines in different colors (green, red, purple, blue, orange) representing different countries. The interface includes a 'VISUALS' panel on the right with sections for X Axis (year), Y Axis (sum(population)), Colors (country), Tooltips, Drill, Labels, and Filters. The 'FIELD PROPERTIES' panel is open, showing a list of analytic functions. The 'Next Value' function is highlighted with an orange box. The 'FIELD PROPERTIES' panel also includes sections for Aggregates, Date/Time Functions, Text Functions, Analytic Functions, Change Type, Order and Top K, Aggregate Display, Alias, Description, and Duplicate.

3. In the Analytic Function: Next Value modal window:

- In Select aggregate for this measure, leave the default Sum. The other aggregation options are Count, Minimum, Maximum, and Average. In addition, some connection types also support Approximate Distinct Count and Exact Distinct Count.
- Under Select entities over which to take next value, select state. This is the grouping that applies when calculating next values.
- Under Select the sort order, click the (Plus) icon, select year, and then set it to ascending sort order. This order specifies what determines what the next row is, and the row where the value is taken.



Note: You can add additional sort parameters, by clicking the (Plus) icon.

- Under Offset, specify the number of next rows over which the function executes.

Analytic Function : Next Value ✕

1

Select aggregate for this measure

✓ Sum

Count

Approx Distinct Count

Exact Distinct Count

Minimum

Maximum

Average

2

Select entities over which to take next value (optional)

year

country

3

* Select the sort order

1 year ⌵ -

2 country ⌵ -

(If you do not select a sort order, the aggregate of the current column will be applied)

+

4

Offset: rows

CANCEL

APPLY

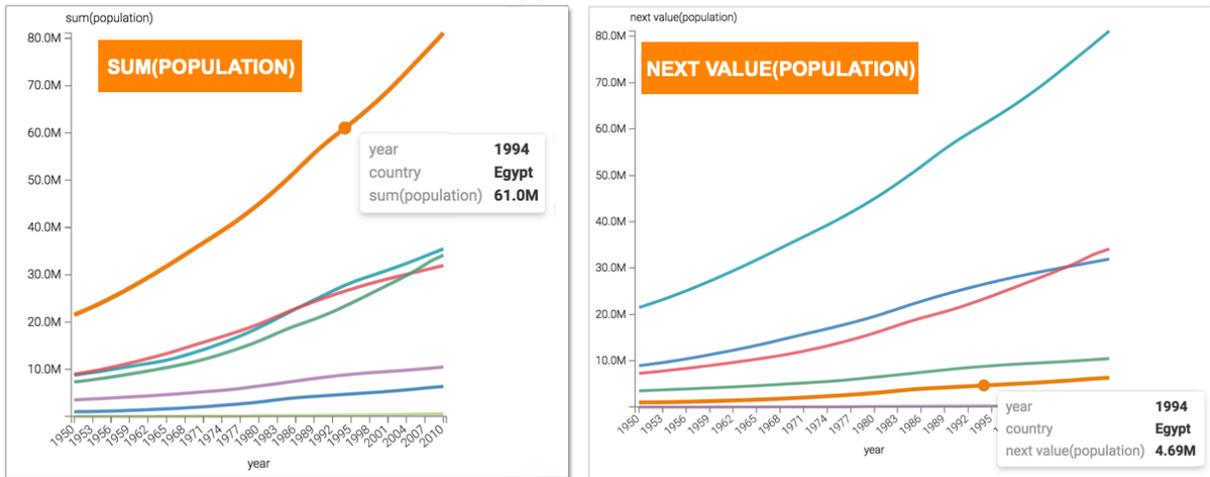
4. Click APPLY.

5. If you examine the expression in the Enter/Edit Expression interface, it shows that the function has the following form:

```
LAG(sum([population]), 1) over (partition by [year] order by [year], [country])
      as 'next value(population)'
```

6. Click REFRESH VISUAL.

You can see the change to the appearance of the visual, and the change in values that appear on the tooltip.



Difference from next value analytic function

This article describes how to use the Difference from Next Value analytic function in CDP Data Visualization.

Procedure

1. Click the field on the measurement shelf.

In this example, the sum(population) field on the Y Axis shelf.

2. In the FIELD PROPERTIES menu, expand Analytic Functions, and select Difference from Next.

The screenshot displays the Cloudera Data Visualization interface. On the left, a line chart titled 'enter title...' shows 'sum(population)' on the y-axis (ranging from 0 to 80.0M) against 'year' on the x-axis (ranging from 1950 to 2010). The chart contains several lines representing different data series. On the right, the 'FIELD PROPERTIES' panel is open, showing a tree view of functions. Under 'Analytic Functions', the 'Difference from Next' option is highlighted with an orange box. Other options include 'Previous Value', 'Difference from Previous', '% Difference from Previous', 'Next Value', and '% Difference from Next'. The 'VISUALS' panel on the far right shows the current configuration: X Axis is '# year', Y Axis is '# sum(population)', and Colors is 'country'. A 'REFRESH VISUAL' button is located at the bottom of the VISUALS panel.

3. In the Analytic Function: Difference from Next Value modal window,
 - a. In Select aggregate for this measure, leave the default Sum. The other aggregation options are Count, Minimum, Maximum, and Average. In addition, some connection types also support Approximate Distinct Count and Exact Distinct Count.
 - b. Under Select entities over which to take difference from next value, select year. This is the grouping that applies when calculating the difference from next values.
 - c. Under Select the sort order, you can specify the order of sorting. This order specifies what the next row is, and the row where the value is taken. If your visual already specifies sorting order, these values appear in the modal.



Note: You can add additional sort parameters, by clicking the (Plus) icon.

In this example, leave the default order: year ascending, followed by country ascending.

- d. Under Offset, specify the number of next rows over which the function executes.

Analytic Function : Difference from Next Value ×

1 Select aggregate for this measure

- ✓ Sum
- Count
- Approx Distinct Count
- Exact Distinct Count
- Minimum
- Maximum
- Average

2 Select entities over which to take difference from next value (optional)

year

country

3 * Select the sort order

1 year ↕ -

2 country ↕ -

(If you do not select a sort order, the aggregate of the current column will be applied)

+

4 Offset: rows

CANCEL
APPLY

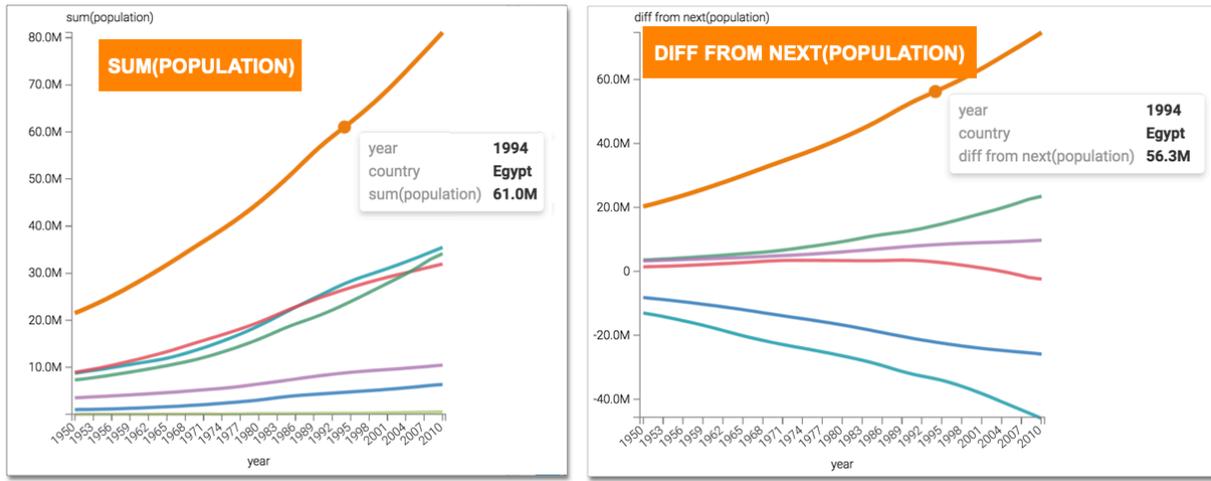
4. Click APPLY.
5. If you examine the expression in the Enter/Edit Expression interface, it shows that the function has the following form:

```
sum([population]) - LEAD(sum([population]), 1) over (partition by [year]
```

```
order by [year], [country]
as 'diff from next(population)'
```

6. Click REFRESH VISUAL.

You can see the change to the appearance of the visual, and the change in values that appear on the tooltip.



Percentage difference from next value analytic function

This article describes how to use the Percent Difference from Next Value analytic function in CDP Data Visualization.

Procedure

1. Click the field on the measurement shelf.

In this example, the sum(population) field on the Y Axis shelf.

2. In the FIELD PROPERTIES menu, expand Analytic Functions, and select % Difference from Next.

The screenshot displays the Cloudera Data Visualization interface. On the left, a line chart titled "US Population by Year and State" shows population data from 1950 to 2010. The Y-axis is labeled "sum(population)" and ranges from 0 to 80.0M. The X-axis is labeled "year" and ranges from 1,950 to 2,010. The chart features several lines representing different states or regions, with the highest line (orange) showing a significant upward trend.

On the right, the "FIELD PROPERTIES" panel is open, showing a list of analytic functions. The function "% Difference from Next" is highlighted with an orange border. Other functions listed include "Difference from Previous", "Next Value", "Difference from Next", "Row Number", "Rank", "Dense Rank", "Moving Average", "% of Group", and "Running Total".

The "VISUALS" panel on the left shows the following configuration:

- Visual Type: Lines
- X Axis: # year
- Y Axis: # sum(population)
- Colors: A country
- Tooltips: drag fields to add here
- Drill: drag fields to add here
- Labels: drag fields to add here
- Filters: # ((year >= 1950) AND (yea...), A un_subregion in ('Norther...)

Buttons for "SAVE", "CLOSE", and "REFRESH VISUAL" are visible at the top and bottom of the interface.

3. In the Analytic Function: Percentage Difference from Next Value modal window:
 - a. In Select aggregate for this measure, leave the default Sum. The other aggregation options are Count, Minimum, Maximum, and Average. In addition, some connection types also support Approximate Distinct Count and Exact Distinct Count.
 - b. Under Select entities over which to take % difference from next value, select country. This is the grouping that applies when calculating the % difference from next values.
 - c. Under Select the sort order, you can specify the order of sorting. This order specifies what the next row is, and the row where the value is taken. If your visual already specifies sorting order, these values appear in the modal.



Note: You can add additional sort parameters, by clicking the (Plus) icon.

In this example, edit the sort order to be: country ascending, followed by year ascending.

- d. Under Offset, specify the number of next rows over which the function executes.

Analytic Function : Percentage Difference from Next Value ×

1 Select aggregate for this measure

✓ Sum

Count

Minimum

Maximum

Average

2 Select entities over which to take % difference from next value (

year

country

3 * Select the sort order

1 country ↕ -

2 year ↕ -

(If you do not select a sort order, the aggregate of the current column will be applied)

+

4 Offset: rows

CANCEL
APPLY

4. Click APPLY.
5. If you examine the expression in the Enter/Edit Expression interface, it shows that the function has the following form:

```
(sum([population]) - LAG(sum([population]), 1)
```

```

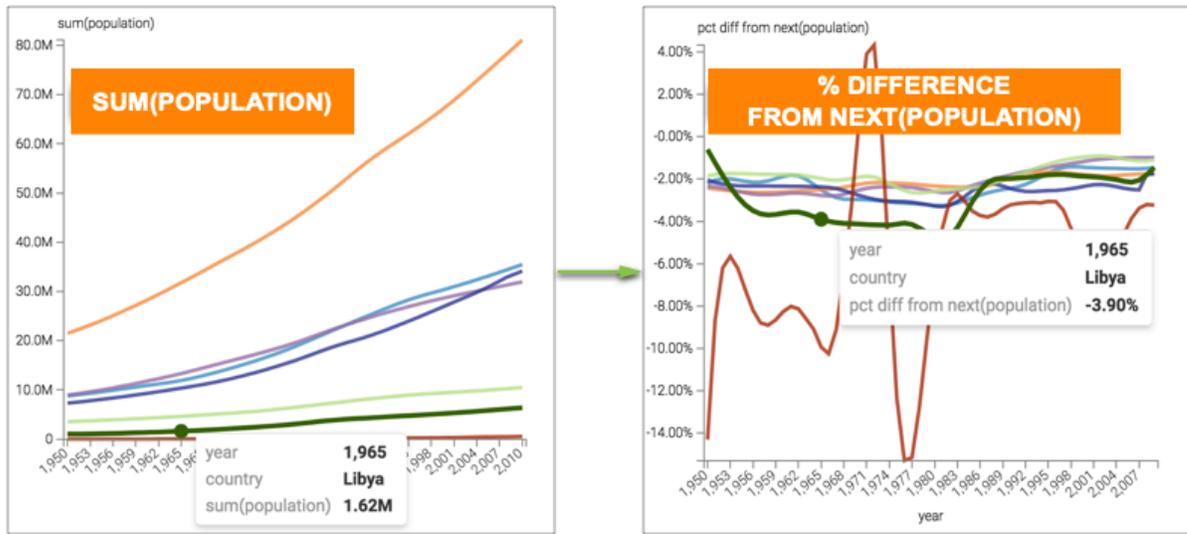
over (partition by [country] order by [country], [year])) /
LAG(sum([population]), 1)
over (partition by [country] order by [country], [year])
as 'pct diff from next(population)'

```

6. Click REFRESH VISUAL.

You can see the change in the Y axis and tooltips, from displaying the sum(population) in millions to population as a percentage difference from next value.

Hover over any country to view the percentage difference of population from the next year. In the following example, the tooltip shows a decrease in the population of Libya from 1965 to next year 1966 by 3.90%.



Row number function

This article describes how to use the Row Number analytic function in CDP Data Visualization.

Procedure

1. Click a field on the dimension shelf. In this case, choose the population field.
2. In the FIELD PROPERTIES menu, select Duplicate.
3. Click the second population field.

4. In FIELD PROPERTIES, expand Analytic Functions, and select Row Number.

The screenshot displays the Cloudera Data Visualization interface. On the left, a table shows data for various countries. On the right, the 'FIELD PROPERTIES' panel is open, and the 'Row Number' function is selected under the 'Analytic Functions' category.

country	life expectancy	GDP	population	population
Mayotte	78.5	\$0	204,114	204,114
Tunisia	75.1	\$80,334,260,876	10,480,934	10,480,934
Libya	74.7	\$75,259,780,831	6,355,112	6,355,112
Cape Verde	74.1	\$1,813,273,168	495,999	495,999
Mauritius	73.1	\$16,002,161,612	1,299,172	1,299,172
Seychelles	72.7	\$1,413,868,538	86,518	86,518
Algeria	70.6	\$223,470,994,789	35,468,208	35,468,208
Egypt	70.5	\$491,593,744,800	81,121,080	81,121,080

The 'FIELD PROPERTIES' panel on the right shows the following structure:

- DATE/TIME FUNCTIONS
- TEXT FUNCTIONS
- ANALYTIC FUNCTIONS
 - Previous Value
 - Difference from Previous
 - % Difference from Previous
 - Next Value
 - Difference from Next
 - % Difference from Next
 - Row Number** (highlighted)
 - Rank
 - Dense Rank
 - Moving Average
 - % of Group
 - Running Total
- CHANGE TYPE
- ORDER AND TOP K
 - [] Enter/Edit Expression
- DISPLAY FORMAT
- ALIAS

5. In the Analytic Function: Row Number modal window:

- a. Under Select entities over which to calculate row number, you can select any of the fields: country, life expectancy, or GDP. This is the grouping that applies when calculating the row number.
- b. Under Select the sort order, you can specify the order of sorting. This order specifies what the row is, and the row where the value is taken. If your visual already specifies sorting order, these values appear in the modal.



Note: You can add additional sort parameters, by clicking the (Plus) icon.

In this example, leave the default order: life expectancy descending, followed by country and GDP, both ascending.

Analytic Function : Row Number ✕

1 Select entities over which to calculate row number (optional)

country

life expectancy

GDP

2 * Select the sort order

1 life expectancy ⬇️ -

2 country ⬇️ -

3 GDP ⬇️ -

(If you do not select a sort order, the aggregate of the current column will be applied)

+

CANCEL
APPLY

6. Click APPLY.

7. If you examine the expression in the Enter/Edit Expression interface, it shows that the function has the following form:

```
ROW_NUMBER() over ( order by [life_expectancy], [country], [gdp_per_capi
ta]*[population]) as 'row number'
```

8. Click REFRESH VISUAL.

You can see that initially, the row number numbers are descending, following the order of the primary sorting column, life expectancy. If you sort the table on another column, such as population, the row number numbers appear disordered.

country	life expectancy	GDP	population	row number
Mayotte	78.5	\$0	204,114	52
Tunisia	75.1	\$80,334,260,876	10,480,934	51
Libya	74.7	\$75,259,780,831	6,355,112	50
Cape Verde	74.1	\$1,813,273,168	495,999	49
Mauritius	73.1	\$16,002,161,612	1,299,172	48
Seychelles	72.7	\$1,413,868,538	86,518	47
Algeria	70.6	\$223,470,994,789	35,468,208	46
Egypt	70.5	\$491,593,744,800	81,121,080	45

country	life expectancy	GDP	population	row number
Nigeria	51.3	\$363,945,588,339	158,423,184	9
Ethiopia	61.5	\$72,315,411,447	82,949,544	36
Egypt	70.5	\$491,593,744,800	81,121,080	45
Congo	49.0	\$24,644,821,788	65,965,796	5
South Africa	54.5	\$467,348,147,107	50,132,816	19
Kenya	59.6	\$59,606,308,062	40,512,680	29
Algeria	70.6	\$223,470,994,789	35,468,208	46
Sudan	61.5	\$88,655,959,426	34,136,520	35

Rank analytic function

This article describes how to use the Rank analytic function in CDP Data Visualization.

Procedure

1. Click a field on the dimension shelf. In this case, choose the population field.
2. In the FIELD PROPERTIES menu, select Duplicate.
3. Click the second population field.
4. In FIELD PROPERTIES, expand Analytic Functions, and select Rank.

The screenshot shows the Cloudera Data Visualization interface. On the left, a table displays data for various countries, including Mayotte, Tunisia, Libya, Cape Verde, Mauritius, Seychelles, Algeria, and Egypt. The table has columns for country, life expectancy, GDP, and two population columns. On the right, the 'FIELD PROPERTIES' panel is open, showing the 'Analytic Functions' section with the 'Rank' function selected and highlighted in a red box. The 'Measures' section is empty, and the 'Filters' section contains 'year in (2010)' and 'un_region in (Africa)'. A 'REFRESH VISUAL' button is visible at the bottom of the panel.

5. In the Analytic Function: Rank modal window:

- a. Under Select entities over which to calculate rank, you can select any of the fields: country, life expectancy, or GDP. This is the grouping that applies when calculating the rank.
- b. Under Select the sort order, you can specify the order of sorting. This order specifies what the row is, and the row where the value is taken. If your visual already specifies sorting order, these values appear in the modal.



Note: You can add additional sort parameters, by clicking the (Plus) icon.

In this example, leave the default order: life expectancy descending, followed by country and GDP, both ascending.

Analytic Function : Rank ✕

1

Select entities over which to calculate rank (optional)

country

life expectancy

GDP

2

* Select the sort order

1 life expectancy ⬇️ -

2 country ⬇️ -

3 GDP ⬇️ -

(If you do not select a sort order, the aggregate of the current column will be applied)

+

CANCEL
APPLY

6. Click APPLY.

7. If you examine the expression in the Enter/Edit Expression interface, it shows that the function has the following form:

```
RANK() over ( order by [life_expectancy], [country], [gdp_per_capita]*[p
opulation]) as 'rank'
```

8. Click REFRESH VISUAL.

You can see that initially, the rank numbers are descending, following the order of the primary sorting column, life expectancy. If you sort the table on another column, such as GDP, the rank numbers appear disordered.

country	life expectancy	GDP	population	rank
Mayotte	78.5	\$0	204,114	52
Tunisia	75.1	\$80,334,260,876	10,480,934	51
Libya	74.7	\$75,259,780,831	6,355,112	50
Cape Verde	74.1	\$1,813,273,168	495,999	49
Mauritius	73.1	\$16,002,161,612	1,299,172	48
Seychelles	72.7	\$1,413,868,538	86,518	47
Algeria	70.6	\$223,470,994,789	35,468,208	46
Egypt	70.5	\$491,593,744,800	81,121,080	45

country	life expectancy	GDP	population	rank
Egypt	70.5	\$491,593,744,800	81,121,080	45
South Africa	54.5	\$467,348,147,107	50,132,816	19
Nigeria	51.3	\$363,945,588,339	158,423,184	9
Algeria	70.6	\$223,470,994,789	35,468,208	46
Morocco	70.2	\$136,215,265,879	31,951,412	44
Angola	50.7	\$104,904,721,275	19,081,912	8
Sudan	61.5	\$88,655,959,426	34,136,520	35
Tunisia	75.1	\$80,334,260,876	10,480,934	51

Dense rank analytic function

This article describes how to use the Dense Rank analytic function in CDP Data Visualization.

Procedure

1. Click a field on the dimension shelf.
In this example, choose the population field.
2. In the FIELD PROPERTIES menu, select Duplicate.
3. Click the second population field.
4. In FIELD PROPERTIES, expand Analytic Functions, and select Dense Rank.

The screenshot shows the Cloudera Data Visualization interface. On the left, a table displays data for various countries, with columns for country, life expectancy, GDP, and two population columns. The table is sorted by life expectancy in descending order. On the right, the 'FIELD PROPERTIES' panel is open, showing the 'Analytic Functions' section with 'Dense Rank' selected. The 'DIMENSIONS' shelf contains 'country', 'life expectancy', 'GDP', and 'population'. The 'MEASURES' shelf is empty. The 'FILTERS' shelf contains 'year in (2010)' and 'un_region in (Africa)'. The 'LIMIT' is set to 100. A 'REFRESH VISUAL' button is visible at the bottom of the interface.

5. In the Analytic Function: Dense Rank modal window:
 - a. Under Select entities over which to calculate rank, you can select any of the fields: country, life expectancy, or GDP. This is the grouping that applies when calculating the dense rank.
 - b. Under Select the sort order, you can specify the order of sorting. This order specifies what the row is, and the row where the value is taken. If your visual already specifies sorting order, these values appear in the modal.



Note: You can add additional sort parameters, by clicking the (Plus) icon.

In this example, leave the default order: life expectancy descending, followed by country and GDP, both ascending.

Analytic Function : Rank ✕

1

Select entities over which to calculate rank (optional)

country

life expectancy

GDP

2

* Select the sort order

1 life expectancy ⌵ -

2 country ⌵ -

3 GDP ⌵ -

(If you do not select a sort order, the aggregate of the current column will be applied)

+

CANCEL
APPLY

6. Click APPLY.
7. If you examine the expression in the Enter/Edit Expression interface, it shows that the function has the following form:

```
DENSE_RANK() over ( order by [life_expectancy], [country], [gdp_per_capi
ta]*[population]) as 'dense rank'
```

8. Click REFRESH VISUAL.

You can see that initially, the dense rank numbers are descending, following the order of the primary sorting column, life expectancy. If you sort the table on another column, such as population, the dense rank numbers appear disordered.

country	life expectancy	GDP	population	dense rank
Mayotte	78.5	\$0	204,114	52
Tunisia	75.1	\$80,334,260,876	10,480,934	51
Libya	74.7	\$75,259,780,831	6,355,112	50
Cape Verde	74.1	\$1,813,273,168	495,999	49
Mauritius	73.1	\$16,002,161,612	1,299,172	48
Seychelles	72.7	\$1,413,868,538	86,518	47
Algeria	70.6	\$223,470,994,789	35,468,208	46
Egypt	70.5	\$491,593,744,800	81,121,080	45

country	life expectancy	GDP	population	dense rank
Seychelles	72.7	\$1,413,868,538	86,518	47
Sao Tome and Principe	65.9	\$303,586,194	165,397	42
Mayotte	78.5	\$0	204,114	52
Cape Verde	74.1	\$1,813,273,168	495,999	49
Western Sahara	66.8	\$0	530,500	43
Equatorial Guinea	51.5	\$9,374,026,767	700,401	10
Comoros	60.2	\$751,355,332	734,750	30
Djibouti	60.3	\$1,924,781,156	888,716	31

Moving average analytic function

This article describes how to use the Moving Average analytic function in CDP Data Visualization.

Procedure

1. Click the field on the measurement shelf.

In this example, the sum(population) field on the Y Axis shelf.

2. In the FIELD PROPERTIES menu, expand Analytic Functions, and select Moving Average.

The screenshot displays the Cloudera Data Visualization interface. On the left, a line chart titled "enter title..." shows population data from 1950 to 2010. The Y-axis is labeled "sum(population)" and ranges from 0 to 80.0M. The X-axis is labeled "year" and ranges from 1950 to 2010. The chart contains several lines representing different countries, with a prominent green line showing exponential growth. At the top of the interface are "SAVE" and "CLOSE" buttons. On the right, the "FIELD PROPERTIES" panel is open, showing a tree view of analytic functions. Under "Analytic Functions", the "Moving Average" option is highlighted with an orange box. Other options include "Previous Value", "Difference from Previous", "Next Value", "Difference from Next", "Row Number", "Rank", "Dense Rank", "% of Group", and "Running Total". The "VISUALS" panel on the left of the Field Properties shows the current visualization settings: X Axis is "# year", Y Axis is "# sum(population)", and Filters include "((year >= 1950) AND (year <= ...))" and "un_subregion in ('Northern Af...')". A "REFRESH VISUAL" button is located at the bottom of the VISUALS panel.

3. In the Analytic Function: Moving Average modal window:

- a. In Select aggregate for this measure, leave the default Sum. The other aggregation options are Count, Minimum, Maximum, and Average. In addition, some connection types also support Approximate Distinct Count and Exact Distinct Count.
- b. Under Select entities over which to calculate moving averages, select year. This is the grouping that applies when calculating the moving average.
- c. Under Select the sort order, you can specify the order of sorting. This order specifies what the row is, and the row where the value is taken. If your visual already specifies sorting order, these values appear in the modal.



Note: You can add additional sort parameters, by clicking the (Plus) icon.

In this example, leave the default order: year ascending, followed by country ascending.

- d. Under Rows to include in the average (relative to current row), the options for leading rows are: All previous rows, Current row, or Previous row(s) (default). Select the last option, then specify the number of preceding rows for the calculation; the default is 5, but we changed it to 2.

Similarly, the options for following rows are: All forward rows, Current row (default), or Forward row(s). Select the last option; the default is 5, but we changed it to 2.

Analytic Function : Moving Average ×

1 Select aggregate for this measure

✓ Sum

Count

Approx Distinct Count

Exact Distinct Count

Minimum

Maximum

Average

2 Select entities over which to create moving averages (optional)

year

country

3 * Select the sort order

1 year ⌵ -

2 country ⌵ -

(If you do not select a sort order, the aggregate of the current column will be applied)

+

4 * Rows to include in the average (relative to current row)

Previous row(s)... 2 to Forward row(s)... 2

CANCEL
APPLY

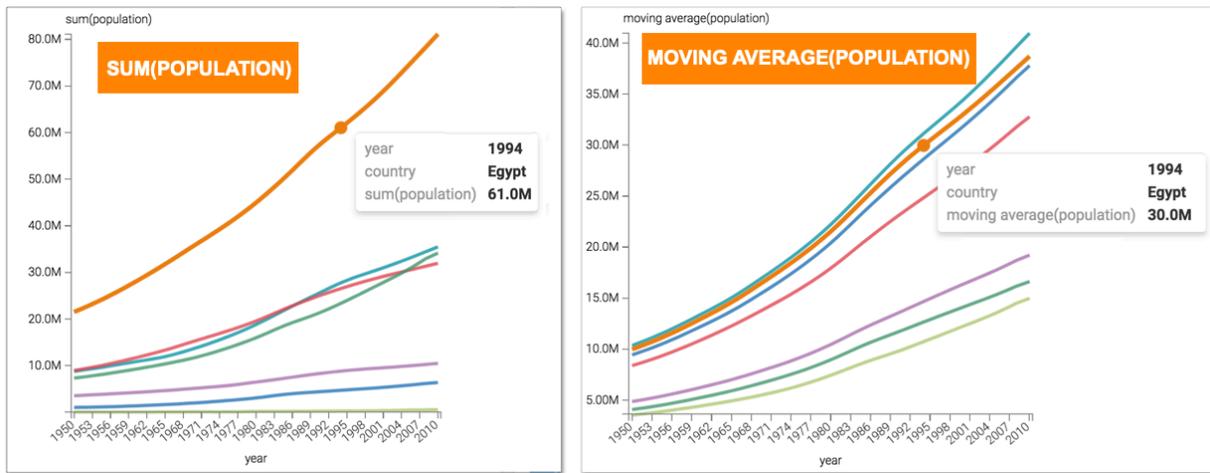
4. Click APPLY.

- If you examine the expression in the Enter/Edit Expression interface, it shows that the function has the following form:

```
avg(sum([population])) over (partition by [year] order by [year], [country]
rows between 2 preceding and 2 following) as 'moving average(population)'
```

- Click REFRESH VISUAL.

You can see the change to the appearance of the visual, and the change in values that appear on the tooltip.



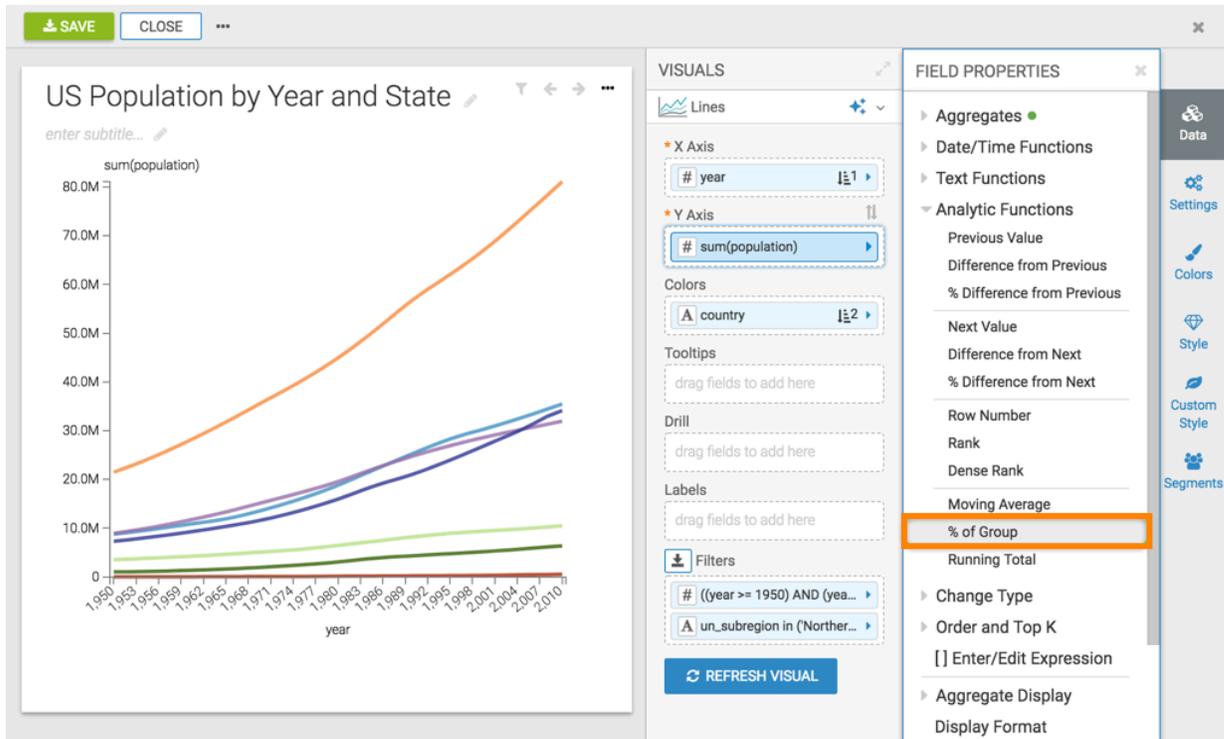
Percent of group analytic function

This article describes how to use the Percent of Group analytic function in CDP Data Visualization.

Procedure

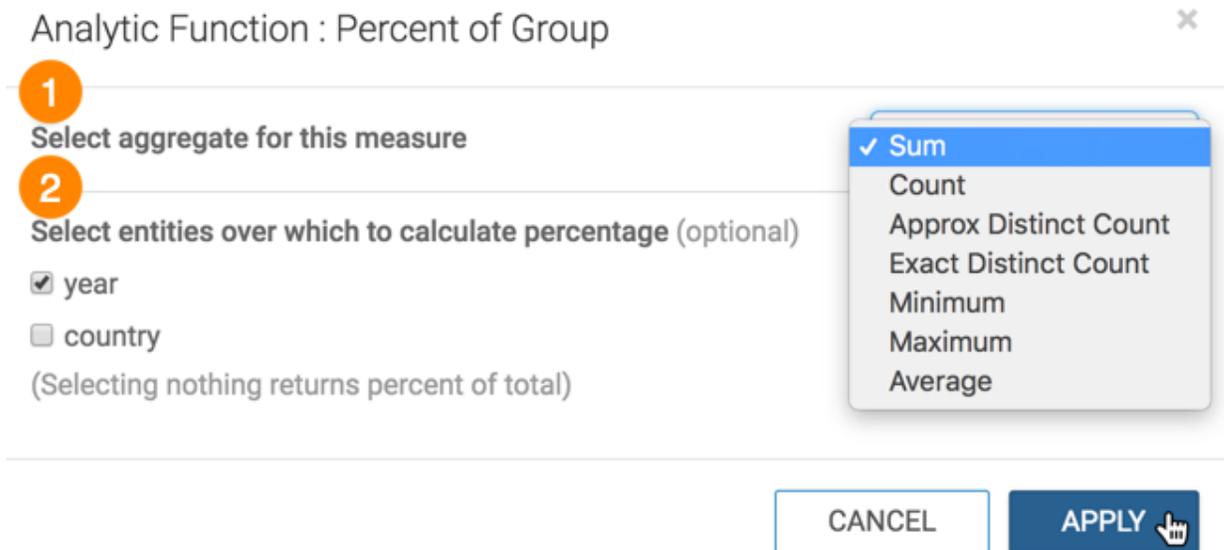
- Click the field on the measurement shelf.
In this example, the sum(population) field on the Y Axis shelf.

2. In the FIELD PROPERTIES menu, expand Analytic Functions, and select % of Group.



3. In the Analytic Function: Percent of Group modal window:

- a. In Select aggregate for this measure, leave the default Sum. The other aggregation options are Count, Minimum, Maximum, and Average. In addition, some connection types also support Approximate Distinct Count and Exact Distinct Count.
- b. Under Select entities over which to calculate percentage, select year. This is the grouping that applies when calculating the percent of group.



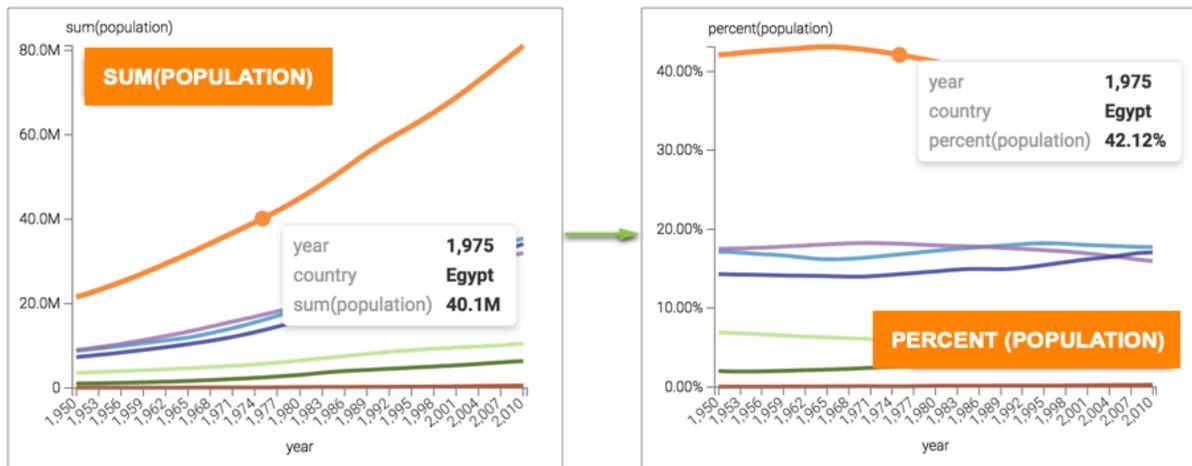
4. Click APPLY.

5. If you examine the expression in the Enter/Edit Expression interface, it shows that the function has the following form:

```
(sum([population]) / sum(sum([population])) over (partition by [year])) as 'percent(population)'
```

6. Click REFRESH VISUAL.

You can see the change in the Y axis and tooltips, from displaying the sum(population) in millions to population as a percent.



Running total analytic function

This article describes how to use the Running Total analytic function in CDP Data Visualization.

Procedure

1. Click the field on the measurement shelf.
In this example, the sum(population) field on the Y Axis shelf.

2. In the FIELD PROPERTIES menu, expand Analytic Functions, and select Running Total.

The screenshot displays the Cloudera Data Visualization interface. On the left, a line chart titled "enter title..." shows population data over time. The Y-axis is labeled "sum(population)" and ranges from 0 to 80.0M. The X-axis is labeled "year" and ranges from 1950 to 2010. The chart contains several lines representing different data series. On the right, the "FIELD PROPERTIES" panel is open, showing a tree view of analytic functions. The "Analytic Functions" category is expanded, and "Running Total" is highlighted with an orange box. Other categories include "Aggregates", "Date/Time Functions", "Text Functions", "Previous Value", "Difference from Previous", "% Difference from Previous", "Next Value", "Difference from Next", "% Difference from Next", "Row Number", "Rank", "Dense Rank", "Moving Average", "% of Group", "Change Type", "Order and Top K", "Aggregate Display", "Display Format", "Alias", "Description", and "Duplicate". The "VISUALS" panel on the left shows the chart configuration, including X Axis (year), Y Axis (sum(population)), Colors (country), Tooltips, Drill, Labels, and Filters.

3. In the Analytic Function: Running Total modal window:
 - a. In Select aggregate for this measure, leave the default Sum. The other aggregation options are Count, Minimum, Maximum, and Average. In addition, some connection types also support Approximate Distinct Count and Exact Distinct Count.
 - b. Under Select entities over which to calculate running totals, select year. This is the grouping that applies when calculating the running total.
 - c. Under Select the sort order, you can specify the order of sorting. This order specifies what the row is, and the row where the value is taken. If your visual already specifies sorting order, these values appear in the modal.



Note: You can add additional sort parameters, by clicking the (Plus) icon.

In this example, leave the default order: year ascending, followed by country ascending.

Analytic Function : Running Total ✕

1 Select aggregate for this measure

2 Select entities over which to take running total (optional)

year

country

3 *Select the sort order

1 year ⌵ -

2 country ⌵ -

(If you do not select a sort order, the aggregate of the current column will be applied)

+

CANCEL

APPLY

4. Click APPLY.
5. If you examine the expression in the Enter/Edit Expression interface, it shows that the function has the following form:

```
sum(sum([population])) over (partition by [year] order by [year], [count
ry]
rows between unbounded preceding and current row) as 'running
total(population)'
```

6. Click REFRESH VISUAL.

You can see the change to the appearance of the visual, and the change in values that appear on the tooltip.

