..

Using Analytic Functions

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

CLOUDƎRA

Legal Notice

Contents

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.

VISUALS

📈 Lines

* 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…

🔄 **REFRESH VISUAL**

FIELD PROPERTIES ✕

▸ Aggregates ●

▸ 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
  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

🔷 Data

⚙ Settings

🖌 Colors

◈ Style

🍃 Custom Style

👥 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 filed 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.

4. Click REFRESH VISUAL to see the new line visual.
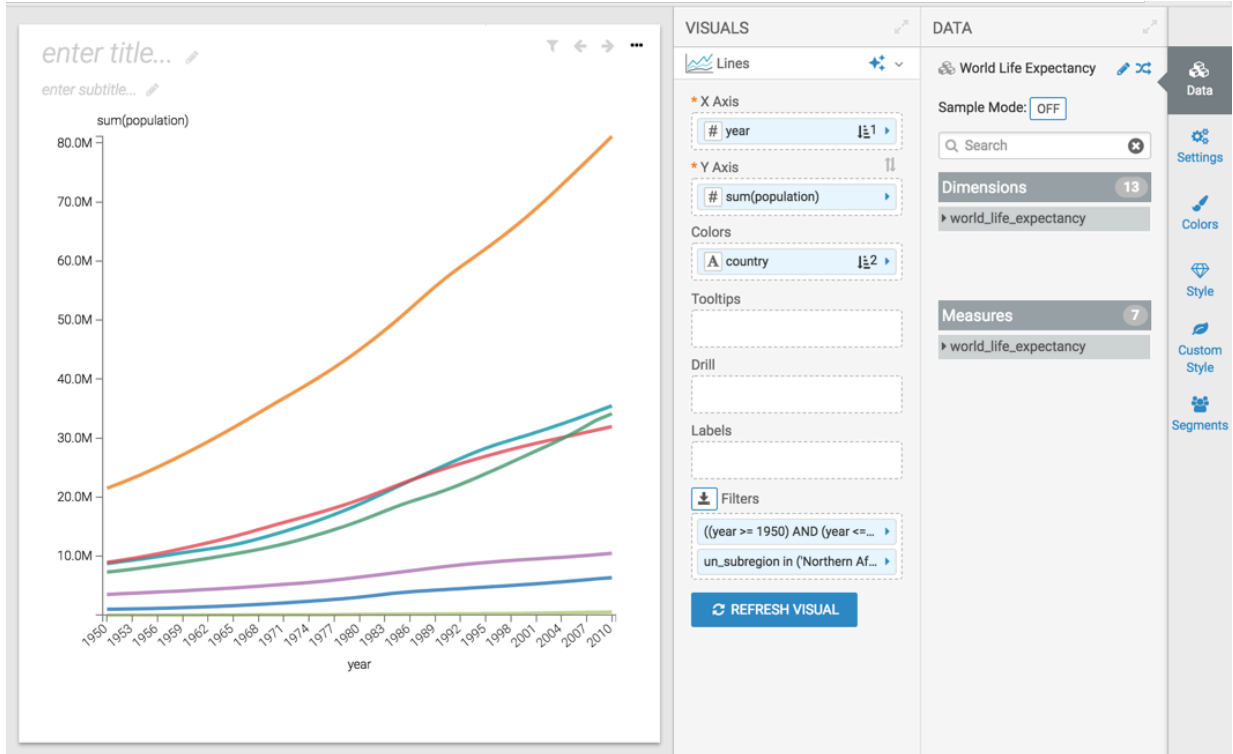
5. Name the visual Basic Lines.

6. Click SAVE.

Setting up a basic visual for single values

Procedure

1. Open a new visual.

2. In the Visuals menu, select the Table visual type.

3. 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.

4. Specify descending order on life_expectancy.

5. In the Enter/Edit Expression editor, change the gdp_per_capita calculation on the shelf, and rename it:[gdp_per
_capita]*[population] as 'GDP'

6. Change the Display Format options for the fields on the shelves to remove extra decimals, specify currency, and
so on.

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

8. Name the visual Basic Table.

9. Click Save.

Previous value analytic function

This article describes how to use the 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 Previous Value.
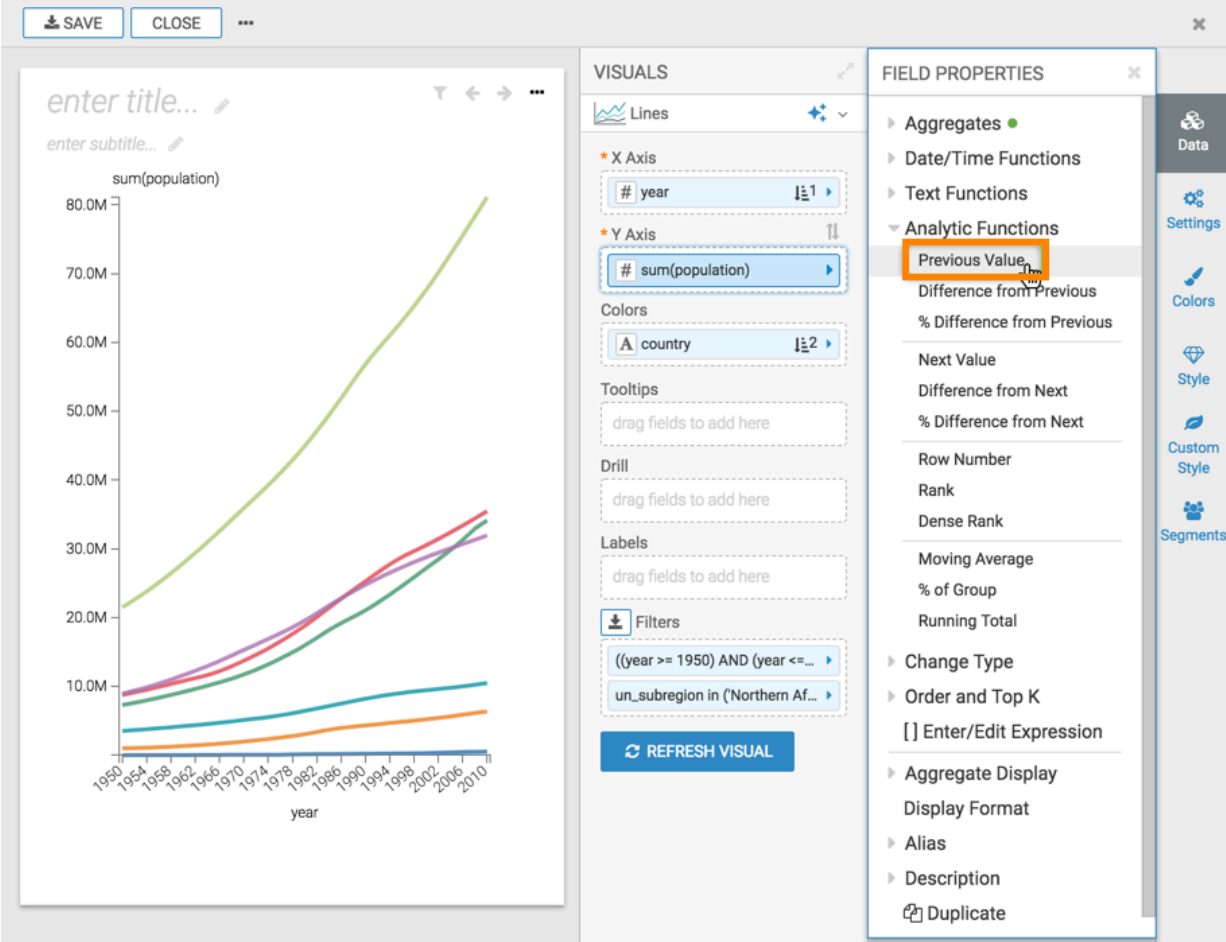
3. In the Analytic Function: Previous Value modal window:

    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 tale previous value, select state. This is the grouping that applies when calculating previous values.

    c. 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.

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

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 '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.

3. In the Analytic Function: Difference from Previous 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 previous value, select year. This is the grouping that applies when calculating the difference from previous values.

   c. 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.

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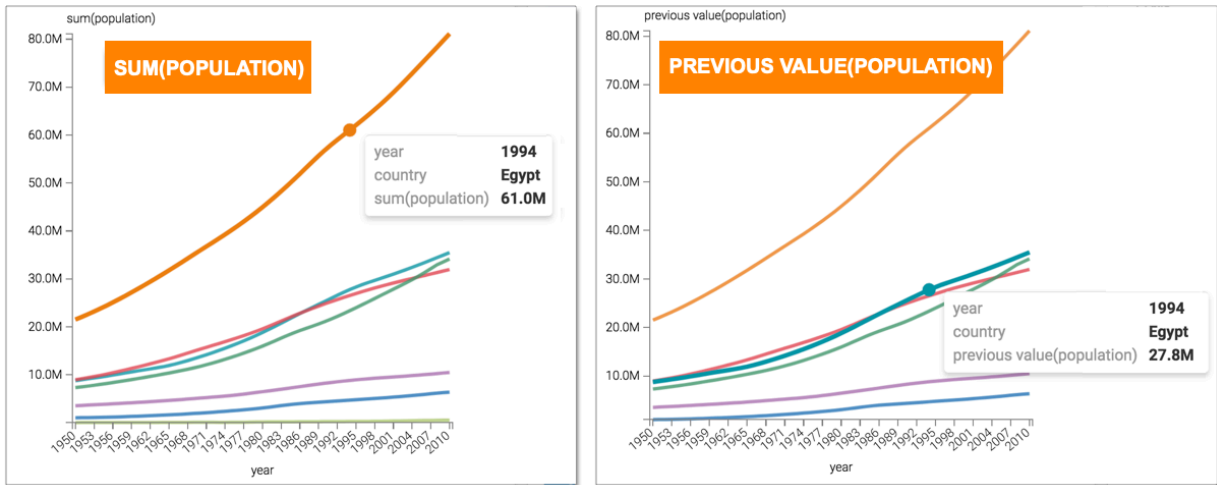

4. Click APPLY.

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 [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.

3. In the Analytic Function: Percentage Difference from Previous 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 previous value, select country. This is the grouping that applies when calculating the % difference from previous values.

    c. 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.

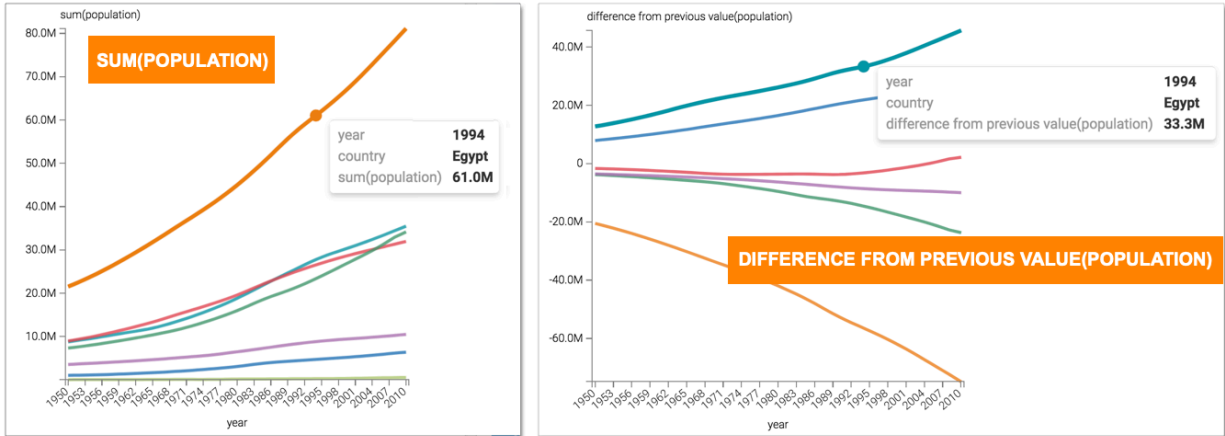   d. Under Offset, specify the number of previous rows over which the function executes.

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([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%.
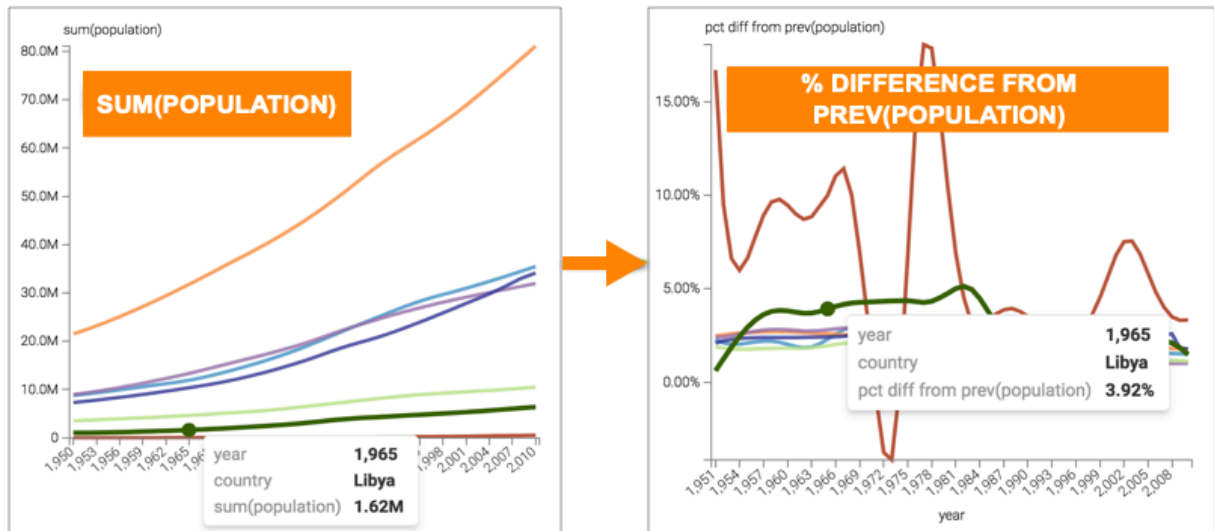
Next value analytic function

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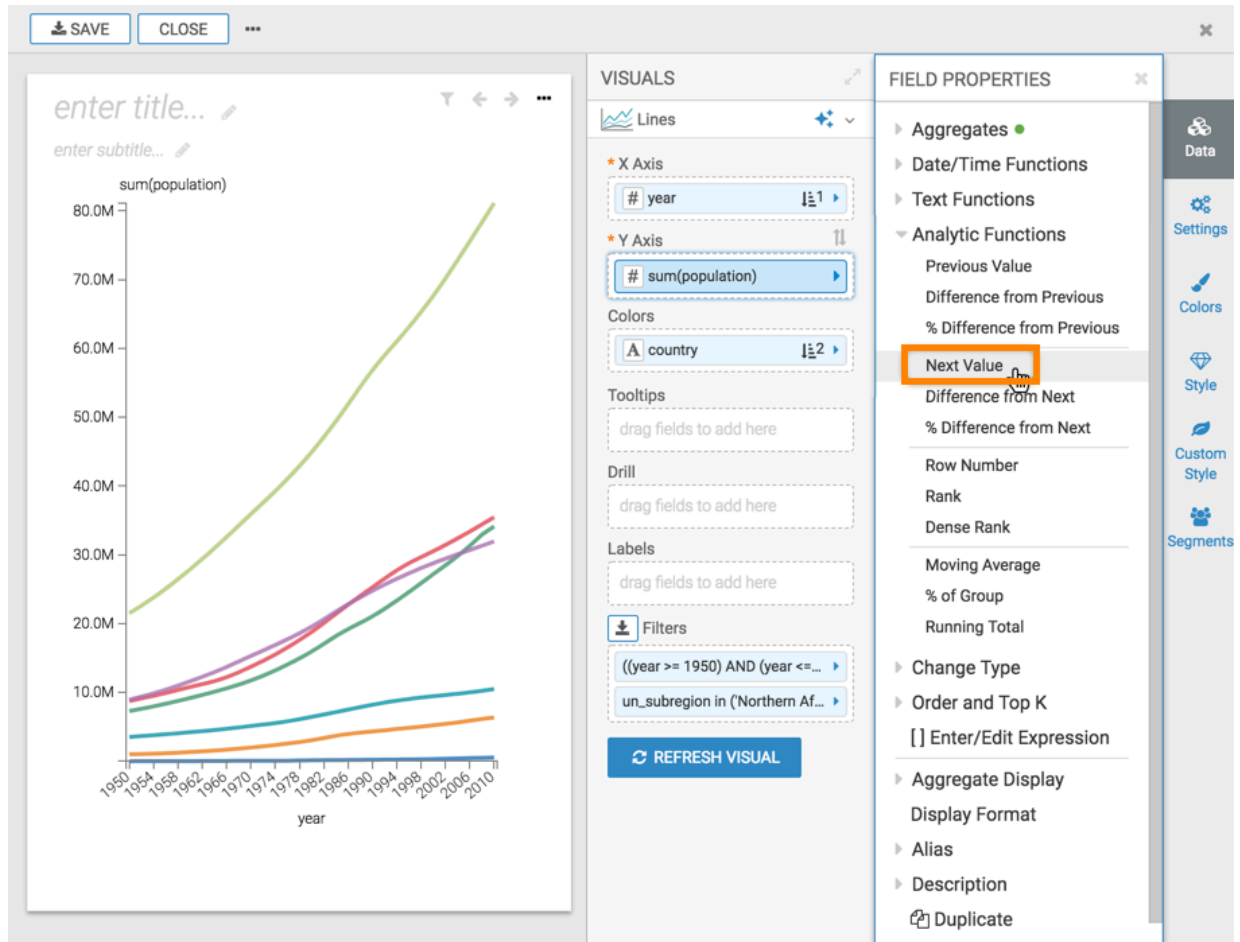
3. In the Analytic Function: Next Value modal window:

    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 tale next value, select state. This is the grouping that applies when calculating next values.

    c. 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.

    d. 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: [1] 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], [co
untry])
            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.

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.

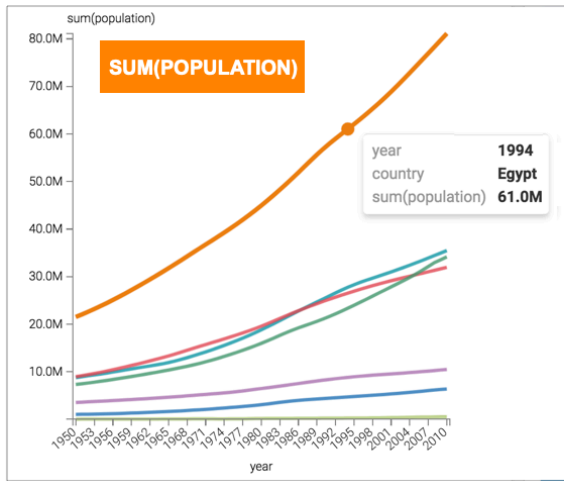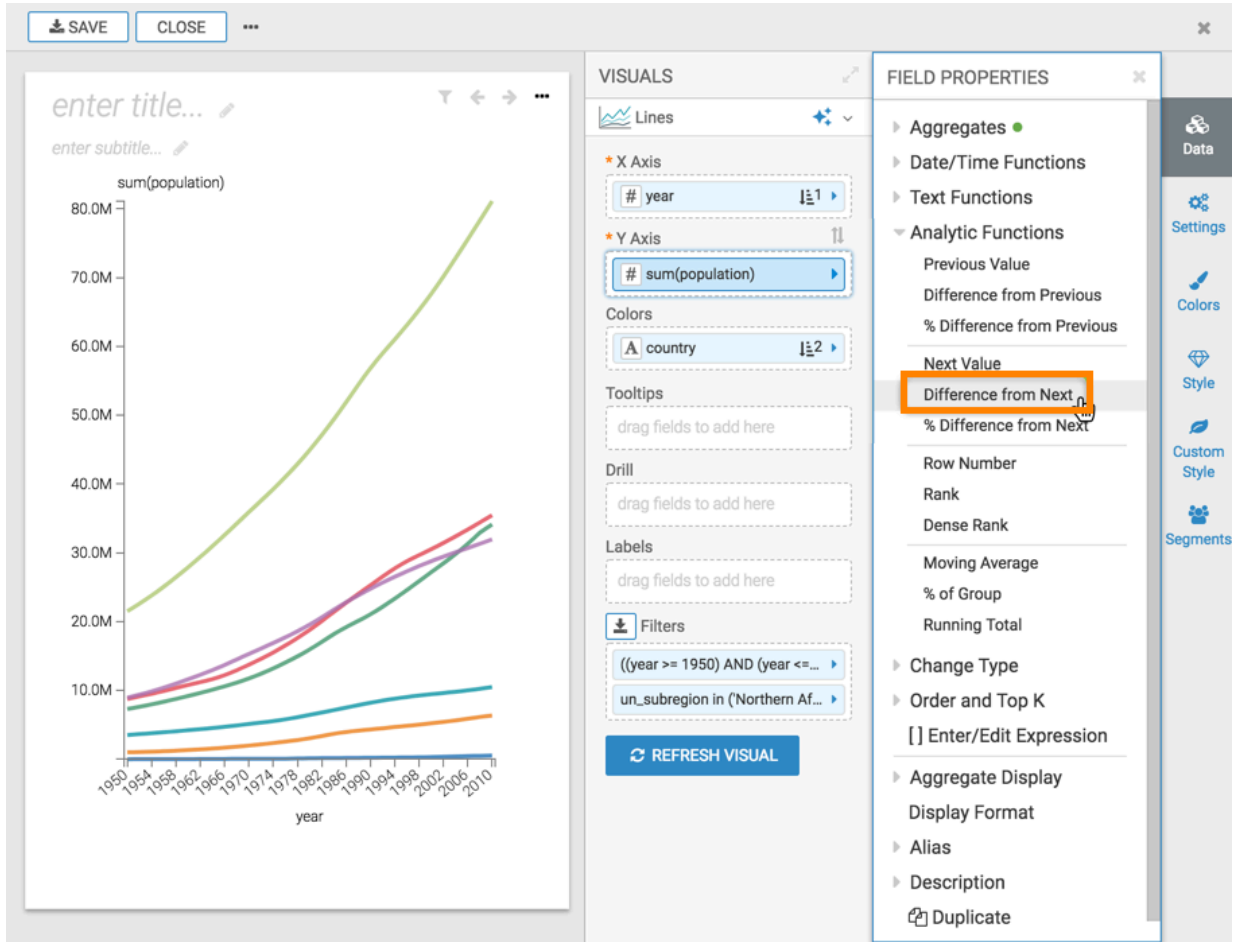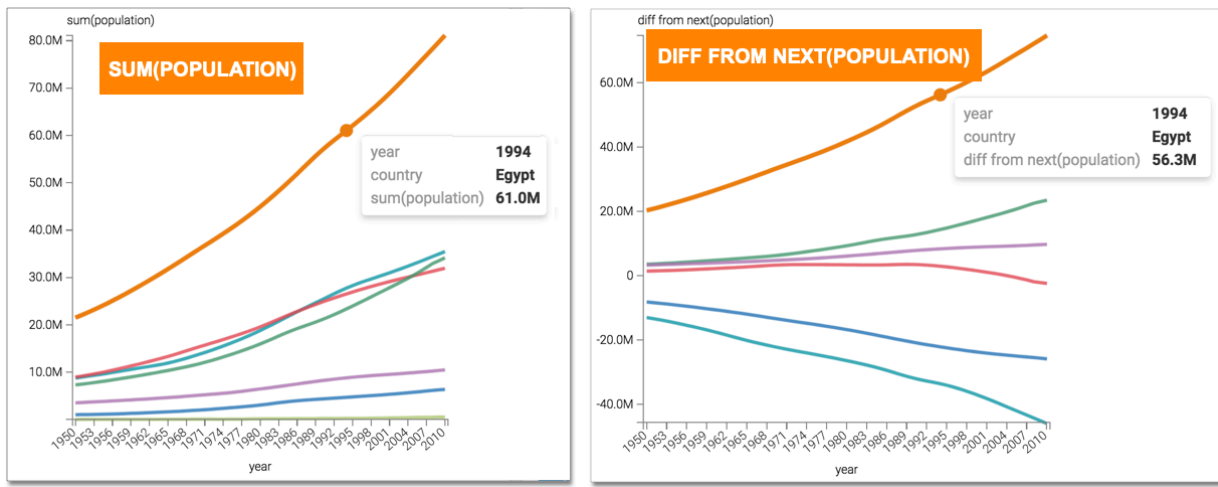   d. Under Offset, specify the number of next rows over which the function executes.

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.

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.

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%.
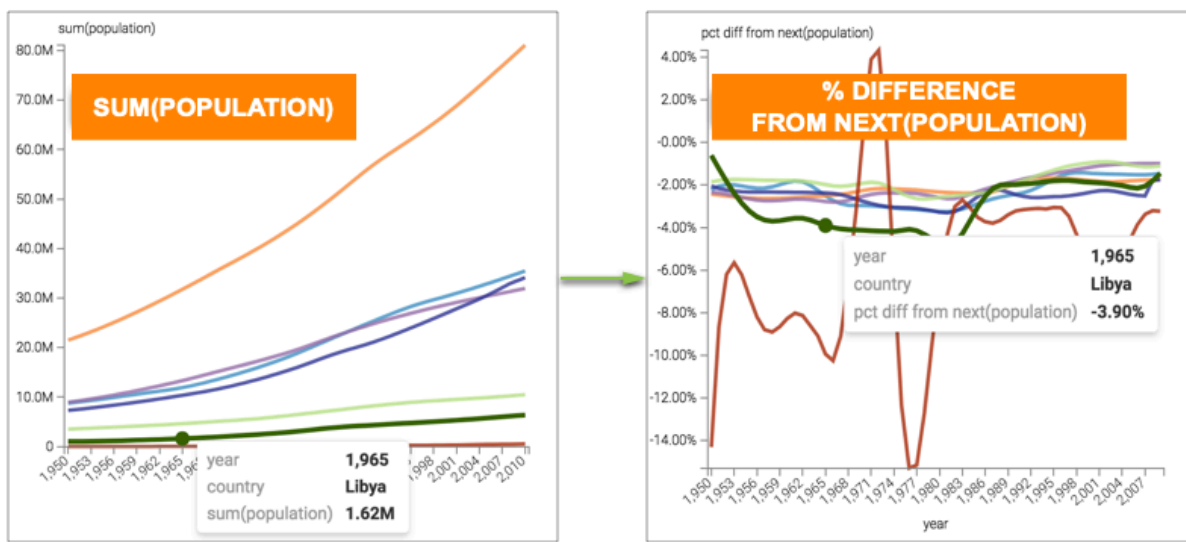
Row number function

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.

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.

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.

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

   | ↓F 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.

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.

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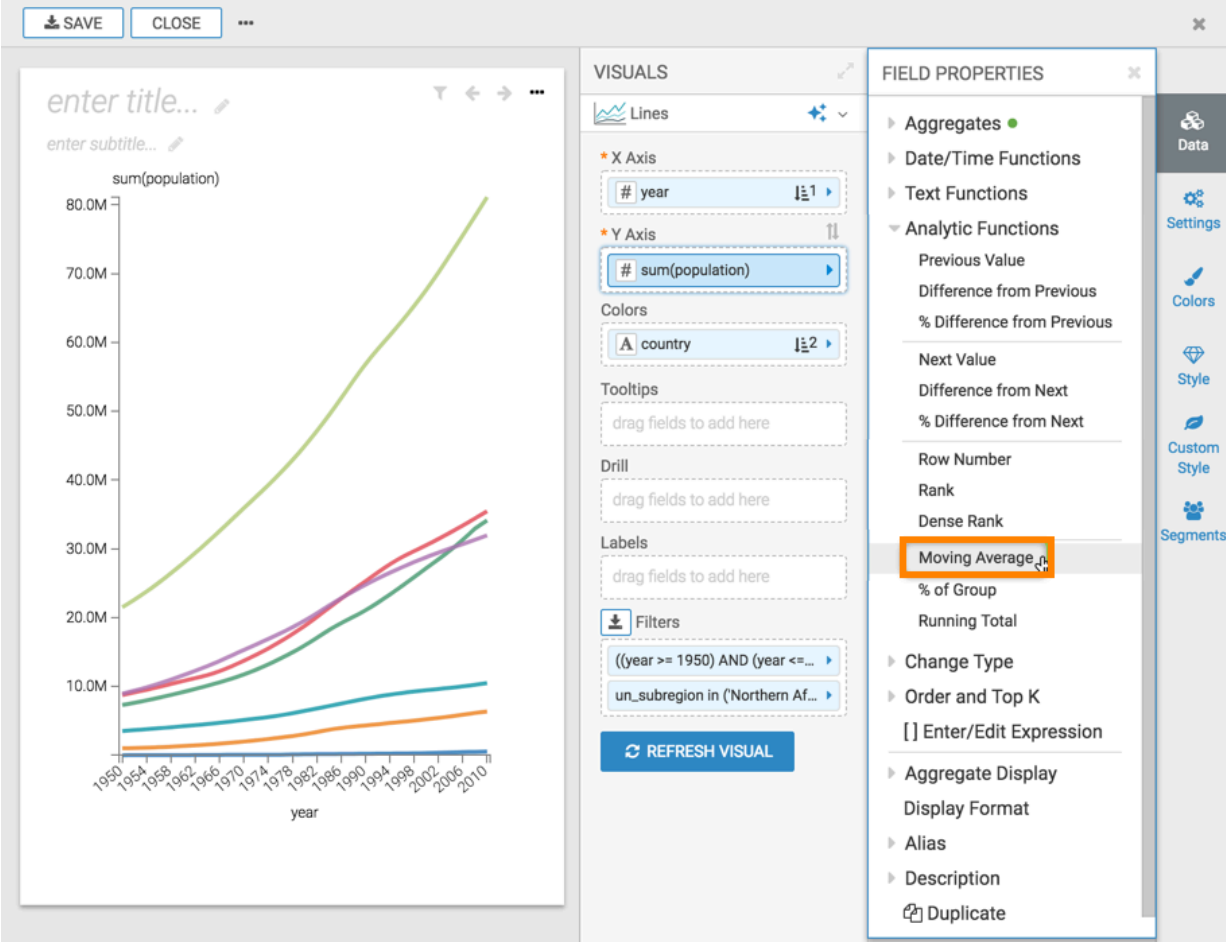
3. In the Analytic Function: Moving Average modal window:

   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.


4. Click APPLY.

5. 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], [count
ry]
             rows between 2 preceding and 2 following) as 'moving averag
e(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.
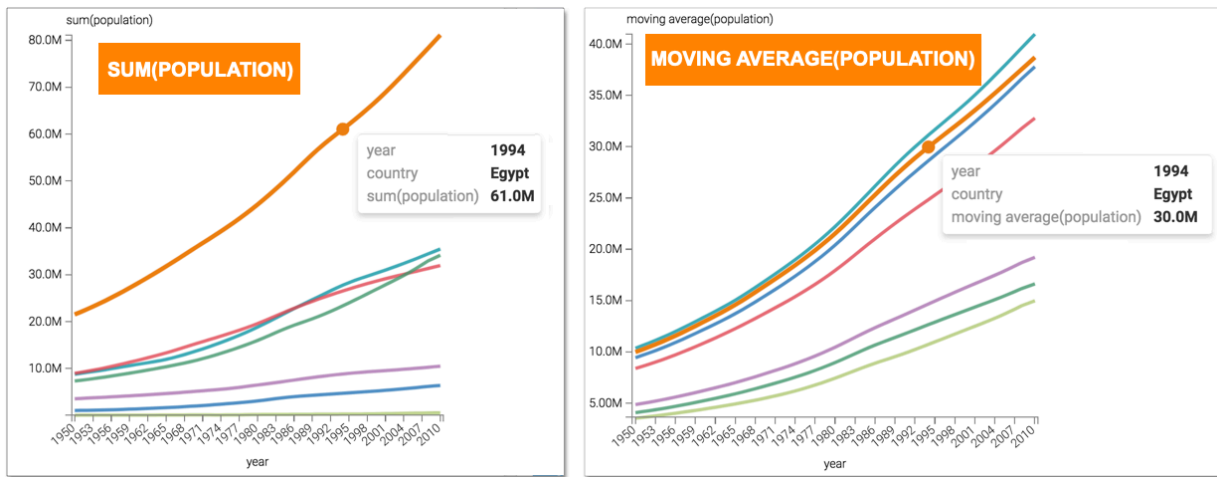
Percent of group analytic function

This article describes how to use the Percent of Group 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 % of Group.
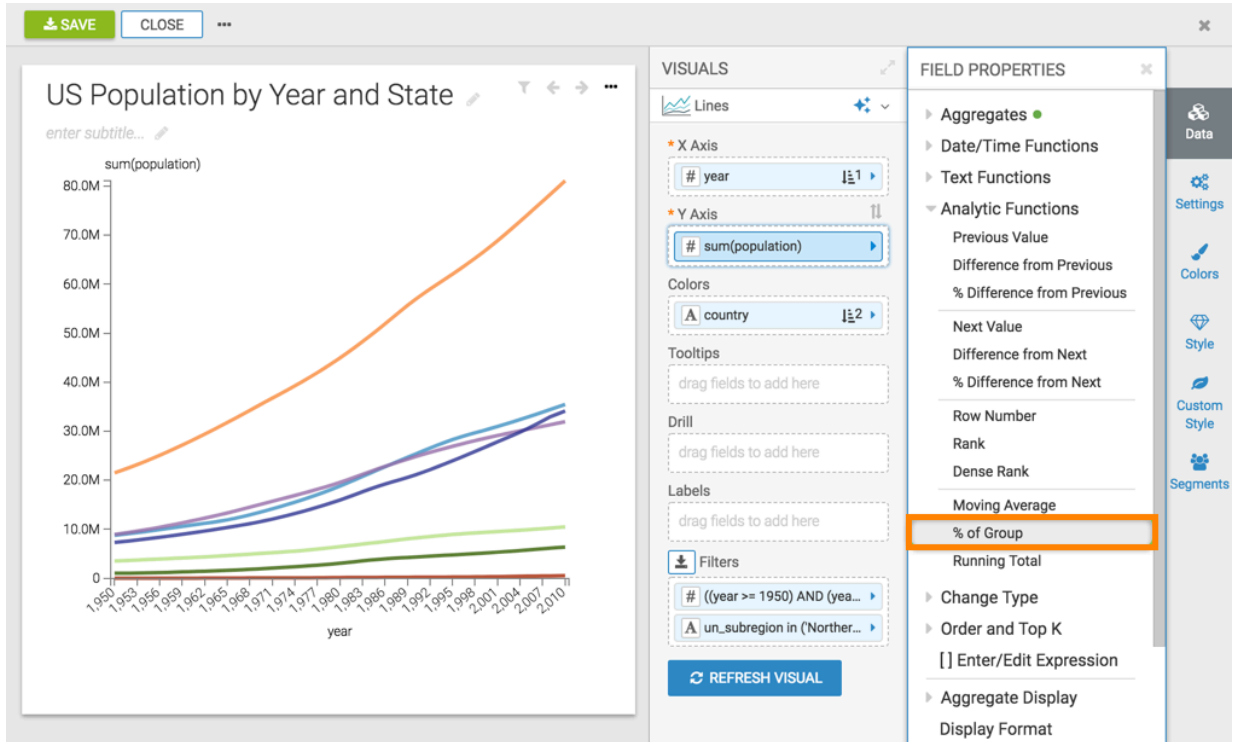
3. In the Analytic Function: Percent of Group modal window:

    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.
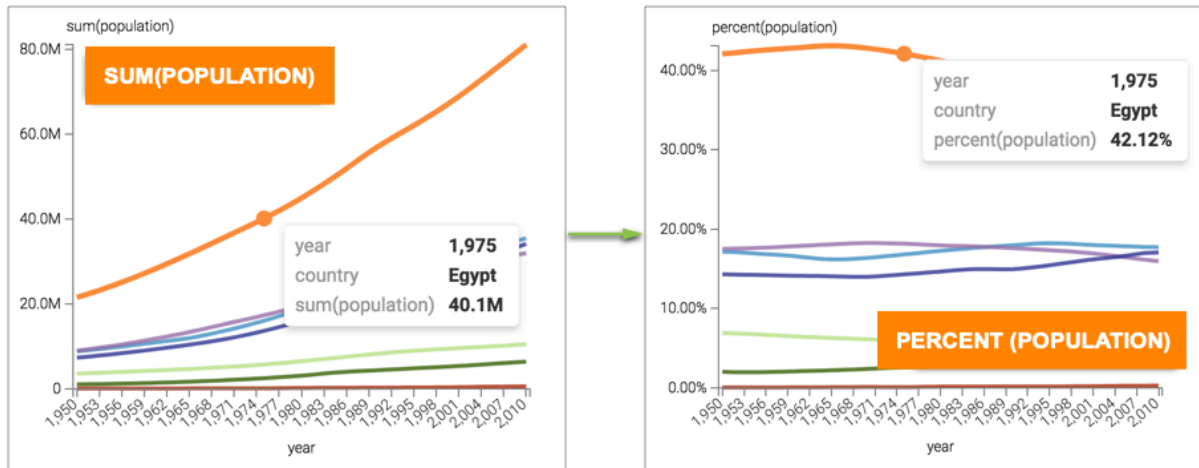

4. Click APPLY.

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.
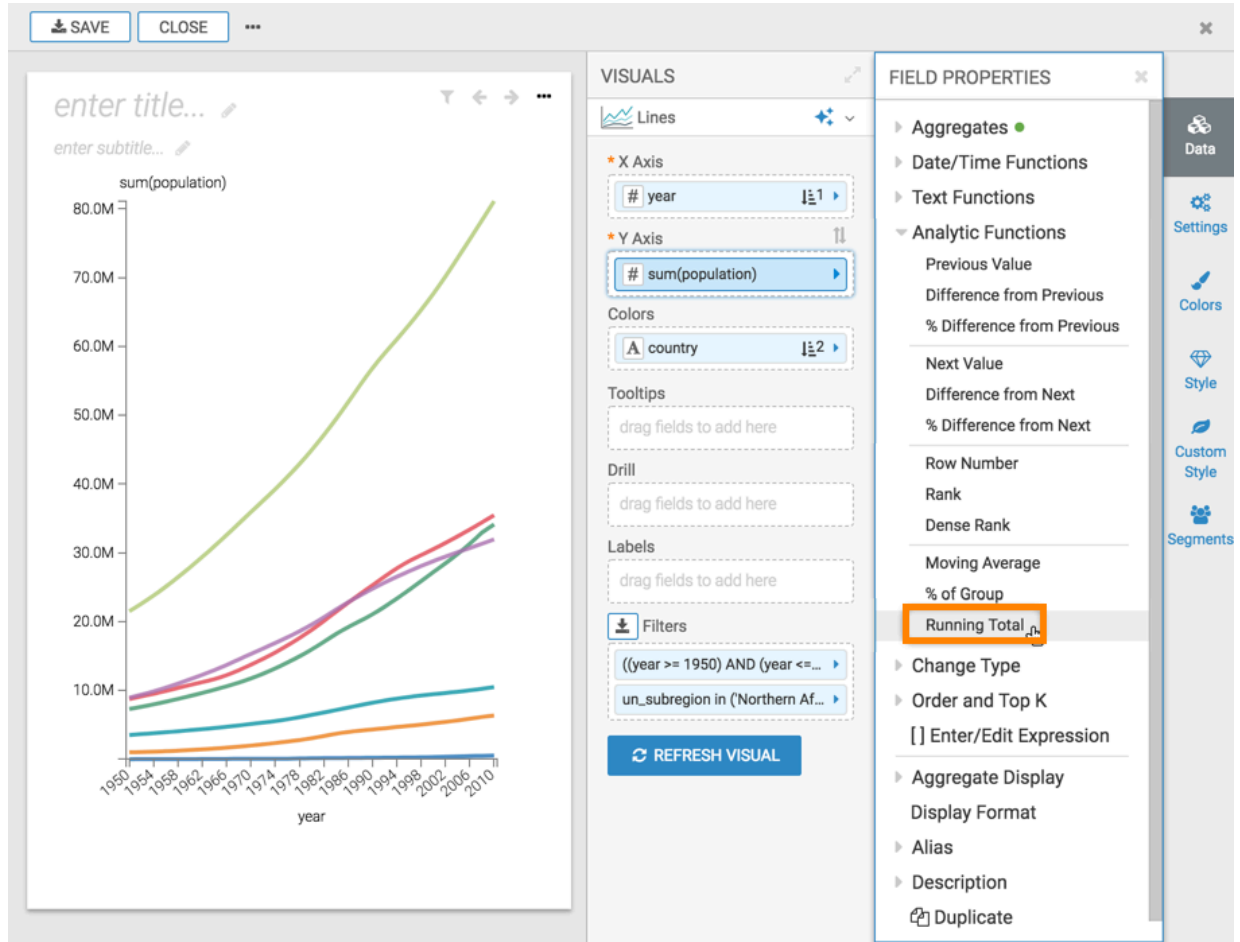

Running total analytic function

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.

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.

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.