Cloudera Data Visualization 7.2.8

Advanced Visualization Techniques

Date published: 2020-10-30
Date modified: 2024-10-30

https://docs.cloudera.com/

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

Trellis on dimensions.....4

Trellis on measures.....7

Multiple measure analysis..... 9

Line visuals with dual axes.....11

Working with dimension hierarchies..... 13

Hierarchical cross tabulation visuals.....17

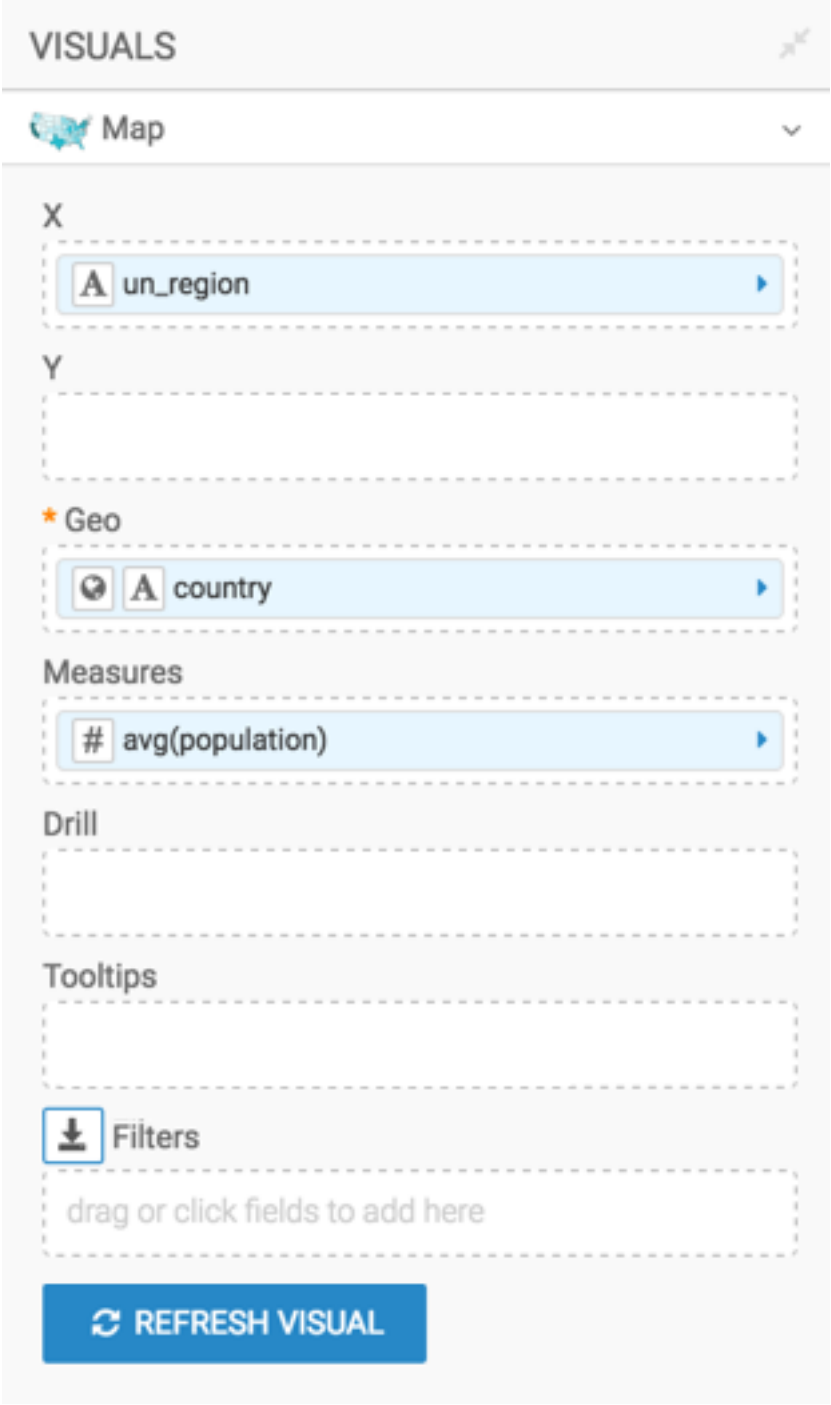
Adding highlights to visuals..... 26

Trellis on dimensions

Trellising charts on dimensions produces multiple charts that focus on a distinct section of data.

Trellising on one dimension

In this example, the field un_region becomes a trellising dimension when it is placed on the X shelf of a map visual.



This setup produces a trellised version of a map visual, on one dimension.

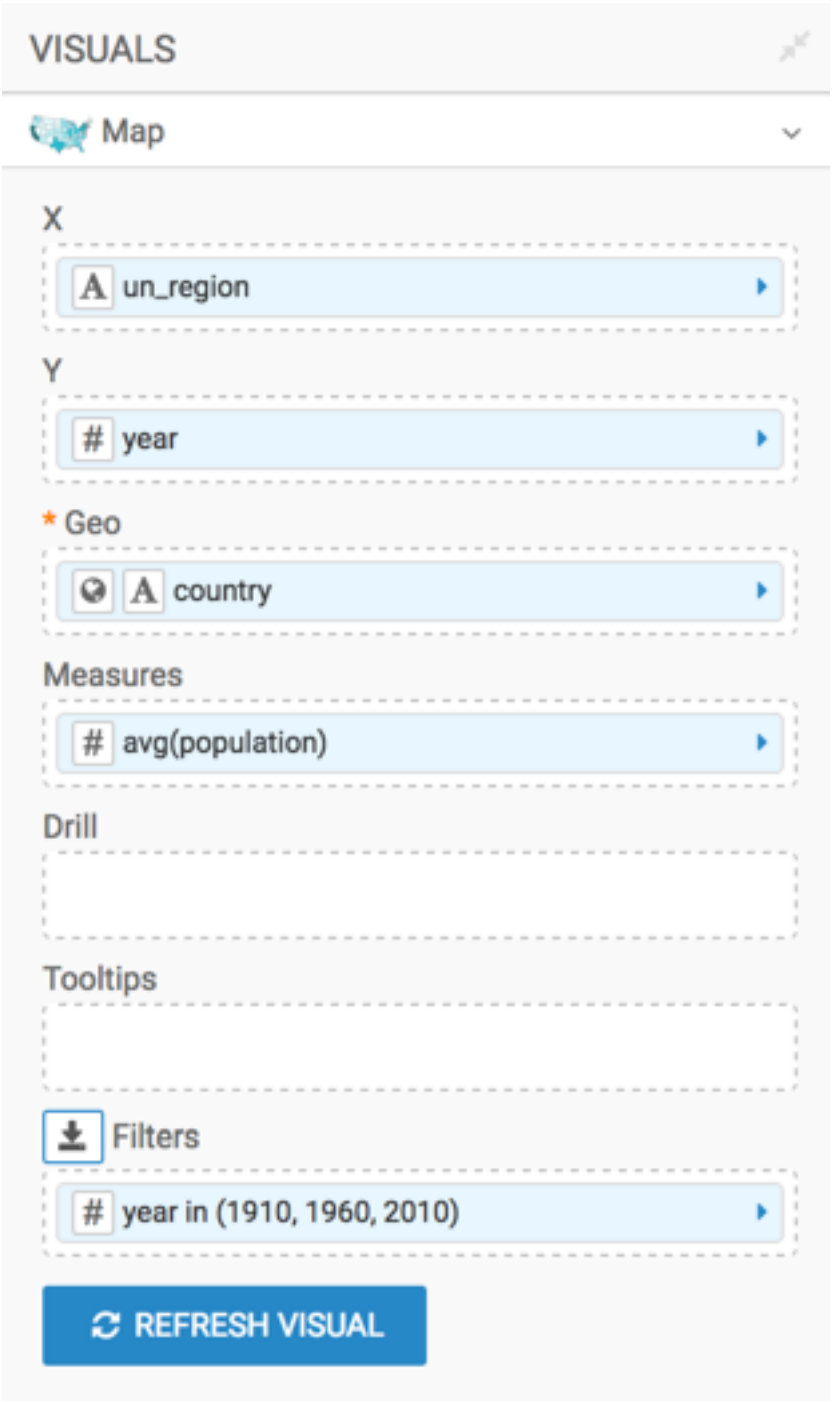
You can trellis many of the charts where the X and Y shelves are optional, by placing the field that represents the partition there.

When using mandatory shelves, such a X Axis and Y Axis on bar visuals, place two dimensions on the shelf. Cloudera Data Visualization uses the first field to trellis the visual, and the second field in the visual itself.

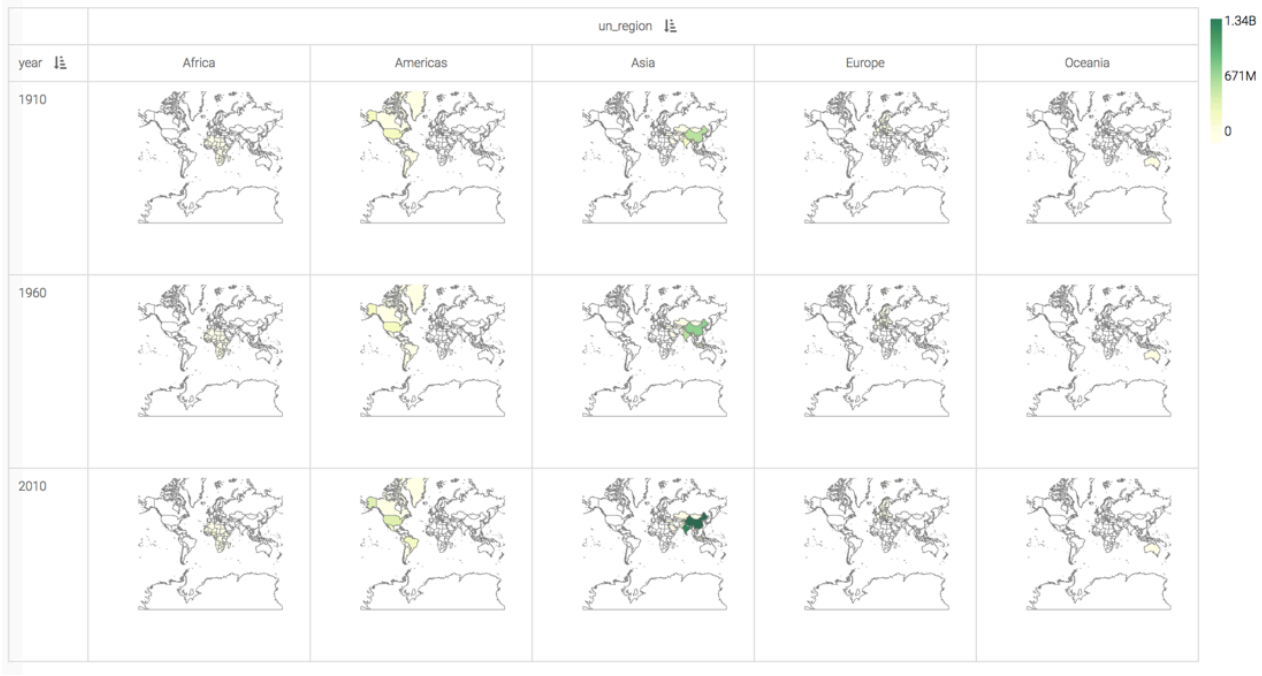
Trellising on two dimensions

You can take it one step further, and define a secondary trellis for your visuals by placing yet another dimension on the Y shelf.

Here, the year field is used both on the Y shelf and on the Filters shelf, selecting years 1910, 1960, and 2010 to demonstrate the comparable populations across continents, and in 50-year increments.



This setup produces a map visual that is trellised on two dimensions, UN Region and Year.



Trellis on measures

Some visual types (bars, lines, areas, and grouped bars) may be trellised on measures, simply by adding more fields to the measurement shelf.

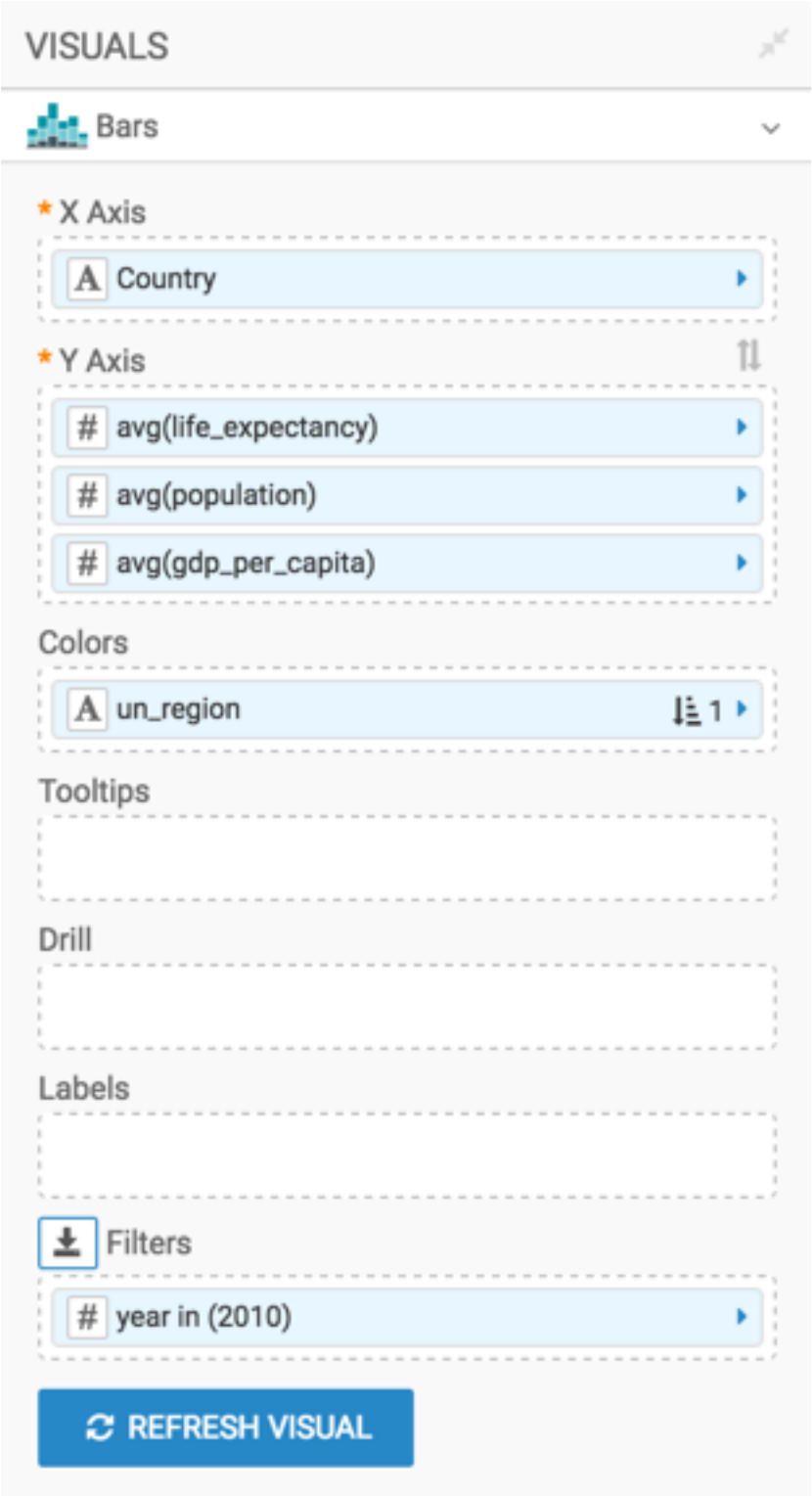
About this task

You can change the chart developed in *Bars* (Country / Life Expectancy) into a visual trellised on three measures: Life Expectancy, Population, and GDP per Capita. You can accomplish this by placing all three measurements on the Y Axis shelf.

Procedure

1. Click the Y Axis shelf. Life Expectancy is already on the shelf.
2. With the shelf active, click the population field in the DATA menu.
3. Click the gdp_per_capita field in the DATA menu.

4. Click REFRESH VISUAL.



Note: As the magnitudes of the measurements vary, Cloudera recommends that you specify independent measurements and then click REFRESH VISUAL again.

Results

This setup produces a measurement-trellised bar chart visual.

To compare this approach with plotting two measures onto the same line chart, see *Line visuals with dual axes*.

Related Information

Line visuals with dual axes

Bars

Multiple measure analysis

Cloudera Data Visualization supports measures in the trellis, for comparing multiple measures at the same time. This feature works in bar, line, area, and grouped bar visual types.

About this task

To implement this feature, simply add all the required measurement fields to the trellising shelves of the visual.

Follow these steps to create a multi-measure representation of data in the World Life Expectancy dataset.

Procedure

1. Start a new visual based on dataset World Life Expectancy [data source samples.world_life_expectancy].
   For more information, see Creating a visual.
2. In the visuals menu, choose the Lines visual type.
3. Populate the shelves from the available fields (Dimensions, Measures, and so on) in the Data menu.
   - Under Dimensions, select year and place it on the X Axis shelf. Specify ascending order, and alias the field as 'Year'.
   - Under Measures, select gdp_per_capita and place it on the Y Axis shelf.
   - Under Measures, select life_expectancy and place it on the Y Axis shelf.
   - Under Measures, select population and place it on the Y Axis shelf.
   - Under Measures, select un_region and place it on the Colors shelf. Alias the field as 'UN Region'.

4. Change the calculations and alias for each of the fields on the Y Axis using the Expression Editor:
   - Change gdp_per_capita to the following expression, and click Validate & Save.

     ```
     sum([gdp_per_capita] * [population]) / sum([population]) as 'GDP per Cap
     ita'
     ```

   - Change life_expectancy to the following expression, and click Validate & Save.

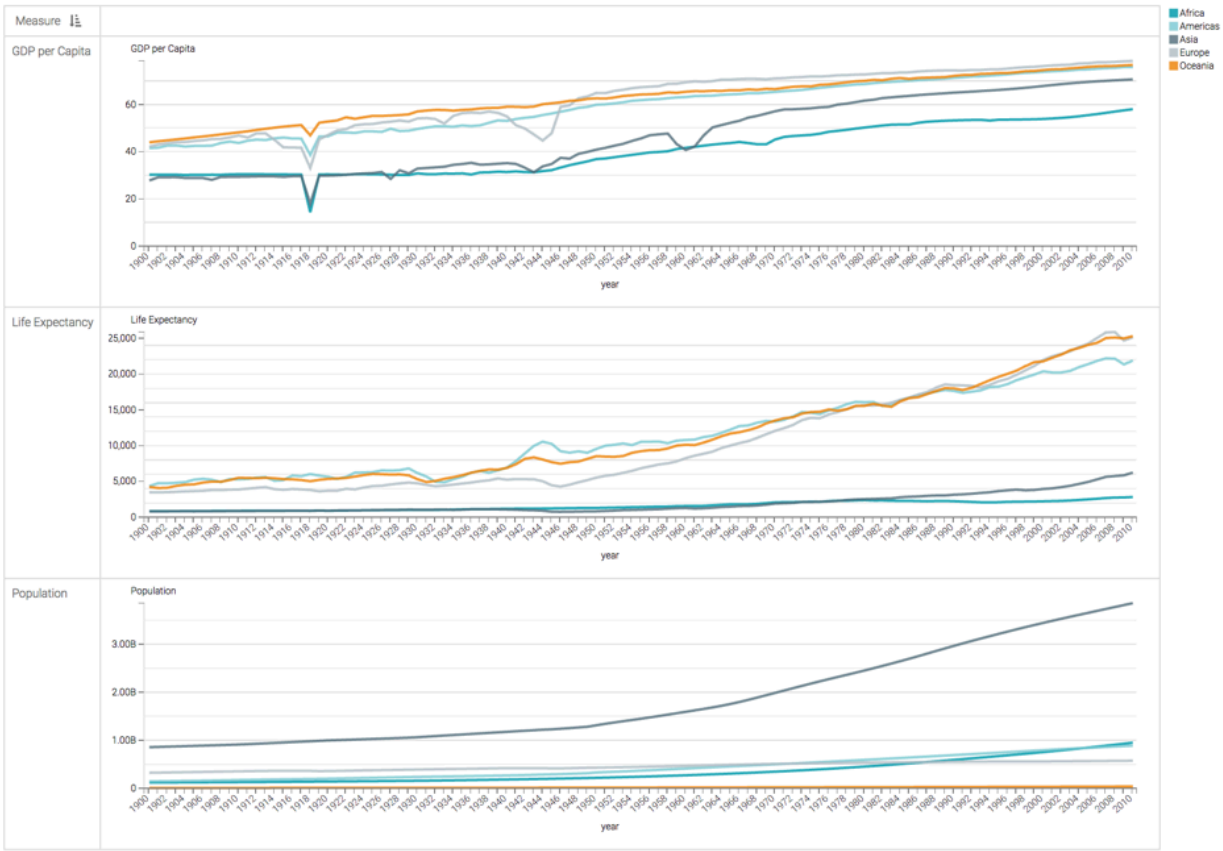
     ```
     sum([life_expectancy] * [population])/sum([population]) as 'Life Expecta
     ncy'
     ```

   - Change population to the following expression, and click Validate & Save.

     ```
     sum([population]) as 'Population'
     ```

5. In the Settings menu, open the Axes group and select the Independent Aggregate Scale option.
6. In the Settings menu, open the Axes group and select the Gridlines Enables option.
7. In the Settings menu, open the Axes group and deselect the Include Y Axis Label option.
8. Click Refresh Visual.

   The trellised line visual appears. This approach makes it easy to see the values of all measurements along the same horizontal dimension (in this case, year).



Related Information

Creating a visual

Line visuals with dual axes

It is often enlightening to see related dimensions on the same line chart; it clearly shows both correlation and pattern. However, values may be measuring completely different things, or the range of values (scale) is too different to show both lines clearly. Cloudera Data Visualization offers a dual axis option for situations like these.

About this task

The following steps demonstrate how to create a line visual that uses a double axis; it is based on dataset World Life Expectancy [data source samples.world_life_expectancy].

Procedure

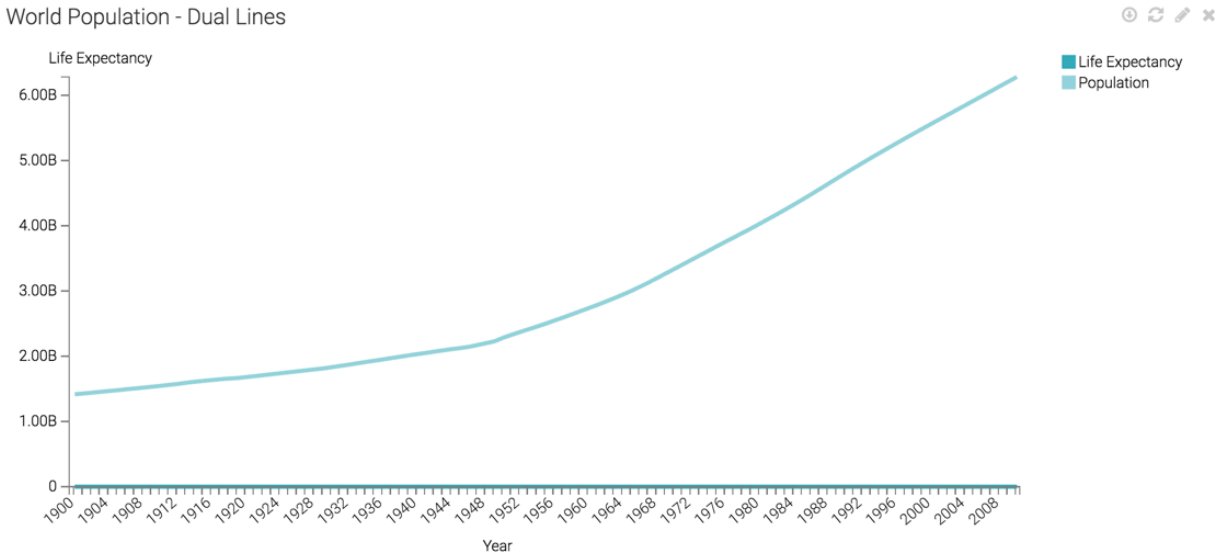
1. Start a new visual based on dataset World Life Expectancy [data source samples.world_life_expectancy].

   For more information, see Creating a visual.
2. In the visuals menu, choose the Lines visual type.
3. Populate the shelves from the available fields (Dimensions, Measures, and so on) in the Data menu.
   - Under Dimensions, select year and place it on the X Axis shelf. Specify ascending order, and alias the field as 'Year'.
   - Under Measures, select life_expectancy and place it on the Y Axis shelf. Alias it.
   - Under Measures, select Population and place it on the Colors shelf. Alias it, too.
4. Click Refresh Visual.

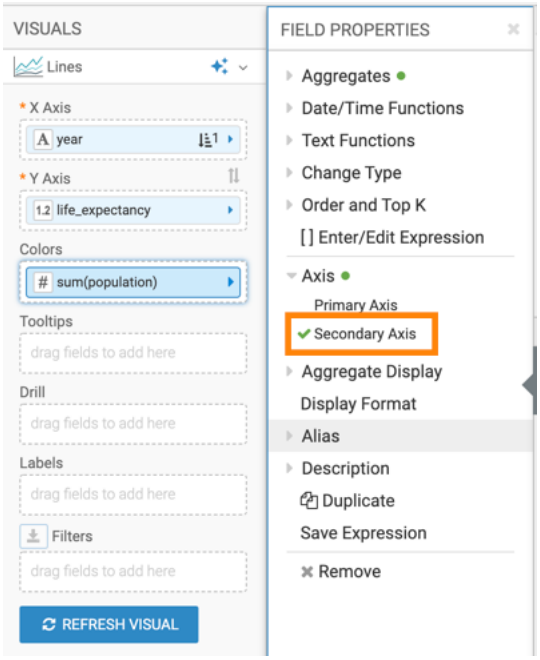
   The line visual appears.

   It is impossible to see the values on the life expectancy line because the population line sets the scale for the Y axis.

   In this case, the label does not correspond to the scale shown, because this information comes from the Y axis shelf.



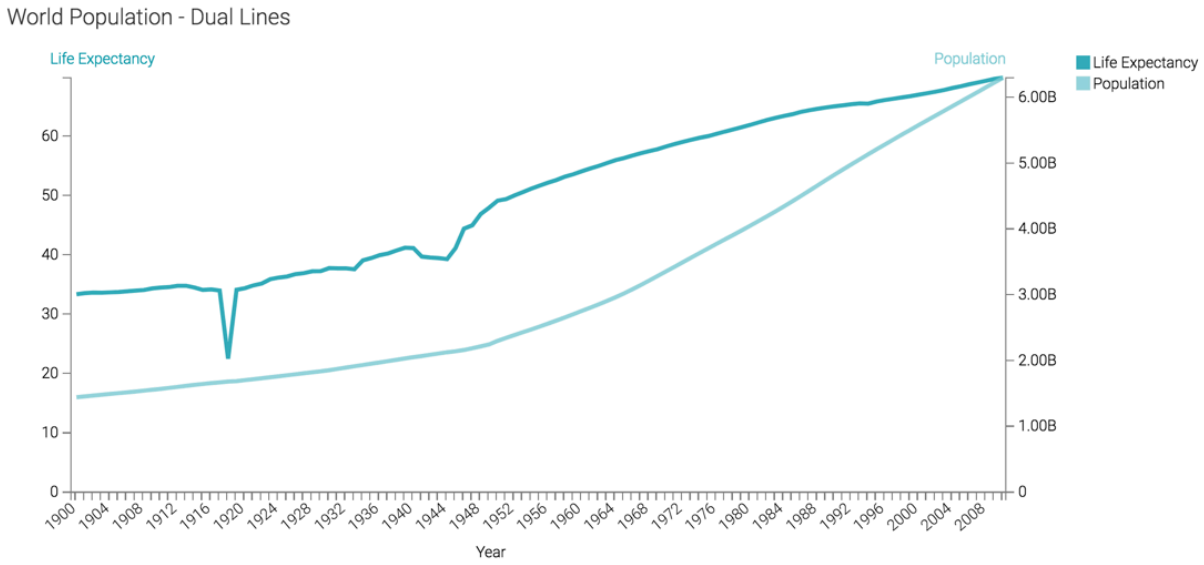
5. On the Colors shelf, on sum(population) field, click the Down arrow icon, select Axis, and then select Secondary Axis.



6. Click Refresh Visual.

   The two lines can now be seen clearly.

   There is now a vertical axis both on the left and the right side of the visual. For convenience, the color of the axis labels match the line.



7. Save the visual as World Population - Dual Lines.

   To compare this approach with plotting multiple measures as trellised charts, see Trellis on measures.

Related Information

Trellis on measures

Creating a visual

Working with dimension hierarchies

Dimension hierarchies enable you to create complete visualizations with powerful drill down capability within a single visual, instead of relying on a complex interaction of two or more visuals inside an app.

About this task

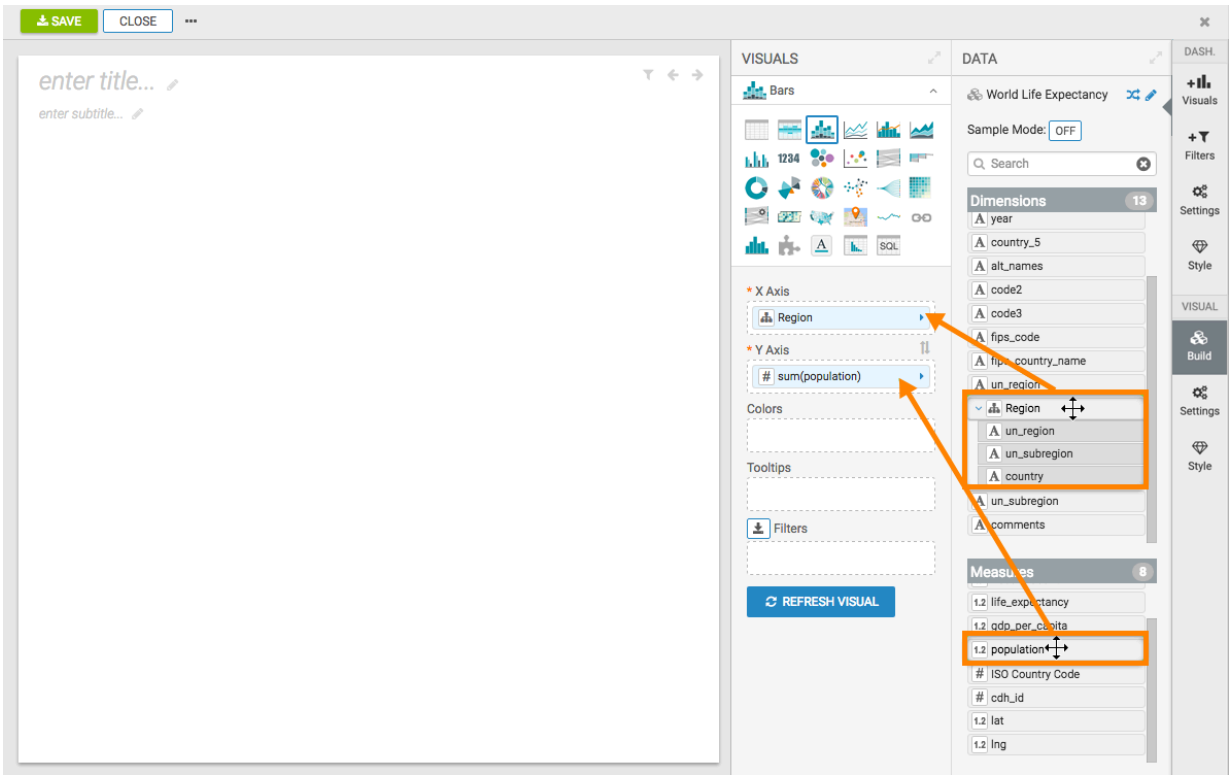
Dimension hierarchies can be used with bar, line, area, and grouped bar visual types. Cloudera Data Visualization also provides an alternate solution for hierarchical dimension representation in cross tabulation visuals. For more information, see *Working with hierarchical cross-tabulation visuals*.

The following steps demonstrate how to use Region, a dimensional hierarchy based on dataset World Life Expectancy [data source samples.world_life_expectancy], to build a bar chart that can switch between representations at three different levels of data granularity. To create this hierarchy, see Creating new dimension hierarchies.

Procedure

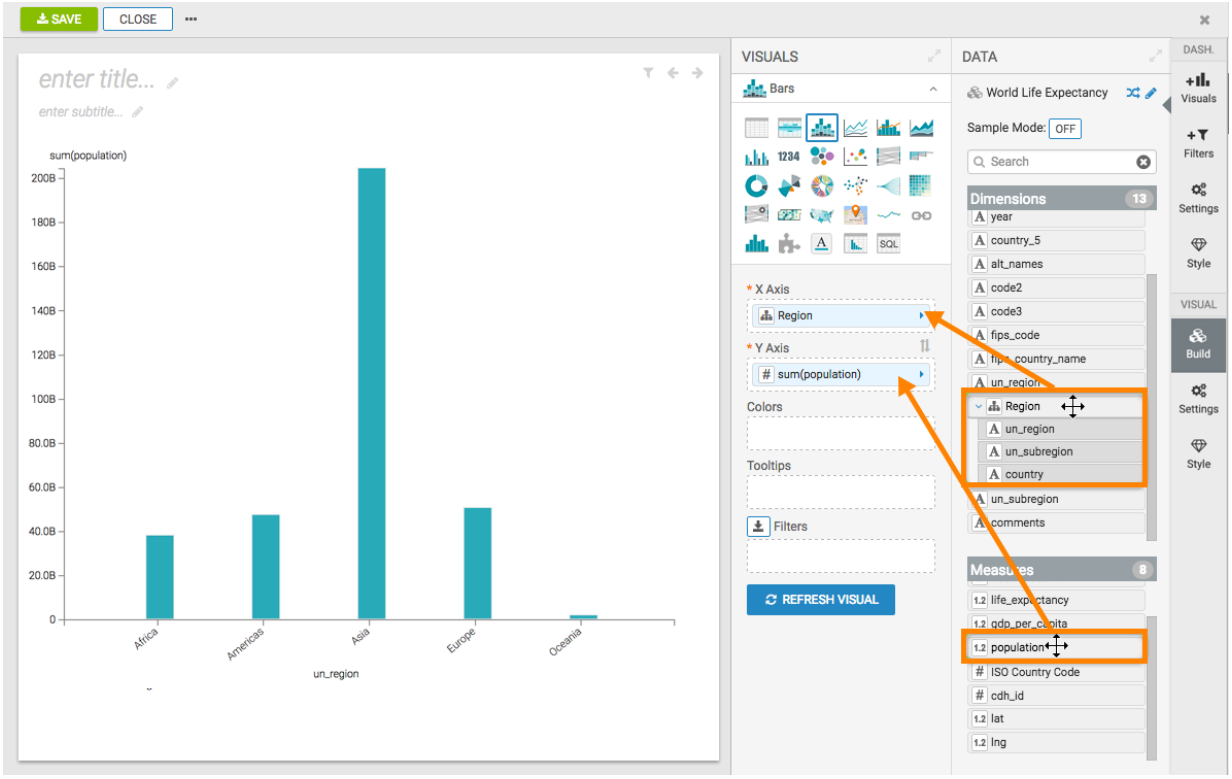
1. Start a new bar visual on the World Life Expectancy dataset.

   For more information, see *Bars*.
2. In the Visual Designer interface, populate the shelves of the visual.
   - Under Dimensions, select the field Region (which is a dimensional hierarchy) and drag it onto the X Axis shelf.
   - Under Measures, select population and drag it onto the Y Axis shelf.



3. Click REFRESH VISUAL.

   The bar chart appears. The legend along the X axis corresponds to the top level of the Region hierarchy, un_region.

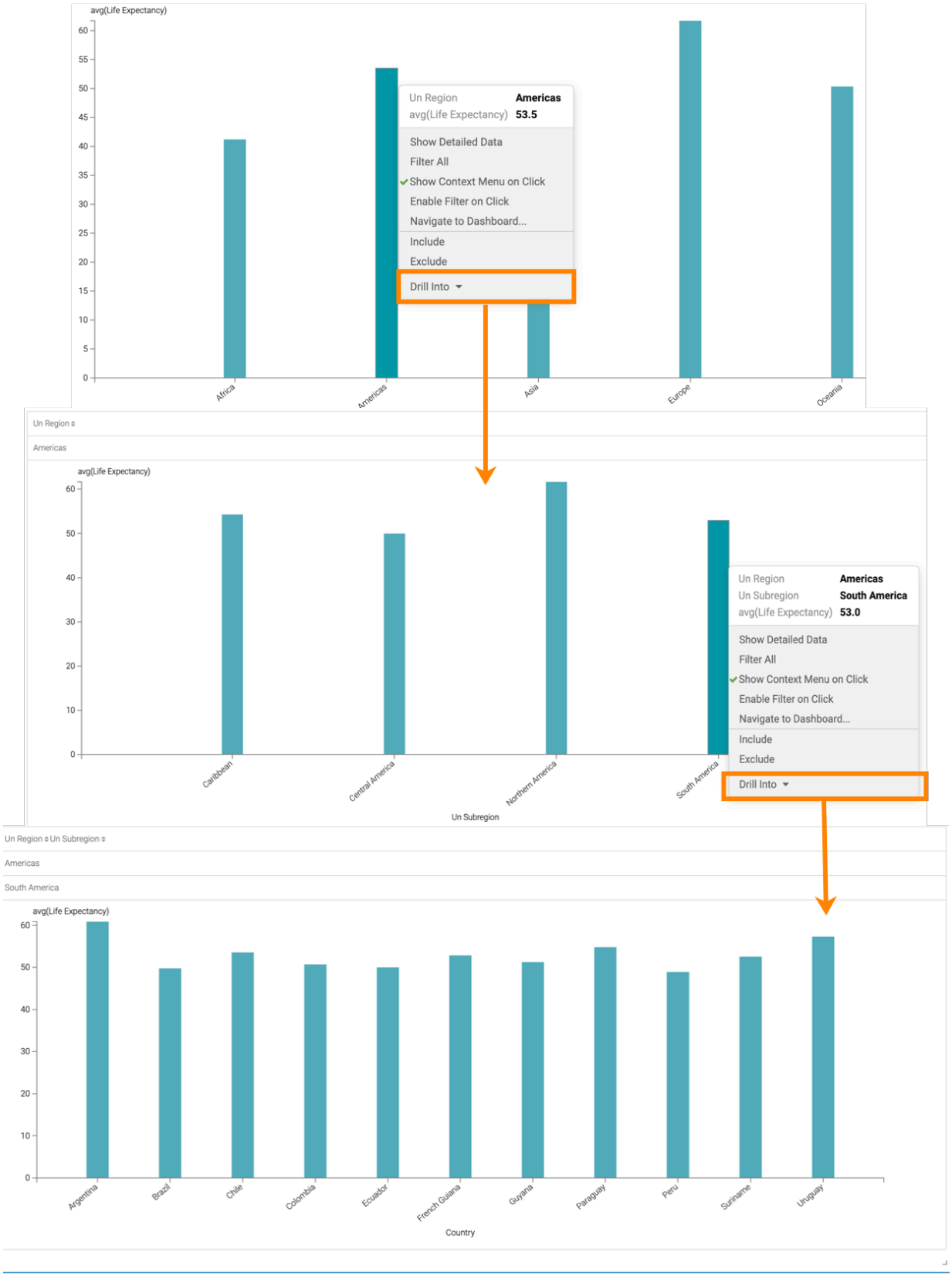


4. Change the title to World Population - Hierarchy.
5. Click Close to switch to View mode, so you can examine the operation of the dimensional hierarchy in the visual.

6. Put the pointer over one of the data labels along the horizontal axis, and note that it changes color; this is because the data labels are link-enabled. Click one of the choices to drill down to a lower level of the hierarchy.

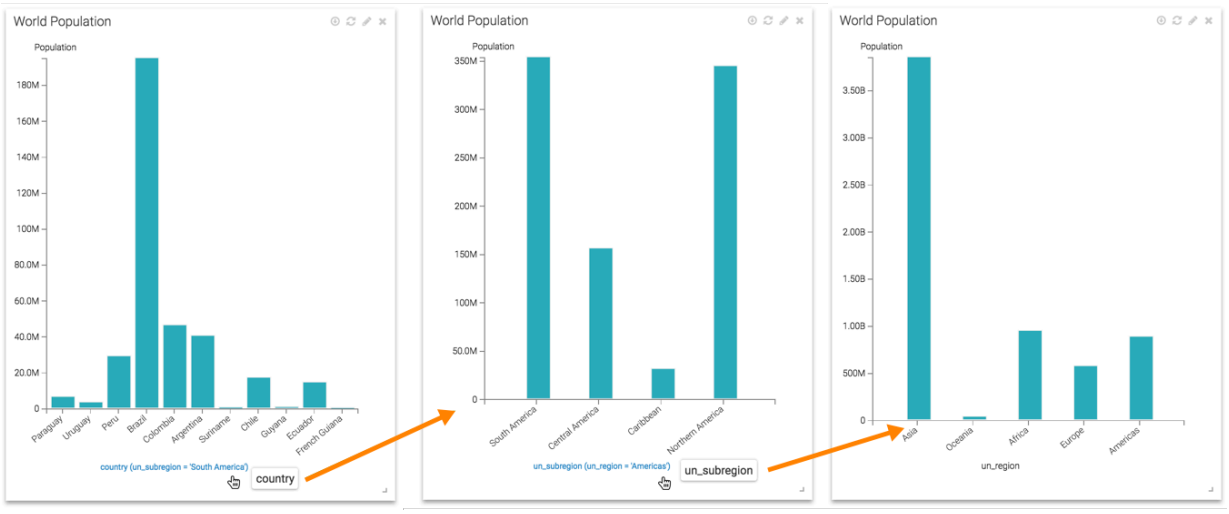
   In the following graphic, we drilled down from Level 1: un_region = 'Americas', to Level 2: un_subregion = ' South America', and then to Level 3: country.

   Every time you move between the hierarchy levels, the scale of the vertical axis adjusts automatically.


avg(Life Expectancy)
Un Region Americas
avg(Life Expectancy) 53.5
Show Detailed Data
Filter All
Show Context Menu on Click
Enable Filter on Click
Navigate to Dashboard...
Include
Exclude
Drill Into
Africa
Americas
Asia
Europe
Oceania
Un Region
Americas
Un Subregion
Un Region Americas
Un Subregion South America
avg(Life Expectancy) 53.0
Caribbean
Central America
Northern America
South America
Un Region Un Subregion
South America
Country
Argentina
Brazil
Chile
Colombia
Ecuador
French Guiana
Guyana
Paraguay
Peru
Suriname
Uruguay

7. Put the pointer over the label of the horizontal axis, and note that it change color; it is also link-enabled. You can click the horizontal axis label to go up the hierarchy levels.

   In the following graphic, we moved from Level 3: country where un_region = 'America' and un_subregion = 'South America' to Level 2: un_subregion = 'South America' where un_region = 'America', and then to Level 1: un_region.



Related Information

Hierarchical cross tabulation visuals

Creating new dimension hierarchies

Bars

Hierarchical cross tabulation visuals

Cloudera Data Visualization has an alternate solution for hierarchical dimension representation in cross tabulation visuals.

About this task

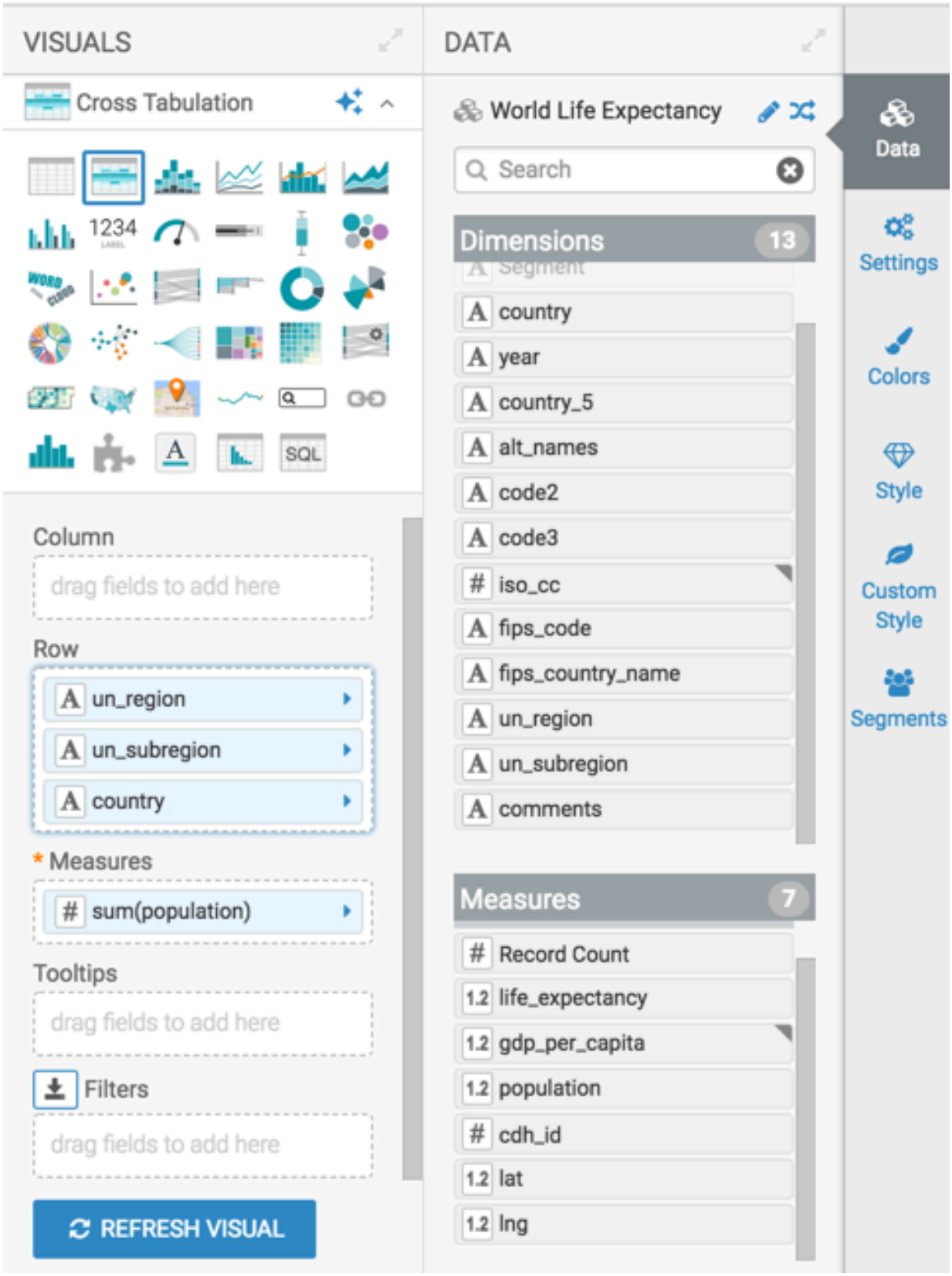
The following steps demonstrate how to create and configure a de facto dimensional hierarchy on the shelves of a cross tabulation visual. We are using dataset World Life Expectancy [data source samples.world_life_expectancy], to build a cross tabulation table that can switch between representations at three different levels of data granularity.

Procedure

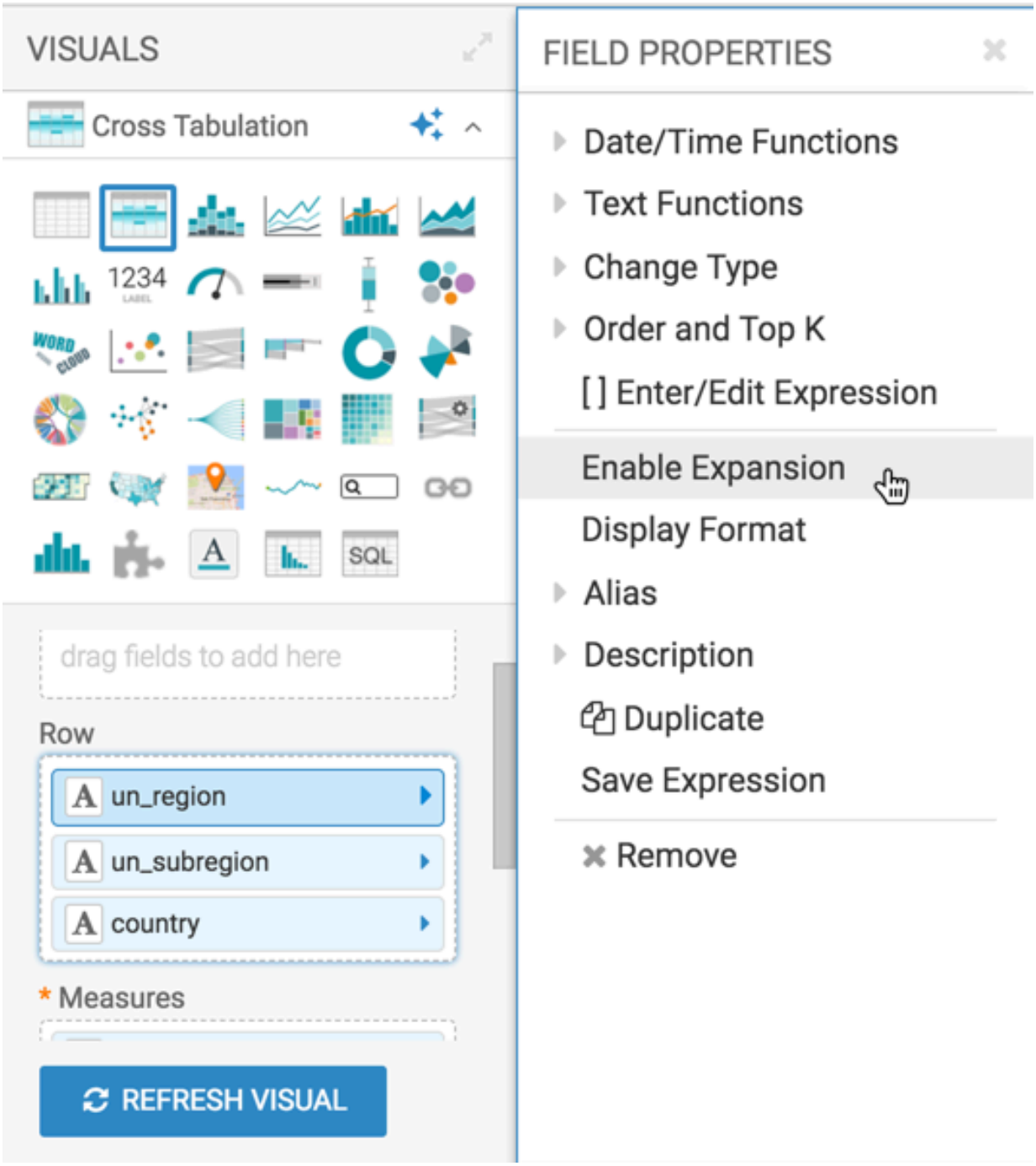
1. Start a new cross tabulation visual on the World Life Expectancy dataset.

   For more information, see *Cross tabulation*.

2. Populate the shelves from the available Fields under Dimensions and Measures in the Data menu.
   - Under Dimensions, select un_region and drag it over the Dimensions shelf. Repeat with un_subregion and country, adding them under un_region. Note that your placement of these fields is from the largest to the smallest, top to bottom.
   - Under Measures, select population and drag it over Measures shelf. Ensure that the default aggregation is sum(population).


VISUALS
DATA
Cross Tabulation
World Life Expectancy
Data
Search
Settings
Dimensions 13
Colors
country
year
country_5
alt_names
code2
code3
iso_cc
fips_code
fips_country_name
un_region
un_subregion
comments
Style
Custom Style
Segments
Column
drag fields to add here
Row
un_region
un_subregion
country
* Measures
sum(population)
Tooltips
drag fields to add here
Filters
drag fields to add here
REFRESH VISUAL
Measures 7
Record Count
life_expectancy
gdp_per_capita
population
cdh_id
lat
lng

3. On the Row shelf, select the un_region field and select Enable Expansion option in the Field Properties panel.



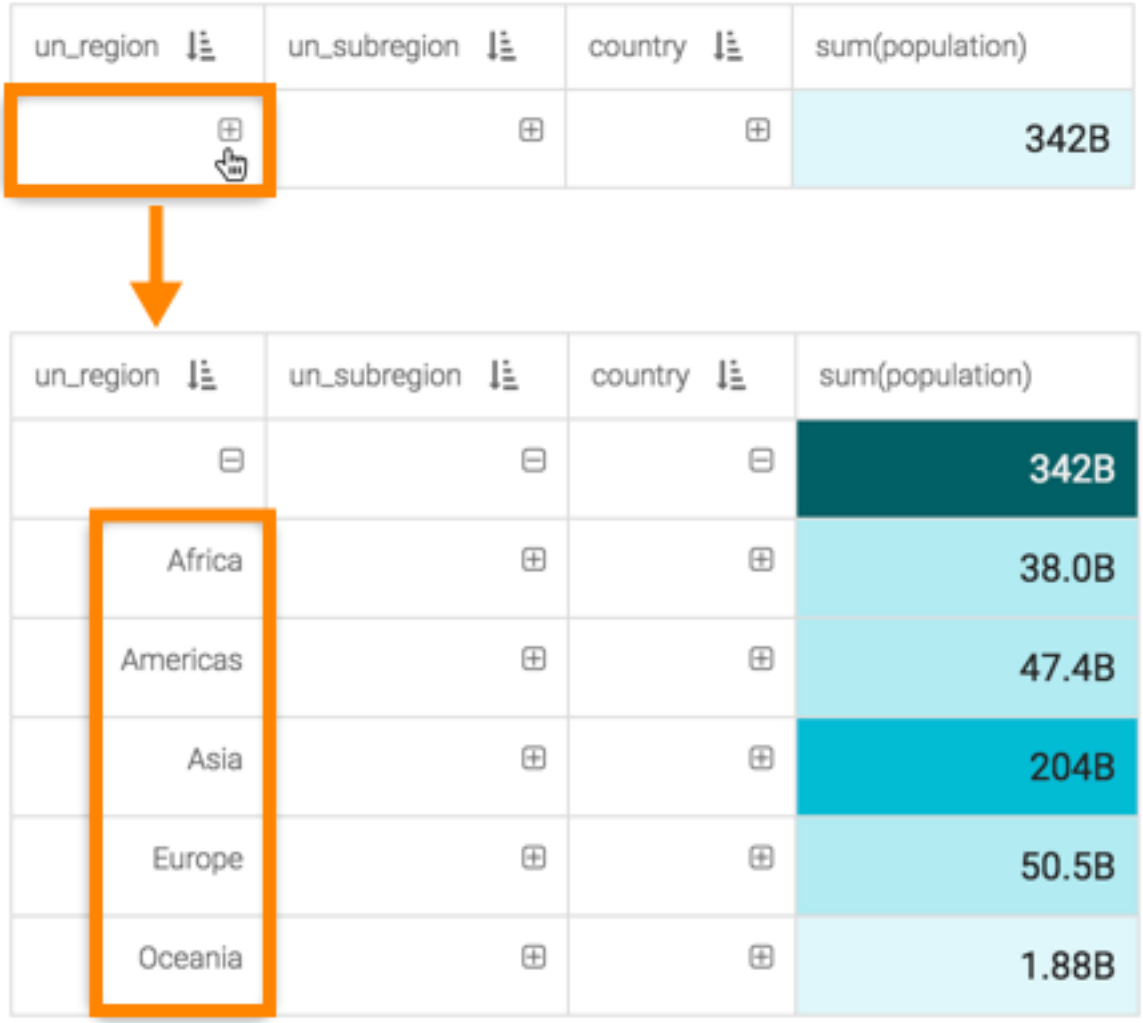
4. On the Row shelf, on the un_subregion and the country field, click the Down arrow icon, and select Enable Expansion.

5. Click REFRESH VISUAL.

   The cross tabulation table appears with a Plus icon in the un_region, un_subregion, and country columns. This indicates that you can expand the data in these columns.

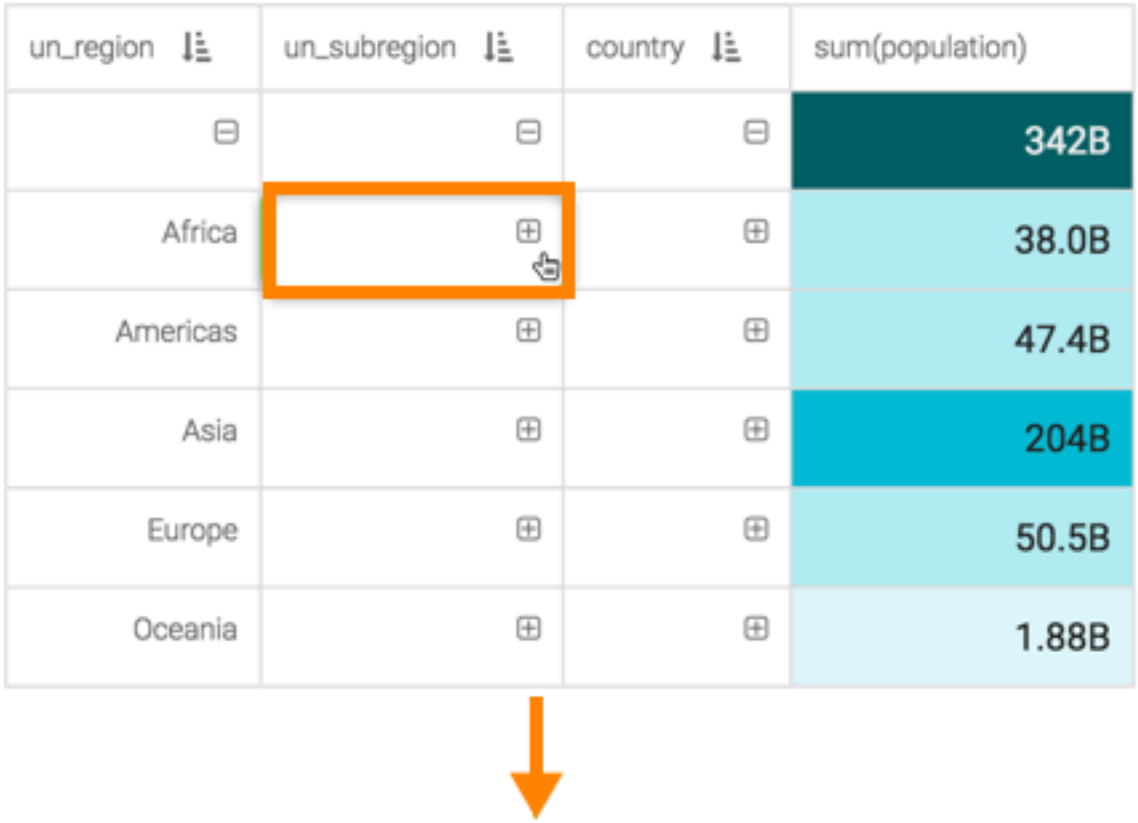
   The sum(population) column shows the combined population for all un_regions.

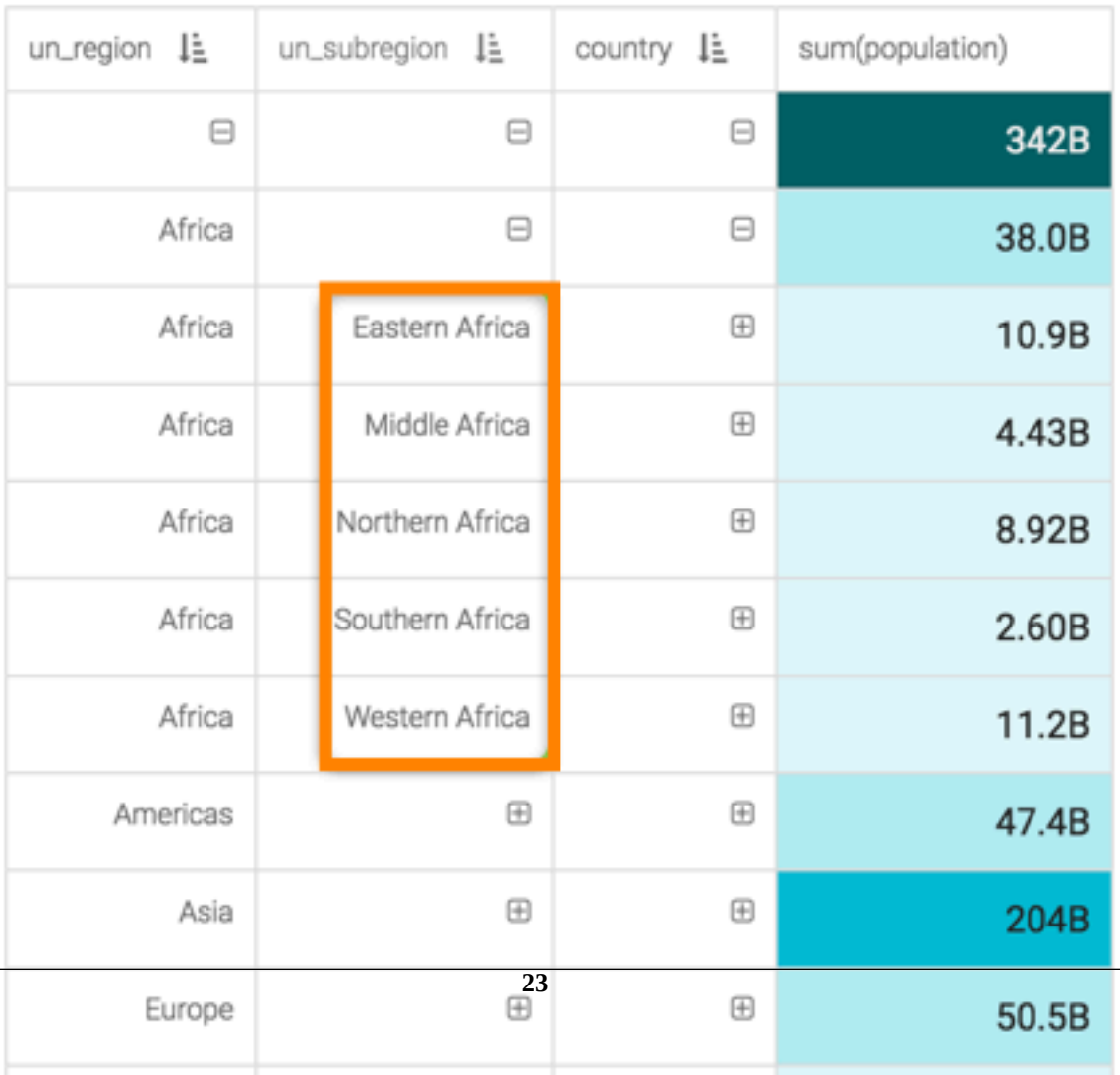
6. To see more details, expand the un_region column of the visual by clicking the Plus icon under the column header.



7. Expand the un_subregion column of the visual by clicking the Plus icon in the field next to one of the un_region. We chose to expand Africa.

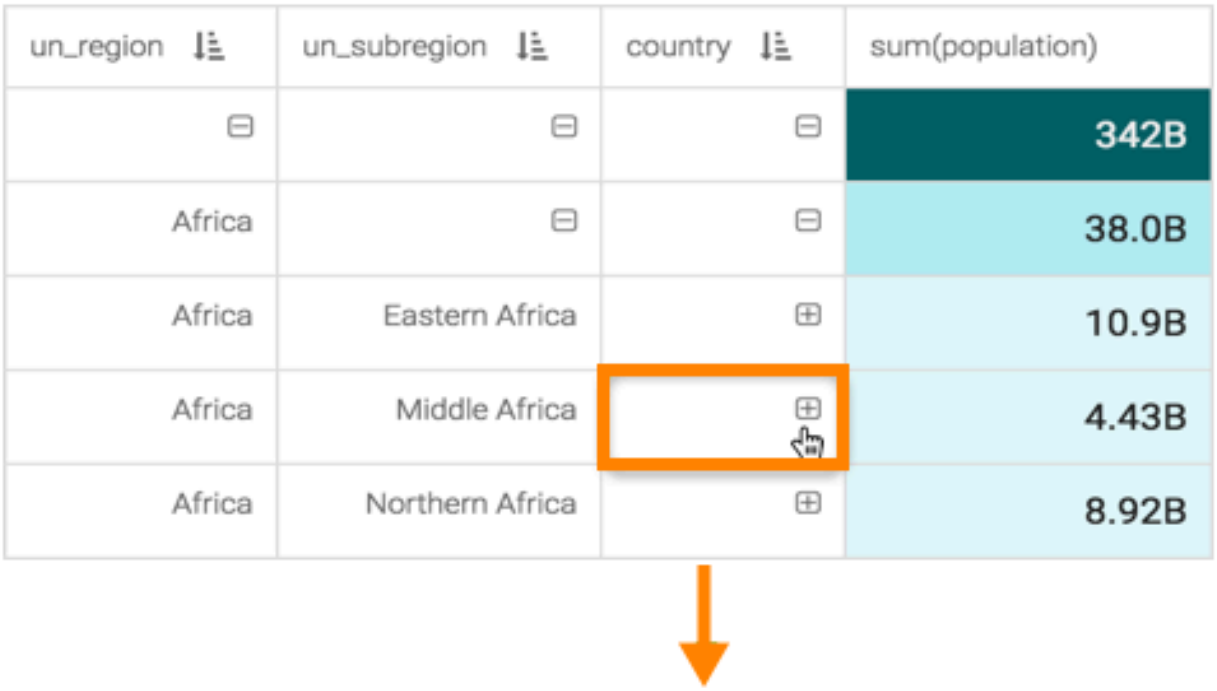
   When you expand Africa, it generates a complete tabulation aggregated by un_subregion.


un_region
un_subregion
country
sum(population)
342B
Africa
38.0B
Americas
47.4B
Asia
204B
Europe
50.5B
Oceania
1.88B

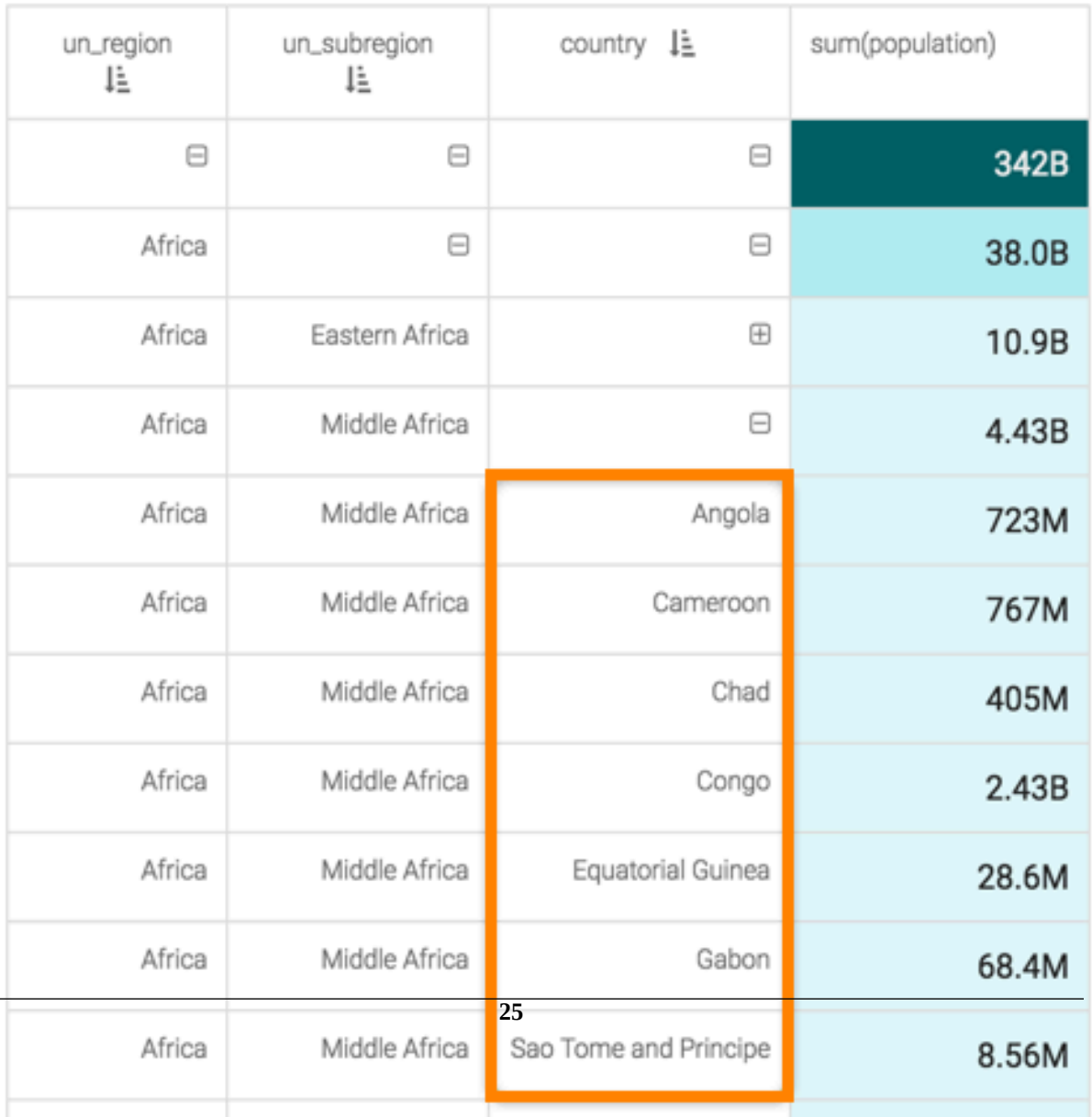

un_region
un_subregion
country
sum(population)
342B
Africa
38.0B
Africa
Eastern Africa
10.9B
Africa
Middle Africa
4.43B
Africa
Northern Africa
8.92B
Africa
Southern Africa
2.60B
Africa
Western Africa
11.2B
Americas
47.4B
Asia
204B
Europe
50.5B

8. Expand the country column of the visual by clicking the Plus icon in the field next to one of the un_subregions. We chose to expand Middle Africa.

   When you expand Middle Africa, it generates a complete tabulation aggregated by country.



un_region	un_subregion	country	sum(population)
			342B
Africa			38.0B
Africa	Eastern Africa		10.9B
Africa	Middle Africa		4.43B
Africa	Northern Africa		8.92B



un_region	un_subregion	country	sum(population)
			342B
Africa			38.0B
Africa	Eastern Africa		10.9B
Africa	Middle Africa		4.43B
Africa	Middle Africa	Angola	723M
Africa	Middle Africa	Cameroon	767M
Africa	Middle Africa	Chad	405M
Africa	Middle Africa	Congo	2.43B
Africa	Middle Africa	Equatorial Guinea	28.6M
Africa	Middle Africa	Gabon	68.4M
Africa	Middle Africa	Sao Tome and Principe	8.56M

9. Change the title to Hierarchical Cross Tabulation of Population.
10. At the top left corner of the Visual Designer, click SAVE.

Related Information

Cross tabulation

Adding highlights to visuals

Highlights are a practical way to bring attention to values that exceed a particular threshold. Cloudera Data Visualization currently supports highlight marks at specific values on the combined bar/line chart type, and enables you to customize these highlights.

About this task

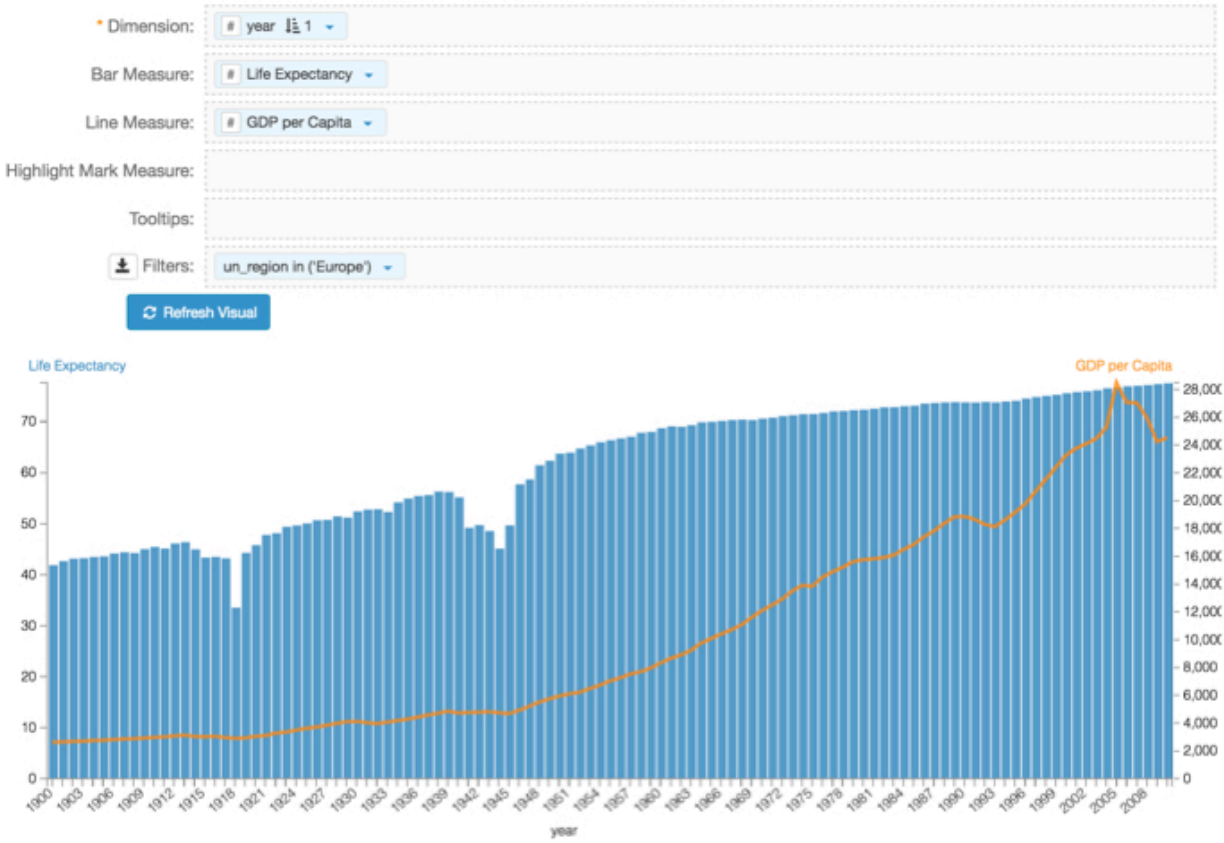
The following steps demonstrate how add highlights to an existing chart. To start, we are using the visual World Population - Dual Bar and Line, developed on dataset World Life Expectancy [data source samples.world_life_expectancy]. For more information, see *Combined bar/line visuals*.

Procedure

1. Open the visual World Population - Dual Bar and Line in Edit mode.
2. Clone the visual.
3. Click the Pencil icon next to the title of the visualization to edit it, and enter the new name.
4. Click the Pencil icon below the title of the visualization to add a brief description of the visual.
5. Change the title to World Population - Highlights on Dual Bar and Line.
6. At the top left corner of the Visual Designer, click Save.
7. Add the un_region field to the Filters shelf, and select Europe from the list of possible values.

   This graph has more variation than the unified World data.

8. Click Refresh Visual.



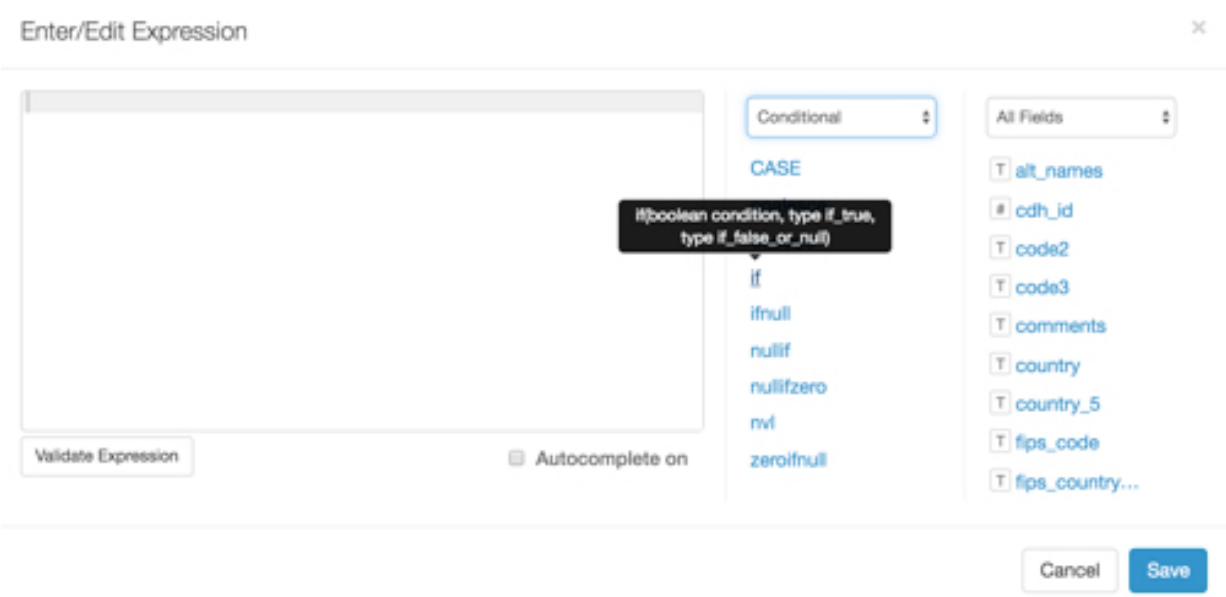
You can see the pronounced peak in the GDP per Capita line, centered around the year 2005.

9. Based on the graph, highlight the values where GDP per Capita is over $25,000.
10. Place the field gdp_per_capita on the shelf Highlight Mark Measure.
11. On the same shelf, click the Down arrow icon on the gdp_per_capita field, and choose Enter/Edit Expression.

12. In the Enter/Edit Expression window modal, delete the text in the default expression, change the function groupings to Conditional, and select the if statement.

Note:

Depending on the SQL dialect of the connection you are using, the syntax of the if statement may be different.



13. Complete the expression template.

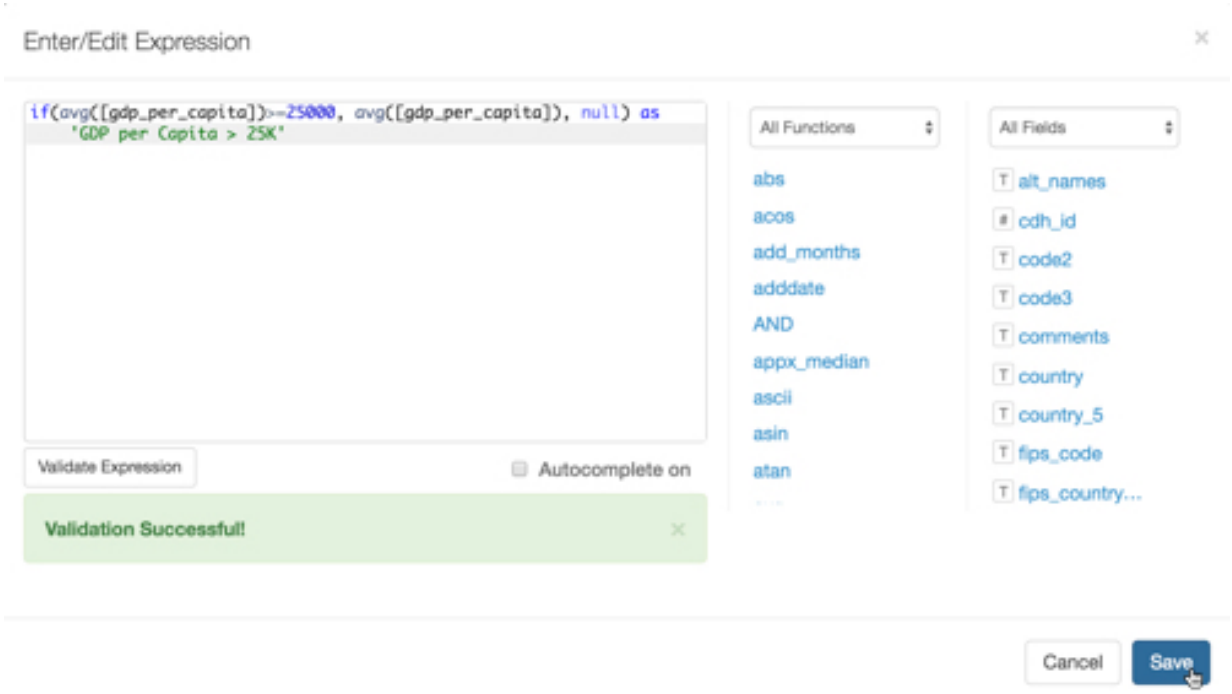
It should look similar to the following expression:

```
if(avg([gdp_per_capita])>=25000, avg([gdp_per_capita]), null)
```

Optionally, you can alias the highlight; see the bold font in the changed expression.

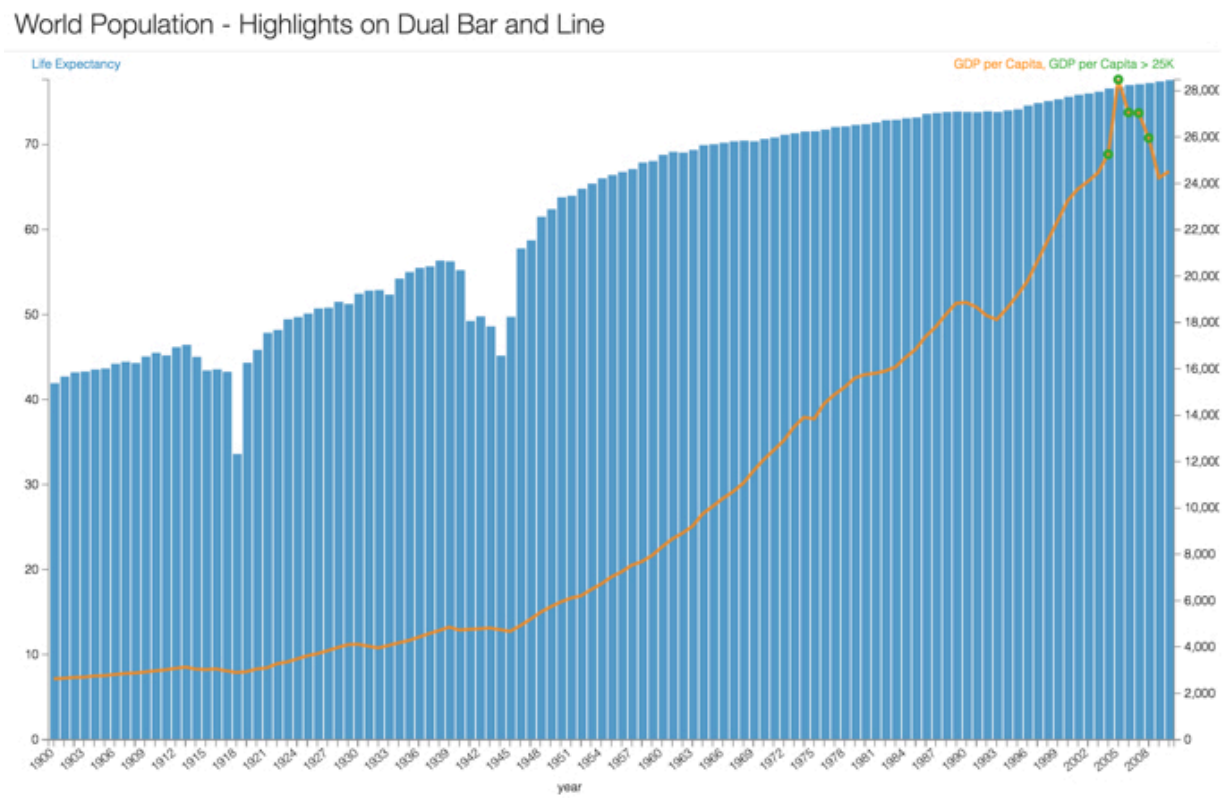
```
if(avg([gdp_per_capita])>=25000, avg([gdp_per_capita]), null)
            as 'GDP per Capita > 25K'
```

14. Click Save.



15. Save the visual, and then click View to change to View mode.

Note: The graph has green circle highlights when GDP per Capita values are equal or greater than 25,000.



16. The default highlight mark is a circle. To convert it to another type, use the visual Settings interface as described in *Change highlight type*. You can also change the relative size of the mark as described in Change highlight type.

Related Information

Combo visuals