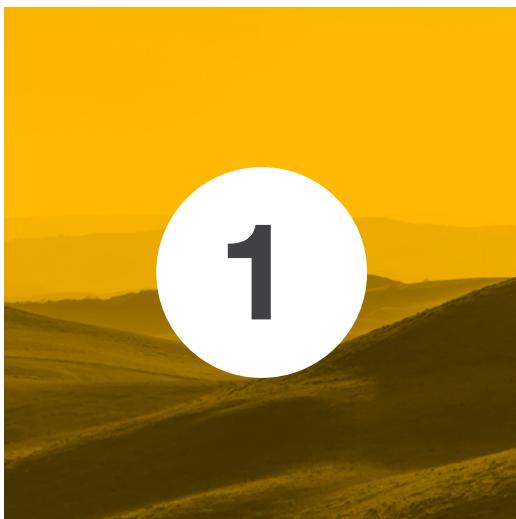


The Data Storytelling Handbook



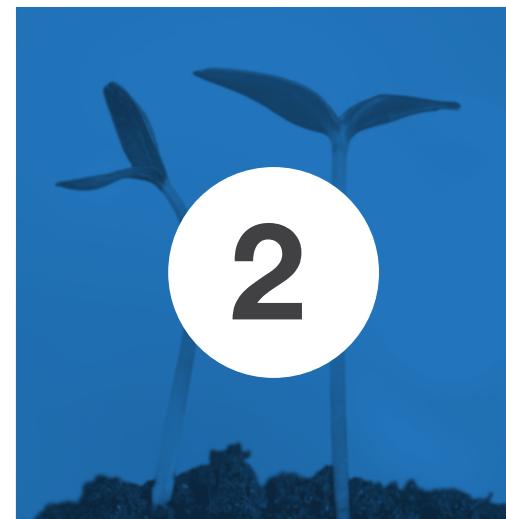
Table of Contents

Introduction page 3



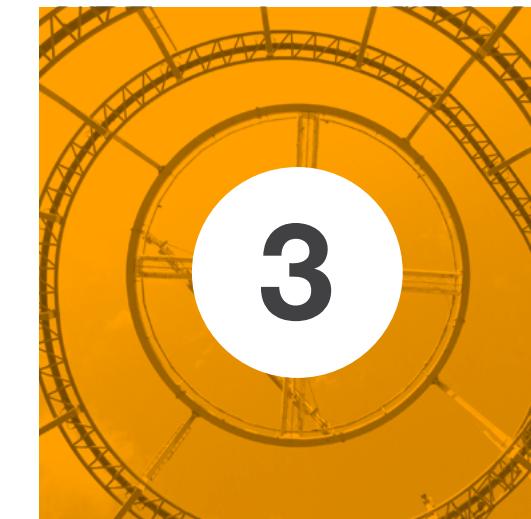
The Art of Data Storytelling

Page 4



Know Your Purpose
and Data

Page 9



Craft Your Message

Page 14



Storytelling Assets

Page 18



Protips

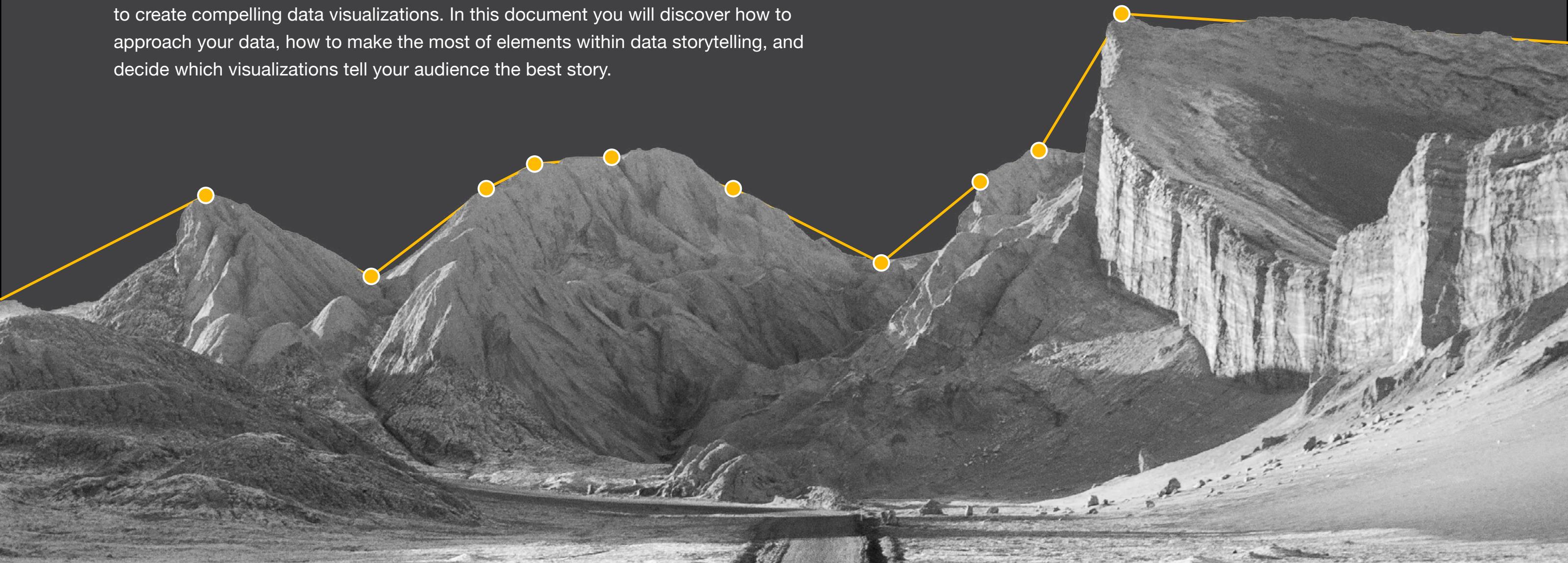
Page 46

Outro page 48

Welcome to **your story.**

Every story is unique. It's an opportunity to inspire, advise or enlighten. We want you to discover the story hidden within your data, so you can become the author that ignites imaginations and turns information into knowledge.

You will be taken beyond your spreadsheets and presentations, and taught how to create compelling data visualizations. In this document you will discover how to approach your data, how to make the most of elements within data storytelling, and decide which visualizations tell your audience the best story.





The Art of Data Storytelling

Why Visualize

Today we receive 5X as much information as we did in 1986. This means how you share your story will drastically determine the size of your audience. Researchers have found that colour visuals increase the willingness to read by 80% and that we get the sense of a visual scene in less than 1/10 of a second. Data visualization makes it easy to recognize patterns and find exceptions while interpreting the data at a faster pace. It allows access to challenging data sets, it allows exploration, can be fun and provides useful information in an efficient way.



THE AMOUNT OF INFORMATION WE RECEIVE COMPARED TO 1985



THE AMOUNT OF TIME IT TAKES TO GET THE SENSE OF A COLOURED VISUAL



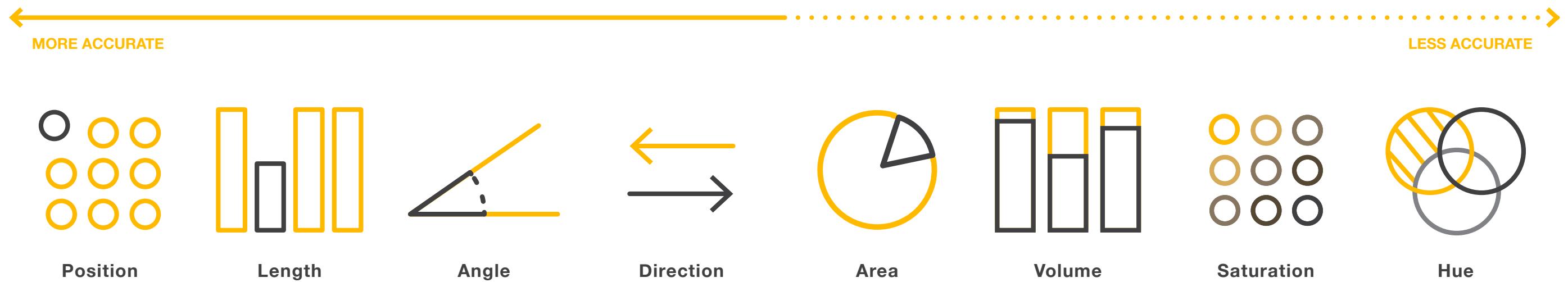
OF THE WORLD'S DATA HAS BEEN CREATED IN THE PAST 2 YEARS

In 2015's IBM Smarter Planet Report it says that, "90% of the data in the world today has been created in the last two years alone." The majority of that data is visual and most people don't know how to present it. The opportunity lies in becoming better visual storytellers and utilizing the data to illuminate the message.

"The more you leave out, the more you highlight what you leave in."

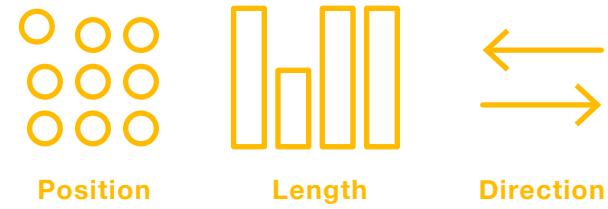
Visual Perception

People are more inclined to perceive certain visual cues (variables) better than non-visual cues.



Visual Perception

Some visual cues, for example:



are better at supporting:



Selection or Grouping

Do changes in the visual cue allow you to distinguish a point from others? Do changes in the cue allow you to group data points?



Measurement

Can you make a numeric observation from a change?



Ordering

Does the visual cue have a perceived order?



Steps

How many distinct “steps” can be perceived in the cue?

Visual Perception

Most quantitative analysis can be performed with charts that use only **four kinds** of objects.

These objects (and their subsequent related charts) work because we immediately and more precisely perceive both position and length.



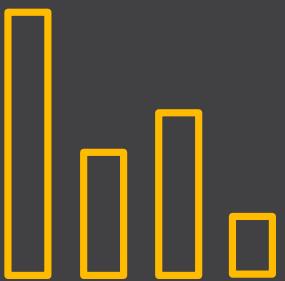
Points

2D position,
example: scatter plot



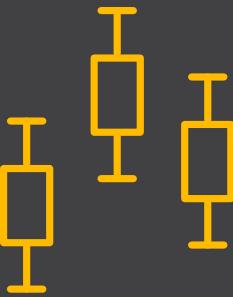
Lines

2D position,
example: line chart



Bars/Columns

2D position + length,
example: bar chart



Boxes

2D position + length (unlike bars, show distribution of an entire set of values)
example: bar chart



2

7.2

3.7

2.1

Know Your Purpose and Data

Know Your Purpose

Before you can begin to create stunning visualizations, you will need to make sense of your data by finding the story that speaks to your audience. Use the data to illuminate that story and the message you are trying to share, and you'll make it unforgettable.



Know Your Data

With SAP Lumira you can measure:

It is important to model your data appropriately, before you explore it, in order to be able to answer your business questions correctly. Data types can be used to model certain characteristics of your data.

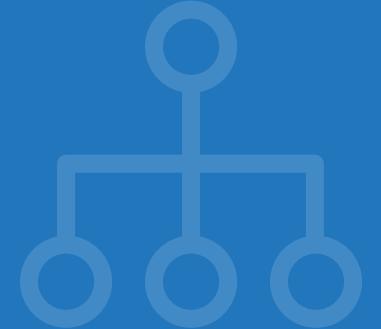
Numerical Data



1 2

AND

Dimensions (Categorical)



Measures

Measures constitute numerical data that are calculated or aggregated – like the sum of Revenue, average Cost, or Profit-per-capita or non-numeric data that are counted.



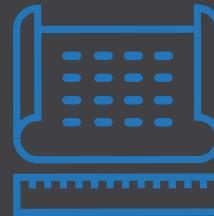
Measures are objects that represent calculations and aggregate functions that are usually applied to numeric data. Aggregating the object must make sense for the column to be a measure.

Sales Revenue is a measure but summing up product list prices isn't. That's a dimension. You can create measures from categories by counting their elements, for example, Number of Countries visited by our Customers.



What do measures represent?

Measures can represent observations in your data or calculated values.



How are they formatted?

Measures have an aggregation type associated with them. By default, SAP Lumira sets this type to sum. For example, if the chart includes Revenue by Country, and sum is associated with Revenue, SAP Lumira allows you to customize the prefix or suffix to indicate data such as, units of measures, like CAD, EUR, and USD.

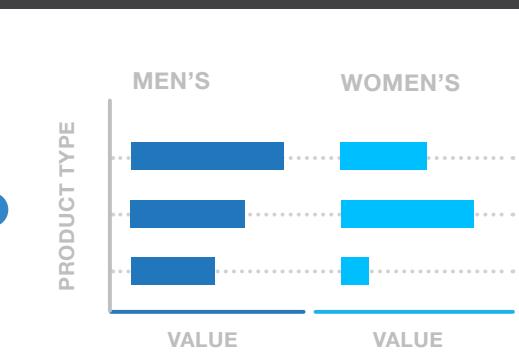
Dimensions

Dimensions constitute categorical data such as **year, product, country and salary range**.

What do they represent?

Categorical

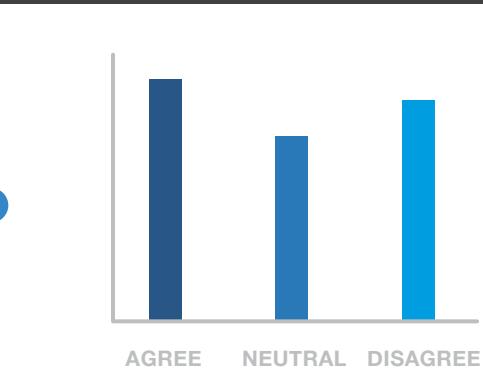
(Also called “nominal”) for discrete values.



The dimension Product Type may include the values Men's Clothing and Women's Clothing.

Ordinal

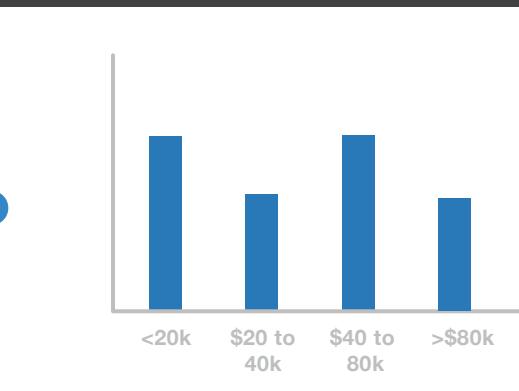
The dimension members have a set default order.



A dimension reflecting the outcome of a survey result may include the values Agree, Neutral, Disagree that have an implicit order.

Interval

Each value in the dimension represents a range of values.



The dimension Salary can be categorized into the following salary ranges: <\$20k, \$20 - 40k, \$40 - 80k, >\$80k



Craft Your Message

Craft Your Message

By exploring your data you now have a better sense of what story to tell your audience. It is time to craft that message and discover which viz best articulates your information.

Keep these questions in mind:



1.

What is your overall goal of your data analysis?



2.

Who is this message intended for? What do you know about your audience?

Then ask yourself:

1. What questions do you want to answer with your data?
2. What kind of relationships exist in your data? What are the best techniques for displaying these? Do you need a chart (overview), a table (details), or maybe both, to convey your message?
3. Can you highlight specific data points to better get your message across?
4. How can you incorporate a summary of your message in your chart titles to emphasize on your overall message?

Know Your Audience



Get to know your audience then use precognitive attributes to create great data visualizations that resonate with them. Precognitive attributes mean the image is being processed without any conscious effort. Communicating in this way means there is no need for explanation on top of the visualizations. It is also important to note that just because you have good visualizations that doesn't necessarily mean you have a good visual story. Reward your audience with the experience and knowledge that led them to your story in the first place.

Every piece needs to be pulled together to create a cohesive story with a beginning, middle, and end. Entertain them.

“It is important to know your audience’s background and the domain of your data.”



Think about the story you want to tell with your data. What insights do you want to bring to light? Keeping your audience in mind, you do not want to include any unnecessary noise to obscure the meaning of your message. It is easy to misrepresent data by choosing the wrong visualization type.

Every little detail matters in the telling of your story and connecting with your audience.



A large white circle containing the number 4 is positioned in the center of the map. The map shows a coastal area with several land parcels outlined in white. The parcels are labeled with their areas in green text: 5.3, 4.1, 4.8, 24.9, 18.1, 8.4, and 4.0.

4

Storytelling Assets



“Don’t forget – no one else sees the world the way you do, so no one else can tell the stories that you have to tell.”

Because every story is unique what you use to tell it should be unique and specific to your story. Choose your tools carefully they will be the plate your knowledge is served on. Here's how to select the right chart type based on the goal of your message.

Selecting the Right Visualizations

Change Over Time

Shows how a measure changes over time, and allows the user to highlight temporal trends



Line Chart: Highlights potential trends in data



Bar Chart: Highlights comparison between individual values

Distribution

Shows how a measure is spread across its domain



Histogram: Column Chart showing the count of binned measure values



Box Plot: shows distributions for different categorical values



Heat and Tree Map: shows the distribution of measure values

Comparison

Shows the comparison of categorical values, where the data does not have any intrinsic order, for example, a list of products



Bar Chart: used for comparing categorical values



Trellis: uses multiple views to show different partitions of a dataset

Ranking

Shows the top or bottom N values to emphasize the largest, or smallest values



Bar Chart: shows categorical values in decreasing or increasing order

Part-To-Whole

Shows how the categories contribute to the whole value



Bar Chart: set to % scale



Pie Chart: Compares percentage values



Stacked Bar Chart: shows overall measure total

Correlation

Shows, whether there is a potential correlation between two measures



Scatter Plot: highlights potential correlation of two measures



Trellis: uses multiple views to show different partitions of a dataset

Overview

Shows the exact values in table format



Table: highlights exact values

Geographical Information and Maps

Shows the geographical distribution of measure values



Choropleth Chart: highlights geographical data by colouring geographical areas according to their measure values



Geo Bubble Chart: highlights geographical data by showing them as bubbles on a map

Change Over Time

Shows how a measure changes over time, and allows the user to highlight temporal trends



Focus Areas



Line and Area Chart



Bar Chart

Line and Area Chart

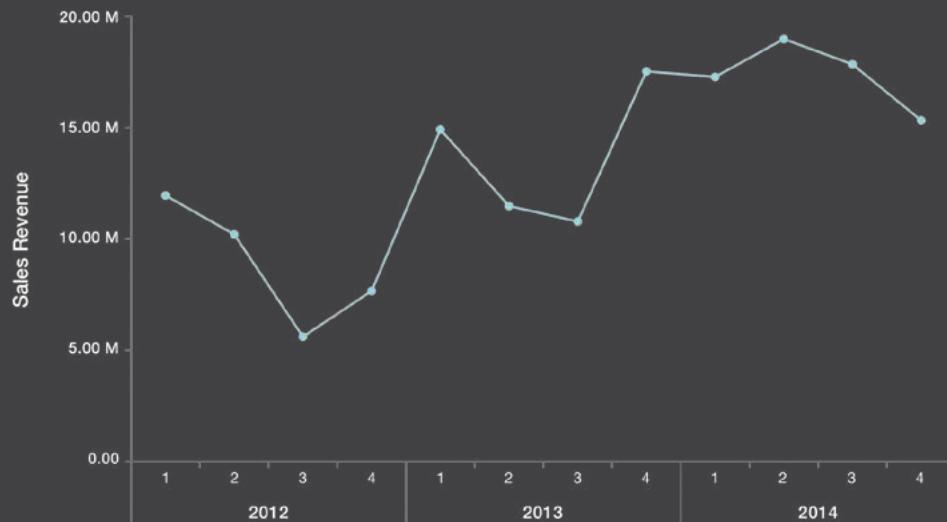
The Line Chart displays measures over a time period.

Line Charts are used frequently to show trends and relationships between them. The Y-Axis always shows a measure value, and the X-Axis denotes a time dimension such as Month, Quarter, or Year.

Used for

- Trends
- Data over time
- Temporal patterns and correlation
- Period-over-period

Sales revenue in 2012 - 2014



Suggestions

1.

Create a time hierarchy to allow drilling up or down to Days, Months, and Years

2.

Add a moving average line to smooth the trend over time

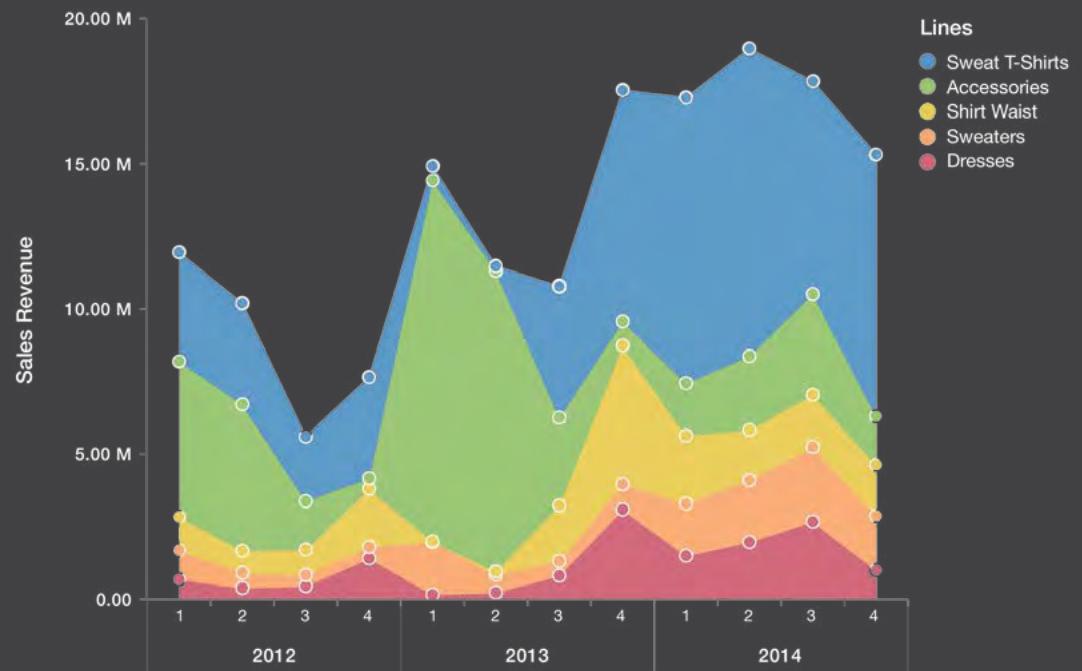
3.

Add a forecast or linear regression to emphasize current or future trends

4.

Consider an Area Chart for showing cumulative totals

The impact of different product lines by sales revenue in 2012 - 2014



Column Line Chart

The Column Line Chart is a combination of a Line Chart and Column Chart. This chart type displays one measure as a column and a secondary measure as a line. The two measures are displayed over a Time Dimension which may include Years, Quarters, or Months. This chart is great for showing the relationship between two measures over a period of time such as Gross Margin and Sales Revenue, or Net Income after Tax and Tax Rates.

Used for

- Trends
- Data over time
- Temporal patterns and correlation



Suggestions

1.

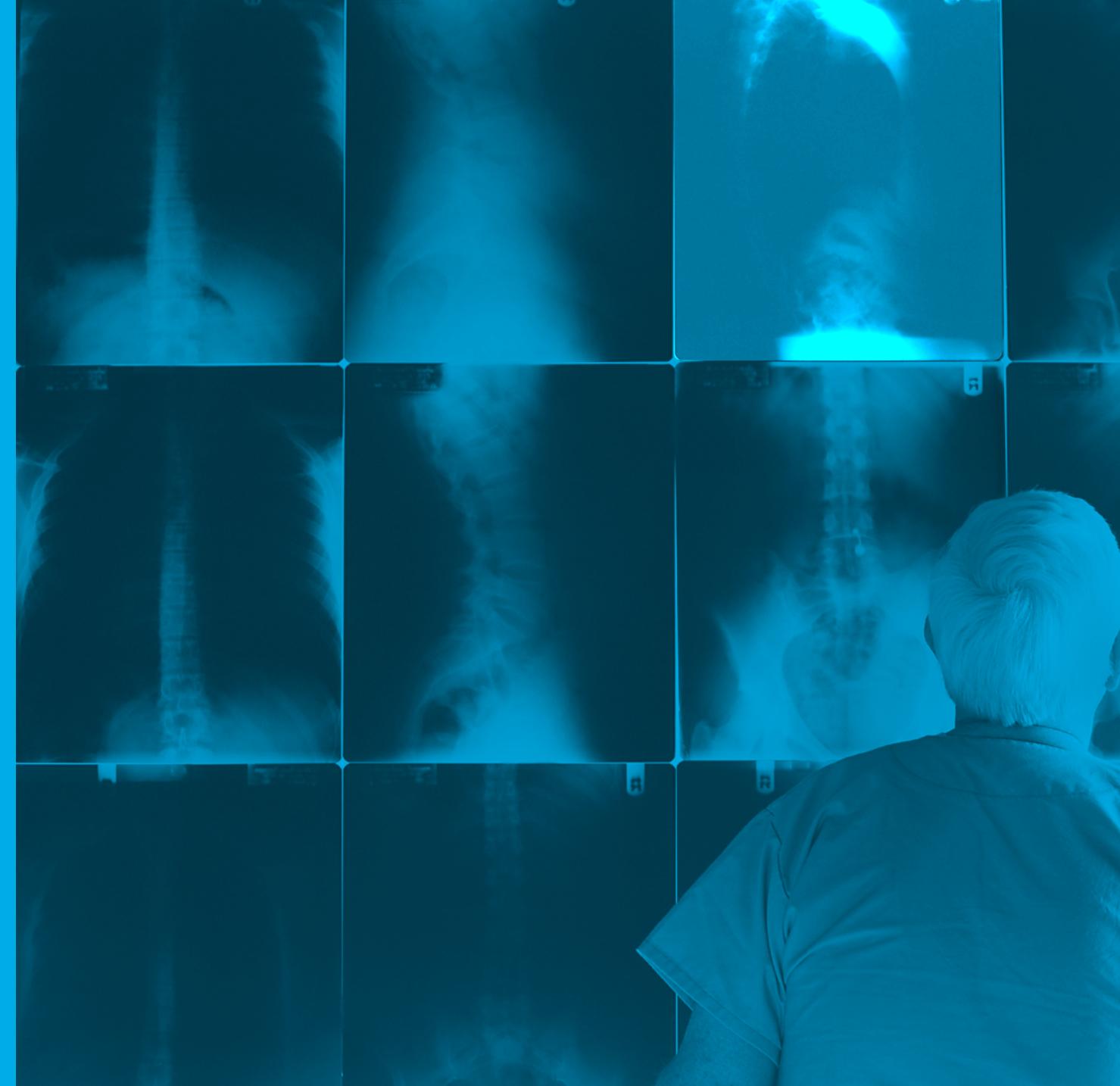
Use this chart type to show two trends of different types (for example, Returning Customers and Sold items) over time

2.

Other options for showing change over time include Bar Charts or Tables.

Comparison

Shows the comparison of categorical values,
where the data does not have any intrinsic order,
for example, a list of products



Focus Areas



Bar Chart



Trellis

Bar Chart and Stacked Bar Chart

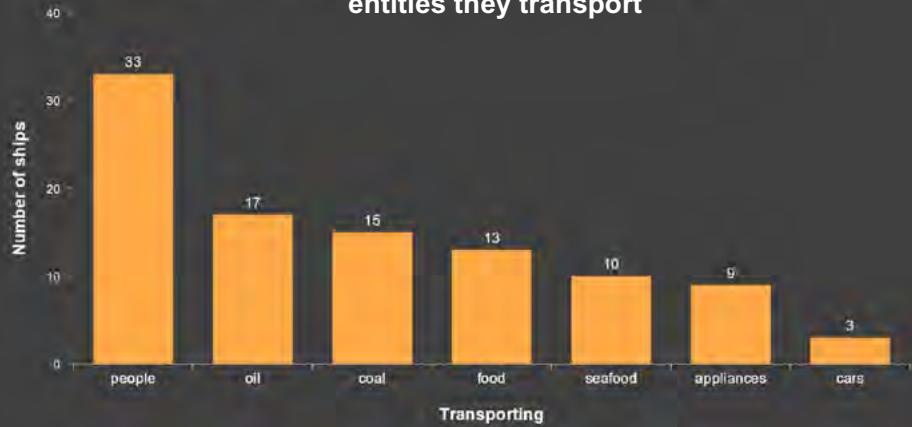
Bar Charts are probably the most frequently used chart type. Focus the attention of your audience to important details by:

- Ranking data from largest to smallest or vice versa
- Filtering out data that isn't important for your message
- Grouping data by combining values in a chart – if there are too many categories, you can group less relevant categorical values together into an Other group (for example, "Other Clothing")

Used for

Comparing different categorical values

Number of ships the company possesses based on entities they transport



Car accidents in US in 2011



Which products are most likely to win deals?



Suggestions

1.

Use data labels, such as Figure 5, to improve the readability of data values

2.

Customize hierarchies to allow drilling from a high-level overview to more specific details; users easily drill up and down

3.

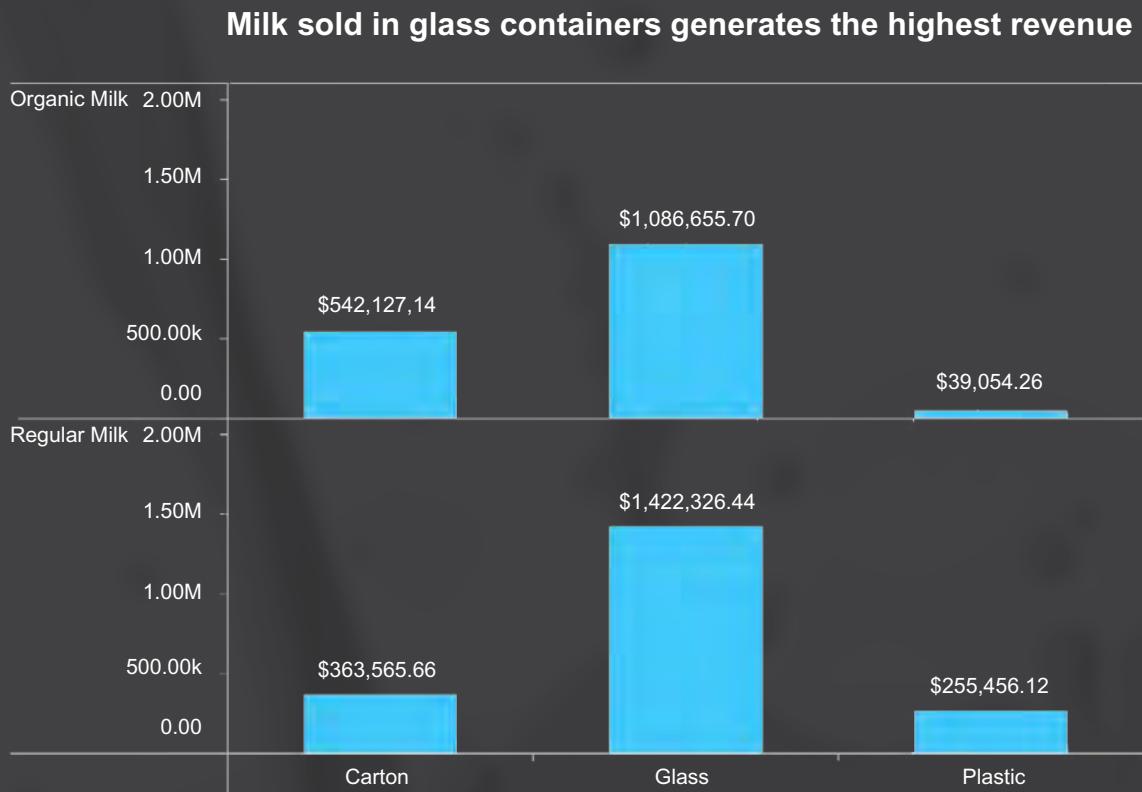
Use Color to clearly differentiate separate categorical values in your dimension

Trellis Layout of Multiple Charts

The Trellis Layout, also known as Small Multiples, contains a set of charts based on the same set of data using the same axes to allow the viewer categorical comparisons of different values within a dimension. Most chart types in SAP Lumira support the Trellis layout.

Used for

Comparison, identifying patterns across multiple categorical values



Suggestions

1.

Used to compare values within a category (such as Trellis by Milk Type to show the Sales Values for each Milk Type in a separate chart).

Ranking

Shows the top or bottom N values to emphasize the largest, or smallest values

Focus Area



Bar Chart

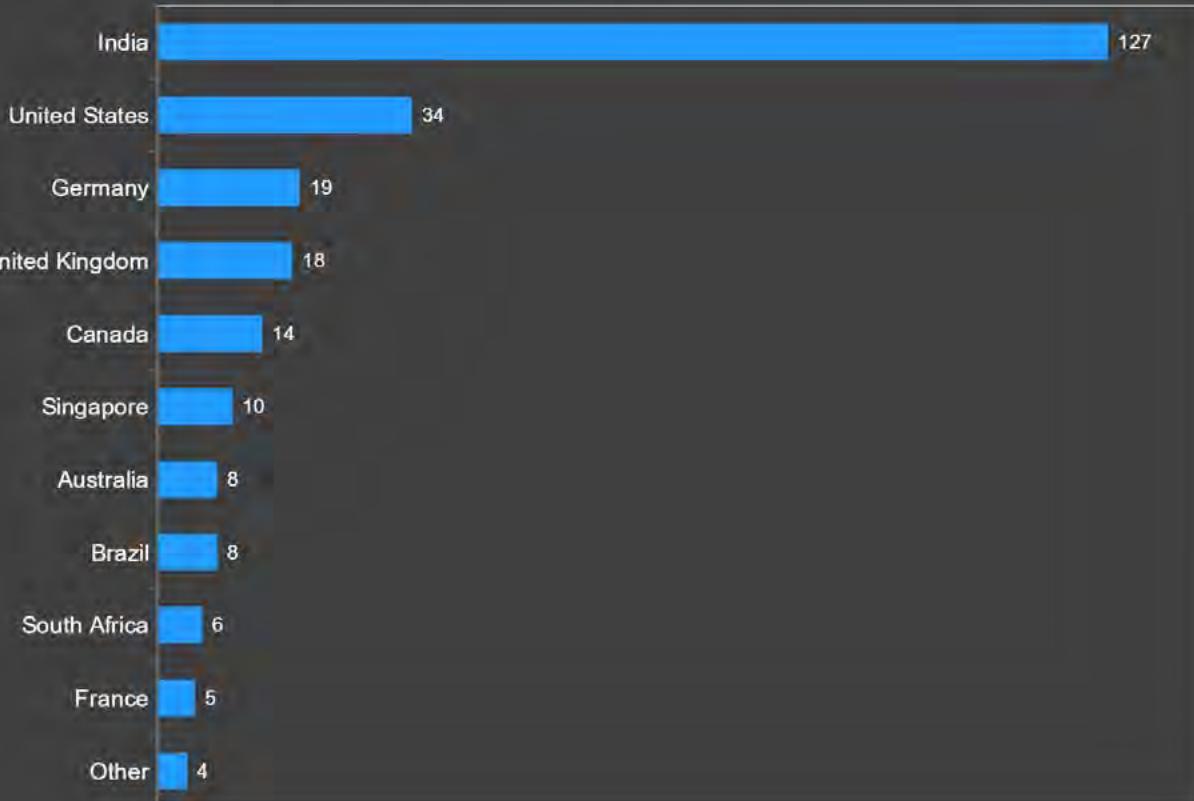
Ranking

The Ranking feature in SAP Lumira allows the user to sort and filter data based on their importance. For example, we may want to sort Countries based on their Number of Participants. The Group by Selection functionality in the Prepare Room can be used in order to group values in SAP Lumira.

Used for

Emphasizing on top or bottom values in a chart

Number of participants from top 10 countries in a design contest



Suggestions

1.

Often categorical values (in this case Countries) that contribute less to the overall measure value might be filtered out or grouped together in another category.

Part-to-Whole

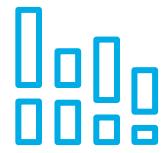
Shows how the categories contribute to the whole value



Focus Areas



Bar Chart



Stacked Bar Chart



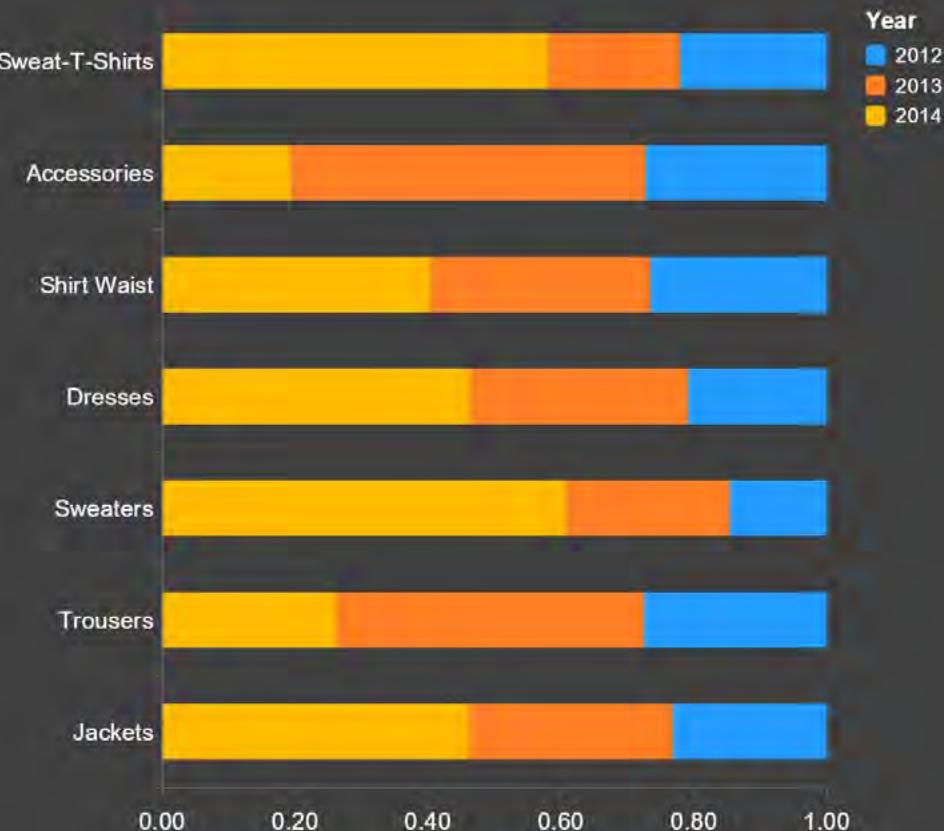
Pie Chart

Part-to-Whole

Used for

A Part-to-Whole relationship shows how measure values that make up the whole of something (for example, Number of containers sold) compare to one another and how they each compare to the whole.

Percentage of items sold for each product line in 2012 - 2014



Suggestions

1.

You can use stacked or side-by-side bars to compare different hierarchy levels (Country → Region) or classifications (Men's Clothing, Women's Clothing).

2.

You can use a 100% Stacked Bar Chart (or Marimekko Chart) to show the portion that each segment makes up in a category.

3.

In addition to Stacked Bar Charts and Marimekko Charts, other charts (such as pie, ring, and funnel charts) can be used to show Part-to-Whole relationships.

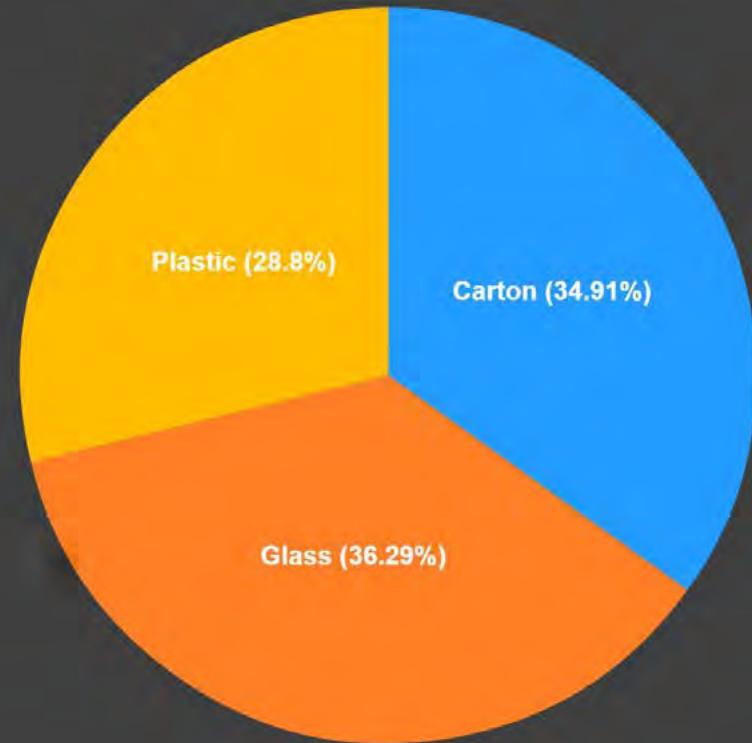
Pie, Ring, and Funnel Charts

Pie, Ring (Donut), and Funnel Charts are used to discern part-to-whole comparisons to either highlight a portion of the data or to compare values for different categorical values. These chart types are generally not recommended if they include too many segments, as the viewer will have a difficult time differentiating between too many different colors.

Used for

Comparing percentage values in proportion to the whole

The percentage of containers sold by container type



Suggestions

1.

Limits use of Pie Charts to a small number of slices (no more than 7 slices)

2.

Consider showing data labels for ease of reading

3.

Highlight only the most important slice if possible

4.

Compare with using a bar chart or ring (donut) chart – the viewer is more likely to perceive the length of a bar over the size of angular slices

Distribution

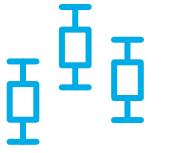
Shows how a measure is spread across its domain



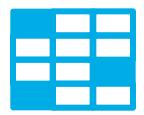
Focus Areas



Histogram



Box Plot



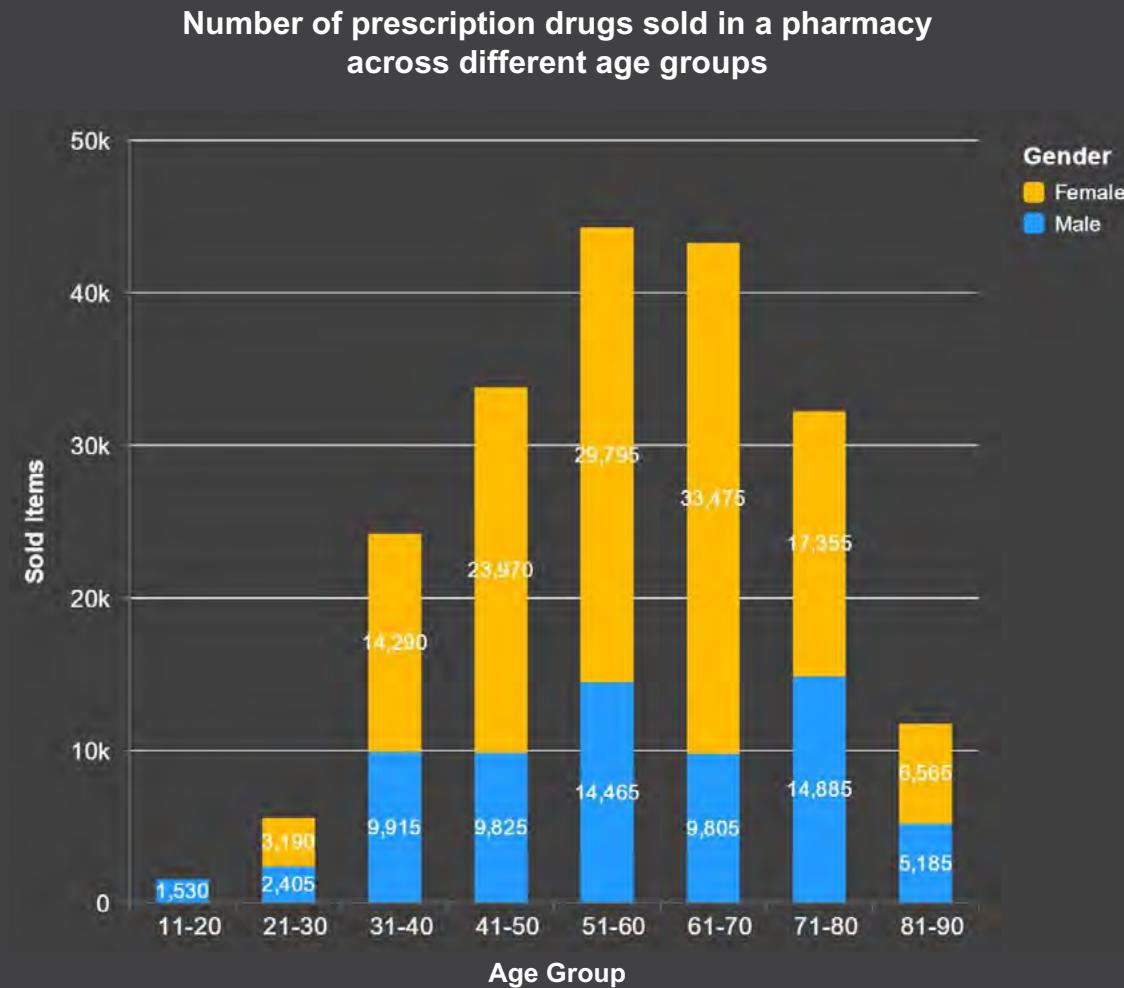
Heat and Tree Map

Histogram and Binning

A Histogram is a type of Column Chart that shows the distribution of measure values, for example, Number of Items Sold. Instead of showing each measure value directly, in a histogram, values are binned first. For example, in Figure 9, instead of creating one column per Age, we binned the values first into the age ranges [11-20], [21-30].....[80-90]. This allowed us to show the distribution of Number of Prescription Drugs Sold in an audience friendly way.

Used for

Distribution of measure values, identifying data issues including outliers



Suggestions

1.

Create bins or ranges of numbers to count the number of occurrences within your data. In SAP Lumira, this can be done either in the Prepare Room using the Group by Range functionality or in the Visualize Room using a Calculated Dimension.

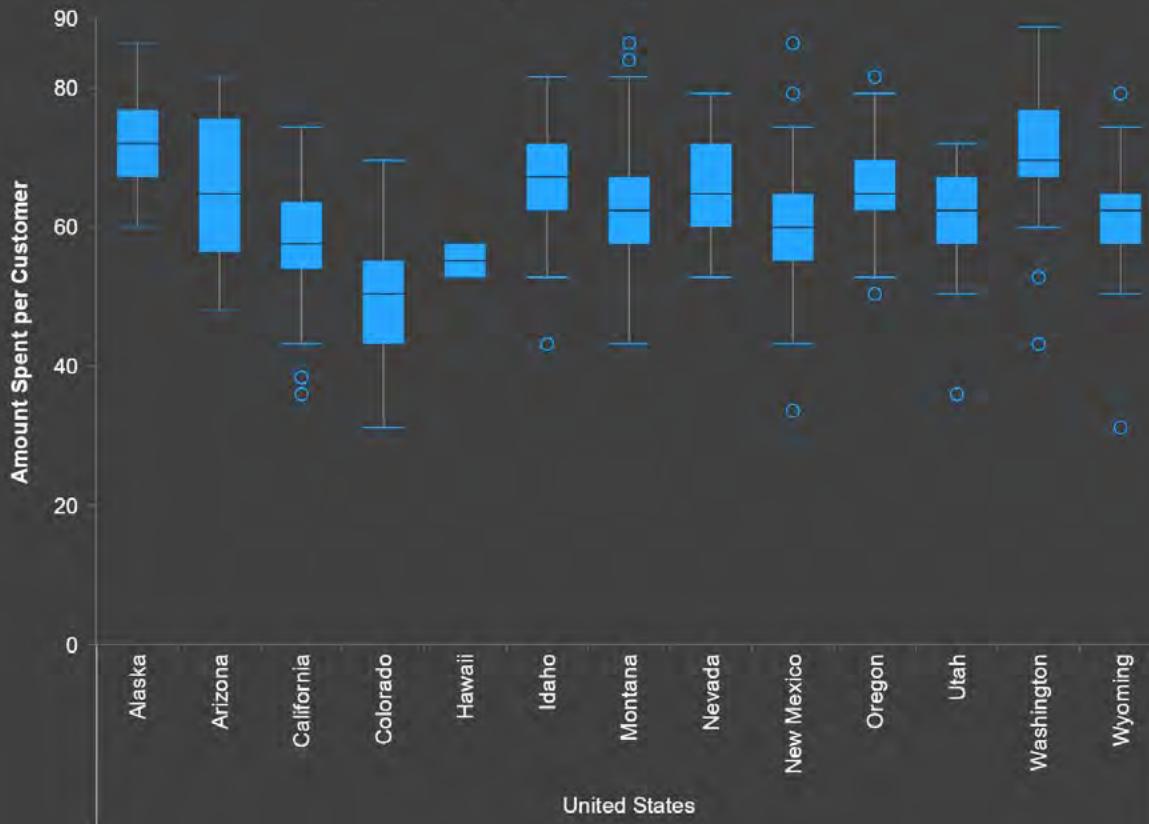
Box Plot

A Box Plot visually displays statistical distribution of a measure within a dataset. It is often used to also show the range in values for each categorical value. Boxplots show the minimum and maximum values, as well as the median and the 25th and 75th quartile. Outliers are visually represented by dots.

Used for

Comparison, distribution of values, identifying outliers

The distribution of dollars spent on our products by customers in the Western United States



Suggestions

1.

Compare data distribution for several categorical values

2.

Show distribution of medians in data

3.

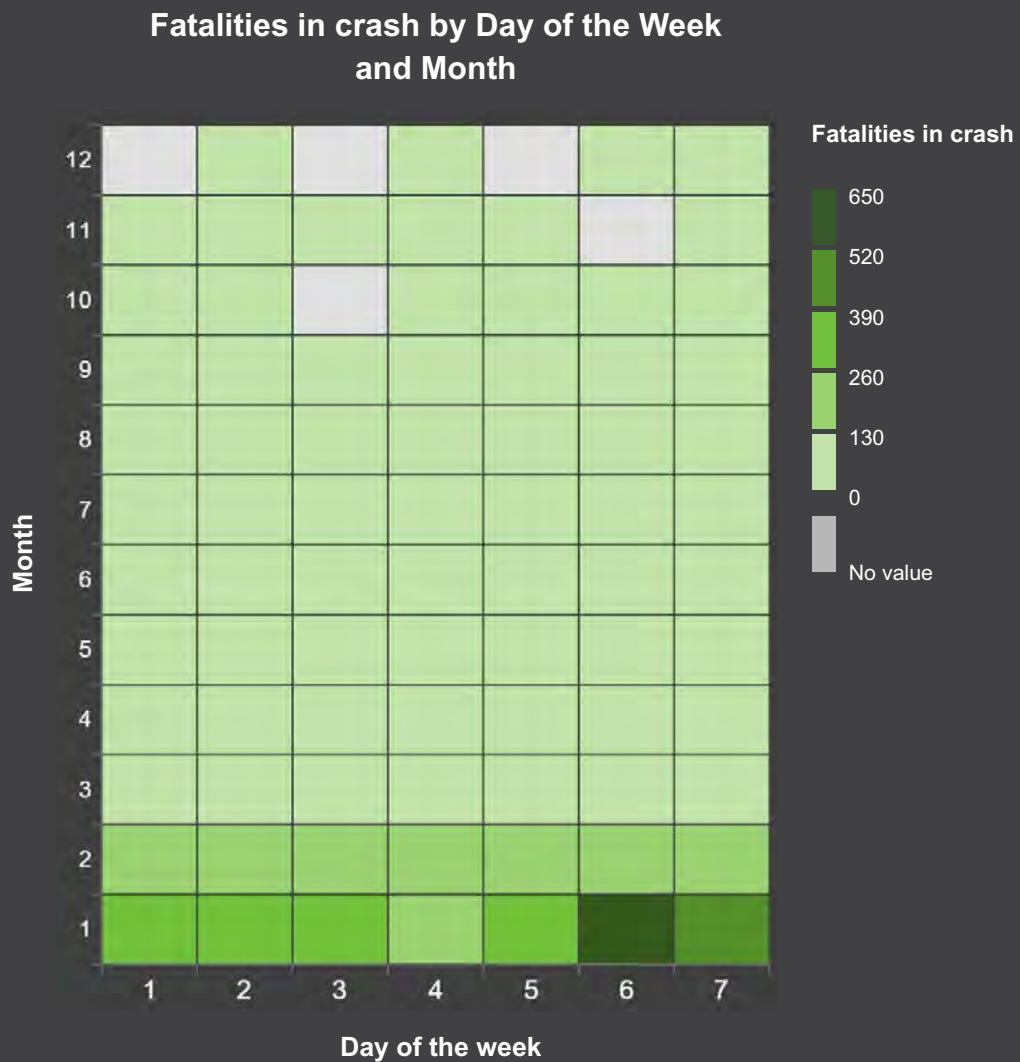
Include a reference line for the overall median in your data

Heat Map and Tree Map

In a Heat Map the categorical values are contained in a matrix of tiles; based on a single measure, these tiles have different shades. In contrast, in a Tree Map, two measures are considered. Larger values are represented by larger tiles and darker shades.

Used for

Showing the distribution of measure values



Suggestions

1.

Only use this if the resulting Heat Map shows visibly different color intensities (it will confuse the viewer if the heat map segments are of similar color intensities).

Correlation

Shows whether there is a potential correlation between two measures



Focus Areas



Scatter Plot



Trellis

Scatter Plot

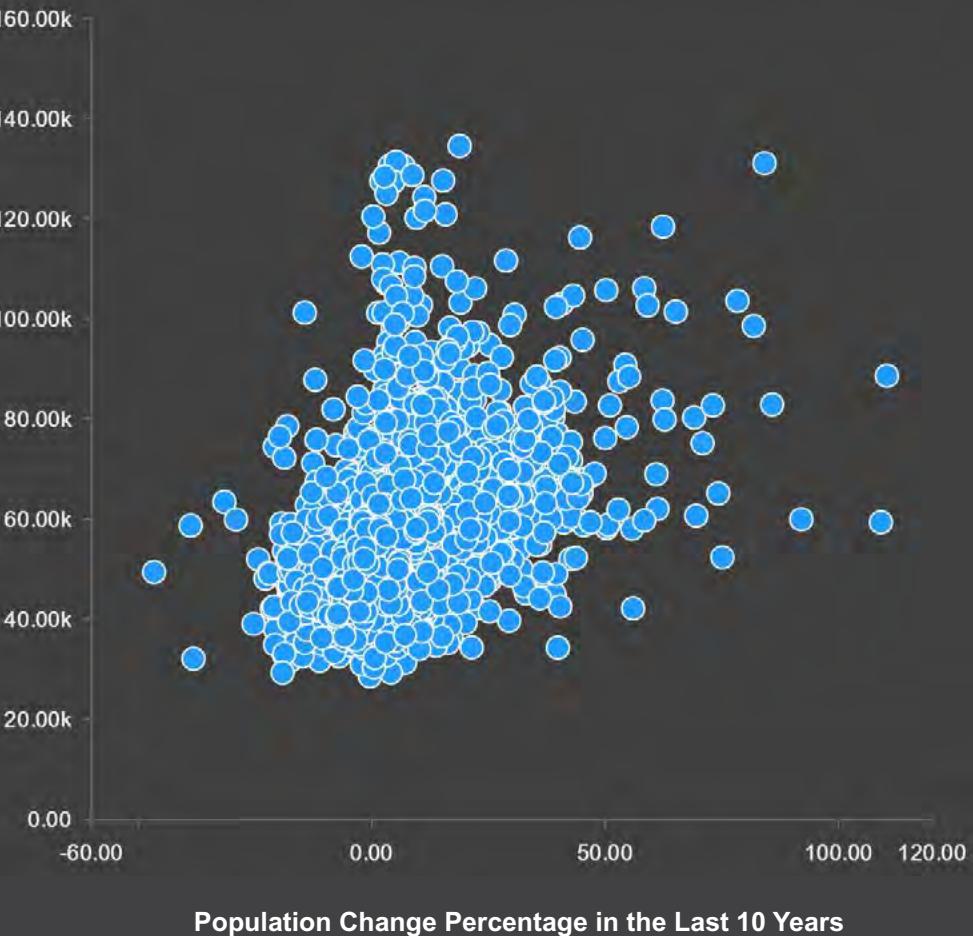
Additional Chart Types Used For Showing Correlation:

- The Scatter Matrix shows several Scatter Plots in a Trellis layout in order to compare several Scatter Plots in one chart.
- A Bubble Chart is similar to a Scatter Plot, but allows visualization of a third measure. The size of the bubbles indicates this third measure. The larger the measure is, the larger the bubble.

Used for

Showing the correlation of two measures

Countries with higher household income have higher population growth



Suggestions

1.

Use the color to show groups of points, but limit the number of colors used; too many colors or shapes will impact the readability of a chart

2.

Keep the aspect ratio square

3.

Create a Geo Hierarchy on top of location data (for example, States, Cities) to enable drilling up to higher levels of geographical detail

Overview

Shows the exact values in table format

Focus Area



Table



Table

You can find the Table as one of the visualization types in the Visualize Room.

Used for

Show multiple measures in one or two categories or hierarchies

Drug and alcohol usage had a significant impact on fatal injuries in the United States in 2011

Injury Severity	Gender	Measures		
		Accidents	Accidents Involving Drugs	Average Blood Alcohol Content
Fatal Injury	Female	716	39	0.05
	Male	1,603	108	0.08
Incapacitating Injury	Female	227	8	0.02
	Male	356	10	0.05
Non-incapacitating Injury	Female	281	7	0.02
	Male	403	13	0.03



Suggestions

1.

Best for showing exact values

2.

Often charts and Tables are shown on the same page, as they emphasize different aspects

3.

Highlight key information with the Conditional Formatting feature

4.

Setting the correct precision (number formatting) for measures included in a Table is paramount in order to not overload the user

Geographical Information and Maps

Shows the geographical distribution of measure values



Focus Areas



Choropleth Chart



Geo Bubble Chart

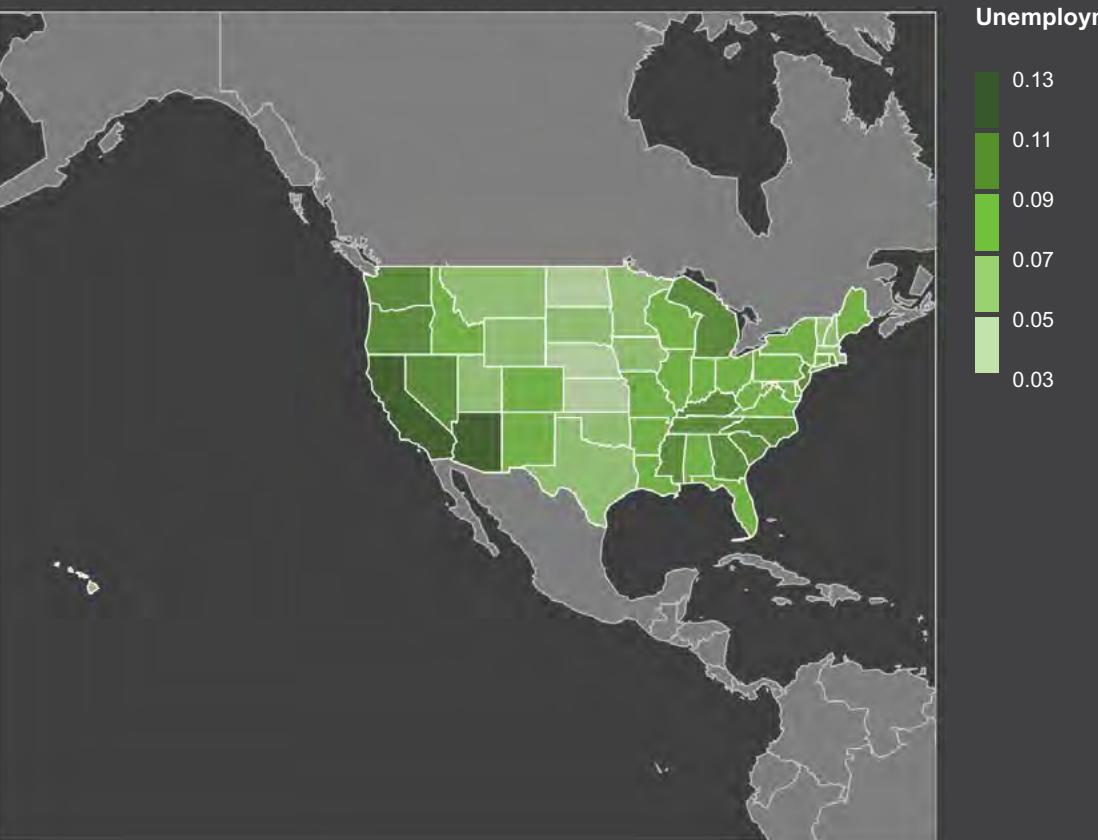
Choropleth Map

A Choropleth Map uses differences in shading, coloring, or the placing of symbols within predefined regions to indicate measure values in those areas.

Used for

Supports location-based comparisons of standardized data such as Rates, Densities, or Percentages

Unemployment percentage across different states in the United States in 2014



Suggestions

1.

Use the Choropleth Map for locations of similar size, as the size of the area coloured may overemphasize larger areas (for example, Canada covers a much larger area than Japan despite being much smaller in terms of population)

2.

Make sure your measure values are normalized by the geographic properties, for example, by the population of a geographic area

3.

Remember that the granularity of your regions (counties, for example) will impact the signal (aggregated measure values) from your data.

Geo Bubble Chart

The Geo Bubble Chart shows measure values in the form of bubbles on a map. The larger the measure, the larger will be the bubble on the map.

Used for

Viewing a measure by Country, Region, or City; comparing measures across different geographic areas

Voyages ending in Americas experience lower delays



Suggestions

1.

Use to show values on a map and to create an animation over time

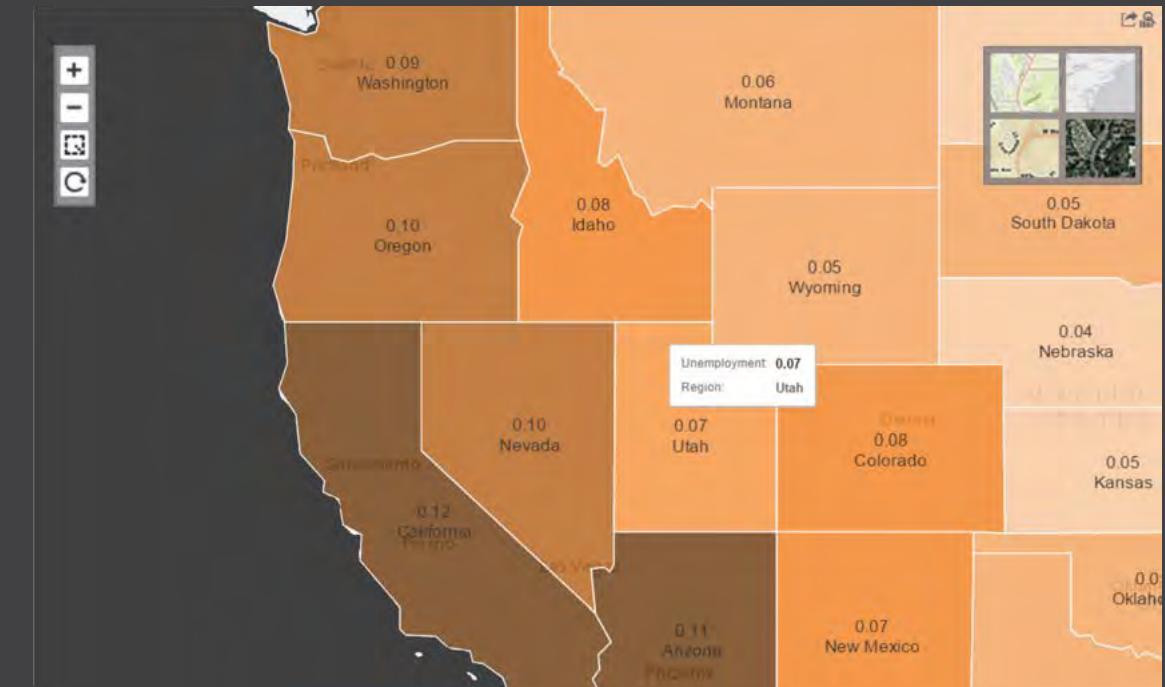
2.

Use Geo Bubble or Pie Charts on maps to show measure values if the relative size of the underlying regions cannot be compared

ESRI

ESRI is an international supplier of Geographic Information System (GIS) software, web GIS and geodatabase management applications. SAP's alliance with ESRI allows for access to their base maps, geo-spatial functions, and data from within SAP Lumira for all users. Customers with an ESRI ArcGIS online account can also create their own custom maps and spatial analysis and bring them into SAP Lumira.

You can get started on a free 60 day ArcGIS Online trial [here](#).

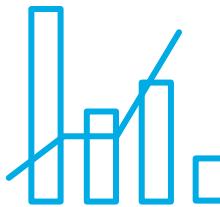


ESRI Maps integration in SAP Lumira connects geospatial assets with business data. Geo-enable your business data to quickly discover new patterns and share your insights across the organisation.

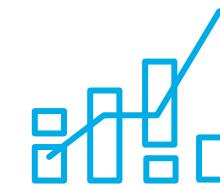
For more information visit
www.esri.com

Other Chart Types

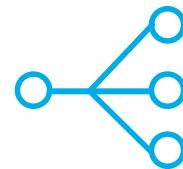
In addition to the chart types mentioned above, SAP Lumira also supports the following chart types:



Combined Stacked Column + Line Chart



Combined Stacked Column + Line Chart with 2 Y-Axes



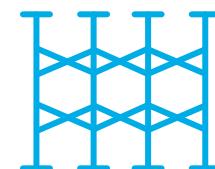
Network Chart

Used to show the relationship between the categorical values of two dimensions, for example, People and Locations



Radar Chart

Compares multiple measures across categorical values



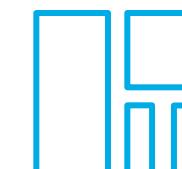
Parallel Coordinates Chart

Used to compare multiple measures for a single category



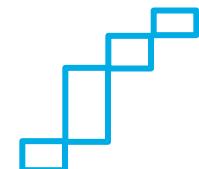
Tag Cloud

Visualization type used for text data. The size and color of the value reflects the measure values



Tree Chart

Often used to show the hierarchical relationship between dimensions



Waterfall Chart

Used to show the cumulative effect of temporal (or other sequential) data

42K

Numeric Point

Often used in Stories to reflect an important key figure

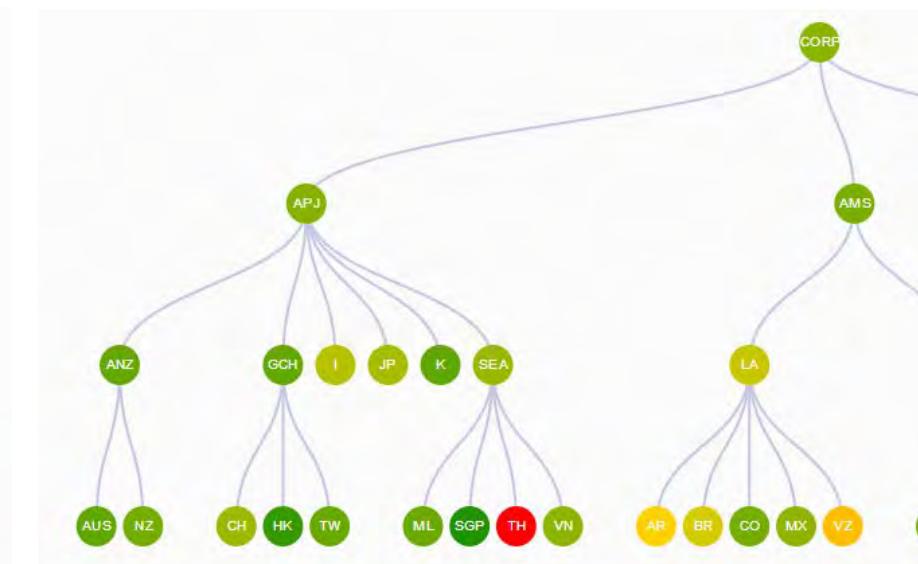
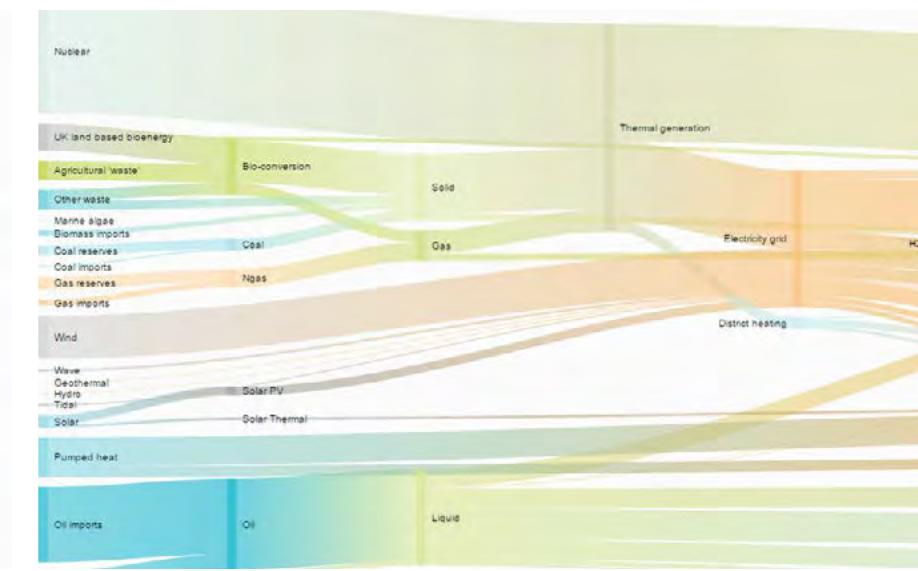
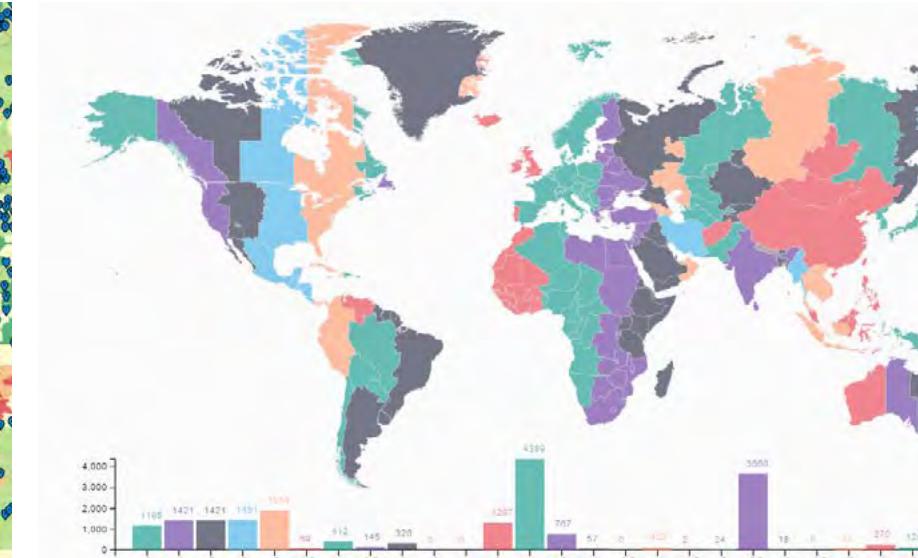
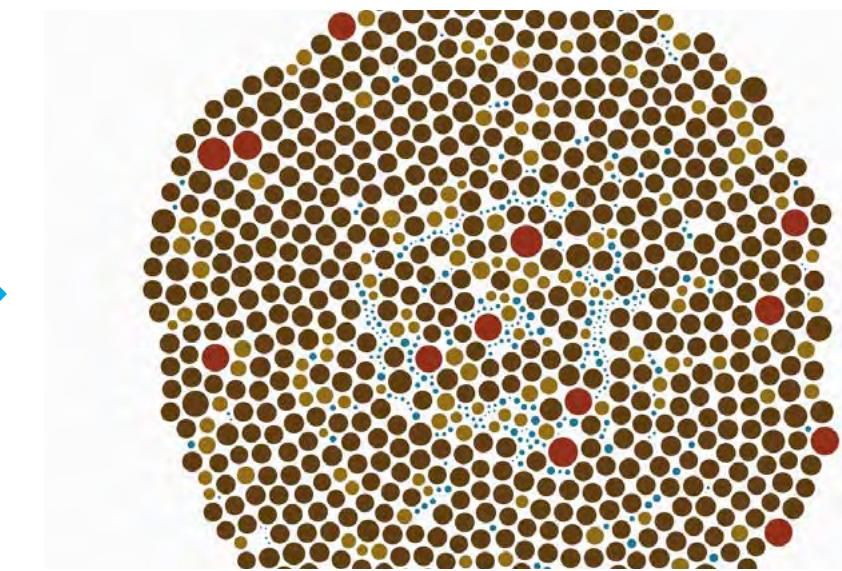
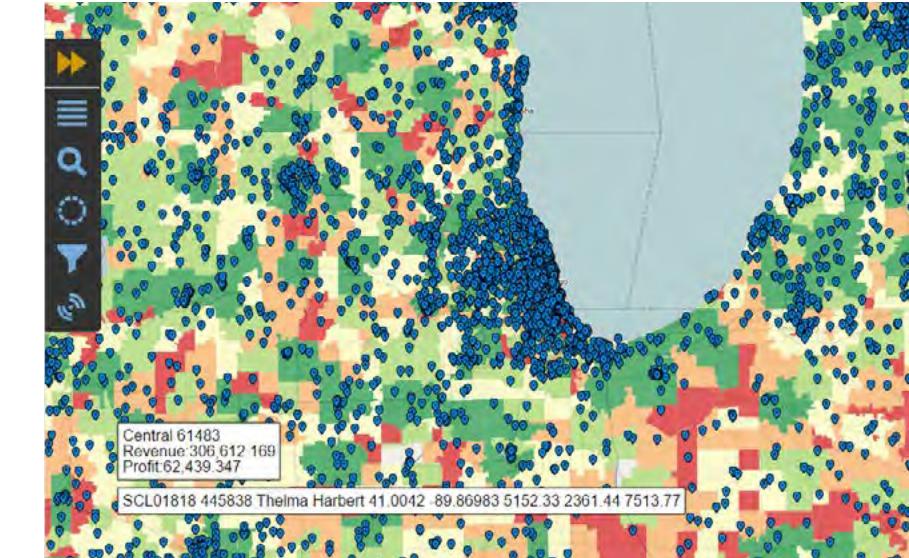
Custom Extensions

Custom Chart Extensions – Designed with a Developer in Mind

We know that there's always a case where you need to create custom visualizations crafted just for your story. We also know that data can be found in a number of different sources – and that you might have a very special one to report on. In both cases, we have you covered - our extensibility framework lets you build your custom charts and connectors.

To build your custom extension, you can start from scratch, adopt D3 charts you find on the web, or use any of the open source visualizations and data access extensions.

Explore open source SAP Lumira custom extensions [here](#).





5

71%

57%

42%

38%

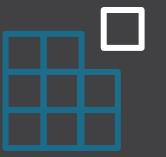
19%

Protips

Protips

A great visual design standard will accelerate understanding and consumptions of your data. It's that simple. For your business to reap the benefits of data visualizations, organizations need to create a visual design standard that incorporates best practices.

The International Business Communications Standards (IBCS), is a non-profit organization that has established a rather comprehensive and detailed visual standard for designing both reports and presentations. It is highly recommended that anyone who develops reports, either as a data professional or business analyst, should peruse both the IBCS Web site (www.ibcs-a.org) and Rolf Hichert's consulting Web site (www.hichert.com).



1.

Less is more. Make every pixel and word count.



2.

Avoid decorative use of graphics.



3.

Avoid three-dimensional chart types.



4.

Avoid pie charts.



5.

Start bar charts at zero.



7.

Use sparklines to show trends on the X-axis.



8.

Show time going from left to right on the X-axis.



9.

Use color only to highlight or accentuate meaning.

Tell the world your data story.



Your visual mind is a powerful asset and unshakable ally for the discovery, exploration, and presentation of ideas. With simple pictures, a little practice, and solid tools, you can turn data into information, information into insight, and insight into action. If you want to look like a data visualization genius, simply leverage the genius of the visual mind and create stories driven by truth, inspired by curiosity.

“The greatest value of a picture is when it forces us to notice what we never expected to see.” - John Tukey



**Tell better
data stories.**

Download your **free 30 day trial** today at:
www.saplumira.com/30daytrial



BusinessObjects Lumira