Welcome to your **story**.

Every story is unique. It’s an opportunity to inspire, advise and 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. You will discover how to approach your data, how to make the most of elements within data visualizations, and decide which visualizations tell your audience the best story.
1

The Art of Data Visualization
Why Visualize?

Today we receive 5X as much information as we did in 1986. This means how you share your data 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.

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

- Position
- Length
- Direction

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?
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: box plot
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 and find the best way to communicate it to your audience. Use the data to illuminate the message you are trying to share, and you'll make it unforgettable.
Know Your Data

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.

With SAP Analytics Cloud you can measure:

- Numerical Data
- Dimensions
  (Categorical)
Measures

Measures constitute numerical data that are calculated or aggregated – like the sum of Revenue, average Cost, 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 Analytics Cloud sets this type to sum. For example, if the chart includes Revenue by Country, and sum is associated with Revenue, SAP Analytics Cloud allows you to customize the prefix or suffix to indicate data such as, units of measures, like CAD, EUR, and USD.
Dimensions constitute categorical data such as year, product, country and salary range.

What do they represent?

**Categorical**
(Also called “nominal”) for discrete values.

- **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 you want to tell your audience. It is time to craft that message and discover which visualization best articulates your information.

Keep these questions in mind:

1. What is the 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 you 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 what you want to communicate 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 connecting with your audience.
Visualization 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 choose to tell should be unique and specific to your story as well. Select 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: Used for comparing categorical values
- Bullet: Indicates progress towards a goal
- Time Series Chart: Shows trends over time

**Comparison**
Shows the comparison of categorical values, where the data does not have any intrinsic order, for example, a list of products
- Bar Chart: Shows categorical values
- Trellis: Uses multiple views to show different partitions of a dataset
- Merimekko: Communicates through width

**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
- Pie Chart: Compares percentage values
- Stacked Bar Chart: Shows overall measure total

**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

**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
- Radar: Sees the distribution of values across dimensions

**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

**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 you to highlight temporal trends.
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

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 Column Line Chart is a combination of a Line Chart and a 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
A Bullet chart can be used to indicate progress towards a goal. It is used to compare primary and secondary (comparison) values. This type of chart allows you to display data based on pre-defined thresholds.

**Used for**
- Current values
- Level of progress

**Suggestions**
Establish thresholds for the measure you are working with.
Time Series Chart

Show trends over time in a Time Series Chart. Choose the time series chart to easily filter by different time periods. This chart has the ability to forecast future values based on historical data within the chart.

Used for

- Forecasting future sales

Suggestions

Use this chart where the X-axis is time.
Comparison

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

Used for

• Comparing categorical values

Suggestions

1. Use data labels 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
Waterfall Chart

A waterfall chart is used to show the cumulative effect of temporal (or other sequential) data. It is useful to visualize the fluctuation of a value in positive and negative values.

Used for

- Cumulative effect
- Deviations and differences

Suggestions

1. Show how you arrived at a net value
2. Break down the cumulative effect of positive and negative contributions
3. Visualize a starting quantity

Cash Flow from Operational Activities in Million USD
The Trellis Layout, also known as Small Multiples, contains a set of charts based on the same set of data and use the same axes. This allows the viewer to make categorical comparisons of different values within a dimension.

**Used for**

- Comparison
- Identifying patterns across multiple categorical values

**Sales Revenue per Country and Lines**

<table>
<thead>
<tr>
<th>Country</th>
<th>Accessories</th>
<th>Leather</th>
<th>Sweat-T-Shirts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>1,928.93</td>
<td>88.71</td>
<td>2,782.24</td>
</tr>
<tr>
<td>Denmark</td>
<td>314.88</td>
<td>4.00</td>
<td>429.89</td>
</tr>
<tr>
<td>India</td>
<td>4,535.78</td>
<td>46.09</td>
<td>4,894.04</td>
</tr>
</tbody>
</table>

*Suggestions*

Used to compare values within a category.
Merimekko

The Merimekko Chart is a bar style chart type that communicates through width, rather than height alone. It is a two-dimensional stack chart that allows for each bar to communicate two measures using the size of the bar.

Used for

- Analyzing marketing and sales data
Ranking

Shows the top or bottom N values to emphasize the largest, or smallest values.
The Ranking feature in SAP Analytics Cloud allows you 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 DataView section can be used in order to group values in SAP Analytics Cloud.

**Used for**

- Emphasizing top or bottom values in a chart

**Suggestions**

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.
Cluster Bubble

The Cluster Bubble Chart is another way to show the value of a measure based on two dimensions. This is a great way of seeing the relationship between values or to compare values.

Used for

- Showing the value of a measure based on two dimensions
Part-to-Whole

Shows how the categories contribute to the whole value.
Part-to-Whole

A Part-to-Whole relationship shows how to 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.

Used for

• Comparing hierarchy levels
• Showing the portion that each segment makes up in a category

Suggestions

1. Stacked area charts are used for multiple data series with part-to-whole relationships.

2. In addition, they are also used to show a cumulative series of values.
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

---

**Suggestions**

1. Limit use of Pie Charts to a small number of slices (no more than 5 slices). Do not use a pie chart if the slices are of similar size.
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.
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. The lines on the box plot refer to the minimum, first quartile, median, third quartile, and maximum range of variation. The dots on the box plot are visual representations of the outliers.

Used for

- Comparison
- Distribution of values
- Identifying outliers

Suggestions

1. Compare data distribution for several categorical values
2. Show distribution of medians in data

Revenue per product in Million USD
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**
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)
Radar

Radar Charts are useful for seeing the distribution of values across dimensions and what the outliers are. They can be used to see which dimensions are scoring high or low for a particular dimension, making them ideal for displaying performance.

**Used for**

- Identifying which dimensions score high or low for particular dimensions

**Passenger satisfaction rating by age group in USD**

**Suggestions**

1. The data series with the highest values covers the largest area
2. The values of a series are connected with a line
Correlation

Shows whether there is a potential correlation between two measures.

Focus Areas

- Scatter Plot
- Bubble Chart
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.

Used for

- Showing the correlation of two measures

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.

Marketing events and revenue in Million USD
To see the correlation between three measures, use a Bubble Chart. The size of the bubbles indicate the third measure. The larger the measure is, the larger the bubble.

**Used for**

- Displaying values that consist of three variables

**Suggestions**

The difference with the XY chart is that the size of the points in the bubble charts are adjustable.
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
- Rates
- Densities
- Percentages

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

- Comparing measures across different geographical areas

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
Pro-Tips
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](http://www.ibcs-a.org)) and Rolf Hichert’s consulting Web site ([www.hichert.com](http://www.hichert.com)).

### Pro-Tips

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.**
6. **Use bullet graphs instead of gauges to save space.**
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.**
Show the world your data.

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.

“The greatest value of a picture is when it forces us to notice what we never expected to see.” - John Tukey
Chapter 5: Protips

Improve your data visualization.

Start your free 30-day trial today at: discover.sapanalytics.cloud/30daytrial