

Data Visualization and Dashboard Projects

Quickstart Guide

Introduction

So, you're launching a data visualization and dashboards project? Not to worry, this quick guide will help you understand the best ways to visualize your data, plan your dashboards and meet your audience's needs. We've based this content on the extensive experience of Hitachi Vantara's custom user interface (UI) and visualization design team in collaborating with customers on a wide variety of projects. Here's a brief agenda of what we cover:

Get Started and Define the Project	The items you need to consider before you start creating analytics: project objectives, core KPIs, user needs, data readiness and more
Data Visualization Selection	How to identify visualizations that best illustrate different types of metrics, plus useful tips for grouping data by geography, time period and other categories
Dashboard Layout and Design	Best practices for placing different metrics and visual analytics on a dashboard for effective presentation to your audience
User Interaction	What to consider as you look to make your dashboard more engaging for users, including dynamic filters, drilldown capabilities and more

Get Started and Define the Project

Behind every dashboard project is some type of organizational need. It could be as simple as keeping salespeople up to date on weekly performance or as sophisticated as providing physicians with better treatment recommendations via predictive modeling. Addressing these needs with concrete measurable objectives provides the necessary framework for delivering the right information to the analytics consumer in the right fashion. If the project goal is to increase customer service efficiency by at least 10%, for example, that specific objective will help shape and prioritize the type of key performance indicators (KPIs) that are presented to the service representatives using the dashboard.

But how do you focus in on the correct KPIs to communicate? This requires coordinating with strategic project sponsors that are driving the business initiative, as well as working closely with the intended day-to-day users of the dashboards. Determining the KPIs to show (and executing on the entire project) should be an iterative process where you continuously circle back with those stakeholder groups for feedback. As you decide on your first set of requirements and begin the design process, you should schedule regular update discussions with users and sponsors, continuing these meetings until the final project rollout.

While we won't get into detail on all of the project items that should be addressed, you can use the following checklist to make sure you are considering everything you need to before laying out your visual analytics.

Question	Sample Answer
What's the organizational need you are trying to address?	Understand return on investment (ROI) of different marketing campaigns in order to better allocate budget going forward.
Have you agreed on objectives with business and executive sponsors?	Yes, the marketing VP is a driving sponsor and the whole team knows about the project.
Have you met with your potential dashboard users to understand their needs, preferences and context?	I've met with the marketing program managers and gathered requirements, but I still need to meet with our social media and event marketing coordinators.
What are the main data sources you need to access?	We need data from a customer relationship management (CRM) system, marketing automation system, finance databases and social media accounts.
Will you be able to receive updated data in the frequency you need it?	Yes, this analysis will mostly be on a week-to-week basis and my information will be refreshed at least every day.
Is there a measurable goal you want to achieve?	Reduce the average marketing spend per qualified lead by at least 10% in the next year.
What are the core KPIs you want your users to understand?	Cost per lead, cost per sales opportunity, return on campaign investment: All are broken down by marketing channel.
What's the analytical sophistication of your audience?	Medium sophistication: They will sometimes want to drill own into the granular information behind the dashboard.
Is the data integrated, accurate and clean?	It will be soon. Our IT operations team is leveraging a data integration solution to cleanse the data and schedule proper data updates.

When gathering user requirements, it is important to understand your audience and its level of data fluency. Are they executives looking at the big picture or business analysts who will want to filter and customize the data? Maybe they are customer-facing employees who really just want to see a few key operational numbers with simple recommendations on what actions to take next. Your audience could also be external to your organization, in which case the dashboard theming, branding and context may need to fit with existing outward-facing applications.

At the same time, it may help to think of your goal as telling a story with the analytics content, and as such make sure to consider the “what”, “who”, “when”, “where” and “why” of your dashboard project. Overall, it's an exercise in meeting your users' needs and creating a good experience for them, one that will promote adoption and boost usage of the analytic content. Considering multiple dimensions of your analytics will help with this. As you start to gather requirements, creating sketches of possible visualizations and dashboards with your users will help you begin to determine optimal ways to meet their needs.

Finally, you'll want to have your data in order before you identify the key metrics to showcase. This means identifying all of the data sources you are using, whether they are relational

databases, big data stores, or other information systems. You also need to ensure that your data can be properly integrated and blended together for your project. This includes joining data sets and ensuring metadata reflects business rules. At the same time, you need to be sure that the information is clean, accurate and available in the frequency you need it to be updated (weekly, real-time and so forth).

Data Visualization Selection

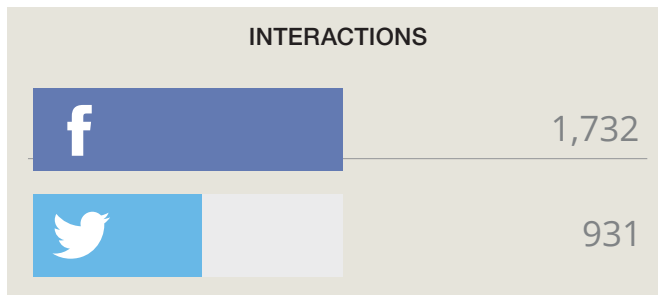
In the large majority of dashboard projects, we've found that KPIs fall into a handful of core categories, which have specific implications for visual analytics. As such, we've mapped visualization guidelines and tips to the different varieties of KPIs, along with examples.

While this section starts out with the basics, it should help guide you as you determine which data visuals are right for you.

Quantities

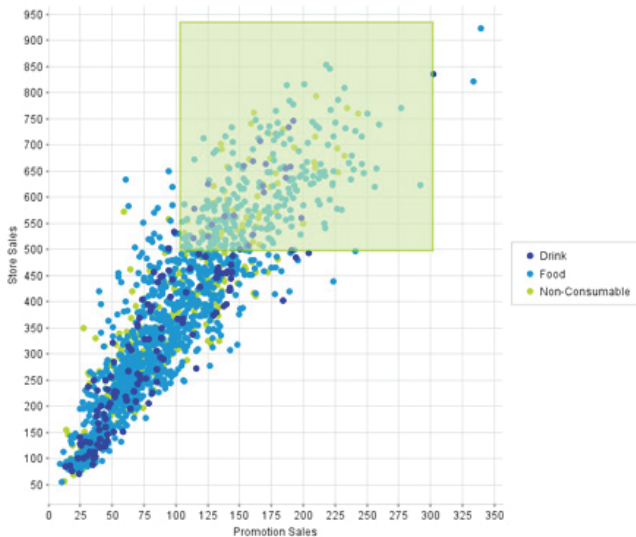
You can think of these as your bread and butter counts or measures, such as the number of units sold, time taken to complete an examination, or tons of freight carried by a rail car. Currency, whether dollars, euros, yen or otherwise, is also an example.

Understandably, there are a wide variety of options for visualizing basic quantities. Bar or column charts are a good place to start, where each bar represents a quantity of something (like units sold) for a specific grouping (like product lines). If there are a limited number of groupings, these types of charts provide a visually intuitive way to compare quantities, as represented by length or height. Line graphs are also a good choice if the quantities are tracked over time.



Simple bar charts are great if the user needs to rapidly compare different quantities.

If you are more interested in the direct relationship between two different quantities, you will probably want to use some type of scatter plot visualization. When a substantial correlation between the variables exists, a linear pattern will form on



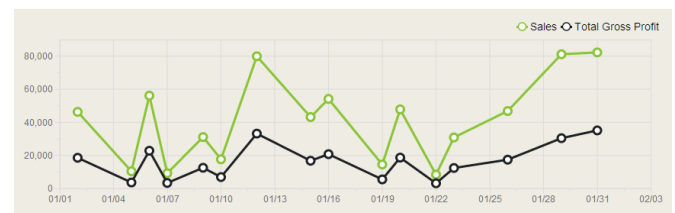
Scatter plots help show correlation between variables. This example shows the relationship between total store sales (vertical) and sales from a specific promotion (horizontal). The colors represent different product types.

the plot. If identifying extreme data points or outliers that don't conform to a normal relationship is important, this could also be a useful way to visualize data for your users.

We will touch on more options for visualizing basic quantities as we navigate through the other types of KPIs and groupings below.

Trends and Changes Over Time

Time series data is a deep analytics topic. However, the crucial point for most visuals is that it adds a new dimension to the analysis, normally where time is displayed on the horizontal axis of your chart. The more lines or series you add to a line graph, the more difficult it becomes to understand and interpret at a glance. So it's a best practice to keep it to less than four series, or else you risk confusing your users.



Basic time series sample graph plots sales and gross profit over a month.

These types of charts give users an intuitive feel for changes over time, such as growth rates and decline rates. When looking at rates of change for a specific measure over time, it can be useful to simply indicate the rate with clear text near the appropriate series on the chart. Showing a directional arrow icon (up, down, or flat) may also provide the user with a clear sense of the change rate.

ENTRANTS



Display a growth rate and icon next to a KPI — in this case entrants in a race.

Relative Shares and Proportions

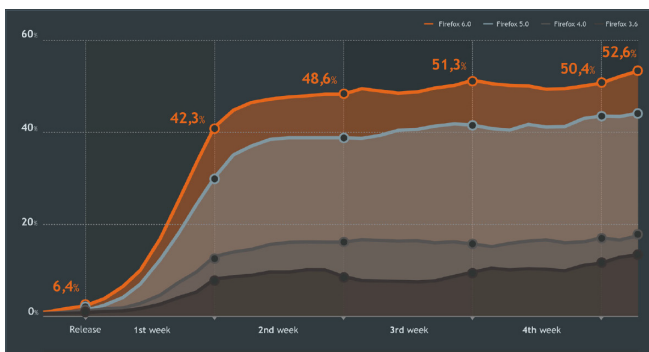
Shares and proportions display a relationship between the parts and the whole, rather than differences over time. Examples here are shares of an investment portfolio allocated to stocks versus bonds, as well as website conversion rates. In other words, they represent the portion of people who opened an email, clicked it or didn't open it.

If the primary interest is to display the composition of an aggregate quantity at one point in time — say the breakdown of the investment portfolio by asset class today — then a pie chart would make sense. Slices would represent different portions that add up to 100%.



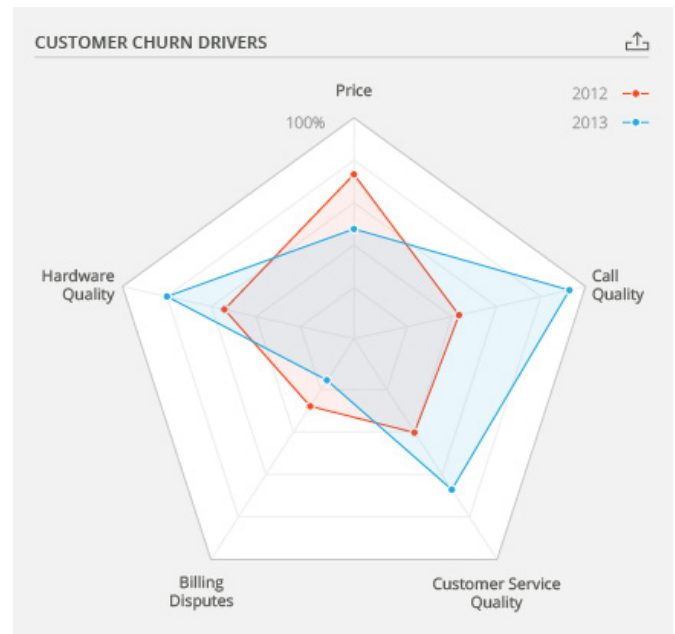
Basic pie chart shows proportions of medical patients who spent time in different types of hospital rooms.

Pie charts on their own, however, do not convey a tremendous amount of information. If we wanted to show proportions over time, we would use a stacked bar or area chart that can be configured to display total quantities over time broken down by categories in either absolute or percentage terms.



Area chart compares the uptake over time for different software versions following their initial release.

A unique alternative for showing proportions could also be a radar chart (also known as a spider or web chart). In this visual, the proximity of the data point to a corner indicates the relative importance of that corner, while several “webs” can be overlaid to illustrate relative results for different periods or other groupings. Indeed, each “web” creates a different shape representing a different scenario to the viewer.



Radar chart captures the relative importance of different customer churn drivers for a telecom company, both in 2013 and in 2012, in a quickly consumable format.

Ranked Lists

Presenting a ranked list of the highest or lowest items from a data set is a common dashboard approach. Though not really data visualization, this is a good way to provide the “need to know” information, assuming there is a way for the user to access greater detail. For instance, a sales management dashboard may include quota attainment for the five best and worst performing sales representatives on the team, with the ability to click through to a view that shows all sales reps with a wider array of metrics for each rep.

RANKING	TRACK	NR STREAMS
	1 Hurry Up We're Dreaming	5,236
	2 Small Things	4,563
	3 Dandelion	4,022
	4 Float	3,785
	5 The Last	3,102
	6 We're Dreaming	4,563
	7 We Stay Together	2,956
	8 One	2,854
	9 Fennesz01	2,820
	10 Passed Me By	2,310

Ranked list of top 10 music tracks by number of streams over a time period.

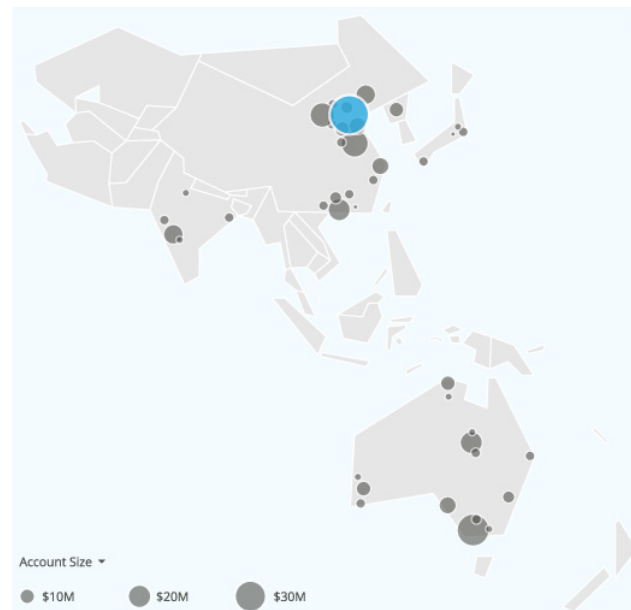
Geography and Location

As you might expect, in this type of visual you are looking at data on a map or physical representation of a real location, which gives the user a richer information experience based on known spatial and physical relationships.

Location can get fairly complex, but there are a few core concepts to remember. First and foremost, don't be tempted to include overlays like streets, terrain, or traffic if they are not relevant to the insight you are trying to provide. If overlays are important, give users the option to turn them on or off according to their preferences. In general, you will want to limit the colors, lines and other detail in your default map view prior to overlaying your specific KPIs on the map (Your metrics are what should stand out!).

Second, there are different options for coding regions, like states, countries, zip codes and so forth. For example, you could use color with borders to represent additional categories or metrics. Alternatively, you could map specific granular data points and look for clusters and patterns independent of regional totals.

Finally, a must-have for geo maps is the ability to zoom in or out to different regions, tracking city versus state or province versus country-level aggregates in the process. Your users are likely used to navigating a service like Google Maps for their own personal geographical needs, so their standard for ease of use will be high.

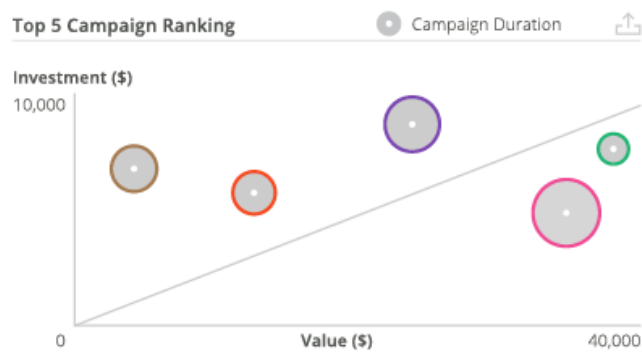


Simplified geographic map with no overlays makes it easier to see data points.

Handle Multiple Dimensions at Once

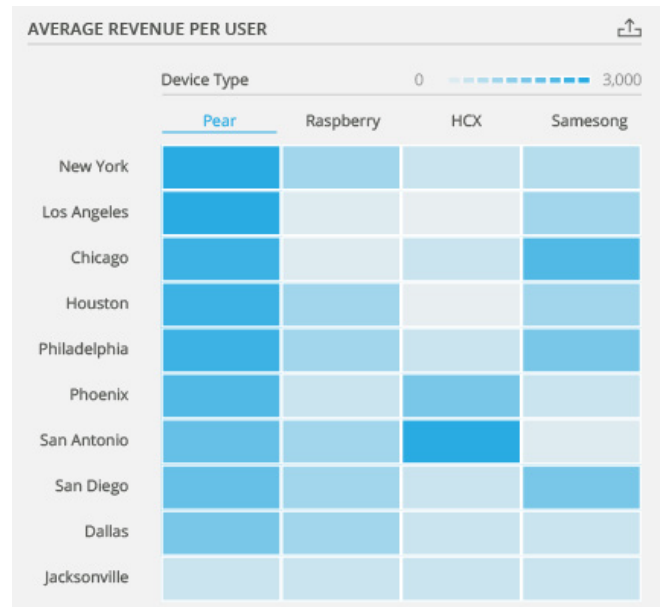
As you might expect, some of the richest visualizations display more than one type of quantity and one type of grouping in the same space. However, as the variety of information you are trying to communicate increases, so does the risk that your users will miss the point. We'll walk through two types of multidimensional visualizations here: bubble-scatter charts and heat grids.

In a bubble-scatter visualization, the scatter plot discussed earlier is augmented with one or more additional dimensions. First would be the size of the data point (hence the "bubble"), where the area of the bubble represents a numeric value associated with that point. This would be followed by color of the data point, to represent another factor.



Bubble-scatter chart shows marketing campaigns by value generated (horizontal) and invested cost (vertical), as well as duration the campaign ran for (bubble size).

In a heat grid, we can leverage color to convey values or "temperatures" of data points on a horizontal and vertical access. This type of visual is more appropriate where we expect multiple Y axis values for every X axis value. Thus, it makes sense for representing values sliced and diced by different general or neutral categories. Put another way, a heat grid can help where a stacked bar chart might fall short: Rather than trying to display values by size stacked on one another, the size of the data points can be left constant with the colors representing the value scale. That said, the size of a data point could be added in as another dimension in some situations.



Heat grid displays average revenue per user (color scheme) across customers, by both geographic market (vertical) and type of mobile device used (horizontal).

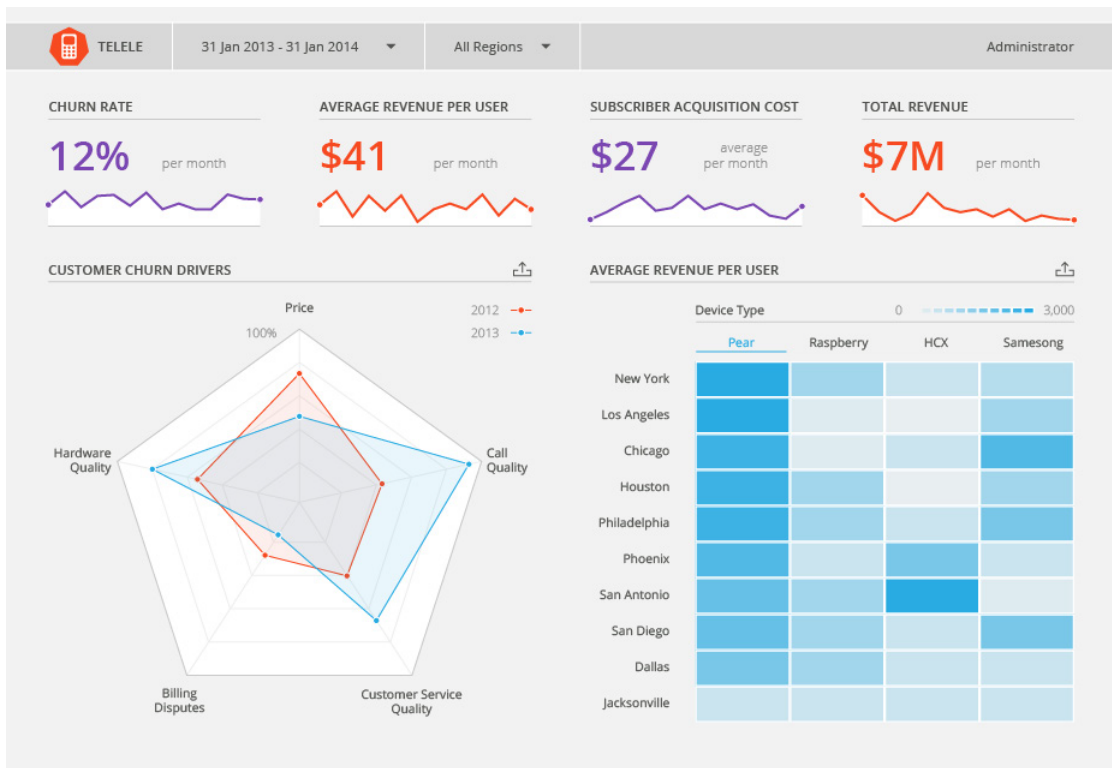
Dashboard Layout and Design

You could have the best data visualizations in the world, but if they aren't organized in a logical and orderly fashion, your dashboard project will miss the mark. While not every dashboard should be designed the same way, there are several guiding principals. As you follow these guidelines, we'd suggest sketching out different possible layouts on paper.

Amount of Content

You will usually want to include no more than 3-4 actual charts or graphs for your audience. Any more than that, and things will quickly get crowded for the user as you start to add headline metrics, titles, text and labels. By the same token, limiting the actual number of visualizations will also minimize the amount of scrolling the user will have to do. If you must have more visuals, then make the others accessible via a button click or toggle that replaces the content shown to the user. It is better to have contained groups of information presented to the user at all times than to try to cram every last data point into one screen.

You could have the best data visualizations in the world, but if they aren't organized in a logical and orderly fashion, your dashboard project will miss the mark.



This telecom dashboard follows best practices by including simply presented aggregate metrics at the top and drill-down visuals beneath. Note also that churn rate and average revenue per user appear first from the left, as they are the most important metrics across the full dashboard.

What Goes Where

A good general pattern to follow in your layout is to have higher level or “headline” metrics displayed near the top part of the dashboard. Sometimes this can be as simple as a few core KPIs in large clear font, perhaps with adjacent spark lines to show trends. Just below should be the data visualizations that show overview data like aggregate KPIs over time. As you move down, more granular visuals should be included. For example, if overall customer growth rate is a key metric, you highlight near the top of dashboard. Then, a breakdown of new customers by segment or demographic that displays actual account names should go further down.

Not surprisingly, you should also follow the general rule of placing the most important information at the top of the screen and to the left, as people start reading in this direction — at least in English-speaking countries. If you are designing for an audience that reads right to left, layout should be adjusted accordingly. It may be obvious, but you should remember that larger visuals and larger text indicate the relative importance of information as well.

Themes and Formats

We couldn’t complete our key guidelines on design without discussing theming and formatting. In choosing color schemes for your dashboard, it is crucial to select a palette with good contrast and balance. For instance, avoid reds overlapping with blues or dark grays. As you want to display more and more different groupings in the same data set (high cardinality), you need to use more colors: Therefore, it becomes harder to tell two shades apart. Ways to deal with this include balancing values for hue, saturation and lightness.

Rather than trying to reinvent the wheel, a good place to start might be to look at your organization’s current branding for its corporate website and collateral. Following color schemes, font types and other visual elements will provide you a good template for the content you are producing.

Finally, we’d note that text such as titles, labels and user instructions should always be present but not in an overwhelming manner. While it makes sense to illustrate the most important KPIs in large clear text, other types of text should appear more for reference than anything else.

User Interaction

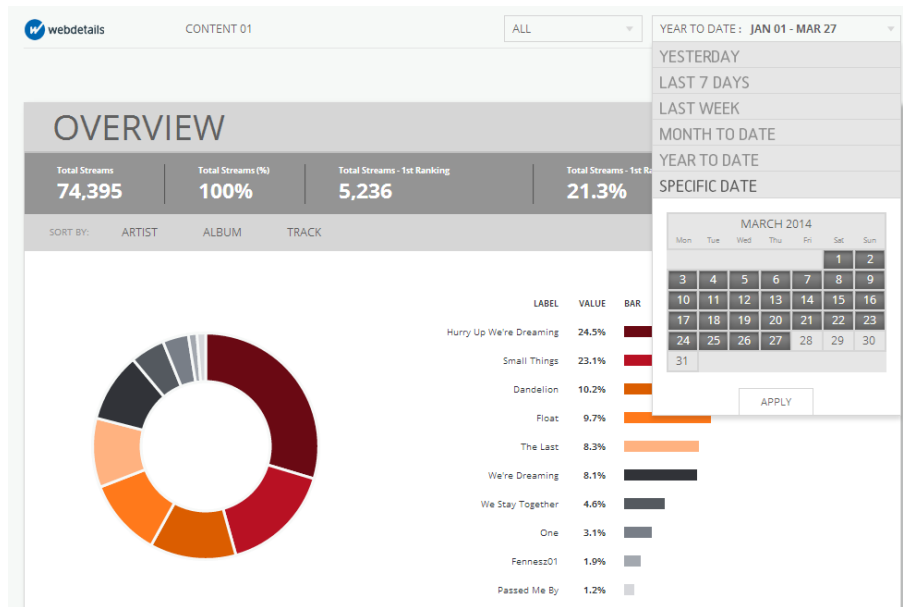
While static visualizations and dashboards have their place, interactive analytics content provides a more engaging experience, with richer informational context and increasing user adoption and satisfaction with the tools. We'll touch on a few key categories below.

Filters and Selectors

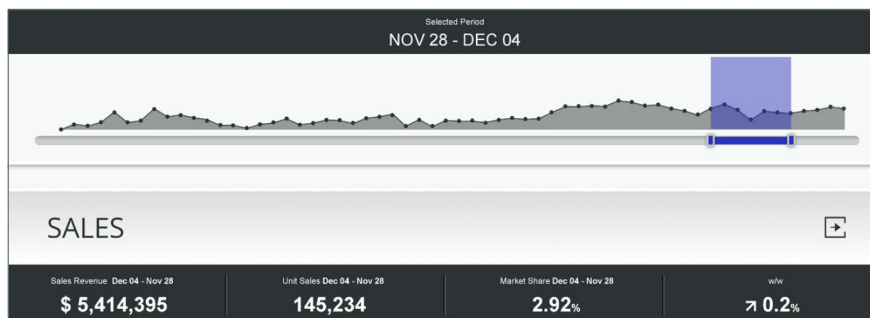
Flexibility is key when it comes to filters and selectors. For instance, when providing a time range selector, you'd want to provide a balanced mix of preformatted date ranges (that is, this week, this quarter, full year) and a full calendar widget for defining a custom range.

Normally, you will want the time selector to drive all relevant metrics and visuals in the dashboard, and if it doesn't, you will have to clearly explain this to your users. At the same time, this selector should be placed somewhere near the top-level titles or menus and be clearly labeled but not obtrusive.

For time series charts, a dynamic period selector that can be expanded and contracted by clicking and dragging an area of the chart can be an easy way for the user to adjust the time-frame as they see fit.



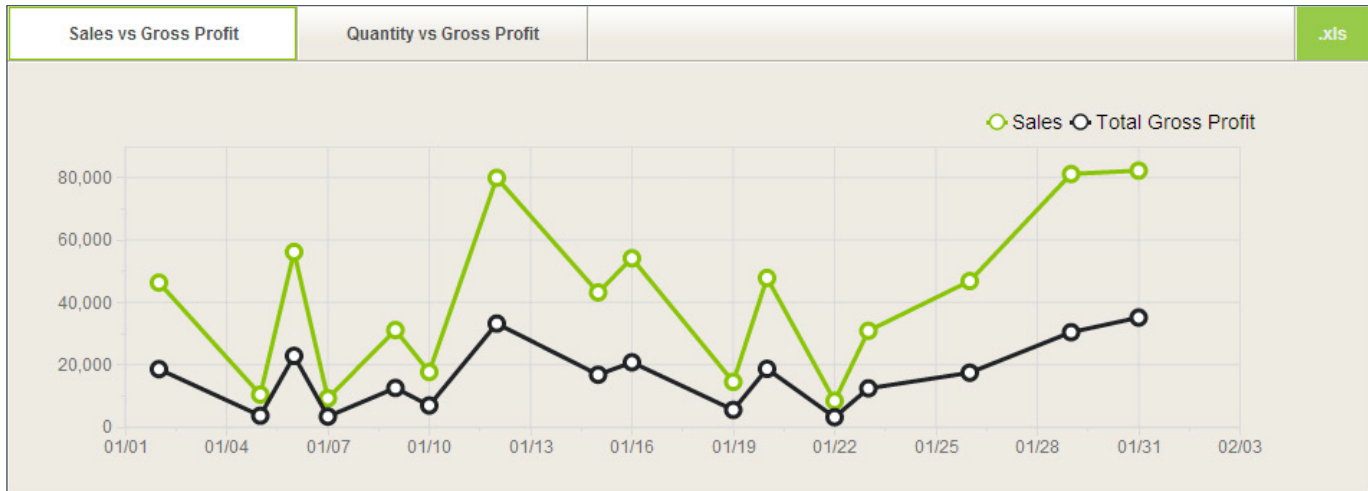
Unobtrusive drop-down selector for time frame with both a predefined period like "last week" and a small calendar for ad hoc date ranges.



A period selector on an overview time series chart drives the timing of the KPIs displayed below.

Hover and Drill Down Interaction

“Mouse overs” and clicks by users can provide great opportunities for interaction. For instance, when the user hovers over an item on the dashboard, such as an interactive menu or a data point on a chart, it can be useful to trigger some sort of highlighting or shading to make that point stand out. At the same time, providing additional information about the data point can be helpful: This could be as simple as labeling the precise value of a single data point on a line graph when the user hovers over it. You could also choose to provide tips and hints upon hover as well: For example, indicate that certain variables can be sorted.



Simple illustration of data point information displayed on hover.

Allowing users to drill down to more granular underlying information can also be helpful, especially for business analysts and other more sophisticated data consumers. The crucial thing is to remember that the data and context presented upon click must be clearly distinct from the original view. This could either be in the form of a “pop-out” window that displays the drill-down information or a zoom-in from the initial view, depending on your requirements. Once the user has accessed the more granular view, there must be an easy way to return to the original visualization, such as a “back” button.

Finally, users often want to be able to export the dashboard data, most often to Microsoft Excel, PDF or CSV format. This is most often enabled by a button or menu for the task.

Going Further

We hope this piece has given you some practical tips to enhance the success of your project. As you finalize your designs and move to implementation, remember that the project doesn’t end with rollout to your users. Requirements will continue to change over time and so will your analytics content. As such, it is absolutely critical for you to anticipate the longer-term requirements of your users and other strategic stakeholders inside and outside the organization. You’ll also want to make sure that the data and analytics software you use is future-proofed. Ensure that it allows you to easily customize and iterate on both the end user analytics and the types of data you want to incorporate in your dashboards.



Learn about Hitachi Vantara's Pentaho platform and its analytics capabilities at HitachiVantara.com



About Hitachi Vantara

Hitachi Vantara, a wholly owned subsidiary of Hitachi, Ltd., helps data-driven leaders find and use the value in their data to innovate intelligently and reach outcomes that matter for business and society. We combine technology, intellectual property and industry knowledge to deliver data-managing solutions that help enterprises improve their customers' experiences, develop new revenue streams, and lower business costs. Only Hitachi Vantara elevates your innovation advantage by combining deep information technology (IT), operational technology (OT) and domain expertise. We work with organizations everywhere to drive data to meaningful outcomes. Visit us at HitachiVantara.com.

Hitachi Vantara



Corporate Headquarters
2535 Augustine Drive
Santa Clara, CA 95054 USA
HitachiVantara.com | community.HitachiVantara.com

Contact Information
USA: 1-800-446-0744
Global: 1-858-547-4526
HitachiVantara.com/contact

HITACHI is a registered trademark of Hitachi, Ltd. Pentaho is a trademark or registered trademark of Hitachi Vantara Corporation. All other trademarks, service marks and company names are properties of their respective owners.

P-026-B BTD April 2019