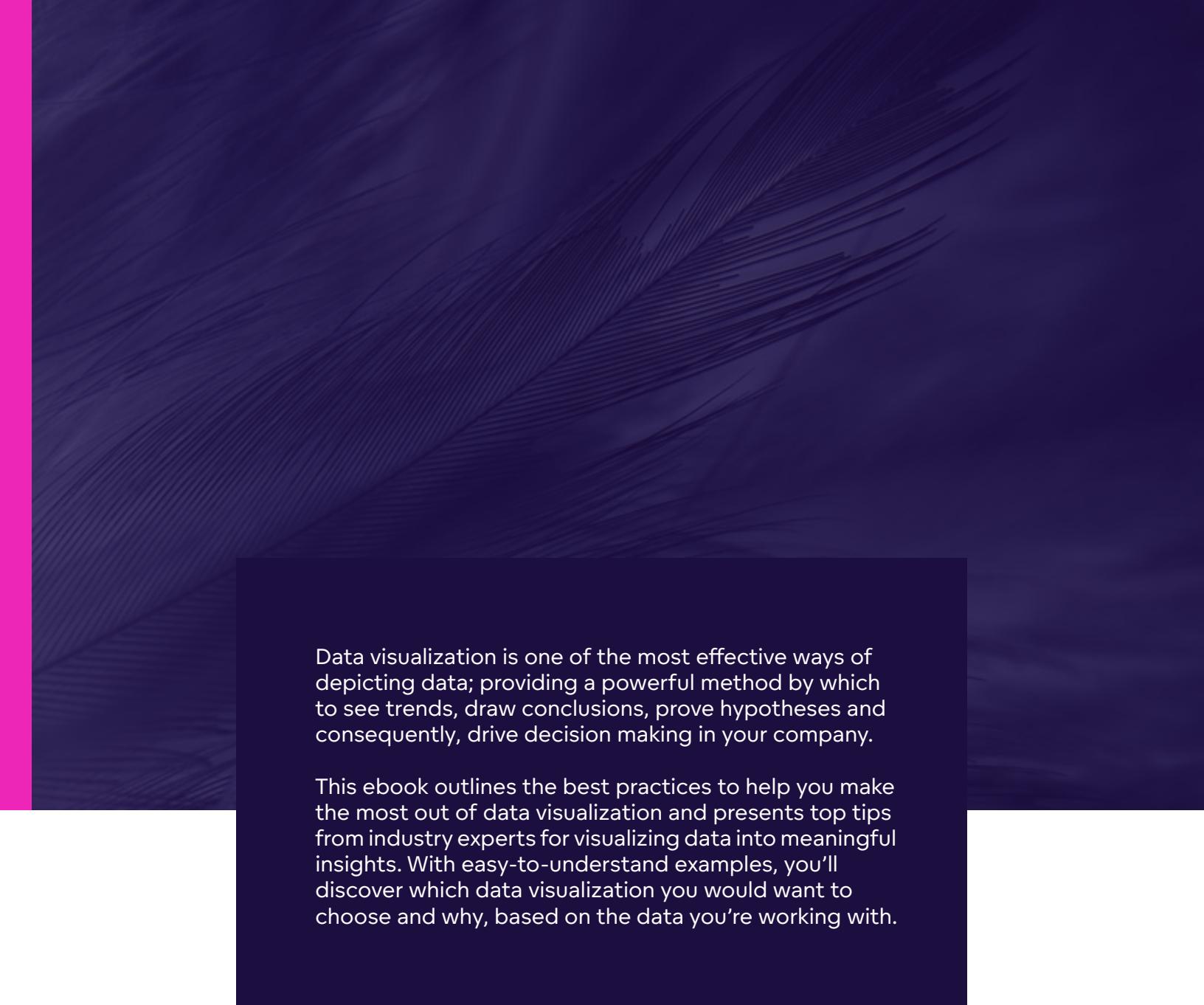


# Best Practices for Data Visualizations





Data visualization is one of the most effective ways of depicting data; providing a powerful method by which to see trends, draw conclusions, prove hypotheses and consequently, drive decision making in your company.

This ebook outlines the best practices to help you make the most out of data visualization and presents top tips from industry experts for visualizing data into meaningful insights. With easy-to-understand examples, you'll discover which data visualization you would want to choose and why, based on the data you're working with.

# Chapter 1: What should you keep in mind when working with data?

The right data visualizations can help you tell a story and help your audience figure out what actions they should take next. Consequently, making sure that your audience understands what the data visualization is communicating is critical.

These best practices are intended not only for analysts and data scientists but also for regular analytics users who are looking to better communicate the results of their ad-hoc analysis. Whether you're using some open source visualization library or mature business intelligence or data analytics tools like the GoodData platform, you can apply these tips to improve your data visualizations.

## Tip 1: Avoid displaying too much data

Displaying too much data can easily overwhelm your users, and I find that tables are one of the visualizations most prone to this issue. Yes, the table is one of the most universally understood visualizations, but take a step back to think critically about whether a table is telling the right story and whether your data analysis tool offers a visualization better suited to communicate the data.

If you chose to go forward with using a table, follow these best practices to help your users make sense of what they're seeing:

- ▶ Filter data only to top/bottom N rows. Some data may be irrelevant for decision making, and a filter can help users determine which data is most important.
- ▶ If you're choosing to use a table to increase trust in the data by showing what's behind the numbers, consider making the detailed data from your sources available only on demand.
- ▶ If you use pivot tables, consider using conditional formatting or heat map visualization to make it easier for users to find hot spots in the presented data set.
- ▶ Think wisely about whether you need to display all the data labels in the table. Too many data labels may spoil your visualization and distract the user. The details can be made accessible using tooltip instead of displaying them directly in the visualization.

## Tip 2: Employ the right visualization for the story

Always think about what role a given visualization is playing in your data analysis. Here are some suggestions for which visualizations to use in some of the most common use cases.

### Trending

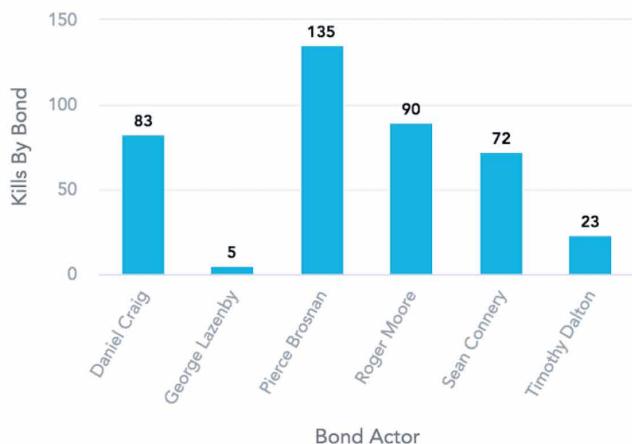
While a column chart is frequently used to display trends with a low number of data points, you can better use a line or area chart if the number of periods in a trend is high. You can also consider displaying forecasts as part of your predictive analytics by using one of the predictive models available in your analysis tool.

### Comparison

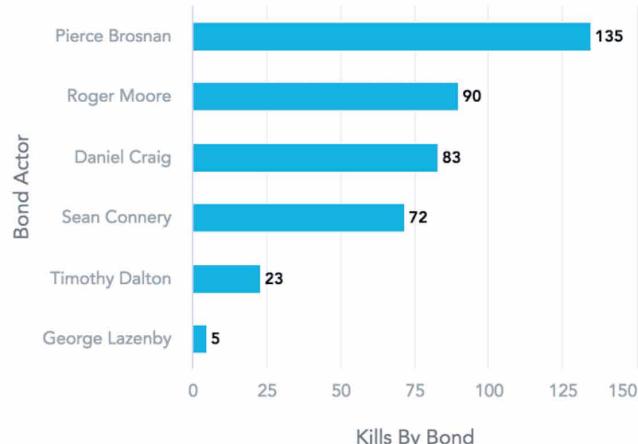
The column chart is often used for comparing data that is ordered by value. If there are too many columns, axis labels may not fit. Many users solve the problem by rotating the label, but consider using a bar chart with horizontal bars instead. This way, there is more space to display the axis label. Ordering by values also helps with comparing the data.

**Example: The chart on the right is easier to read, and the differences are more apparent.**

Kills by James Bond (column chart)



Kills by James Bond (bar chart)



Comparison of measures with different units

Measures with different units in one chart may become invisible due to differences in the ranges of values.

- If you need to compare measures with different units, you can of course use a table, but also consider splitting into multiple charts or using one of the techniques mentioned below.

- ▶ You can use a dual axis chart to compare two measures of different scales.
- ▶ You can also use a combo chart to ensure that data with different axis scales can be more easily differentiated.

**Example: Comparison of metrics with different scales is possible in the right chart.**

James Bond Movies (column chart)



James Bond Movies (combo chart)

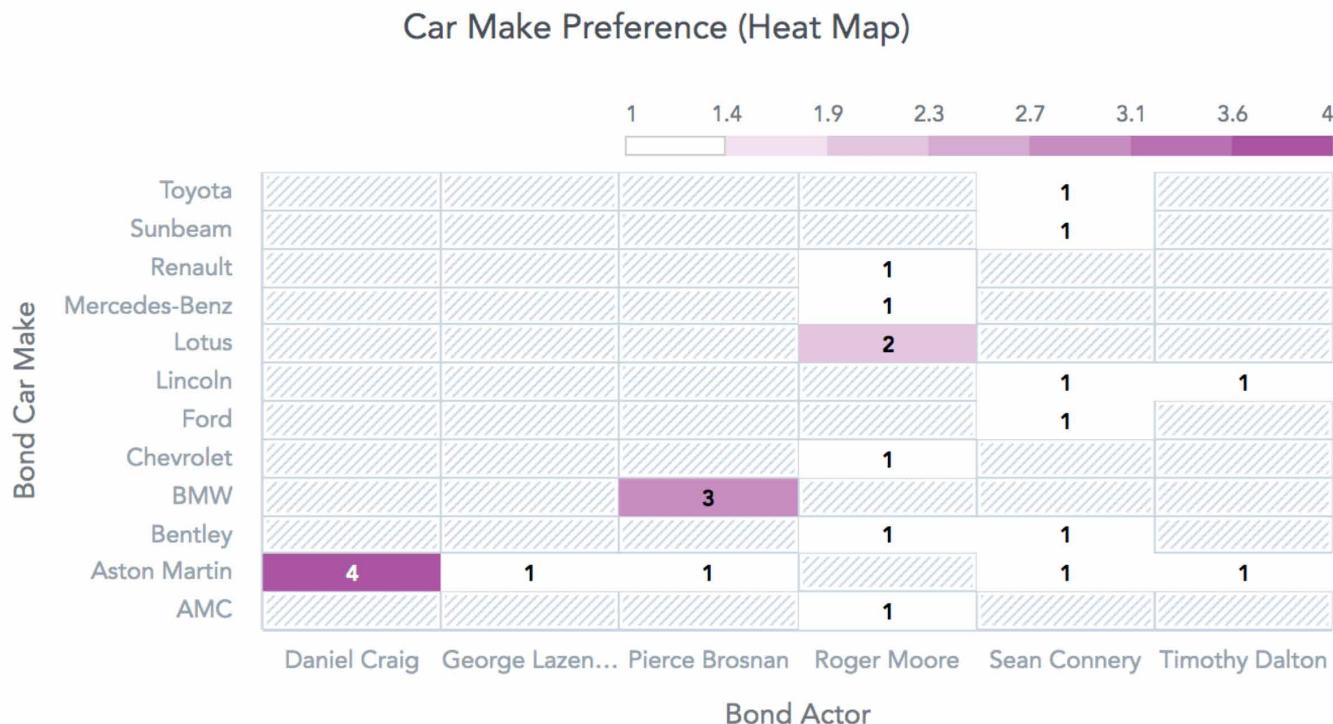


Correlation/Relationship

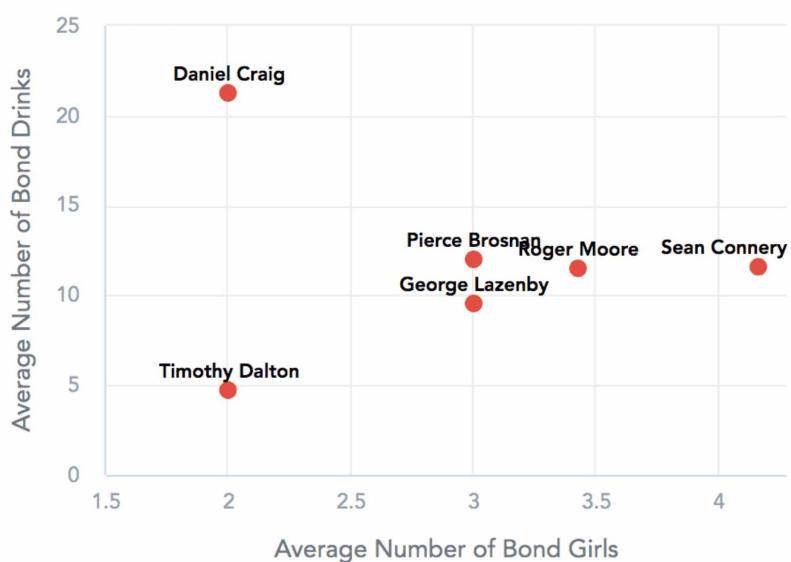
Measures with different units in one chart may become invisible due to differences in the ranges of values.

Depending on the type of data you're presenting, you can also use scatter plots, bullet charts, or heat maps.

**Example: You can see the differences in James Bond's preferences depending on the actor who played him.**



Drinks vs. Romantic Liaisons (Scatter Plot)

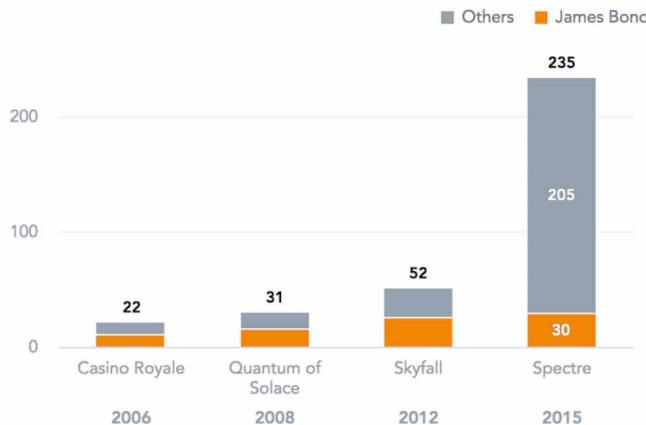


Contribution (parts-to-a-whole relationship)

To convey this kind of relationship, you can use stacked charts, tree maps, or donut charts. See the next tip for additional considerations when doing this type of analysis.

**Example: The chart on the left shows the total number of individual participants and their contributions. Differences between contributions and changes over time are more apparent in the chart on the right.**

Kills in Movies with Daniel Craig (Stacked)



Kills in Movies with Daniel Craig (Percent Stacked)



Contribution (parts-to-a-whole relationship)

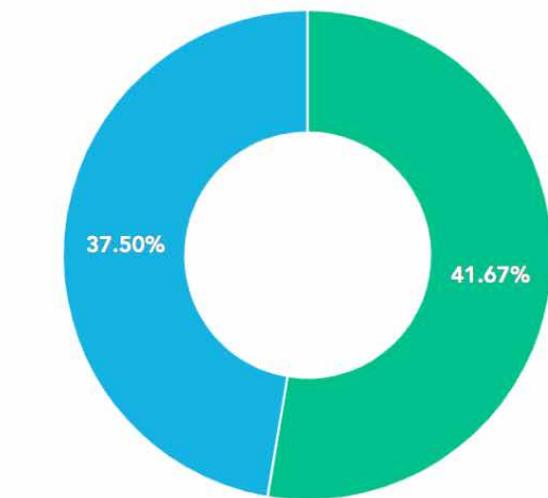
### Tip 3: Avoid common mistakes with non-additive measures

You should be careful when using measures that are not additive or have negative values. For example, adding up median values does not provide meaningful results. Some best practices include:

- ▶ Do not visualize non-additive measures in pie or donut charts.
- ▶ Do not stack non-additive measures.
- ▶ Use waterfall charts to visualize positive and negative increments alike.

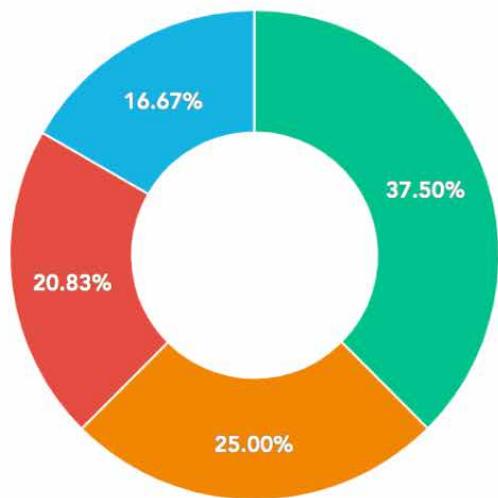
**Example: Some James Bond movies were nominated for both Golden Globe Awards and Academy Awards, and some were not nominated at all. The chart on the left is wrong because the percentage of movies do not add up to 100%. The chart on the right provides a better overview of the nominated movies.**

Percentage of Movies with Nomination



■ % Movies with Golden Globe Nomination  
■ % Movies with Academy Award Nomination

James Bond Movies by Nominations



■ None ■ Only Golden Globe  
■ Only Academy Award ■ Both

#### Tip 4: Make the data easier to read

You can improve the readability of big numbers by using commas or by skipping unnecessary decimal points.

Within the GoodData platform, you can also shorten numbers by truncating values and adding letters like K, M, and B (thousands, millions, and billions) to make data shorter and easier to read.

**Example: Take a look at the world box office earnings for the last three James Bond movies using three different formats. Which one is easiest to read?**

Quantum of Solace

**\$591692078.00**

Skyfall

**\$1,108,561,108**

Spectre

**\$879.6 M**

#### Tip 5: Avoid using pie charts

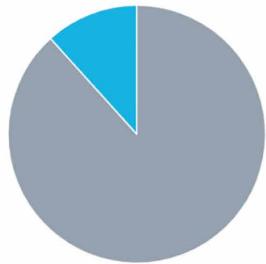
Pie charts are often used to display how segments contribute to the whole. However, the problem with this type of visualization is that the user needs to compare areas or angles, which can be difficult.

The center of a pie chart looks messy and overwhelming if the number of segments is too high or some of the segments are too small. You can try using some alternatives, such as:

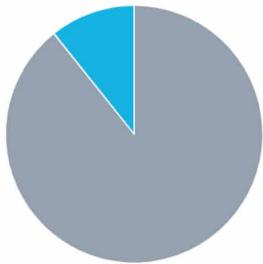
- ▶ Donut charts to eliminate the problem of comparing angles by focusing more on the length of the arc.
- ▶ Line charts to more clearly display trend of ratio than multiple pie chart visualizations can.

**Example: The kills share of Bond in Goldfinger and You Only Live Twice cannot be easily compared using a pie chart. The chart on the right communicates the information much more clearly.**

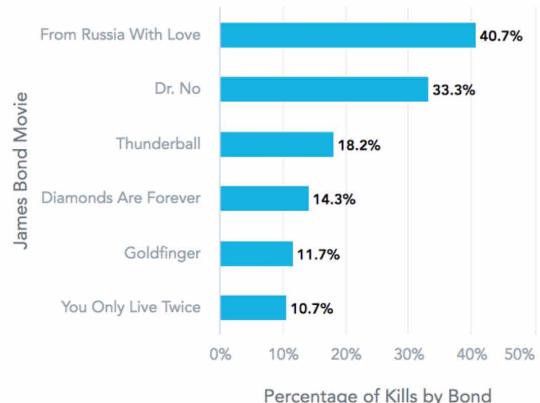
Kills in Goldfinger



Kills in You Only Live Twice



Percentage of Kills by Bond



## Tip 6: Match series colors

Colors in a chart are usually assigned from a common palette without any logic. I suggest you assign colors by following this methodology:

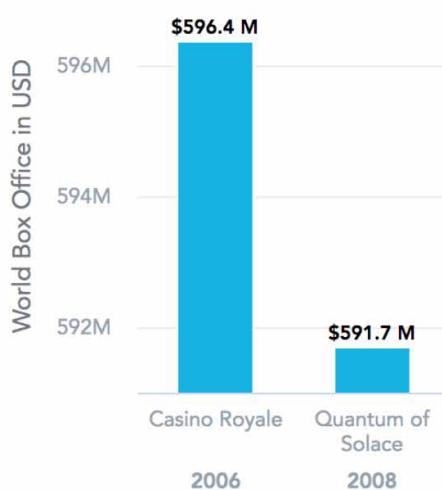
- ▶ Match series colors on multiple charts so that the legend does not need to be included in each chart and it is easier to look at the same data from different angles through multiple charts.
- ▶ Color data with intention. For example, if a series represents a brand and there is some well-understood color for that brand, you can use it to color the related data.
- ▶ Do not use a legend with too many colors, or users will not be able to differentiate the data. Keep the number of colors low to tell a good story clearly.

## Tip 7: Be careful when changing the axis scale

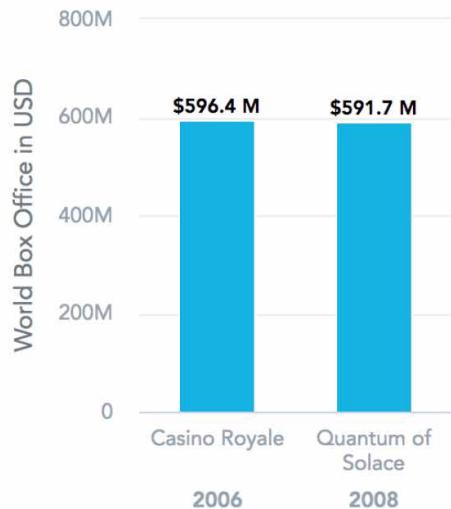
While it is easy to change the axis scale so that axis does not start at zero, it can confuse readers or make it appear as if you're attempting to intentionally mislead the reader. I suggest always starting with zero in a column and area chart to avoid giving the wrong signal to the reader. Change the axis scale only if you really need to show detail which would otherwise be invisible or difficult to spot.

**Example: The world box office earnings from Quantum of Solace seem to be very low in the chart on the left. In reality, its earnings were very similar to the earnings from Casino Royale, as can be clearly seen in the chart on the right.**

Axis scale does not start at zero



Axis scale starts at zero



## Chapter 2: Make it clear what you want to show

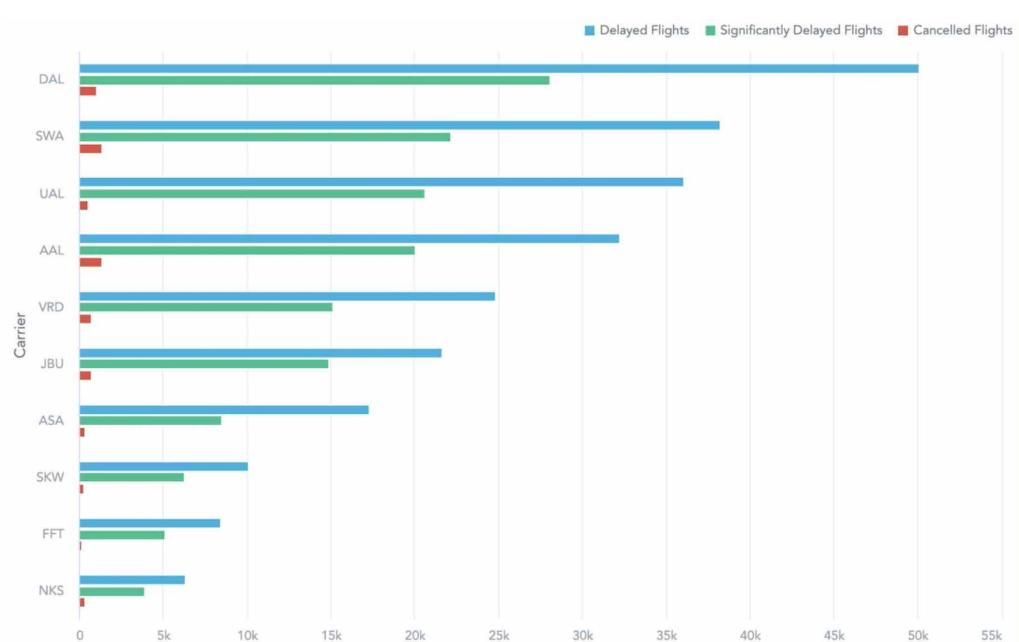
The same numbers can be interpreted and understood in completely different ways; it's up to you to choose the right visualization for the message you want to convey. In this section, we'll focus on how to determine the right visualization for your data based on the characteristics of the dataset you are working with.

## You want to show comparison

Comparison is a common reason for using visualizations. Nevertheless, there are many reasons why you would need to compare something, and each of those is best suited for a particular visualization.

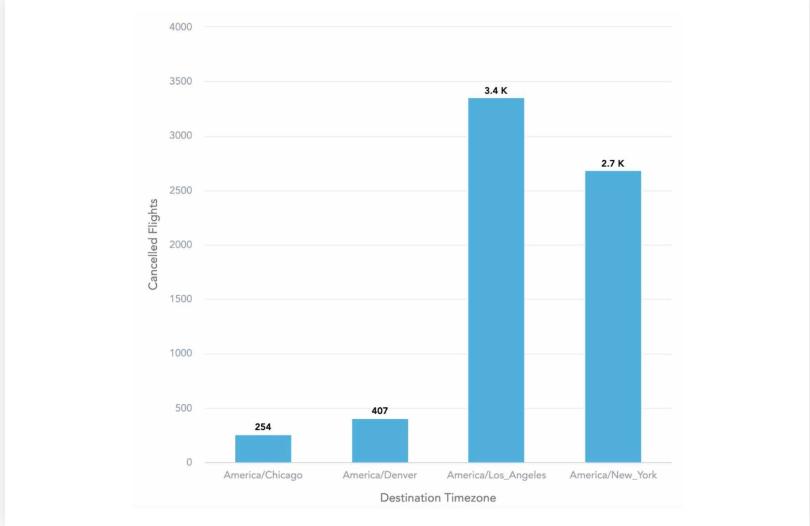
### You are comparing values among items

One of the most common reasons for comparison is comparing one variable per item in a few categories. If you have many items to compare, use a **bar chart**. In the example below, we compare flight carriers based on their delayed and cancelled flight stats.



Bar charts allow for better readability when many items are placed on the chart, and you can add multiple bar charts next to each other, such as on our **KPI Dashboard**, which leverages BI dashboard best practices.

On the other hand, if you are only comparing **a few items**, it's more appropriate to use a **column chart**, which provides better readability when the chart doesn't need to be too wide, such as showing the number of cancelled flights per time zone in the United States.



Other examples include:

- ▶ Comparison of stock items by value of storage costs
- ▶ Number of patients with a pollen allergy by city, focusing on the environment where they live

### You are comparing over time

Do you want to compare values over time? The key question to ask yourself is whether you need to compare over many periods or just a few periods.

For just a **few periods**, it's a smart idea to continue with a **column chart**, when you are comparing a single or couple of categories over time. However, if you want to compare many categories, it's recommended to use a **line chart**, which will be easier to read than multiple columns. For instance, take a look at this chart showing the trend of delayed flights over 2019.



Also, if you need to compare data over many periods, a line chart is the simplest solution to go with if you are comparing non-cyclical data.

Other examples include:

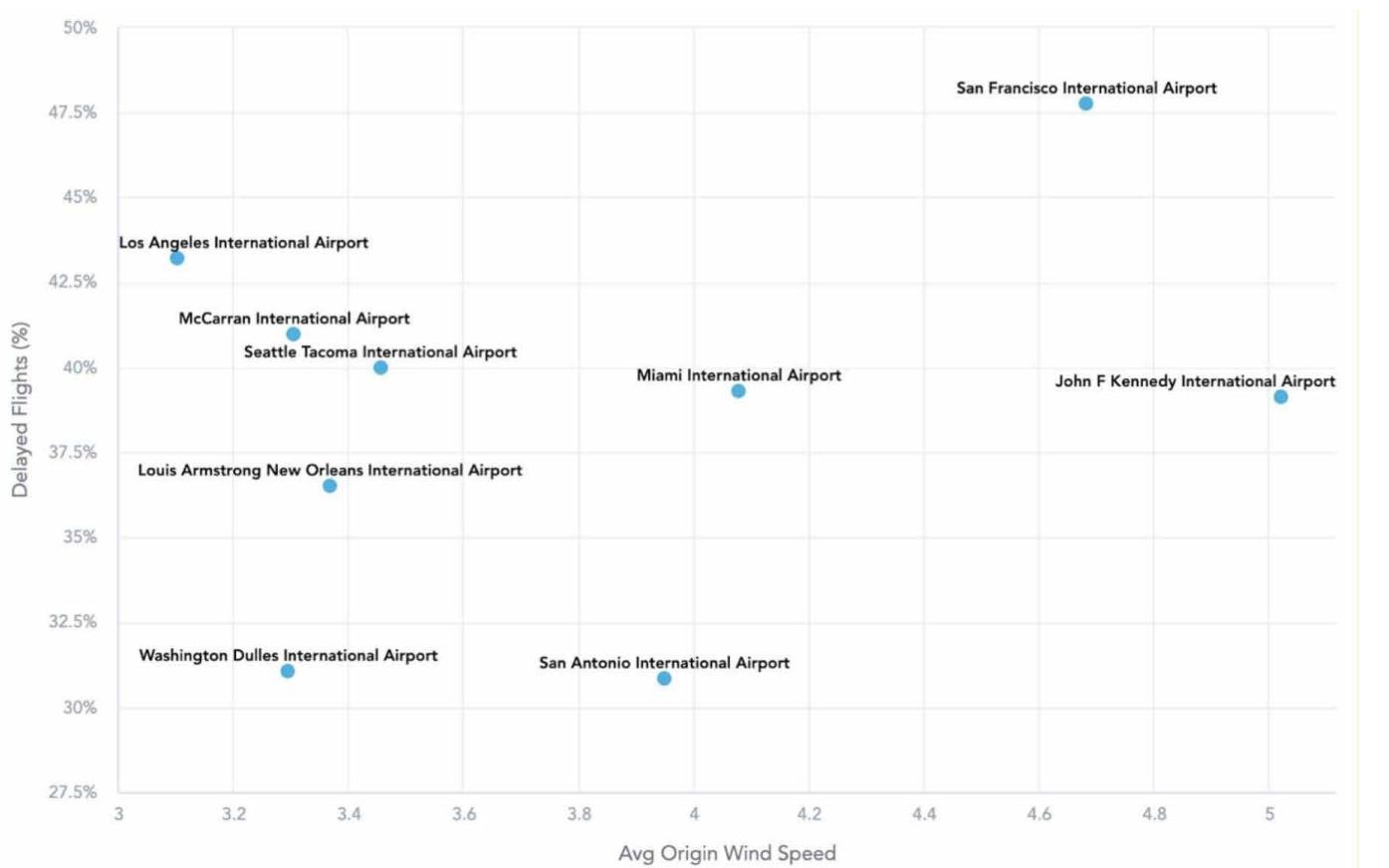
- ▶ Company revenue per product category compared to the previous year to identify weak points
- ▶ Number of cinema visitors compared by week, grouped by age of visitors

## You want to show a relationship

When you're trying to illustrate a relationship between two or three variables, choosing the right chart is much simpler. Let's take a look at data visualization use cases for relationships.

### A relationship between two variables

To visualize a relationship between two variables, a **scatter plot** (or scatter chart) is the way to go. It is a type of diagram that is based on Cartesian coordinates, where the data is displayed as a collection of data points (therefore "scatter" plot or scatter chart) having value on a horizontal and vertical axis based on the value of a variable. On the scatter plot below, you can see the relationship between average wind speed at the origin airport and percentage of delayed flights there.



Other examples include:

- ▶ Estimate how fuel costs to salary affect the overall traffic in countries
- ▶ Determine whether salary depends on education level

### A relationship among three variables

As with a scatter plot, creating **bubble charts** is a good choice when visualizing relationships. In this case, we have three variables to compare. As with a scatter plot, the two variables are represented by Cartesian coordinates corresponding with a value on the horizontal or vertical axis. The third variable is visualized by the size of the bubble. Hence, we can think of bubble charts as a variation of scatter plots, where the particular point was transformed into a bubble by adding the third variable.

Other examples include:

- ▶ Campaign costs spent on a specific day with the revenue based on the campaign (bubble size will be the revenue)
- ▶ Calories consumed and time spent at the gym with weight loss

### You want to explain composition

The goal of composition is to visualize particular parts in the whole as a sum of their combination. Another reason for using composition is to explain the total value split and emphasize the importance of individual parts to the total value.

### A composition which changes over time

The first question you should ask when deciding which chart to use is whether the composition changes over time or not. If yes, the second question is whether it changes over a few periods or many periods, and whether only relative or both relative and absolute differences matter. Let's take a look at the examples below where we show the share of departure airports per state over the year, in terms of either time spent in the air or total flight numbers.

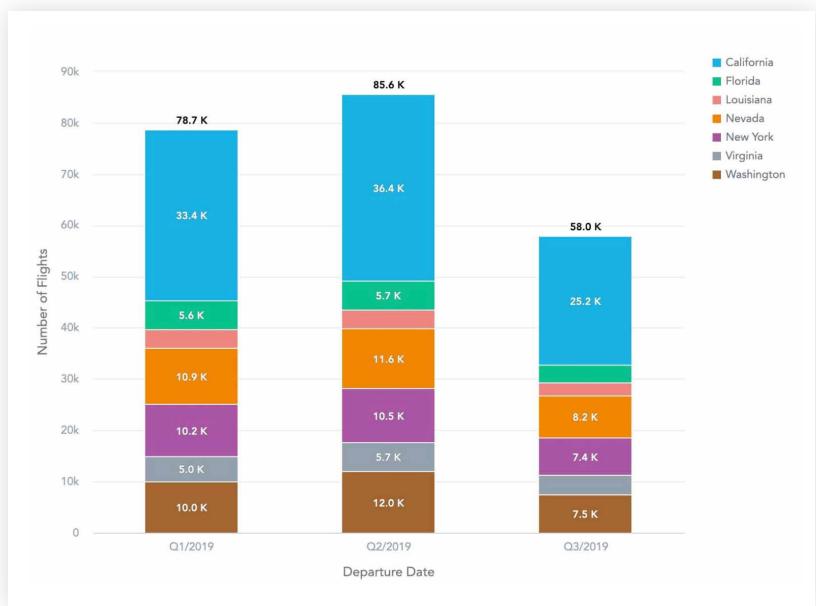
To ensure better readability, a **stacked column chart** is great when visualizing a few periods.

The difference whether a stacked column chart should be stacked to 100% or not is:

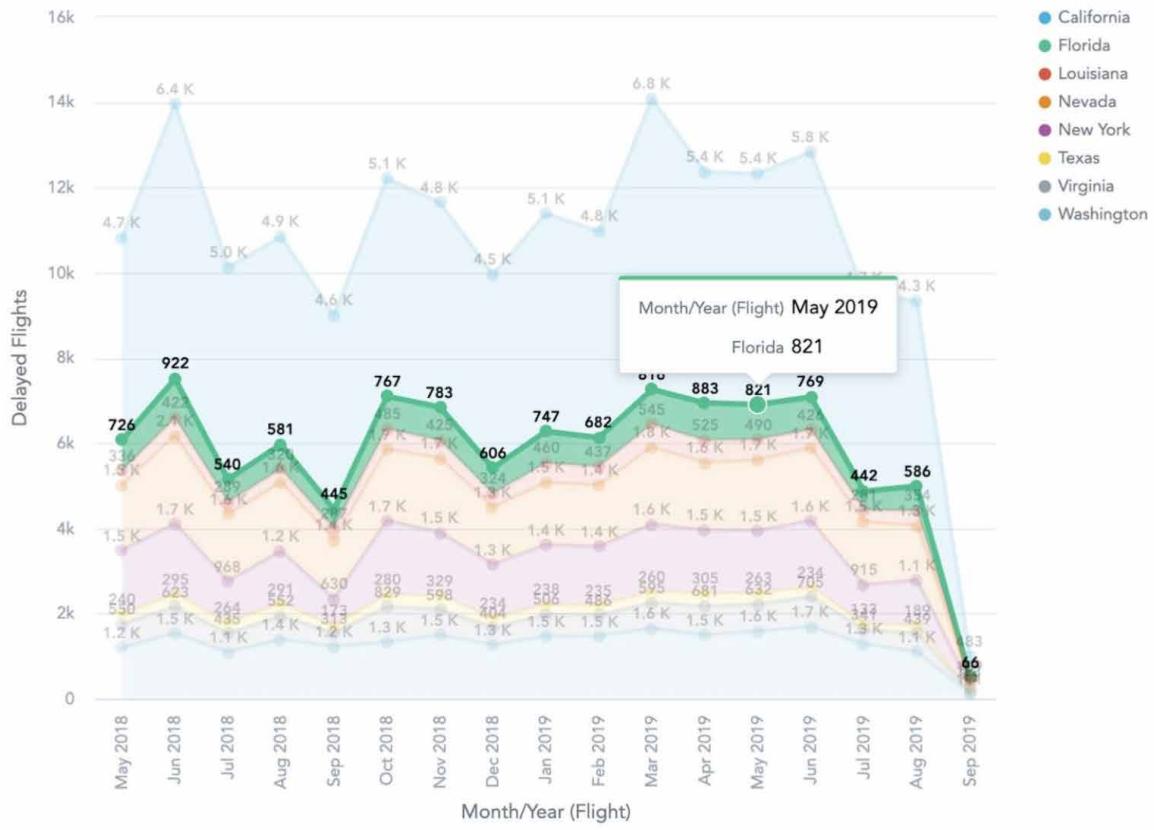
- ▶ If only relative differences matter, use a stacked column chart to 100%.



- If both relative and absolute differences matter (i.e. also specific numbers), use a standard stacked column chart.



When you visualize changing composition over many periods, a **stacked area chart** is what you need. Again, the decision of whether to keep it to 100% or not depends on if only relative difference matters or if the absolute difference matters as well.

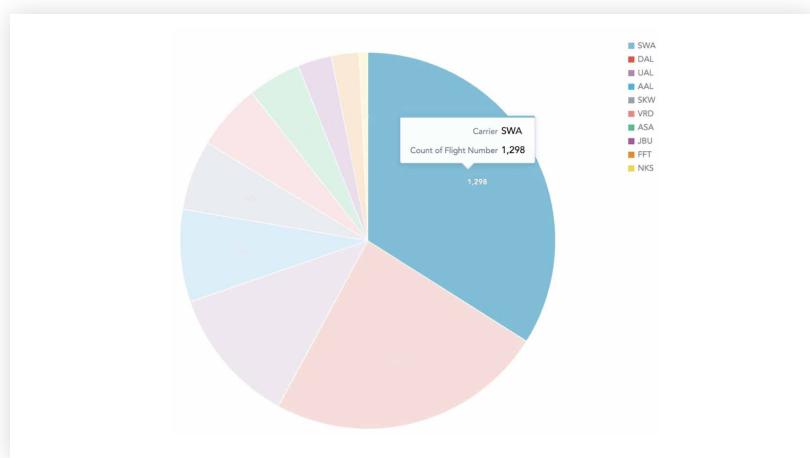


Other examples include:

- ▶ Company market share in a specific region over time
- ▶ Company revenue by product categories over time

### A composition which is static

A static composition, which doesn't change over time, is the perfect match for the famous **pie chart**, where you show the simple share of the total result. In the example below, you can see the total number of departing flights operated by a given airport. However, as we already mentioned in the previous chapter, you have to act carefully when using pie charts. Pie charts are not suitable when showing accumulation or subtraction to total. In this situation, a waterfall chart would be a better fit.



Other examples include:

- ▶ Company revenue by region
- ▶ World population by religion
- ▶ Market share with a few competitors

## You want to describe distribution

Describing distribution depends on the number of variables you're using.

- ▶ For two variables, stick with a **scatter chart**, which we have already used for visualizing relationships.
- ▶ For three variables, a 3D area chart is the way to go.

Examples include:

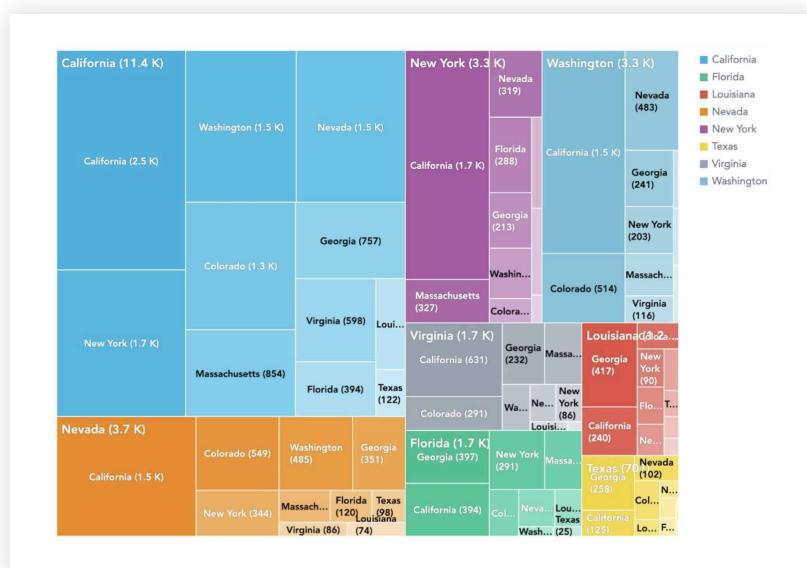
- ▶ Visualizing the distribution of car age in the car market
- ▶ Localizing outliers in production line control results

## Treemaps and heatmaps

If you have already tried GoodData, such as through [GoodData Free](#), you have noticed some charts we haven't talked about yet: treemaps and heatmaps.

Whenever you want to analyze data that is in a hierarchy, **treemaps** are the right choice. Treemaps are typically structured to a set of nested rectangles. The rectangle has a proportional size to a dimension of the data, colored to show separate dimensions of the data.

In this example, you can see the structure of target destinations per state and the total number of flights. When you see California under California, it means that 2.5K flights have been operated within California only.



Other examples include:

- ▶ Product inventory in stock by product group, then product category, and then product time by level of space required to store them
- ▶ Distribution of animal types in a zoo by class, order, and species with their count

To use a **heatmap** as a visualization, you typically have a table with different values to which you want to give graphical emphasis. The higher the value, the darker the color. There are multiple forms of heatmaps used in analytics, such as a heatmap over a map or over a website to track the behavior of users.

In the GoodData platform, you'll see the most common type: a heatmap table, which is a click map. In the example below, you can see the heatmap of cancelled flights over nine months based on the state where the planes departed from.



Other examples include:

- ▶ Sales heatmap to visualize best sales figures over a week at specific time
- ▶ Most delayed flights by carrier over a specific time period



---

## Learn more about GoodData

GoodData's API-first analytics platform enables you to build analytics solutions like any other software and provide consistent metrics to all end users and data tools. To learn more about headless BI, GoodData, or how we can help you with your analytics use cases, you can try [GoodData's fully managed, API-first analytics platform for free](#) or [schedule a commitment-free demo](#) with a GoodData specialist.



[Start Trial](#)



[Article](#)

Interactive Data Visualization