**Pre-attentive attributes: marks and channels**

Creating effective visuals means leveraging what we know about how the brain works, and then using specific visual elements to communicate the information effectively. **Pre-attentive attributes** are the elements of a data visualization that people recognize automatically without conscious effort. The essential, basic building blocks that make visuals immediately understandable are called marks and channels.

**Marks**

**Marks** are basic visual objects like points, lines, and shapes. Every mark can be broken down into four qualities:

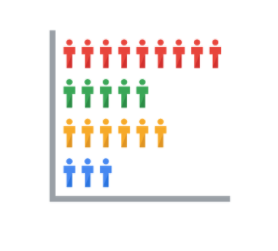
1. **Position** - Where a specific mark is in space in relation to a scale or to other marks



2. **Size** - How big, small, long, or tall a mark is



3. **Shape** - Whether a specific object is given a shape that communicates something about it



4. **Color** - What color the mark is

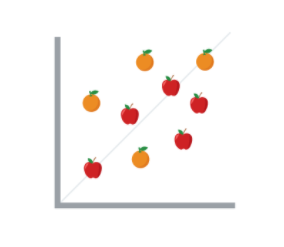


**Channels**

**Channels** are visual aspects or variables that represent characteristics of the data. Channels are basically marks that have been used to visualize data. Channels will vary in terms of how effective they are at communicating data based on three elements:

1. **Accuracy** - Are the channels helpful in accurately estimating the values being represented?

For example, color is very accurate when communicating categorical differences, like apples and oranges. But it is much less effective when distinguishing quantitative data like 5 from 5.5.



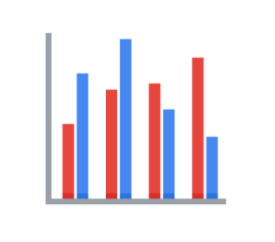
2. **Popout** - How easy is it to distinguish certain values from others?

There are many ways of drawing attention to specific parts of a visual, and many of them leverage pre-attentive attributes like line length, size, line width, shape, enclosure, hue, and intensity.



3. **Grouping** - How good is a channel at communicating groups that exist in the data?

Consider the proximity, similarity, enclosure, connectedness, and continuity of the channel.



But, remember: the more you emphasize different things, the less that emphasis counts. The more you emphasize one single thing, the more that counts.

# The wonderful world of visualizations

As a data analyst, you will often be tasked with relaying information and data that your audience might not readily understand. Presenting your data visually is an effective way to communicate complex information and engage your stakeholders. One question to ask yourself is: “what is the best way to tell the story within my data?” This reading includes several options for you to choose from (although there are many more).

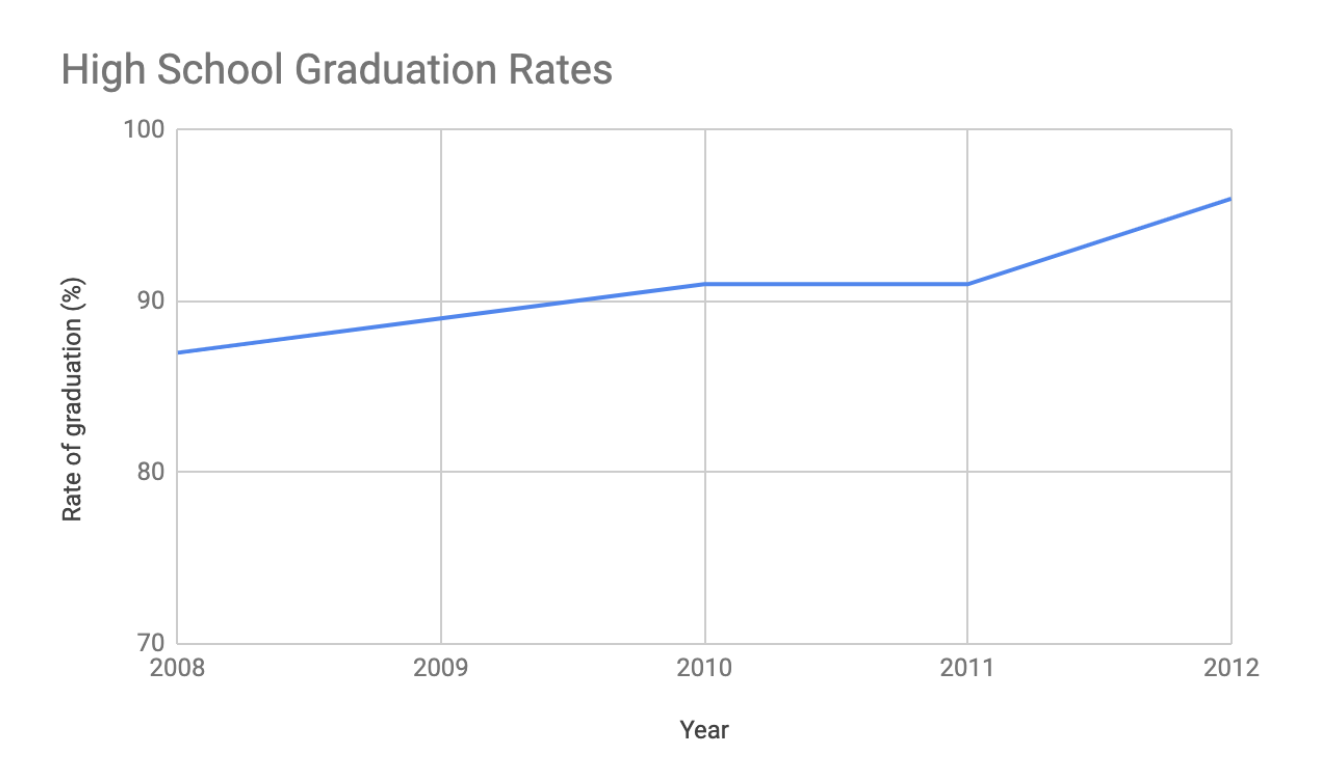
## Line chart

A **line chart** is used to track changes over short and long periods of time. When smaller changes exist, line charts are better to use than bar graphs. Line charts can also be used to compare changes over the same period of time for more than one group.

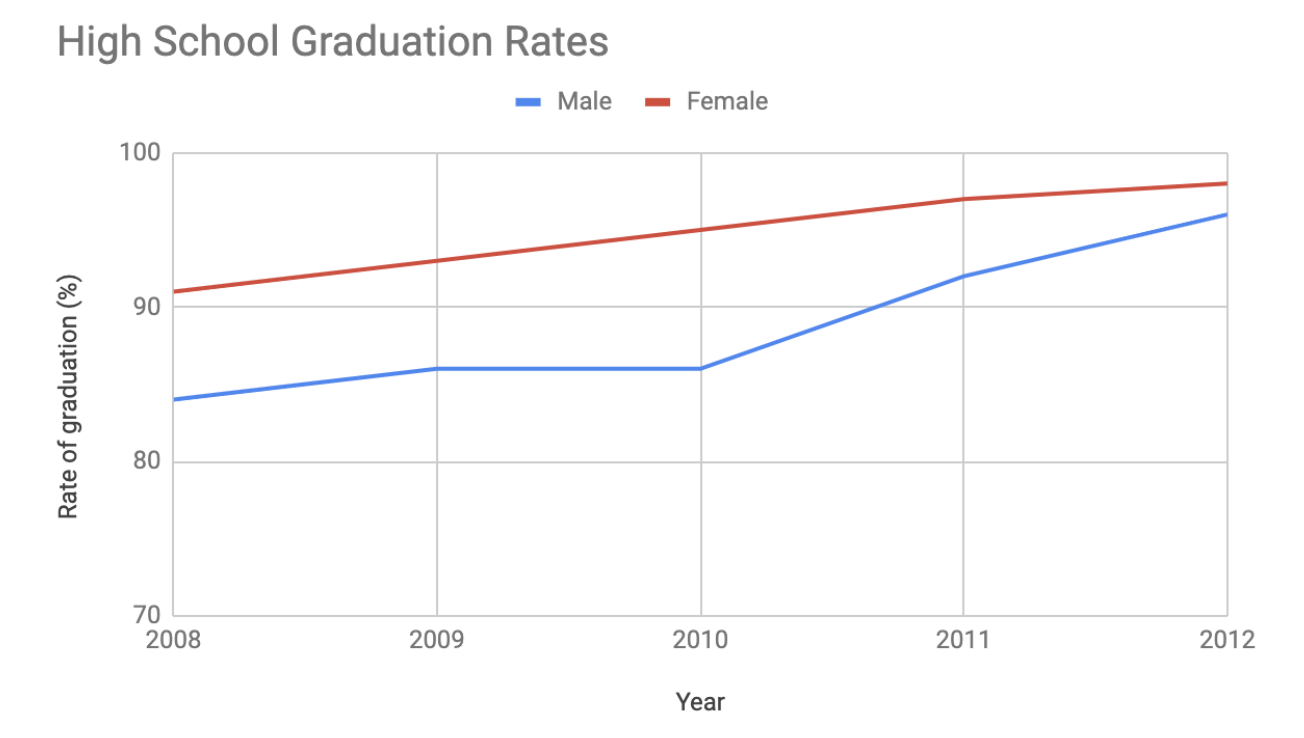
Let’s say you want to present the graduation frequency for a particular high school between the years 2008-2012. You would input your data in a table like this:

| **Year** | **Graduation rate** |
| --- | --- |
| 2008 | 87 |
| 2009 | 89 |
| 2010 | 92 |
| 2011 | 92 |
| 2012 | 96 |

From this table, you are able to present your data in a line chart like this:



Maybe your data is more specific than above. For example, let’s say you are tasked with presenting the difference of graduation rates between male and female students. Then your chart would resemble something like this:



## Column chart

**Column charts** use size to contrast and compare two or more values, using height or lengths to represent the specific values.

The below is example data concerning sales of vehicles over the course of 5 months:

| **Month** | **Vehicles sold** |
| --- | --- |
| August | 2,800 |
| September | 3,700 |
| October | 3,750 |
| November | 4,300 |
| December | 4,600 |

Visually, it would resemble something like this:

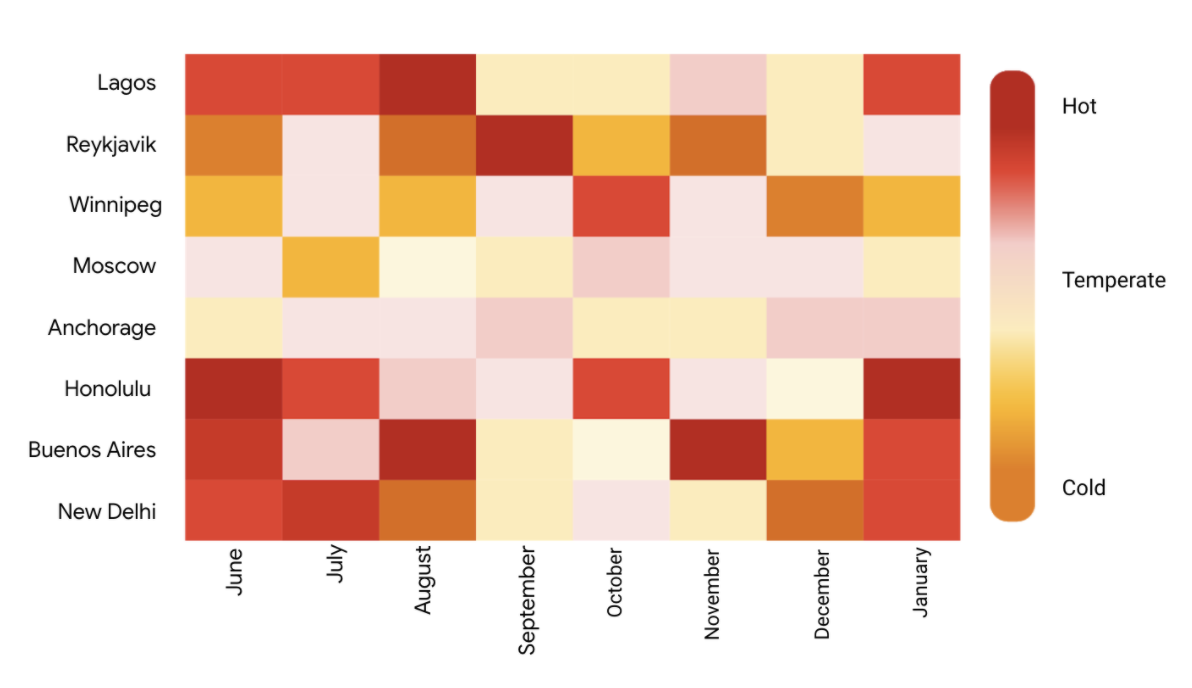


What would this column chart entail if we wanted to add the sales data for a competing car brand?



## Heatmap

Similar to bar charts, **heatmaps** also use color to compare categories in a data set. They are mainly used to show relationships between two variables and use a system of color-coding to represent different values. The following heatmap plots temperature changes for each city during the hottest and coldest months of the year.



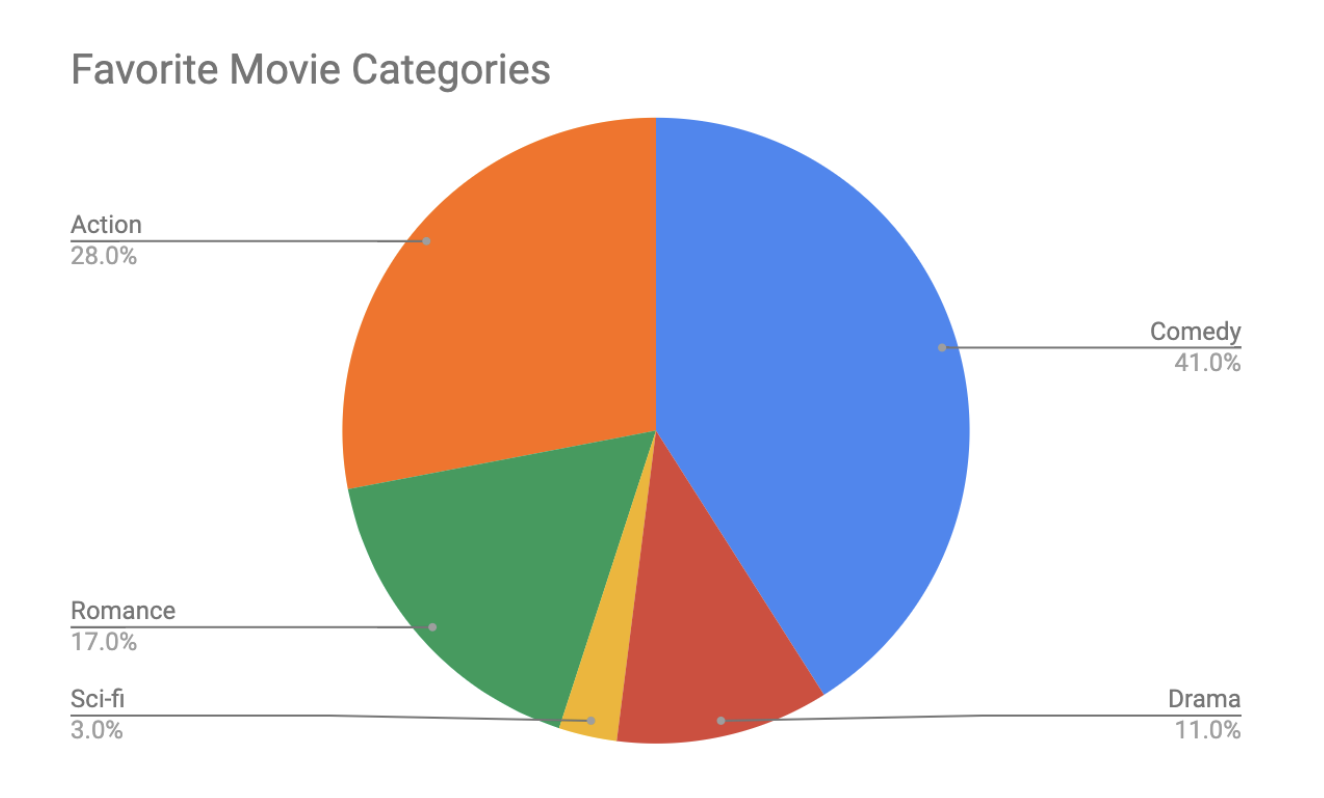
## Pie chart

The **pie chart** is a circular graph that is divided into segments representing proportions corresponding to the quantity it represents, especially when dealing with parts of a whole.

For example, let’s say you are determining favorite movie categories among avid movie watchers. You have gathered the following data:

| **Movie category** | **Preference** |
| --- | --- |
| Comedy | 41% |
| Drama | 11% |
| Sci-fi | 3% |
| Romance | 17% |
| Action | 28% |

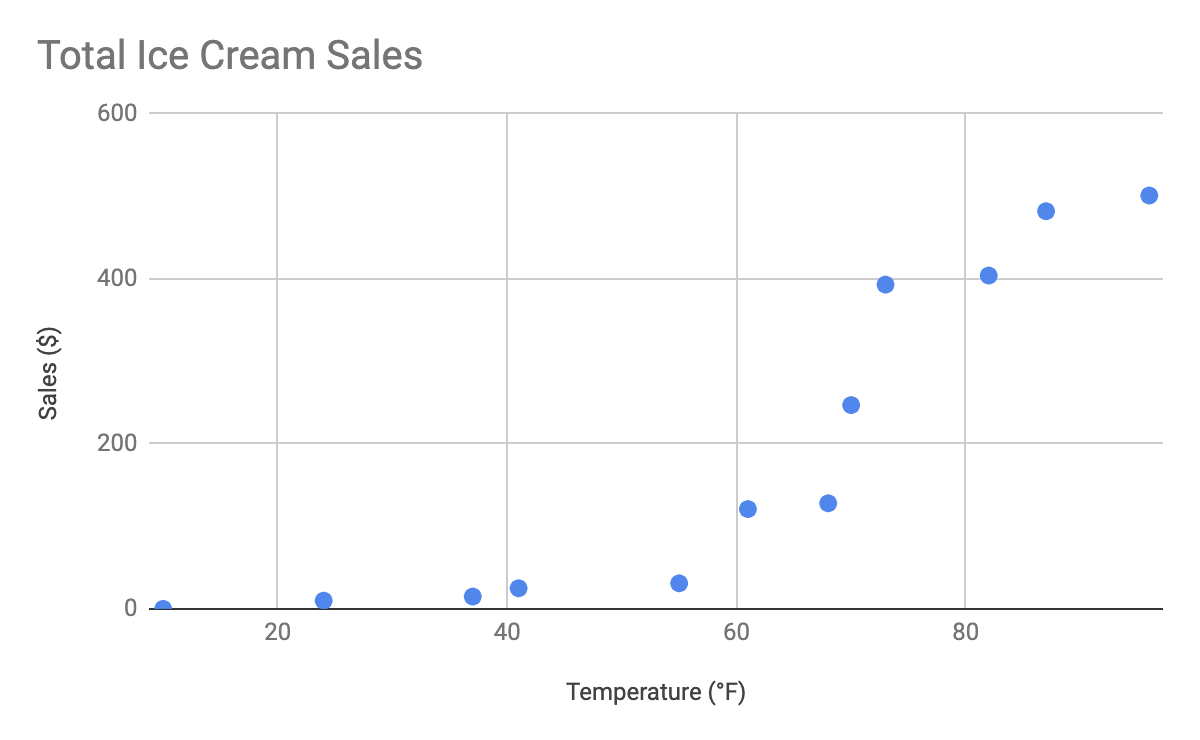
Visually, it would resemble something like this:

Action- 28% Comedy- 41% Romance- 17% Sci-fi- 3% Drama- 11%

## Scatterplot

**Scatterplots** show relationships between different variables. Scatterplots are typically used for two variables for a set of data, although additional variables can be displayed.

For example, you might want to show data of the relationship between temperature changes and ice cream sales. It would resemble something like this:

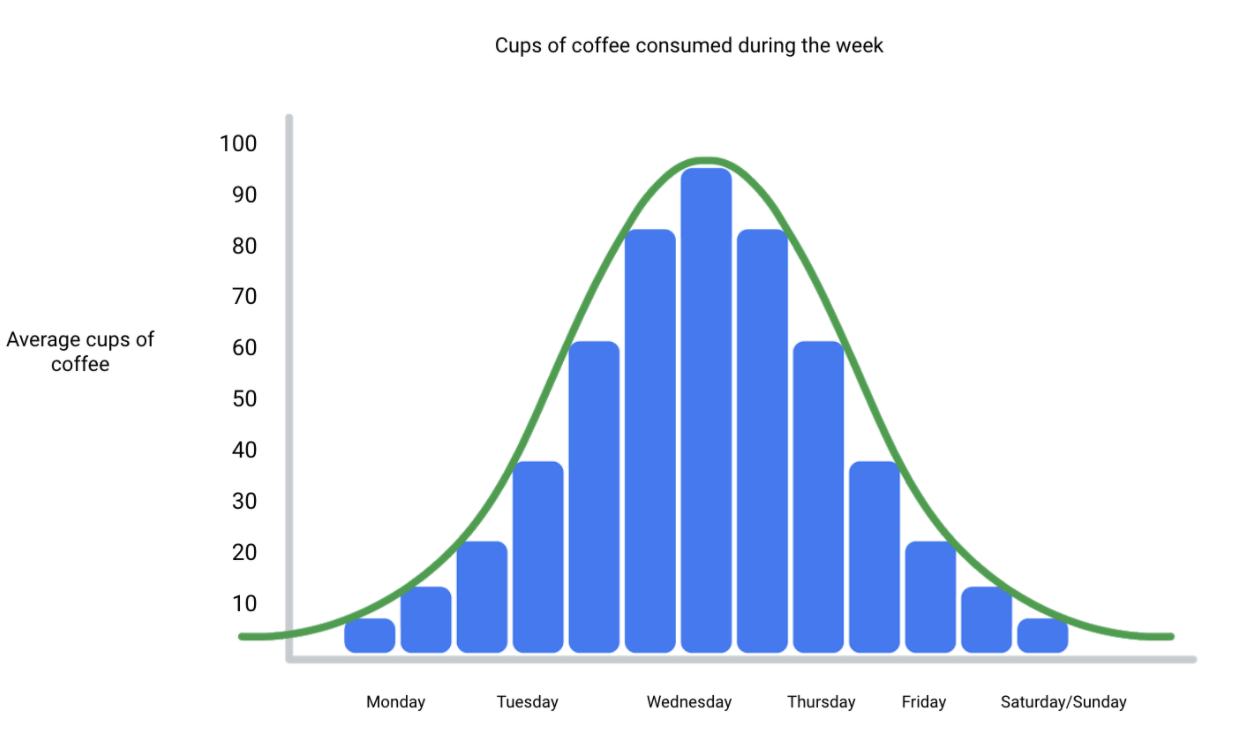


As you may notice, the higher the temperature got, the more demand there was for ice cream – so the scatterplot is great for showing the relationship between the two variables.

## Distribution graph

A **distribution graph** displays the spread of various outcomes in a dataset.

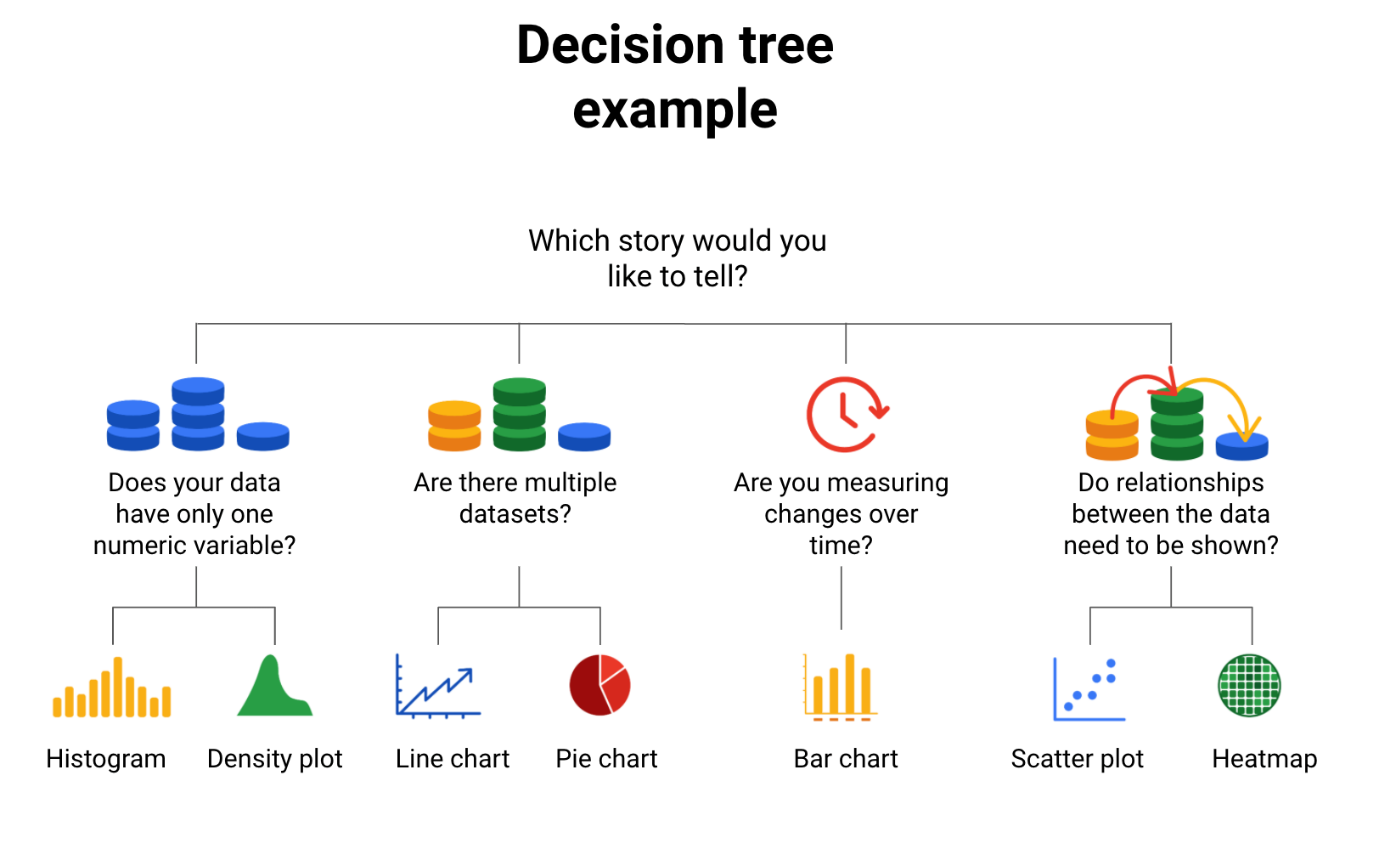
Let’s apply this to real data. To account for its supplies, a brand new coffee shop owner wants to measure how many cups of coffee their customers consume, and they want to know if that information is dependent on the days and times of the week. That distribution graph would resemble something like this:



From this distribution graph, you may notice that the amount of coffee sales steadily increases from the beginning of the week, reaching the highest point mid-week, and then decreases towards the end of the week.

If outcomes are categorized on the x-axis by distinct numeric values (or ranges of numeric values), the distribution becomes a **histogram**. If data is collected from a customer rewards program, they could categorize how many customers consume between one and ten cups of coffee per week. The histogram would have ten columns representing the number of cups, and the height of the columns would indicate the number of customers drinking that many cups of coffee per week.

A **decision tree** is a decision-making tool that allows you, the data analyst, to make decisions based on key questions that you can ask yourself. Each question in the visualization decision tree will help you make a decision about critical features for your visualization. Below is an example of a basic decision tree to guide you towards making a data-driven decision about which visualization is the best way to tell your story. Please note that there are many different types of decision trees that vary in complexity, and can provide more in-depth decisions.



**Guidelines and pro tips**

Refer to the following table for recommended guidelines and style checks for headlines, subtitles, labels, and annotations in your data visualizations. Think of these guidelines as guardrails. Sometimes data visualizations can become too crowded or busy. When this happens, the audience can get confused or distracted by elements that aren’t really necessary. The guidelines will help keep your data visualizations simple, and the style checks will help make your data visualizations more elegant.

| **Visualization components** | **Guidelines** | **Style checks** |
| --- | --- | --- |
| Headlines | - **Content**: Briefly describe the data - **Length**: Usually the width of the data frame - **Position**: Above the data | - Use brief language - Don’t use all caps - Don’t use italic - Don’t use acronyms - Don't use abbreviations - Don’t use humor or sarcasm |
| Subtitles | - **Content**: Clarify context for the data - **Length**: Same as or shorter than headline - **Position**: Directly below the headline | - Use smaller font size than headline - Don’t use undefined words - Don’t use all caps, bold, or italic - Don’t use acronyms - Don't use abbreviations |
| Labels | - **Content**: Replace the need for legends - **Length**: Usually fewer than 30 characters - **Position**: Next to data or below or beside axes | - Use a few words only - Use thoughtful color-coding - Use callouts to point to the data - Don’t use all caps, bold, or italic |
| Annotations | - **Content**: Draw attention to certain data - **Length**: Varies, limited by open space - **Position**: Immediately next to data annotated | - Don’t use all caps, bold, or italic - Don't use rotated text - Don’t distract viewers from the data |