Data visualization is the graphic representation and presentation of data

**Effective data visualizations**

It can be difficult to understand data insights by examining individual data points or a table of information. Often, insights become more obvious when presented in an effective visual format. You can use data visualization (often called  “data viz”) techniques to help your audience interpret data in a concise, visual manner.

When creating data visualizations, you must strike a balance between presenting enough information for your audience to understand the meaning of the visualization and not overwhelming them with too much detail. In this reading, you’ll learn tips and techniques for crafting visualizations that are both impactful and effective. You’ll explore:

* Two frameworks for organizing data
* Pre-attentive attributes

**Frameworks for organizing your thoughts about visualization**

Frameworks help organize your thoughts about data visualization and give you a useful checklist to reference as you plan and evaluate your data visualization. Here are two frameworks that employ slightly different techniques. Both are intended to improve the quality of your visuals.

[The McCandless method](https://www.informationisbeautiful.net/visualizations/what-makes-a-good-data-visualization/)

You learned about the David McCandless method earlier in the course; as a refresher, the McCandless method lists four elements of good data visualization:

1. **Information:** the data with which you’re working
2. **Story:** a clear and compelling narrative or concept
3. **Goal:** a specific objective or function for the visual
4. **Visual form:** an effective use of metaphor or visual expression

The McCandless method provides terminology that isolates the specific elements of a graphic, allowing the person making a visual the ability to evaluate how well those criteria have been met. The aim when crafting a visualization is to incorporate all four elements effectively. Visualizations that fail to incorporate all four elements can be ineffective at communicating insights in various ways. For example, visual form without a goal, story, or data could be a sketch or even art. Data in visual form without a goal or function is just a pretty picture. Data with a goal but no story or visual form can be boring. All four elements need to be present to create an effective visual.

[Kaiser Fung’s Junk Charts trifecta checkup](https://junkcharts.typepad.com/junk_charts/junk-charts-trifecta-checkup-the-definitive-guide.html)

This approach is a set of questions that can help consumers of data visualization critique what they are consuming and determine how effective it is. You can also use these questions to determine if your data visualization is effective:

1. What is the practical question?
2. What does the data say?
3. What does the visual say?

Each of these questions offers an opportunity to investigate a given problem with a slightly different context. A well-designed visual effectively answers all three of those questions at once. Moreover, this framework helps you think about your data viz from the perspective of your audience.

**Pre-attentive attributes**

In addition to the frameworks mentioned above, several standard building blocks can help you construct your data visualizations. Creating effective visuals means leveraging what is known 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 and without conscious effort. The essential, basic building blocks that make visuals immediately understandable are called marks and channels.

**Marks**

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

1. **Position:** Where is a specific mark in space relative to a scale or to other marks?

For example, if you’re looking at two different trends, position allows you to compare the pattern of one element relative to another.



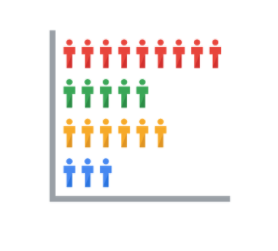
2. **Size:** How big, small, long, or tall is a mark?

The comparison of object sizes can be an easy visual interpretation for humans. This can be very useful for conveying the relationship between categories or data points. However, this also presents a potential problem: The human eye can inadvertently interpret comparisons that aren’t intended to convey meaning. For example, sometimes objects that appear to be the same size when they are not. Controlling the scale of a visual is important even when comparative sizes are not intended to offer information.



3. **Shape:** Does the shape of a specific object communicate something about it?

Rather than using simple dots or lines, a bit of creativity can enhance how quickly people are able to interpret a visual by using shapes that align with a given application. In the example below, it is immediately obvious that numbers of people are represented because the bars are person-shaped.



4. **Color:** What color is a mark?

Colors can be used both as a simple differentiator of groupings or as a way to communicate other concepts such as profitable versus unprofitable, or hot versus cold.

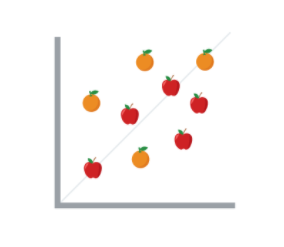


**Channels**

**Channels** are visual aspects or variables that represent characteristics of the data in a visualization. They are basically specialized marks that have been used to visualize data. It’s important to understand that channels 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, such as apples and oranges. But it is much less effective when distinguishing quantitative data, such as 5 from 5.5.



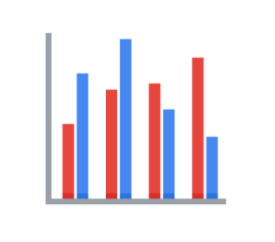
2. **Popout:** How easy is it to distinguish certain values from others?

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



3. **Grouping:** How effective 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 one single thing, the more that counts. Emphasis diminishes with each item you emphasize because the items begin to compete with one another.

**Key takeaways**

Throughout your career as an analyst, you will use different techniques and types of data visualizations to present data and insights in a concise, impactful manner. This will include organizing your data, selecting the right type of data visualizations, and designing them  in such a way that they are easy to understand and highly communicative while avoiding any visuals that are misleading or inaccurate.

Keep in mind that data visualization is an art form, and it takes time to develop these skills. Over your career as a data analyst, you will  learn how to design and evaluate data visualizations. Use these tips to think critically about data visualization—both as a creator and as an audience member.

Types of graphs

* Bar graphs use size contrast to compare two or more values
* Line graphs help your audience understand shifts or changes in your data
* Pie charts show how much each part of something makes up the whole
* Maps help organize data geographically
* Histogram is a chart shows how often data values fall into certain ranges
* Correlation charts show relationships among data

Dynamic visualizations

* Static visualizations do not change over time unless they’re edited
* Dynamic visualizations are visualizations that are interactive or change over time

Tableau is a business intelligence and analytics platform that helps people see, understand, and make decisions with data

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.

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

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.

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.

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

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

Reviewing each of these visual examples, where do you notice that they fit in relation to your type of data? One way to answer this is by evaluating patterns in data. Meaningful patterns can take many forms, such as:

* **Change:** This is a trend or instance of observations that become different over time. A great way to measure change in data is through a line or column chart.
* **Clustering:** A collection of data points with similar or different values. This is best represented through a distribution graph.
* **Relativity:** These are observations considered in relation or in proportion to something else. You have probably seen examples of relativity data in a pie chart.
* **Ranking:** This is a position in a scale of achievement or status. Data that requires ranking is best represented by a column chart.
* **Correlation:** This shows a mutual relationship or connection between two or more things. A scatterplot is an excellent way to represent this type of data pattern.

A diagram of a tree

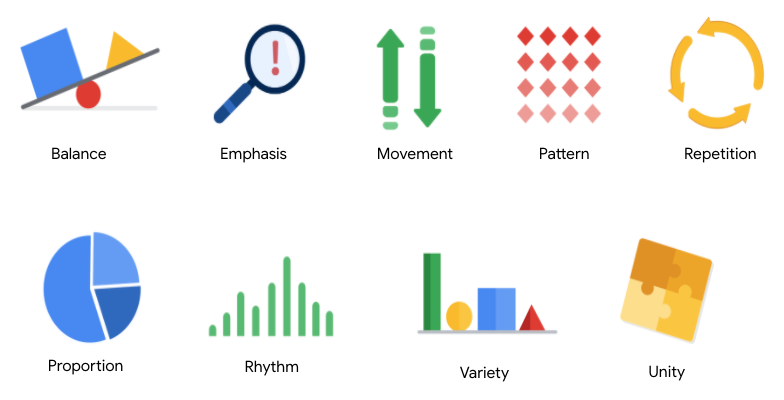
Description automatically generated

The element of art

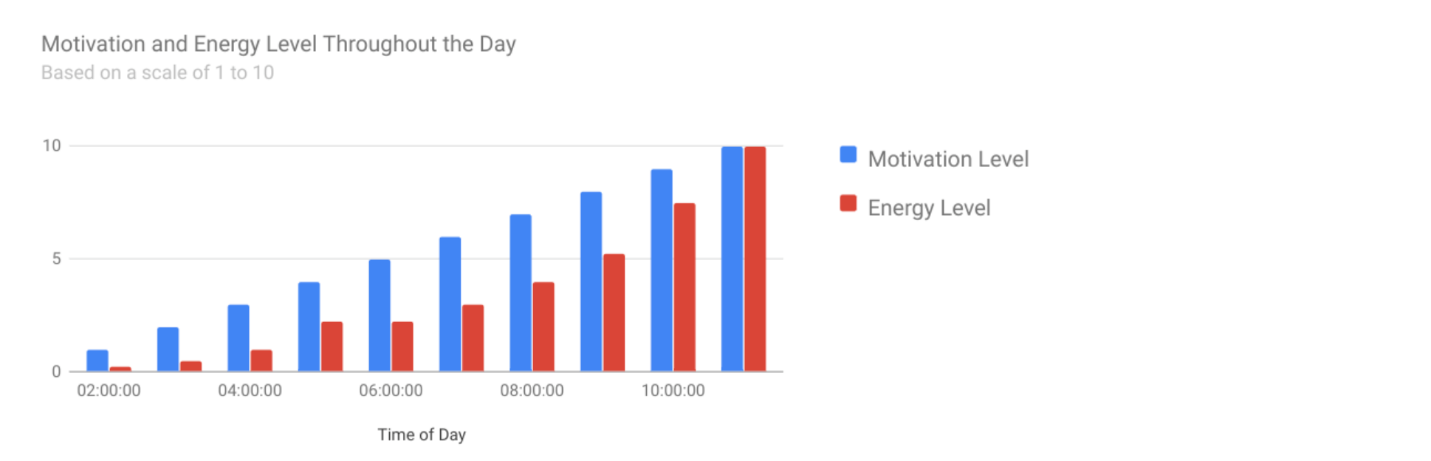
* Line
* Shape
* Color
* Space
* Movement

**Nine basic principles of design**

There are ninebasic **principles of design** that data analysts should think about when building their visualizations.

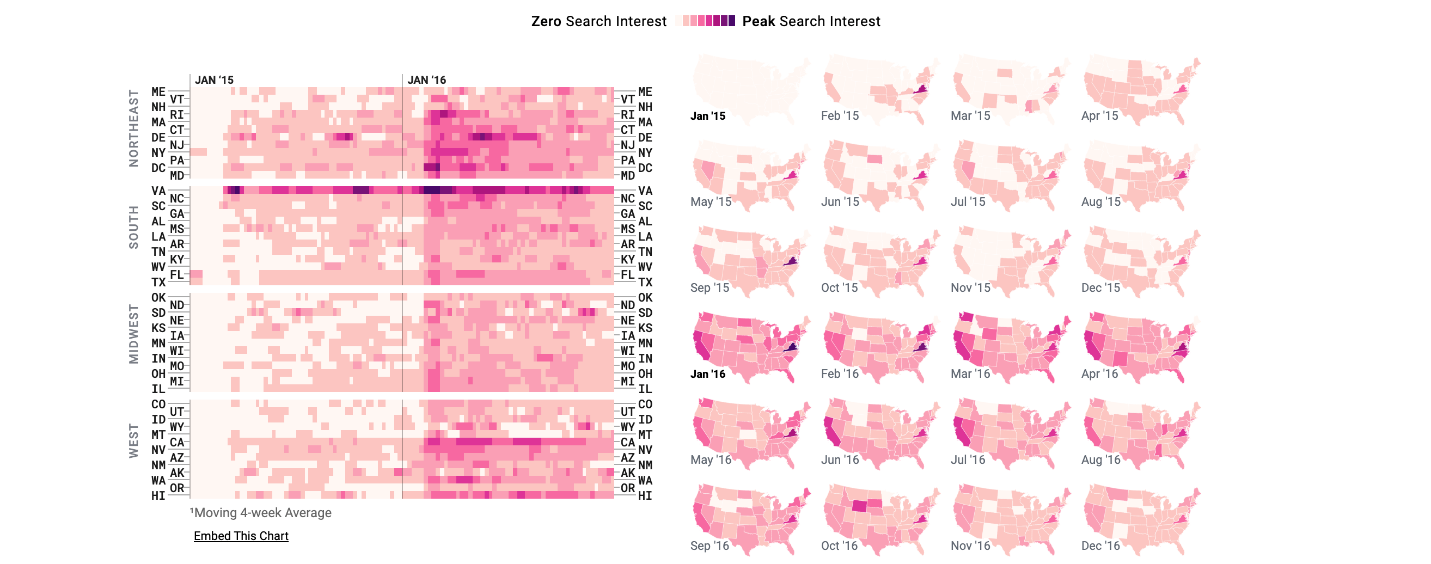


**1. Balance**: The design of a data visualization is balanced when the key visual elements, like color and shape, are distributed evenly. This doesn’t mean that you need complete symmetry, but your visualization shouldn’t have one side distracting from the other. If your data visualization is balanced, this could mean that the lines used to create the graphics are similar in length on both sides, or that the space between objects is equal. For example, [**this column chart**](https://developers.google.com/chart/interactive/docs/gallery/columnchart) (also shown below) is balanced; even though the columns are different heights and the chart isn’t symmetrical, the colors, width, and spacing of the columns keep this data visualization balanced. The colors provide sufficient contrast to each other so that you can pay attention to both the motivation level and the energy level displayed.



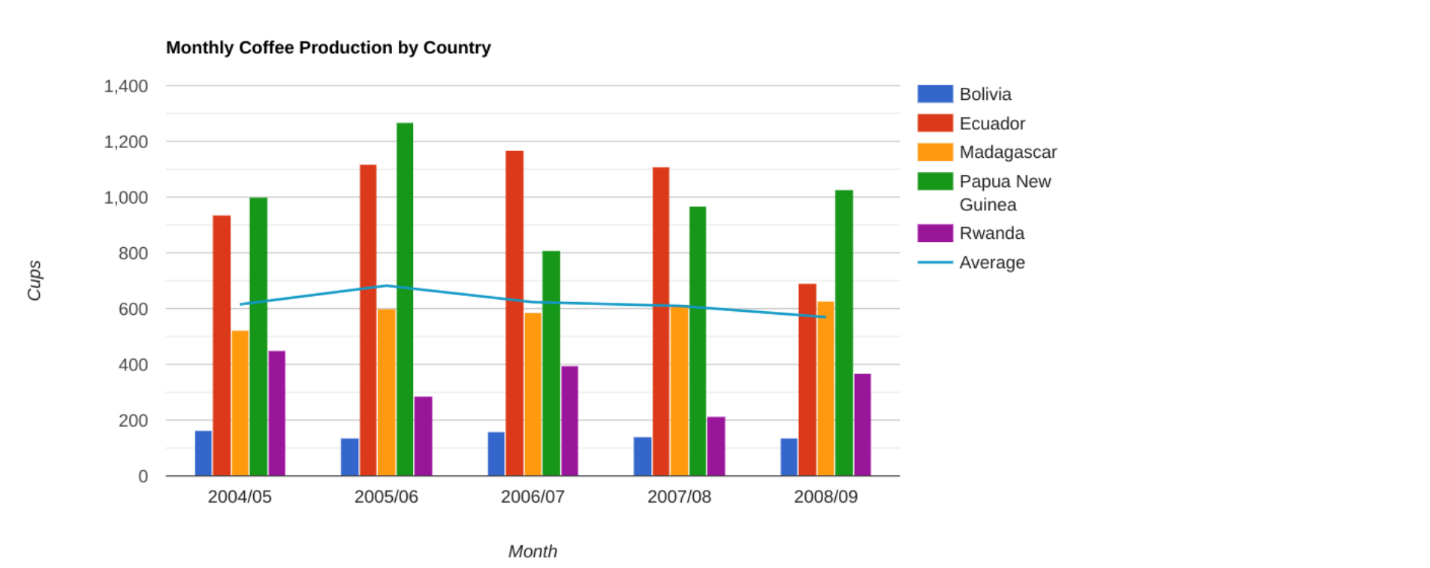
**2. Emphasis:** Your data visualization should have a focal point, so that your audience knows where to concentrate. In other words, your visualizations should emphasize the most important data so that users recognize it first. Using color and value is one effective way to make this happen. By using contrasting colors, you can make certain that graphic elements—and the data shown in those elements—stand out.

For example, you will notice a heat map data visualization below from [**The Pudding’s “Where Slang Comes From"**](https://pudding.cool/2017/02/new-slang/)article. This heat map uses colors and value intensity to emphasize the states where search interest is highest. You can visually identify the increase in the search over time from low interest to high interest. This way, you are able to quickly grasp the key idea being presented without knowing the specific data values.



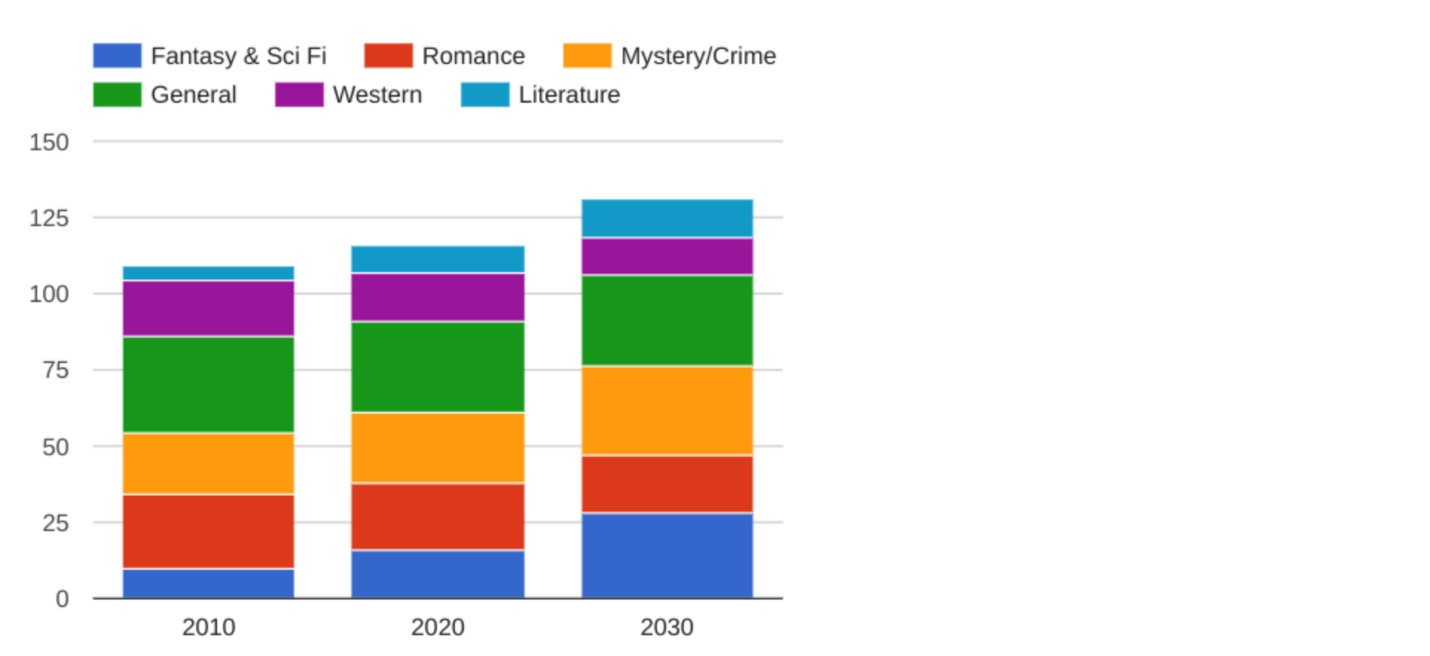
**3. Movement:** Movement can refer to the path the viewer’s eye travels as they look at a data visualization, or literal movement created by animations. Movement in data visualization should mimic the way people usually read. You can use lines and colors to pull the viewer’s attention across the page.

For example, notice how the average line in [**this combo chart**](https://developers.google.com/chart/interactive/docs/gallery/combochart) (also shown below) draws your attention from left to right. Even though this example isn’t moving, it still uses the movement principle to guide viewers’ understanding of the data.



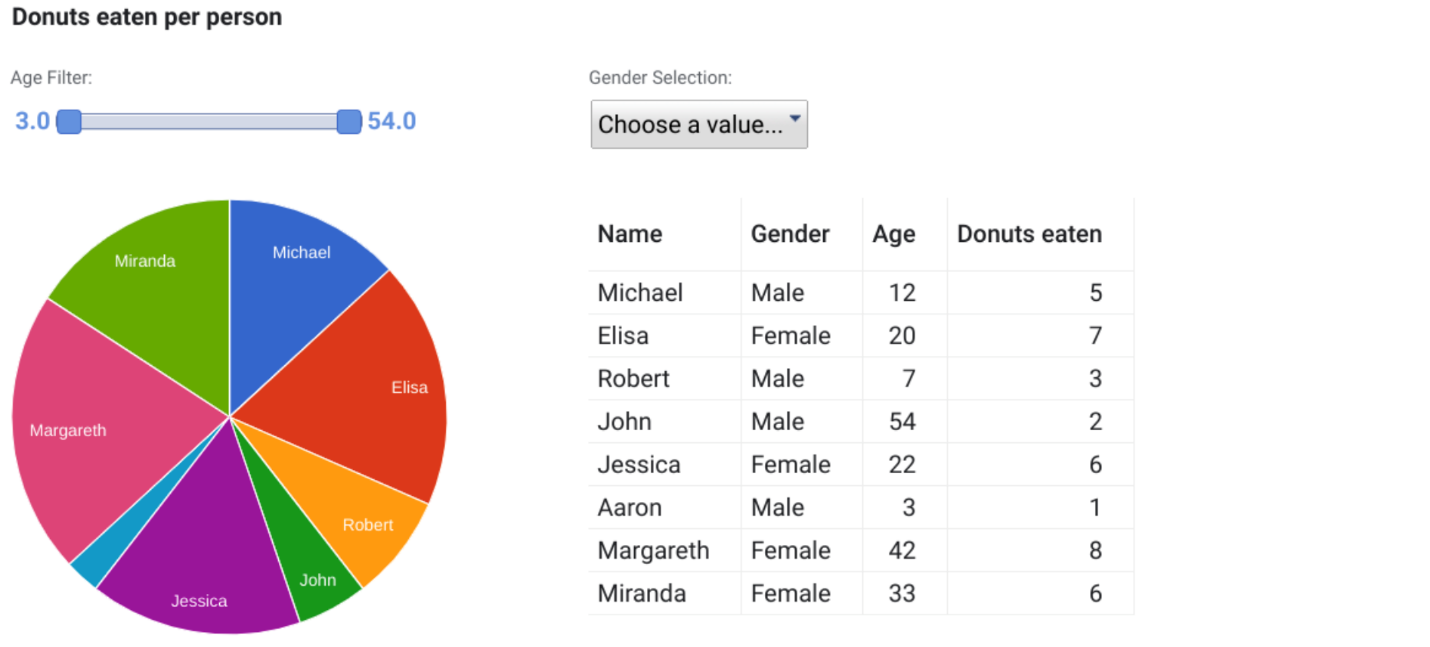
**4. Pattern:** You can use similar shapes and colors to create patterns in your data visualization. This can be useful in a lot of different ways. For example, you can use patterns to highlight similarities between different data sets, or break up a pattern with a unique shape, color, or line to create more emphasis.

In the example below, the different colored categories of [**this stacked column chart**](https://developers.google.com/chart/interactive/docs/gallery/barchart)(also shown below)are a consistent pattern that makes it easier to compare book sales by genre in each column. Notice in the chart that the Fantasy & Sci Fi category (royal blue) is increasing over time even as the general category (green) is staying about the same.



**5. Repetition:** Repeating chart types, shapes, or colors adds to the effectiveness of your visualization. Think about the book sales chart from the previous example: the repetition of the colors helps the audience understand that there are distinct sets of data. You may notice this repetition in all of the examples we have reviewed so far. Take some time to review each of the previous examples and notice the elements that are repeated to create a meaningful visual story.

**6. Proportion:** Proportion is another way that you can demonstrate the importance of certain data. Using various colors and sizes helps demonstrate that you are calling attention to a specific visual over others. If you make one chart in a dashboard larger than the others, then you are calling attention to it. It is important to make sure that each chart accurately reflects and visualizes the relationship among the values in it. In [**this dashboard**](https://developers.google.com/chart/interactive/docs/gallery/controls) (also shown below), the slice sizes and colors of the pie chart compared to the data in the table help make the number of donuts eaten by each person the focal point.



These first six principles of design are key considerations that you can make while you are creating your data visualization. These next three principles are useful checks once your data visualization is finished. If you have applied the initial six principles thoughtfully, then you will probably recognize these next three principles within your visualizations already.

**7. Rhythm:** This refers to creating a sense of movement or flow in your visualization. Rhythm is closely tied to the movement principle. If your finished design doesn’t successfully create a flow, you might want to rearrange some of the elements to improve the rhythm.

**8. Variety:** Your visualizations should have some variety in the chart types, lines, shapes, colors, and values you use. Variety keeps the audience engaged. But it is good to find balance since too much variety can confuse people. The variety you include should make your dashboards and other visualizations feel interesting and unified.

**9. Unity:** The last principle is unity. This means that your final data visualization should be cohesive. If the visual is disjointed or not well organized, it will be confusing and overwhelming.

Being a data analyst means learning to think in a lot of different ways. These nine principles of design can help guide you as you create effective and interesting visualizations.

A close-up of a computer screen

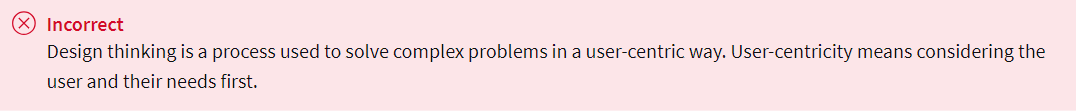
Description automatically generated

Elements for effective visuals

* *Clear meaning*: good visualizations clearly communicate their intended insight
* *Sophisticated use of contrast*: which helps separate the most important data from the rest using visual context that our brains naturally look for
* *Refined execution:* Visuals with refined execution include deep attention to detail, using visual elements like lines, shapes, colors, value, space and movement

Five phases of the design process

* Empathize: you think about the emotions and needs of the target audience of your data viz, whether it's stakeholders, team members or the general public
* Define: elps you to find your audiences needs, their problems, and your insights.
* Ideate: generate your data viz ideas. You'll use all of your findings from the empathize and define phases to brainstorm potential data viz solutions.
* Prototype: Putting visualizations together for testing and feedback
* Test: Showing prototype visualizations to people before stakeholders see them



Ways to make data visualizations accessible:

* Labeling
* Text alternatives provides a textual alternative to non-text content
* Text-based format
* Distinguishing
* **Glossary terms from module 1**
* **Terms and definitions for Course 6, Module 1**
* **Alternative text:** Text that provides an alternative to non-text content, such as images and videos
* **Annotation:** Text that briefly explains data or helps focus the audience on a particular aspect of the data in a visualization
* **AVERAGEIF:** A spreadsheet function that returns the average of all cell values from a given range that meet a specified condition
* **Balance:** The design principle of creating aesthetic appeal and clarity in a data visualization by evenly distributing visual elements
* **Bar graph:** A data visualization that uses size to contrast and compare two or more values
* **Calculus:** A branch of mathematics that involves the study of rates of change and the changes between values that are related by a function
* **Causation:** When an action directly leads to an outcome, such as a cause-effect relationship
* **Channel:** A visual aspect or variable that represents characteristics of the data in a visualization
* **Chart:** A graphical representation of data from a worksheet
* **Cluster:** A collection of data points on a data visualization with similar values
* **CONVERT:** A SQL function that changes the unit of measurement of a value in data
* **Correlation:** The measure of the degree to which two variables change in relationship to each other
* **CREATE TABLE:** A SQL clause that adds a temporary table to a database that can be used by multiple people
* **Data composition:** The process of combining the individual parts in a visualization and displaying them together as a whole
* **Decision tree**: A tool that helps analysts make decisions about critical features of a visualization
* **Design thinking:** A process used to solve complex problems in a user-centric way
* **Distribution graph:** A data visualization that displays the frequency of various outcomes in a sample
* **DROP TABLE:** A SQL clause that removes a temporary table from a database
* **Dynamic visualizations:** Data visualizations that are interactive or change over time
* **Emphasis:** The design principle of arranging visual elements to focus the audience’s attention on important information in a data visualization
* **HAVING:** A SQL clause that adds a filter to a query instead of the underlying table that can only be used with aggregate functions
* **Headline:** Text at the top of a visualization that communicates the data being presented
* **Heat map:** A data visualization that uses color contrast to compare categories in a dataset
* **Histogram:** A data visualization that shows how often data values fall into certain ranges
* **Inner query:** A SQL subquery that is inside of another SQL statement
* **Label:** Text in a visualization that identifies a value or describes a scale
* **Legend:** A tool that identifies the meaning of various elements in a data visualization
* **Line graph:** A data visualization that uses one or more lines to display shifts or changes in data over time
* **Map:** A data visualization that organizes data geographically
* **Mark:** A visual object in a data visualization such as a point, line, or shape
* **MAXIFS:** A spreadsheet function that returns the maximum value from a given range that meets a specified condition
* **Mental model:** A data analyst’s thought process and approach to a problem
* **Movement:** The design principle of arranging visual elements to guide the audience’s eyes from one part of a data visualization to another
* **MINIFS:** A spreadsheet function that returns the minimum value from a given range that meets a specified condition
* **Narrative:** (Refer to story)
* **Ordinal data:** Qualitative data with a set order or scale
* **Pattern:** The design principle of using similar visual elements to demonstrate trends and relationships in a data visualization
* **Pie chart:** A data visualization that uses segments of a circle to represent the proportions of each data category compared to the whole
* **Pre-attentive attributes:** The elements of a data visualization that an audience recognizes automatically without conscious effort
* **Proportion:** The design principle of using the relative size and arrangement of visual elements to demonstrate information in a data visualization
* **R:** A programming language used for statistical analysis, visualization, and other data analysis
* **Ranking:** A system to position values of a dataset within a scale of achievement or status
* **Relativity:** The process of considering observations in relation or proportion to something else
* **Repetition:** The design principle of repeating visual elements to demonstrate meaning in a data visualization
* **Rhythm:** The design principle of creating movement and flow in a data visualization to engage an audience
* **Scatterplot:** A data visualization that represents relationships between different variables with individual data points without a connecting line
* **SELECT INTO:** A SQL clause that copies data from one table into a temporary table without adding the new table to the database
* **Sort range:** A spreadsheet menu function that sorts a specified range and preserves the cells outside the range
* **Sort sheet:** A spreadsheet menu function that sorts all data by the ranking of a specific sorted column and keeps data together across rows
* **Static visualization:** A data visualization that does not change over time unless it is edited
* **Story:** The narrative of a data presentation that makes it meaningful and interesting
* **Subtitle:** Text that supports a headline by adding context and description
* **Tableau:** A business intelligence and analytics platform that helps people visualize, understand, and make decisions with data
* **Unity:** The design principle of using visual elements that complement each other to create aesthetic appeal and clarity in a data visualization
* **Variety:** The design principle of using different kinds of visual elements in a data visualization to engage an audience
* **Visual form:** The appearance of a data visualization that gives it structure and aesthetic appeal
* **X-axis:** The horizontal line of a graph usually placed at the bottom, which is often used to represent time scales and discrete categories
* **Y-axis:** The vertical line of a graph usually placed to the left, which is often used to represent frequencies and other numerical variables