



Data Visualization

Mrs. Sujata Shankar Gawade¹, Mrs. Purnima Abhishek Kamble²

Lecturer, Computer Technology, BVIT, Navi Mumbai, India¹

Lecturer, Computer Technology, BVIT, Navi Mumbai, India²

Abstract: Data visualization is the graphical(visual) representation of data by using visual elements like charts, graphs, and maps. It also provides an excellent way to present data to non-technical audiences without confusion. The first and foremost objective of data visualization is to convey data correctly.

Keywords: Data, percent, graphs, charts, maps

I. INTRODUCTION

Data visualization is the graphical representation of data and information. There is use of visual elements like maps, charts and graphs. In data visualization there are various tools which an accessible way to understand and see patterns, trends and outliers. It is used to provide an excellent way to present data to non-technical audiences. Objective of data visualization is to convey data correctly. Data visualization means take data values and convert it into logical and systematic way by which there is final graphic. There are various types of data visualizations, like a heatmap, a pie chart and a scatterplot. In data visualization data values are turned into blobs of ink on paper or colored pixels on a screen.

II. AESTHETICS AND DATA TYPES IN DATA VISUALIZATION

Aesthetics are used to describe every aspect of a given graphical element.

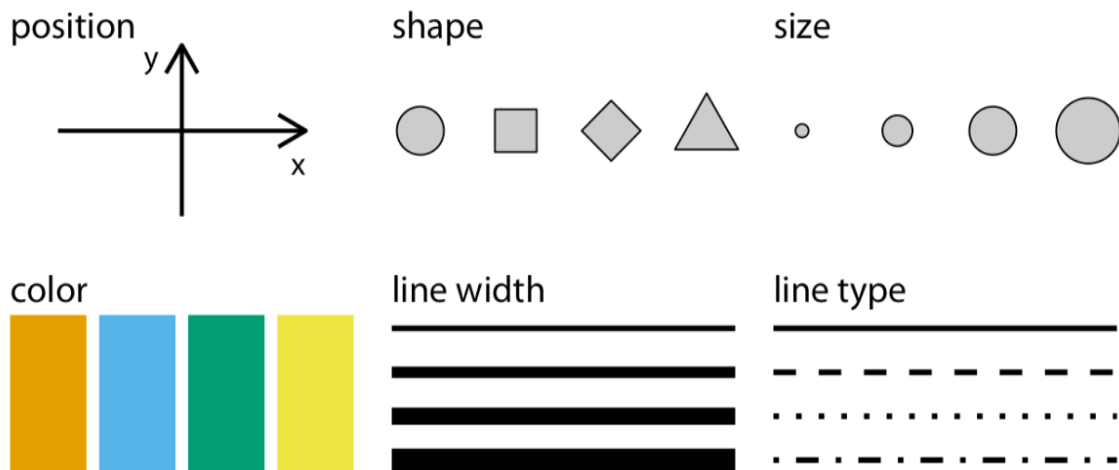


Fig. Commonly used aesthetics in data visualization

If we want to display text, we may have to specify font family, font face, and font size as its Aesthetics.

All aesthetics are categorized into two groups:

1. Continuous data
2. Discrete data

Continuous data values are values for which arbitrarily fine intermediates exist.

For example, percentage of student is a continuous value. Between any two percentages, say 60 percent and 61 percent, there are arbitrarily many intermediates, such as 60.2 percent, 60.5 percent and so on.



Discrete data: For example, number of items in a grocery is a discrete value. A grocery can hold 6 soaps or 7, but not 6.5.

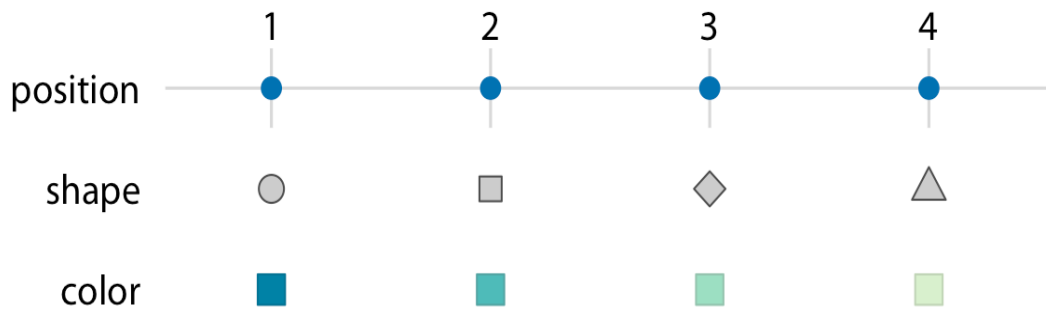
Position, size, color, and line width can represent continuous and discrete data, but shape and line type can usually only represent discrete data. In addition to continuous and discrete numerical values, data can come in the form of discrete categories (in the form of dates or times, and as text). When data is numerical, we also call it as quantitative and when data is categorical, we also call it as qualitative. Variables holding categorical or qualitative data are factors, and factors are classified into two types: ordered factor and unordered factors.

Different categories are called levels.

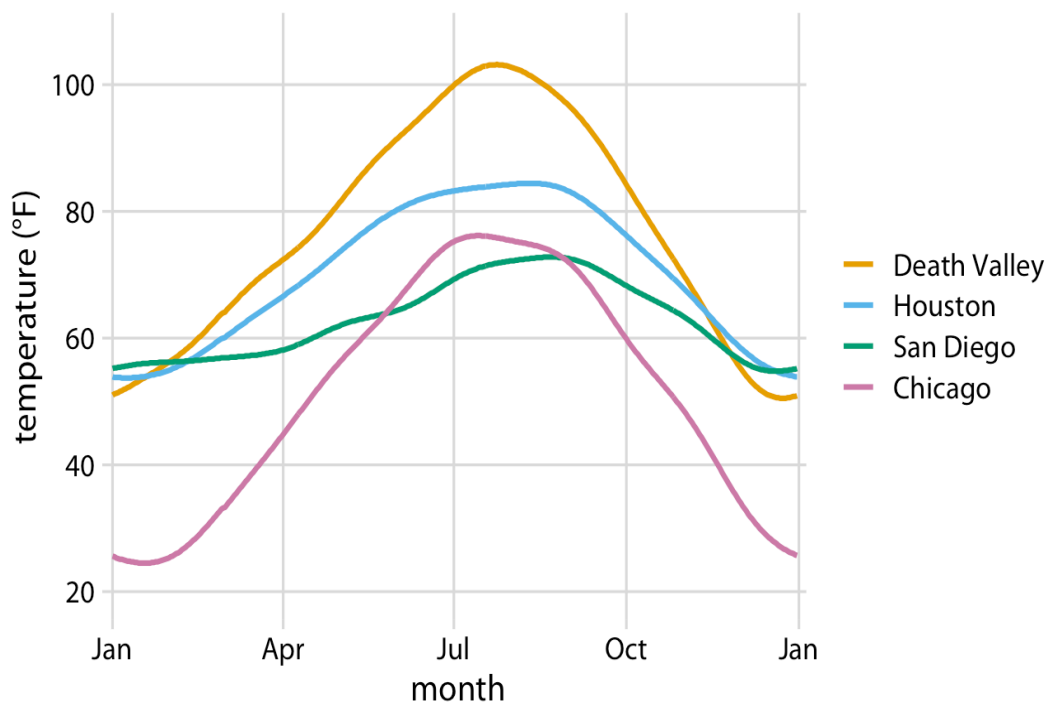
Scales Map Data Values onto Aesthetics

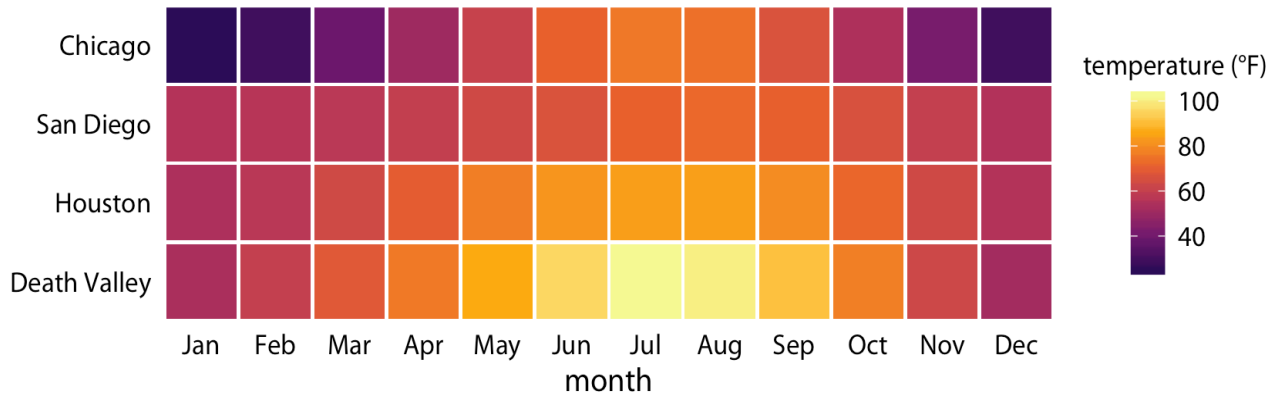
Mapping of data values onto aesthetics, we set the relationship of which data values correspond to which specific aesthetics values

For example, if our graphic has an x axis, then we represent which data values fall onto positions along this axis, shape and colour.



Similarly, we may represent Day of the year to the x axis, temperature is mapped to the y axis, and location to line colour and visualize these aesthetics with solid lines. Variation in temperature is representing by line plot at four locations. Line plot is very useful for comparison.





In above figure uses two position scales (location along the y axis and month along the x axis), location is unordered factor with 4 levels Month is an ordered factor with 12 levels. Therefore, the two position scales are both discrete. If the factor is not ordered then the order is arbitrary, and you can choose any order you want. Ordering the locations from overall coldest (Chicago) to overall hottest (Death Valle). If the factor is ordered (as is here the case for month), then the levels need to be placed in the appropriate order.

- Heat map: it represents specific data value with different color

III. USE OF COORDINATE SYSTEM IN DATA VISUALIZATION

To make any sort of data visualization, it's necessary to define position scales, which determine where in a graphic different data values are located. For 2D visualizations, two numbers are required to uniquely specify a point, and two position scales. These two scales are the x and y axes of the plot. There is relative geometric arrangement of these scales Basically, the x axis runs horizontally and the y axis vertically, but we could choose other arrangements. In this one axis run in a circle and the other run radially. The combination of a set of position scales and their relative geometric arrangement is called a coordinate system. The combination of a set of position scales and their relative geometric arrangement is called a **coordinate system**.

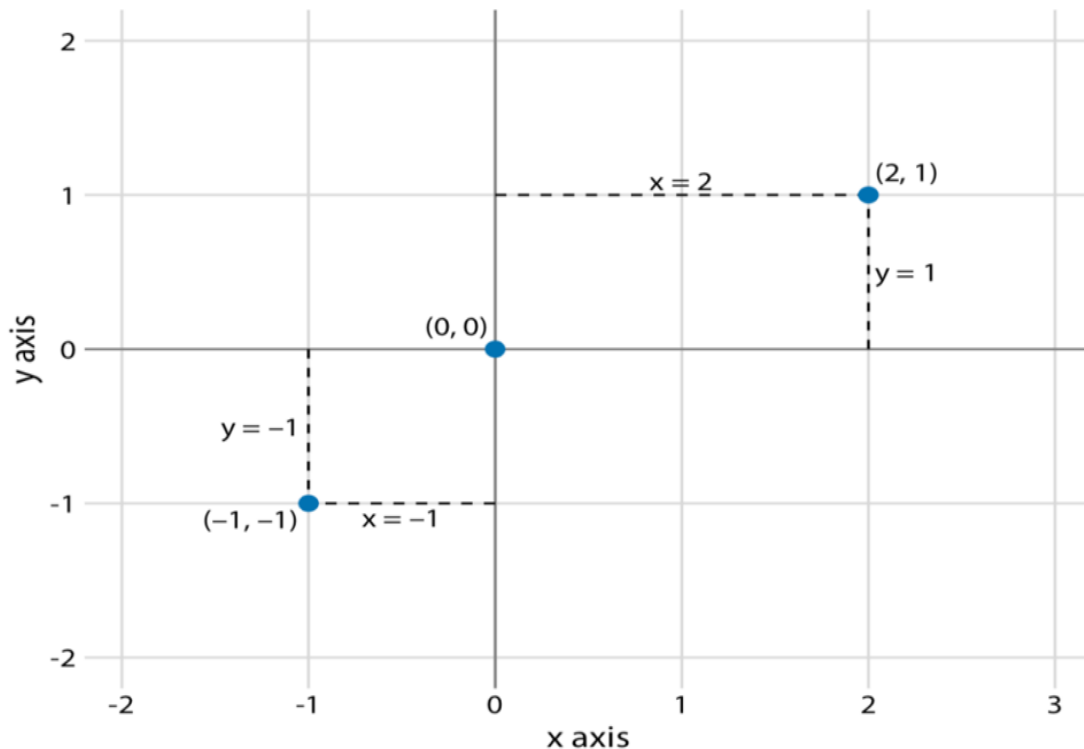
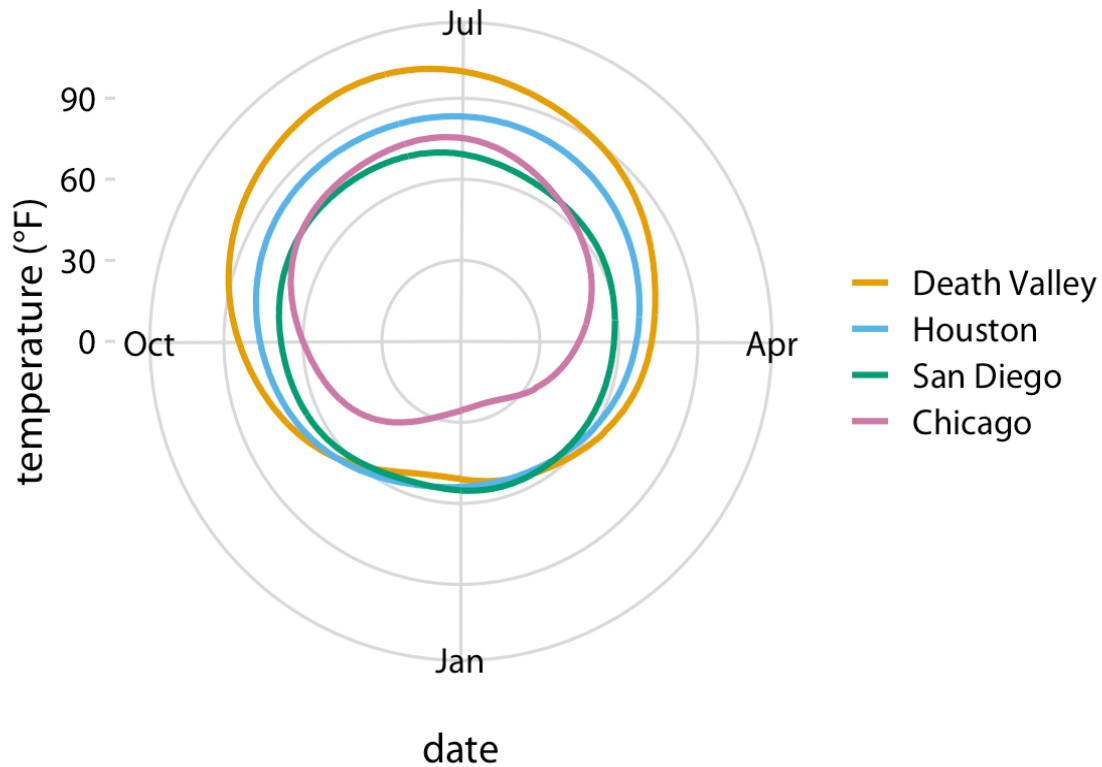


Figure. Standard Cartesian coordinate system



In coordinate system for data visualization is the 2D Cartesian coordinate system, where each point is uniquely specified by an x and a y value.

In Figure, the x axis runs from -2.2 to 3.2 and the y axis runs from -2.2 to 2.2 .



Daily temperature normal for four selected locations in the US, shown in polar coordinates. The radial distance from the center point indicates the daily temperature in Fahrenheit, and the days of the year are arranged counterclockwise starting with Jan. 1st at the 6:00 position.

IV. USE OF COLORS TO REPRESENT DATA VALUES.

There are three fundamental use cases for color in data visualizations:

1. Color as a Tool to distinguish discrete items or groups.
2. Colors to represent data values.

Color as a Tool to Highlight specific elements in the data.

In this case use of color to distinguish discrete items or groups that do not have an intrinsic order, such as different countries on a map or different manufactures of a certain product.

In this case, there is a use a qualitative color scale.

The colors is not create the impression of an order.



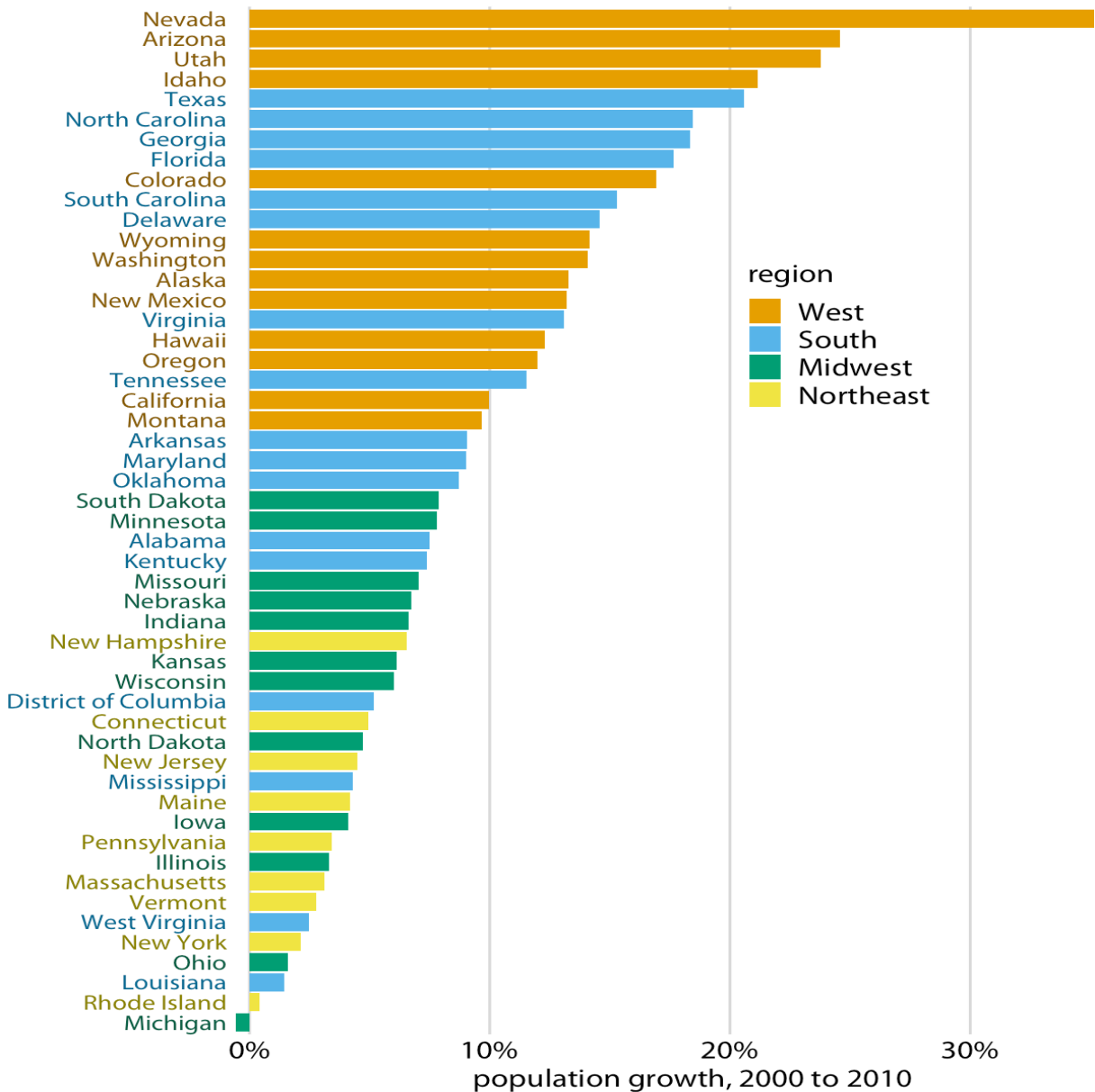
Okabe Ito



ColorBrewer Dark2



ggplot2 hue





There is a use of colors to represent quantitative data values, such as income, temperature, or speed.

In this case, there is a use of a sequential color scale.

ColorBrewer Blues



Heat



Viridis



We can use color also be an effective tool to highlight specific elements in the data at the time of data storytelling.

In this case, we use a accent color scale.

Okabe Ito Accent



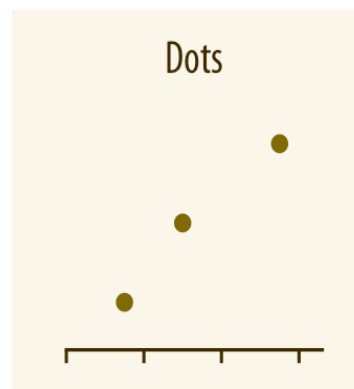
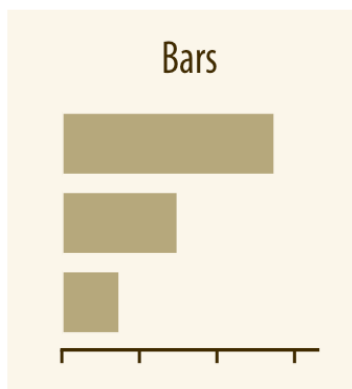
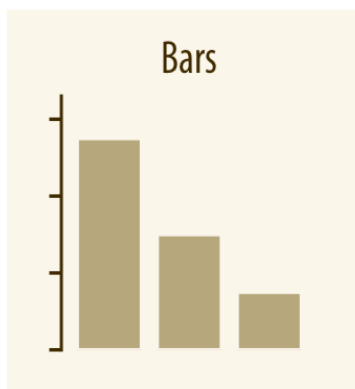
Grays with accents



ColorBrewer Accent



In data visualization there is use of bar chart and dot chart.





- Grouped bars: it represent data that is grouped into different categories.
- Stacked bars: it represents data with different bars and each bar represents sub categories within main categories.
- Heat map : it represents specific data value with different Color.

V. CONCLUSION

Data visualization is used toto provide an excellent way to present data to non-technical audiences without confusion.

Data visualization is used to represent data correctly. Data visualization is the process of representing data in graphical or pictorial manner.

REFERENCES

- [1]. J. L. V. Sancho, J. C. Dominguez, and B. E. M. Ochoa, "An approach to the taxonomy of data visualization," *Revista Latina de Comunicacion Social*, vol. 69, 2014, pp. 486-507.
- [2]. "Data visualization techniques," SAS, http://www.sas.com/en_us/offers/sem/data-visualization-techniques-2332568.html?keyword=data+visualization+techniques&matchtype=p&publisher=google&gclid=COycgoCbutACFcolgQodqwgLiA
- [3]. J. Wolfe, "Teaching students to focus on the data in data visualization," *Journal of Business and Technical Communication*, vol. 29, no. 3, 2015, pp. 344-359.