

Sep 30, 2024

# Data Visualization

Week 2. Visual Aesthetics

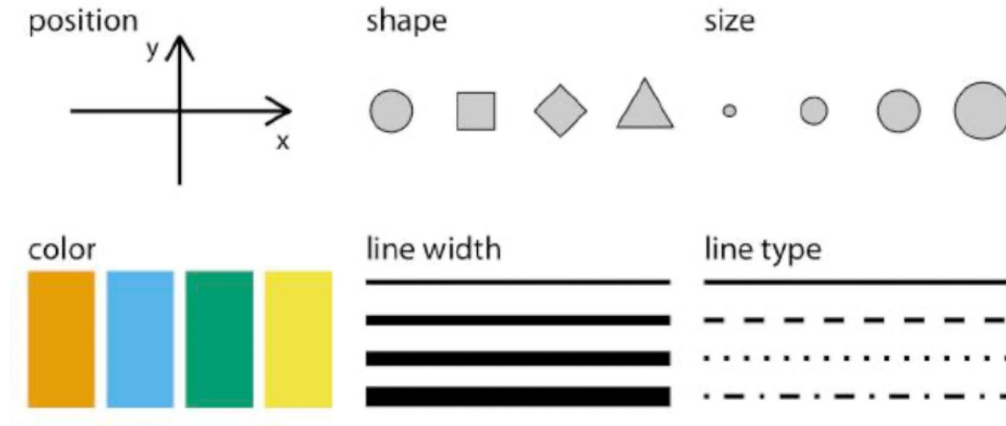
© Mustafa Cavus, Ph.D.

# Variable types

1. Numerical
  - a. Continuous
  - b. Discrete
2. Categorical
  - a. Ordinal
  - b. Nominal
3. Time
4. Text
5. Image

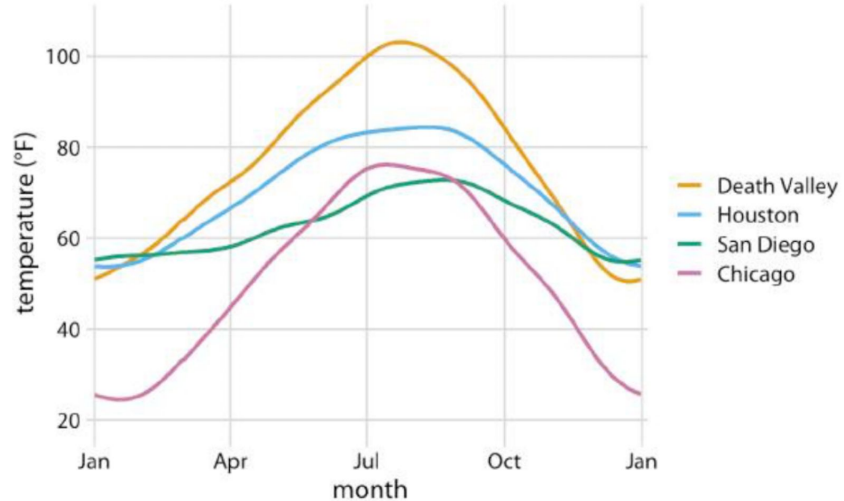
# Visual aesthetics

- position
- shape
- size
- color
- line width
- line type



# Example 1

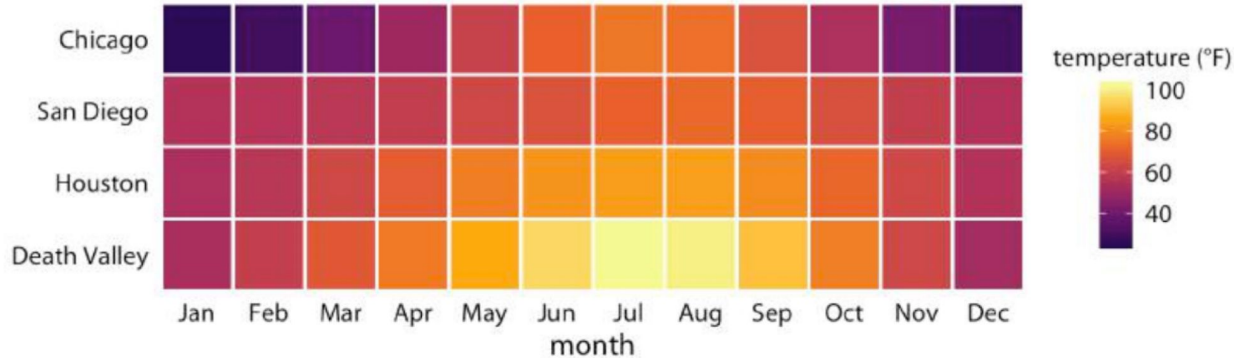
List the aesthetics used in the plot and their corresponding variables.



*Figure 2-3. Daily temperature normals for four selected locations in the US. Temperature is mapped to the y axis, day of the year to the x axis, and location to line color. Data source: NOAA.*

## Example 2

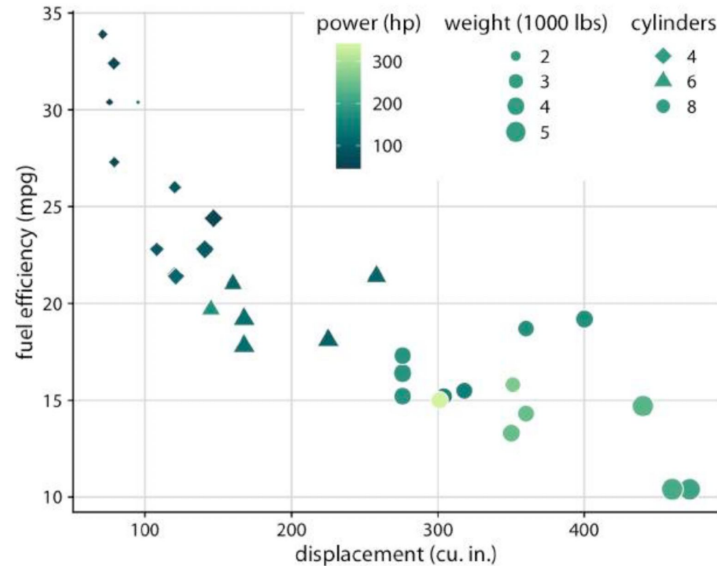
List the aesthetics used in the plot and their corresponding variables.



*Figure 2-4. Monthly normal mean temperatures for four locations in the US. Data source: NOAA.*

# Example 3

List the aesthetics used in the plot and their corresponding variables.



## 2. Axes

- In order to create a data visualization, position scales that determine where the observation values are located on the plot must be defined.
- The relative geometric arrangement of these scales should also be determined.
- The combination of a set of position scales and geometric arrangement is called a coordinate system.

## 2.1. Cartesian coordinates

- The axes are on a continuous scale and can take both positive and negative values.
- The values may be in different units of measurement: for example, Celsius for temperature, kilometers for distance, or hours for time.
- The axes can be in two different units of measurement. In this case, the grids formed by the intersection of the values on the axes do not need to be equally spaced.

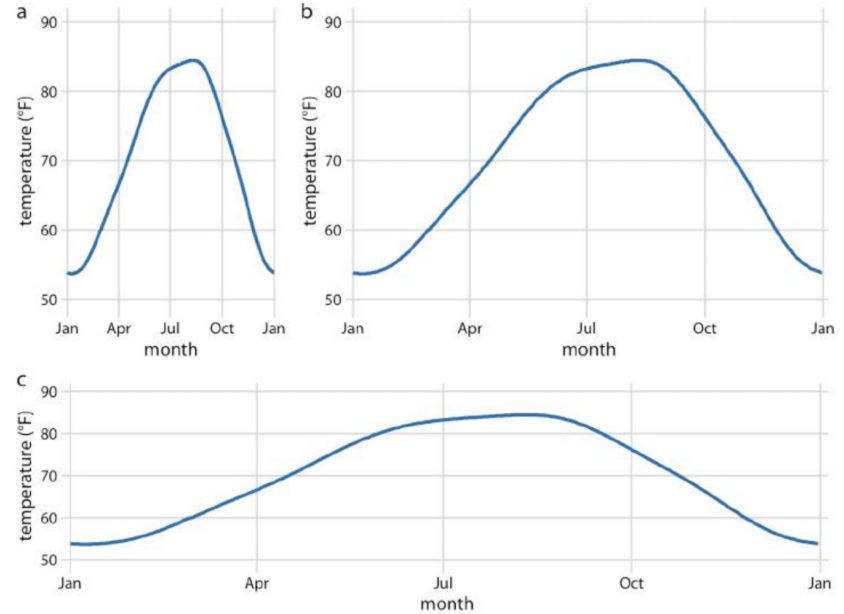
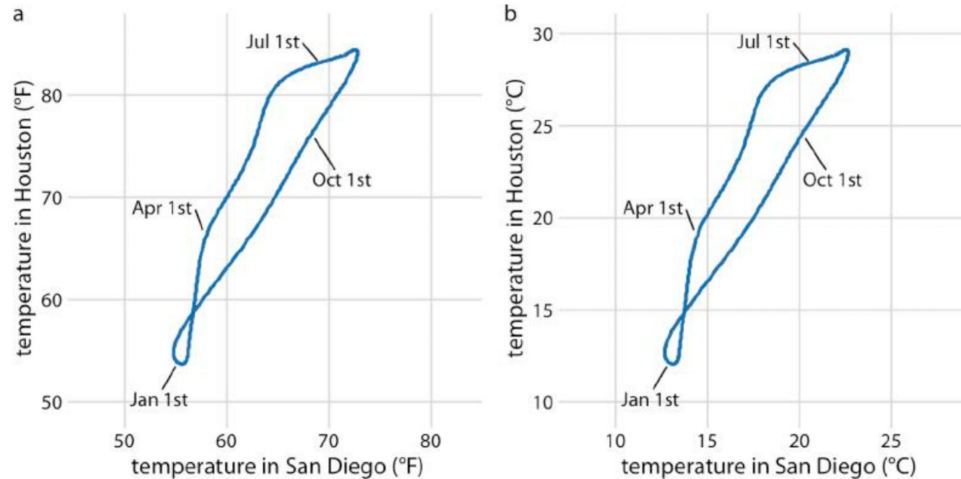


Figure 3-2. Daily temperature normals for Houston, TX. Temperature is mapped to the y axis and day of the year to the x axis. Parts (a), (b), and (c) show the same figure in different aspect ratios. All three parts are valid visualizations of the temperature data. Data source: NOAA.



## 2.1. Cartesian coordinates

Cartesian coordinate systems are invariant under linear transformations. Therefore, when the units of observation values are changed, the appearance of the plots will remain unchanged.



## 2.2. Other types of coordinates

- Non-linear axes
- Curved axes

### 3. Usage of Colors

Colors are used for three main purposes for visualizing data:

- to differentiate observations,
- to represent the value of observations,
- to highlight certain observations.

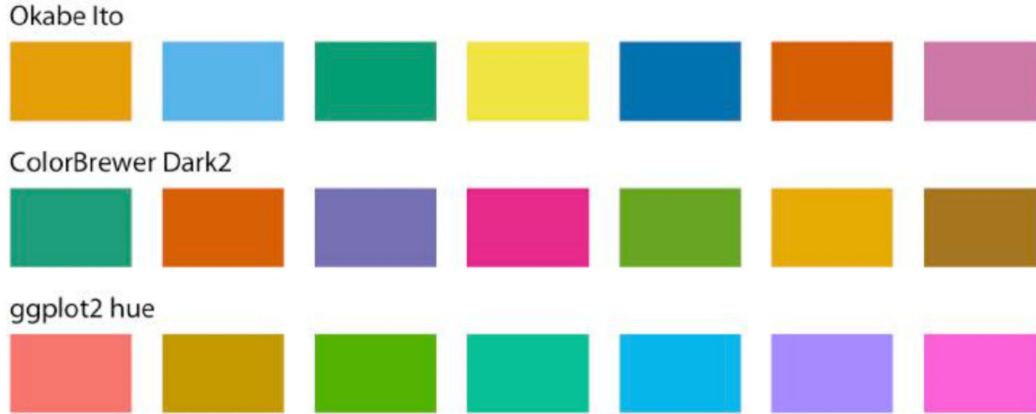
## 3.1. Usage for differentiation

Colors are often used to distinguish separate elements and groups that lack a specific order. In such cases, the key considerations for using colors are:

- they must be clearly distinguishable from one another,
- they should belong to a specific color scale,
- no color should stand out over the others,
- and they should not imply any sense of order.

There are various studies on color scales (see Okabe and Ito, 2018; Brewer, 2017).

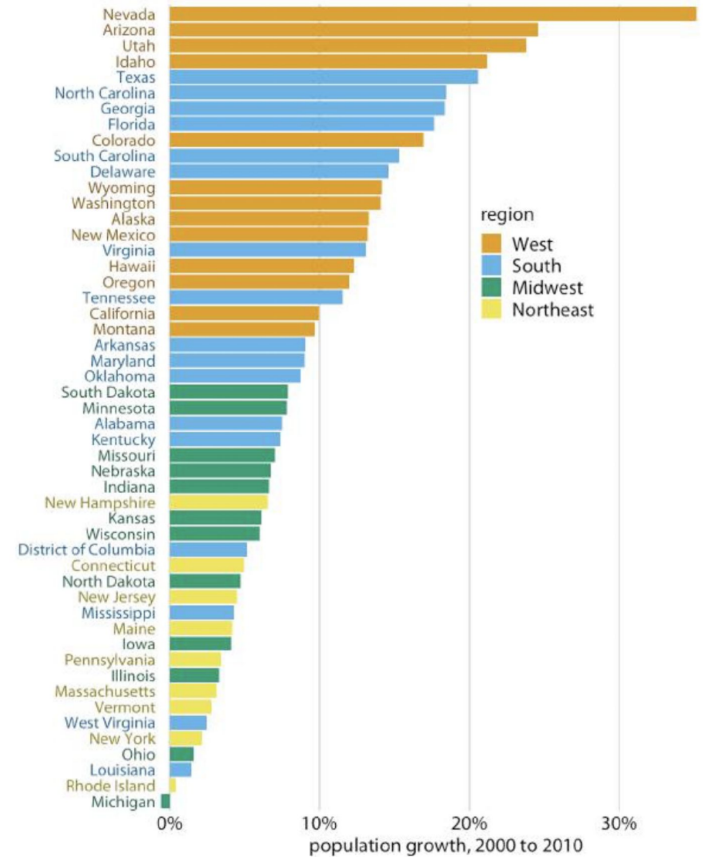
## 3.1. Usage for differentiation



*Figure 4-1. Example qualitative color scales. The Okabe Ito scale is the default scale used throughout this book [Okabe and Ito 2008]. The ColorBrewer Dark2 scale is provided by the ColorBrewer project [Brewer 2017]. The ggplot2 hue scale is the default qualitative scale in the widely used plotting software ggplot2.*

## Example 4

Let's examine the plot by ranking the percentage of population growth in U.S. states from 2000 to 2010 and coloring them according to geographic regions as an example of using qualitative color scales.



## 3.2. Usage for represent the value of observations

Color can be used to indicate an increase or decrease in the values of a variable, meaning it represents the values it takes. In this case, a sequential color scale is necessary. The sequential color scale should:

- allow for the perception of which values are larger or smaller than others,
- include a color series that clearly shows how far apart the values are,
- and change uniformly across the range of variation.

## 3.2. Usage for represent the value of observations

When using multi-tone scales, color gradients commonly found in nature, such as dark red to light yellow or dark purple to light green, should be used.

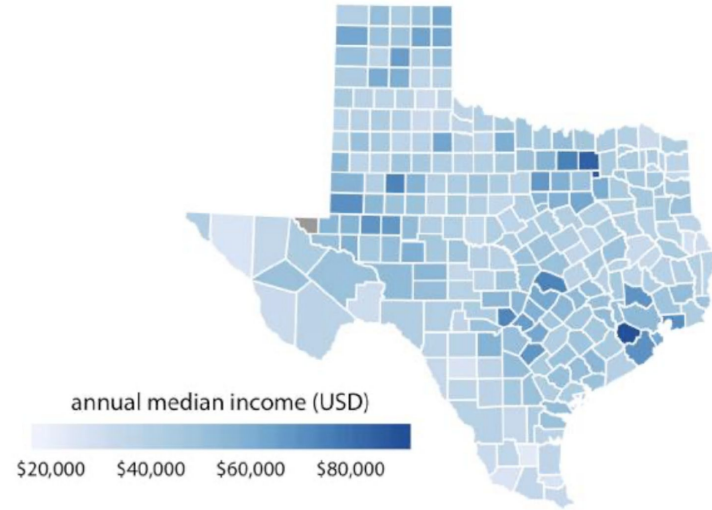


*Figure 4-5. Example diverging color scales. Diverging scales can be thought of as two sequential scales stitched together at a common midpoint color. Common color choices for diverging scales include brown to greenish blue, pink to yellow-green, and blue to red.*



# Example 5

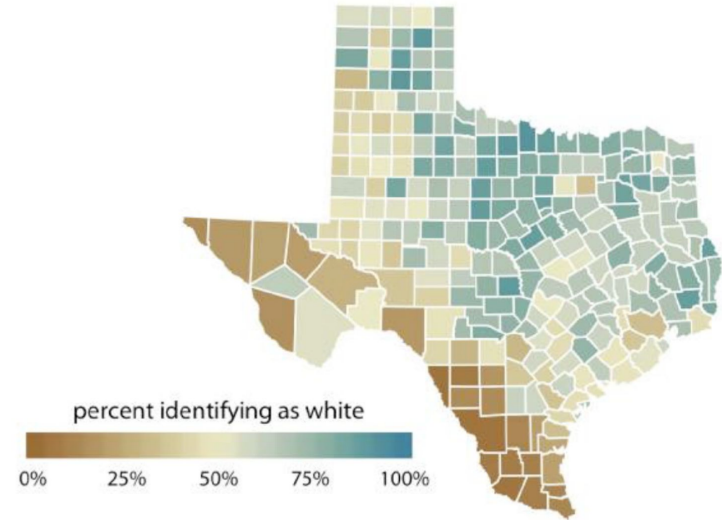
Representing data values with colors is particularly useful when it is desired to show how data values change across geographic regions.



*Figure 4-4. Median annual income in Texas counties. The highest median incomes are seen in major Texas metropolitan areas, in particular near Houston and Dallas. No median income estimate is available for Loving County in West Texas, and therefore that county is shown in gray. Data source: 2015 Five-Year American Community Survey.*

# Example 6

Let's examine the plot as an example of using a color scale composed of different colors,



*Figure 4-6. Percentage of people identifying as white in Texas counties. Whites are in the majority in North and East Texas but not in South or West Texas. Data source: 2010 US Decennial Census.*

## 3.3. Usage for emphasizing

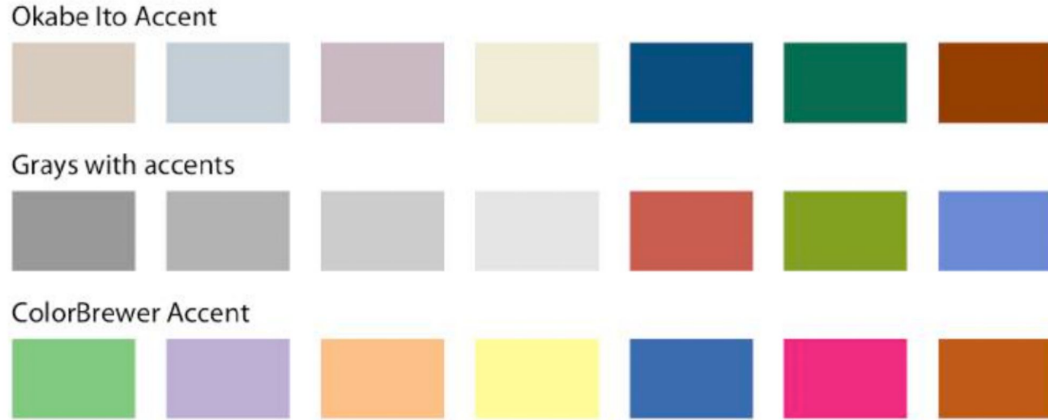
Colors can be used very effectively to highlight specific data points in the observations.

There may be specific categories or values in the dataset that carry key information related to the story we want to tell, and by highlighting them, the narrative can be strengthened.

To make this emphasis:

- some colors in any color scale can be used in darker ,
- gray colors can be replaced with the colors in the scale to make other colors stand out,
- or a specially designed scale can be used for emphasis.

### 3.3. Usage for emphasizing



*Figure 4-7. Example accent color scales, each with four base colors and three accent colors. Accent color scales can be derived in several different ways: (top) we can take an existing color scale (e.g., the Okabe Ito scale, [Figure 4-1](#)) and lighten and/or partially desaturate some colors while darkening others; (middle) we can take gray values and pair them with colors; (bottom) we can use an existing accent color scale (e.g., the one from the ColorBrewer project).*

# Example 7

What information might the blue color have been used to emphasize?

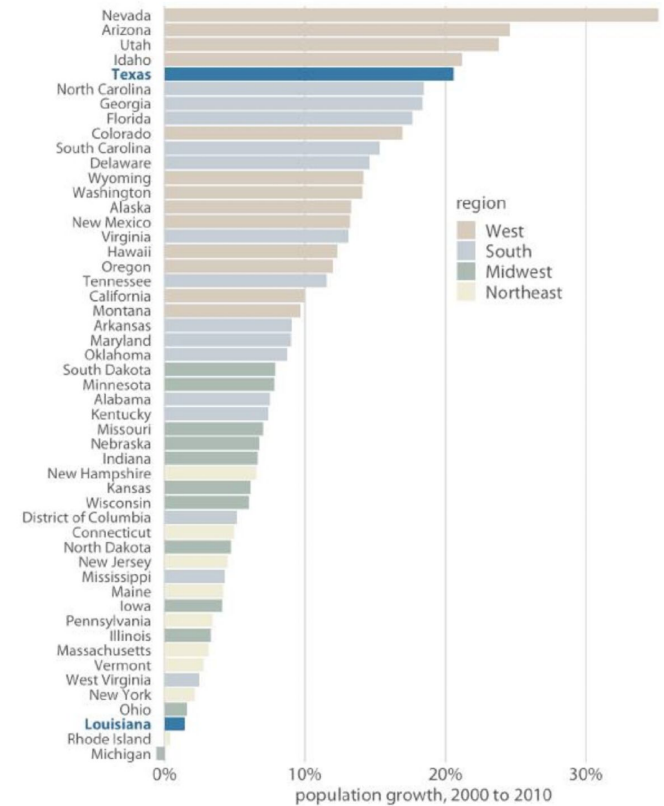


Figure 4-8. From 2000 to 2010, the two neighboring southern states, Texas and Louisiana, experienced among the highest and lowest population growth across the US. Data source: US Census Bureau.

# Example 8

What kind of visual would have been obtained if, instead of using the gray-red color pair to highlight some observation values, different colors were used for each type of sport in the plot below?

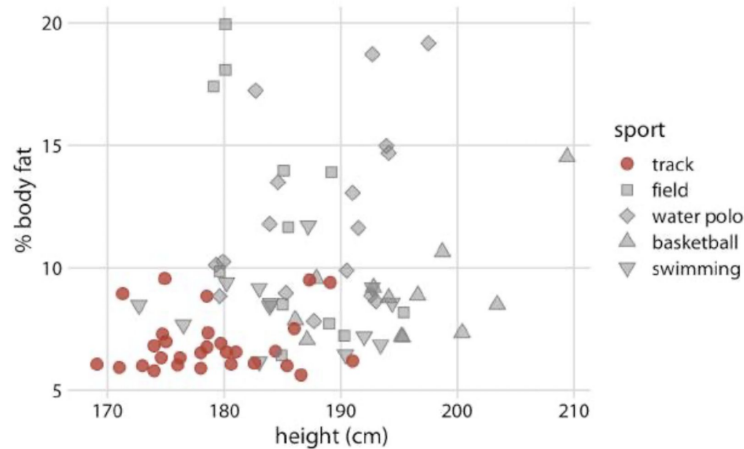


Figure 4-9. Track athletes are among the shortest and leanest of male professional athletes participating in popular sports. Data source: [Telford and Cunningham 1991].

# Reference

The notes and plots in the presentation are compiled from Claus O. Wilke's book, *Fundamentals of Data Visualization*.

