Introduction to Data Visualization

Introduction

- Data visualization is a combination of art and science, aiming to accurately convey data without misleading or distorting information.
- The challenge is to strike a balance between scientific accuracy and aesthetic appeal.
- A good data visualization should enhance the message and avoid distracting elements.

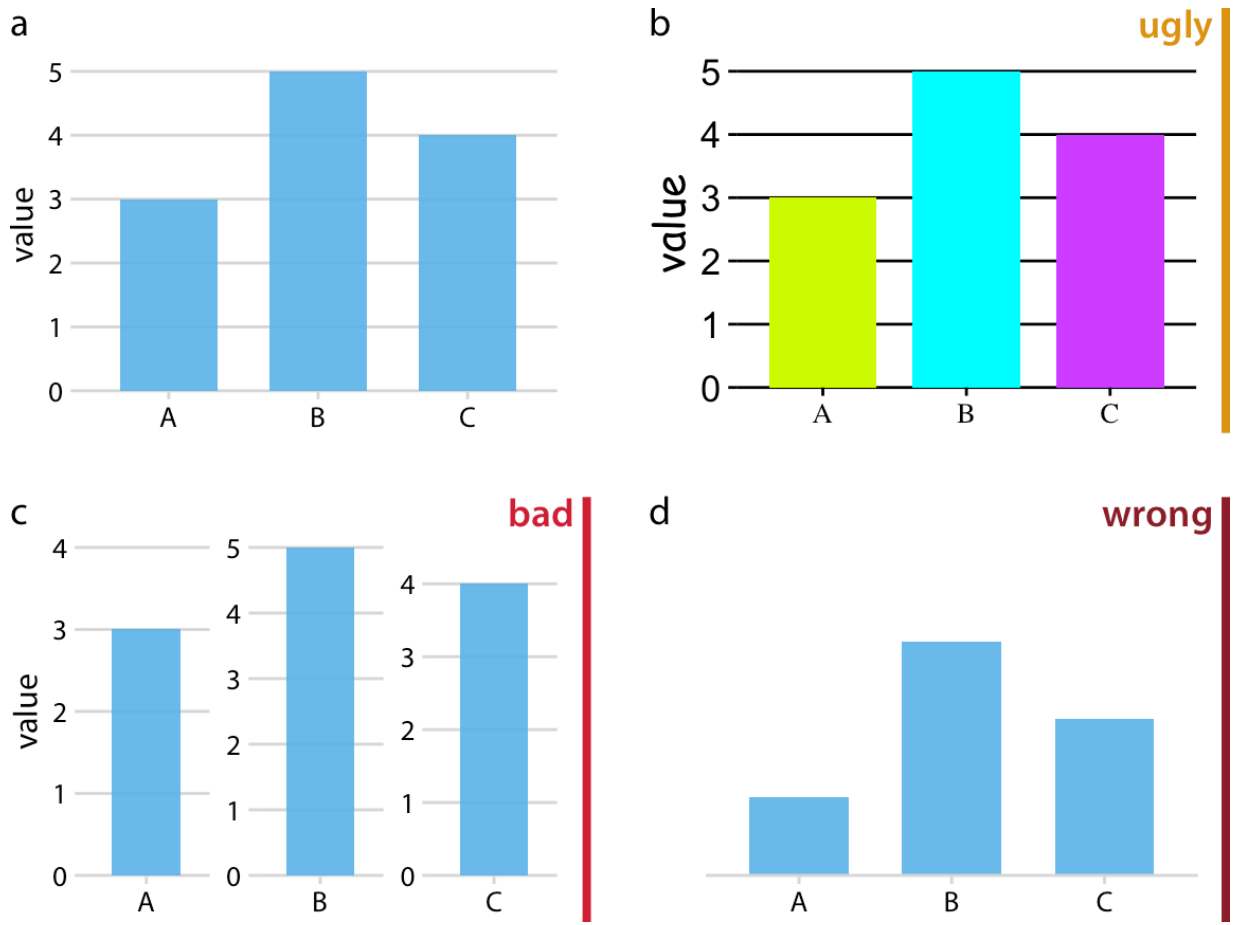
Target Audience

- Scientists may understand data but might lack a sense of visual aesthetics.
- Designers may create visually appealing visuals but may compromise data accuracy.
- The goal is to provide valuable information to both groups.

Principles

- **Categories**
  - **From data to visualization**: types of plots and charts
  - **Principles of figure design**: aesthetic aspects of data visualization
- **Labelling of images**
  - **Ugly**: aesthetic problems
  - **Bad**: perception issues
  - **Wrong**: mathematical inaccuracies

Examples of ugly, bad, and wrong figures



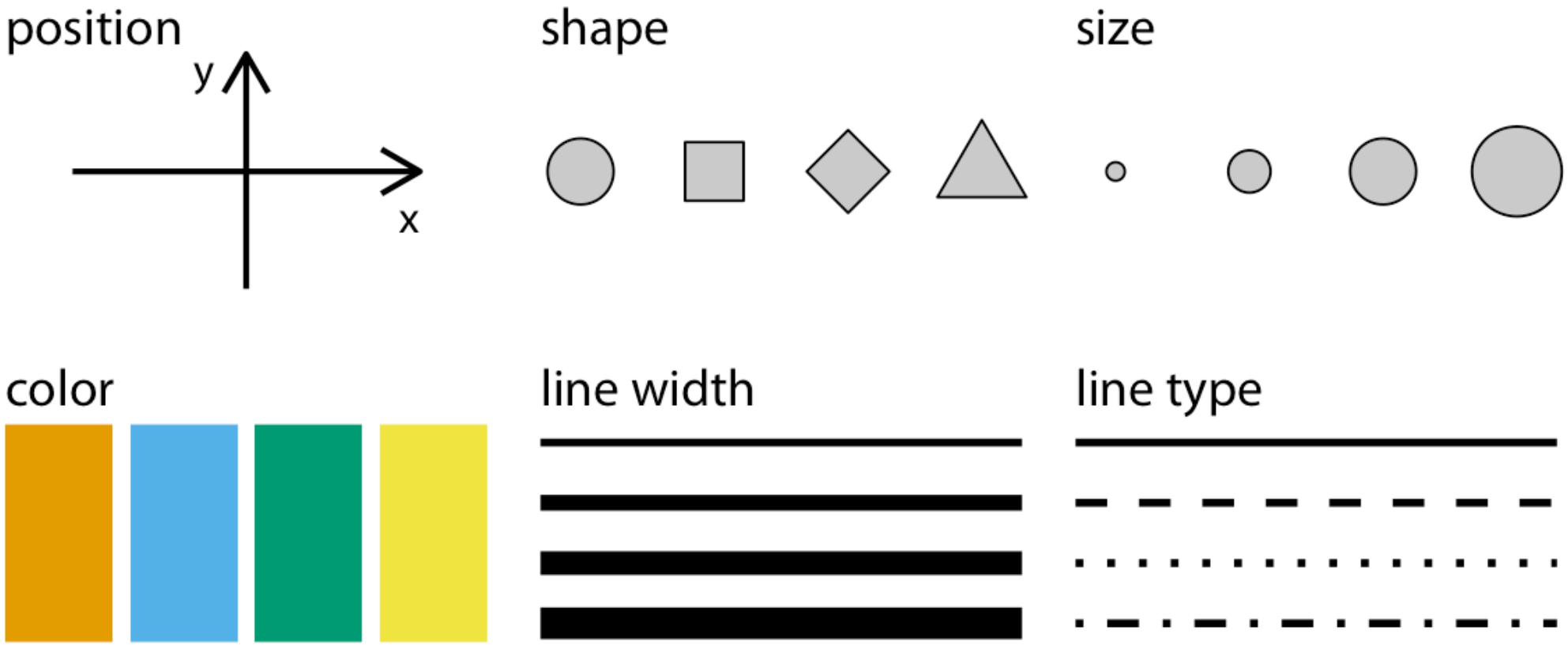
Visualizing data: Mapping data onto aesthetics

- Visualization involves converting data values into visual elements systematically
- Different types of visualizations share a common language in mapping data values to quantifiable features called aesthetics
- Aesthetics include position, shape, size, color, line width, and line type

Types of Aesthetics

- Aesthetics describe aspects of graphical elements, such as position, shape, size, color, line width, and line type
- Different aesthetics represent different types of data, with some suitable for continuous and others for discrete data

Common aesthetics in data visualization



Types of Data

- Data can be quantitative (numerical), qualitative (categorical), or involve dates/times and text
- Variables include quantitative continuous, quantitative discrete, qualitative unordered, qualitative ordered, date/time, and text

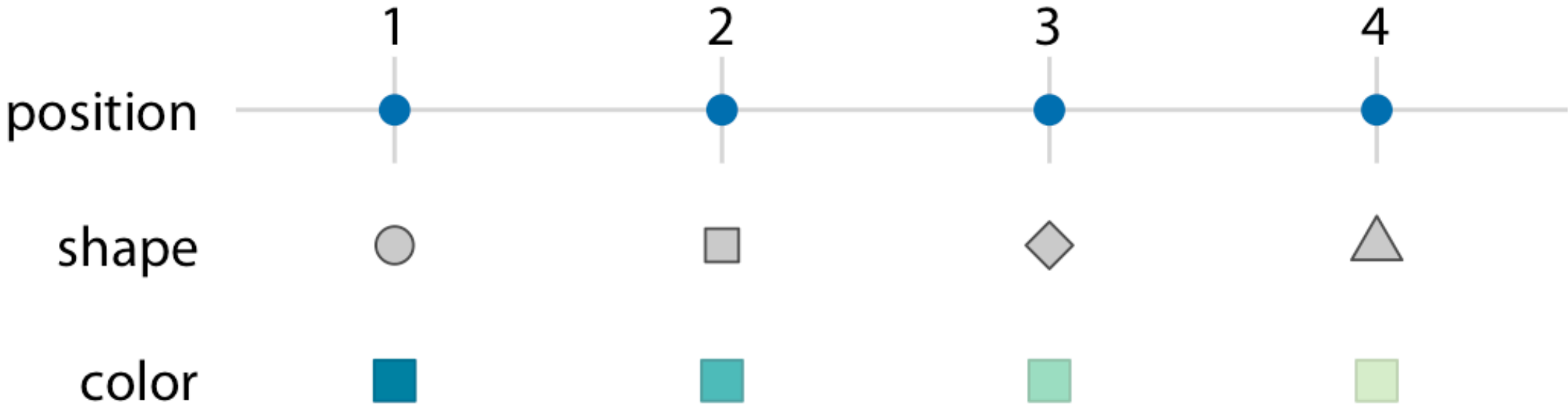
Types of variables in data visualization

Type of variable	Examples	Appropriate scale	Description
quantitative/numerical continuous	1.3, 5.7, 83, $1.5x10^{-2}$	continuous	Arbitrary numerical values. These can be integers, rational numbers, or real numbers.
quantitative/numerical discrete	1, 2, 3, 4	discrete	Numbers in discrete units. These are most commonly but not necessarily integers. For example, the numbers 0.5, 1.0, 1.5 could also be treated as discrete if intermediate values cannot exist in the given dataset.
qualitative/categorical unordered	dog, cat, fish	discrete	Categories without order. These are discrete and unique categories that have no inherent order. These variables are also called *factors*.
qualitative/categorical ordered	good, fair, poor	discrete	Categories with order. These are discrete and unique categories with an order. For example, “fair” always lies between “good” and “poor”. These variables are also called *ordered factors*.
date or time	Jan. 5 2018, 8:03am	continuous or discrete	Specific days and/or times. Also generic dates, such as July 4 or Dec. 25 (without year).
text	The quick brown fox jumps over the lazy dog.	none, or discrete	Free-form text. Can be treated as categorical if needed.

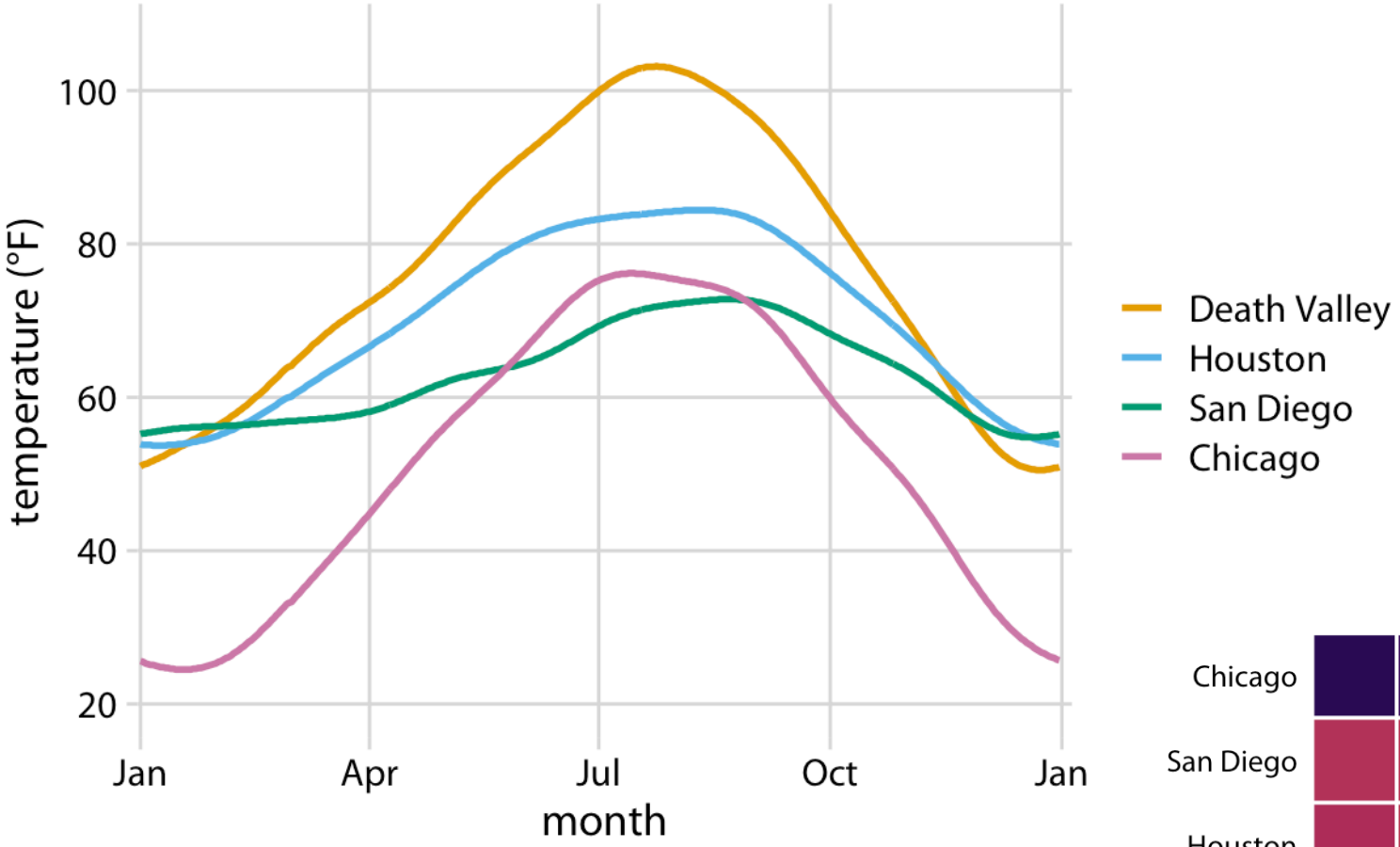
Example Dataset

Month	Day	Location	Station ID	Temperature
Jan	1	Chicago	USW00014819	25.6
Jan	1	San Diego	USW00093107	55.2
Jan	1	Houston	USW00012918	53.9
Jan	1	Death Valley	USC00042319	51.0
Jan	2	Chicago	USW00014819	25.5
Jan	2	San Diego	USW00093107	55.3
Jan	2	Houston	USW00012918	53.8
Jan	2	Death Valley	USC00042319	51.2
Jan	3	Chicago	USW00014819	25.3
Jan	3	San Diego	USW00093107	55.3
Jan	3	Death Valley	USC00042319	51.3
Jan	3	Houston	USW00012918	53.8

Scales Mapping Data to Aesthetics

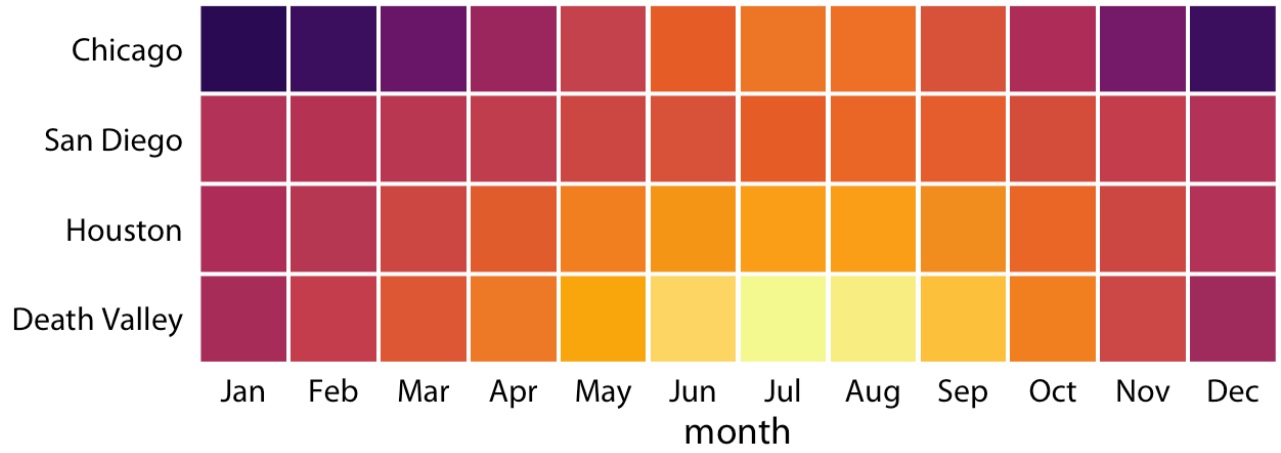


Practical Application

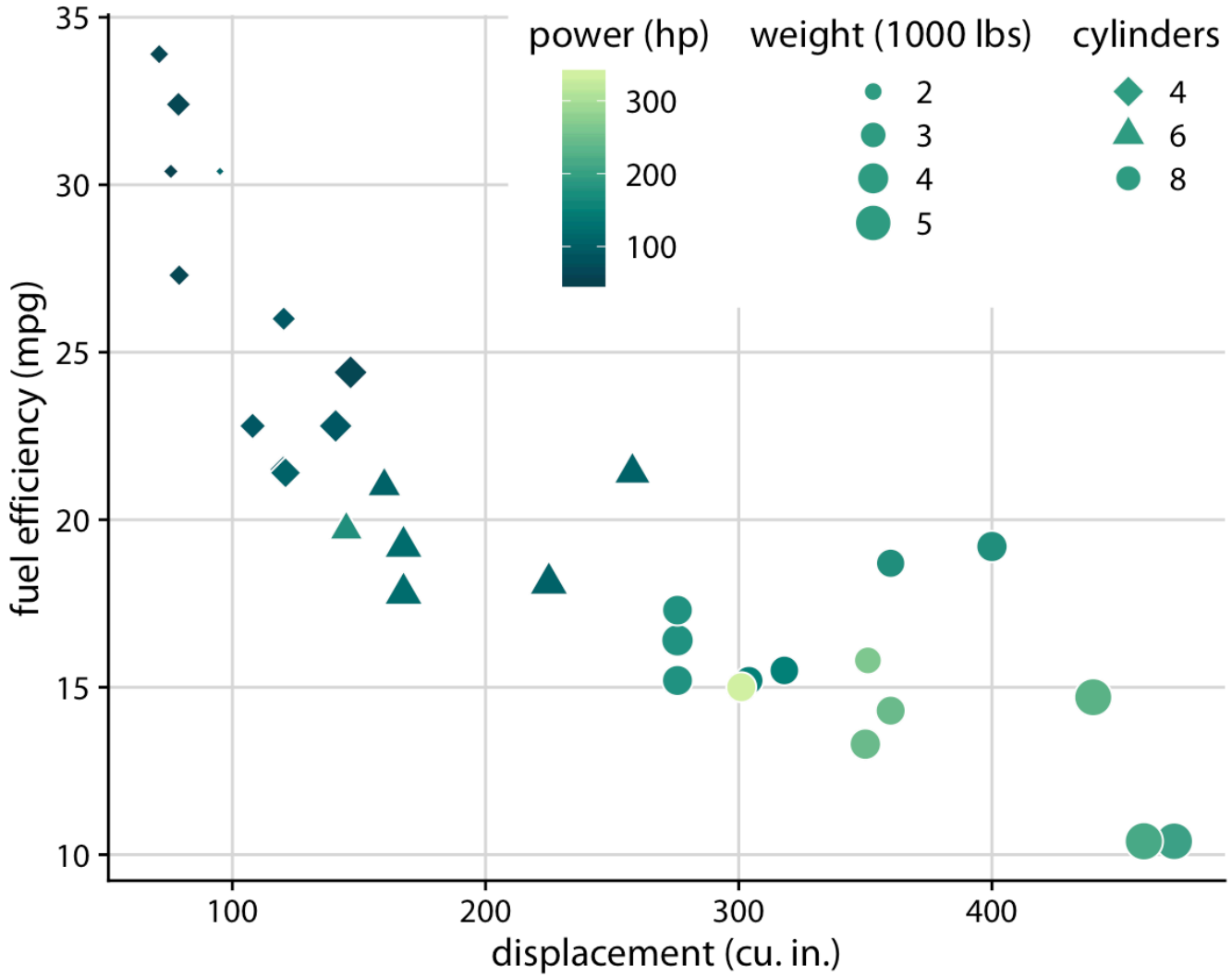


Left: showing daily temperature normals with temperature mapped to the y-axis, day of the year to the x-axis, and location to line color

Right: demonstrates an alternative mapping, using squares and color to represent temperature normals



Multiple Scales in Visualization



- Complex visualization with five scales representing different variables.
- Includes *x* and *y* axes, color, size, and shape scales, each mapping to a specific variable.