CP321 Data Visualization

Practical Visualization & Coordinate systems

Jiashu (Jessie) Zhao

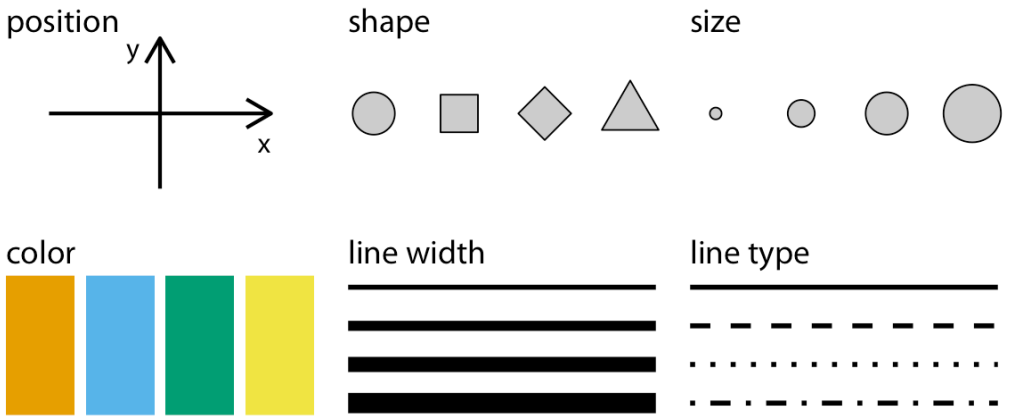
Outline

- Visualizing data: Mapping data onto aesthetics

- Coordinate systems and axis
  - Cartesian coordinates
  - Nonlinear axes
  - Curved axes

Visualizing data: Mapping data onto aesthetics

- Commonly used aesthetics in data visualization:
  - Position
  - Shape
  - Size
  - Color
  - Line width
  - Line type



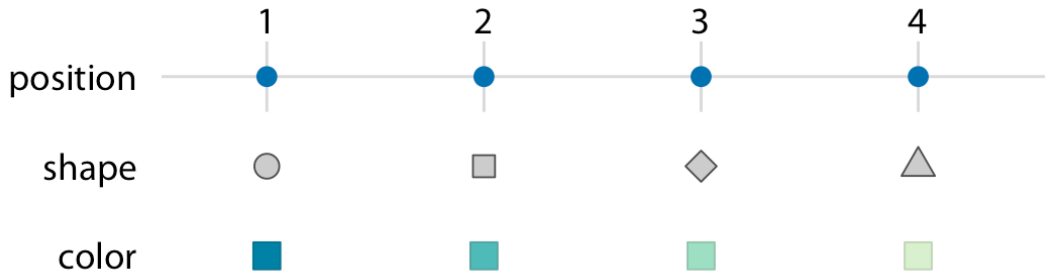
Which one(s) can only represent discrete data??

- Types of data in typical data visualization scenarios

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.

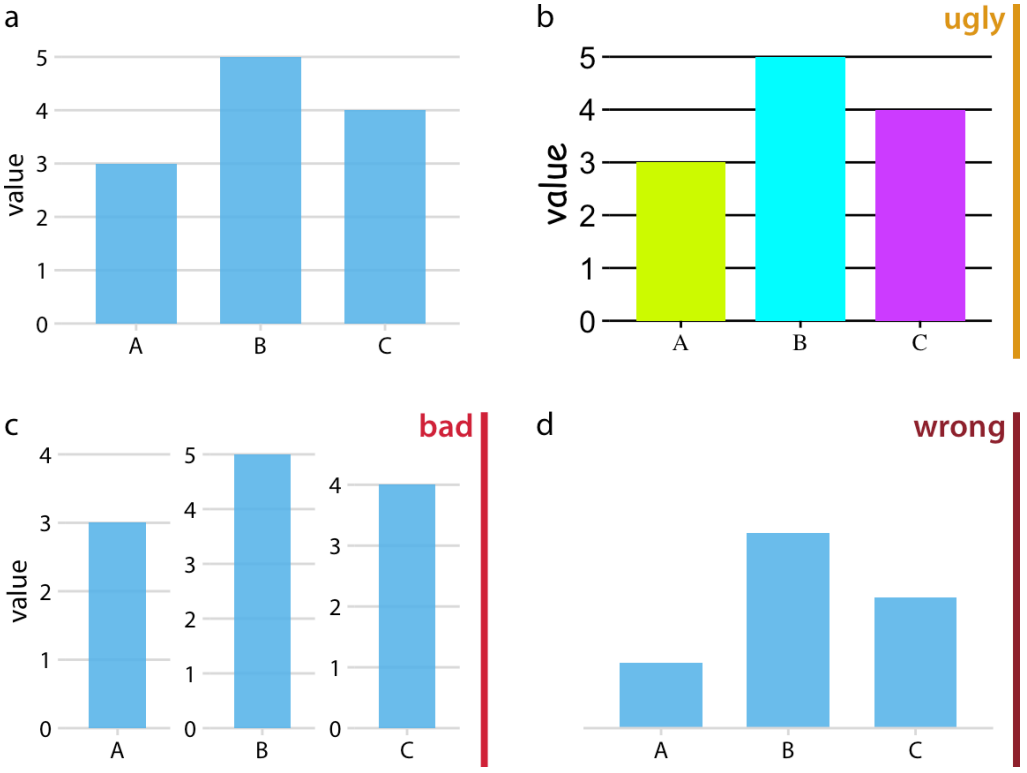
Scales map data values onto aesthetics

- The mapping between data values and aesthetics values is created via ***scales***.



- A scale must be one-to-one

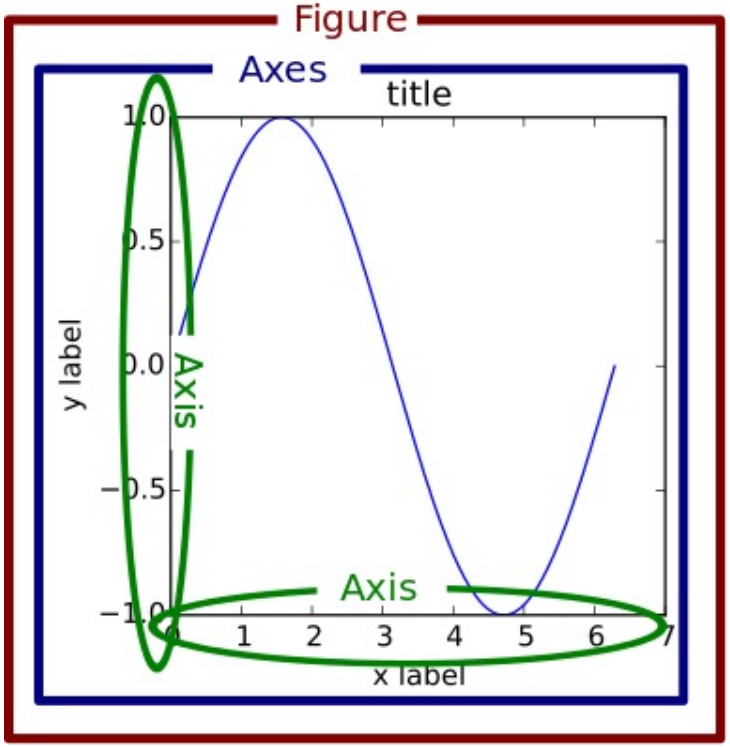
- Data visualization is part art and part science.



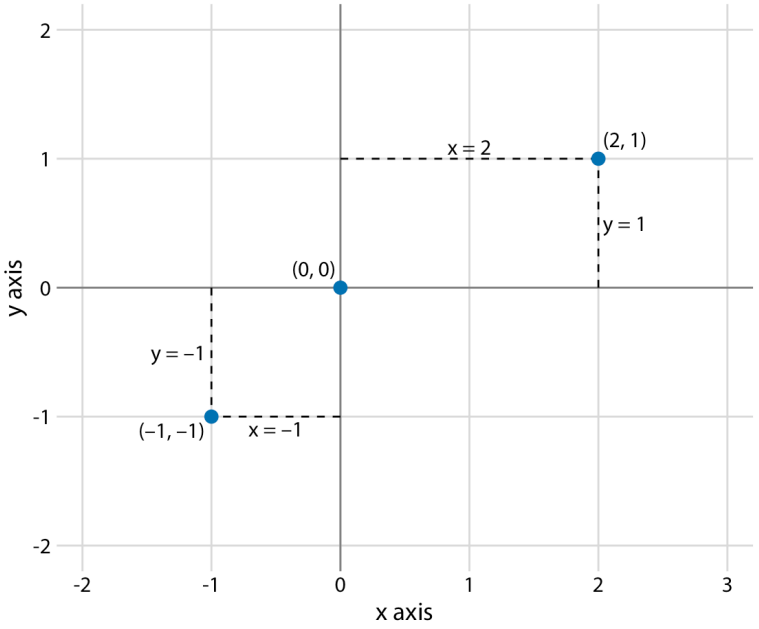
Examples of ugly, bad, and wrong figures

Coordinate systems and axes

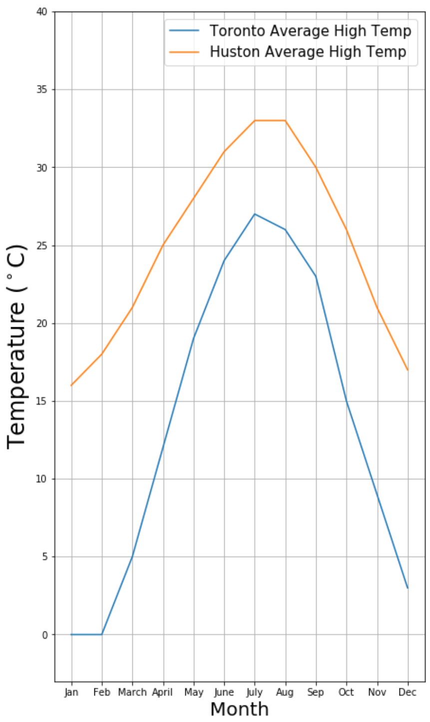
- Position scales - determine where in a graphic different data values are located.
- 2-dimension visualizations - two numbers are required to uniquely specify a point, and therefore we need two position scales.
- 3-d: 3 position scales

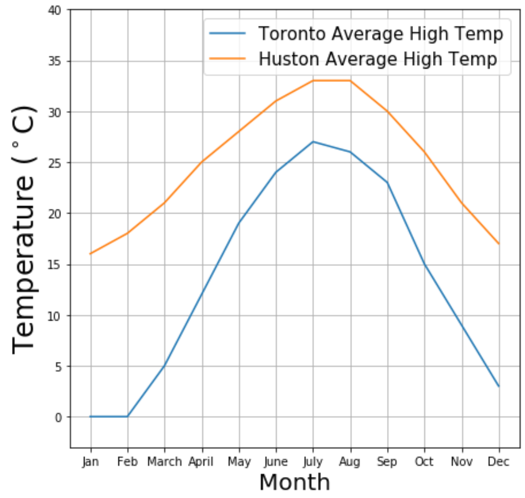


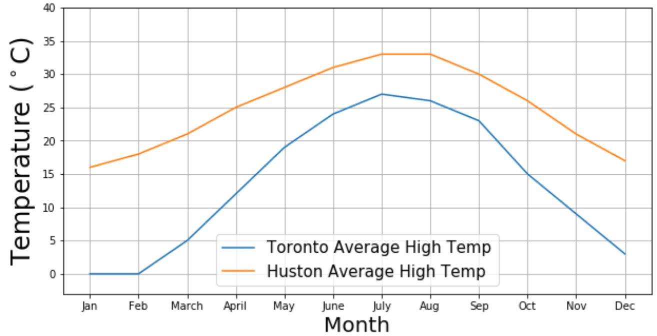
- Cartesian coordinates
  - The most widely used coordinate system
  - The axis run orthogonally to each other.
  - Data values are placed in an even spacing along both axis



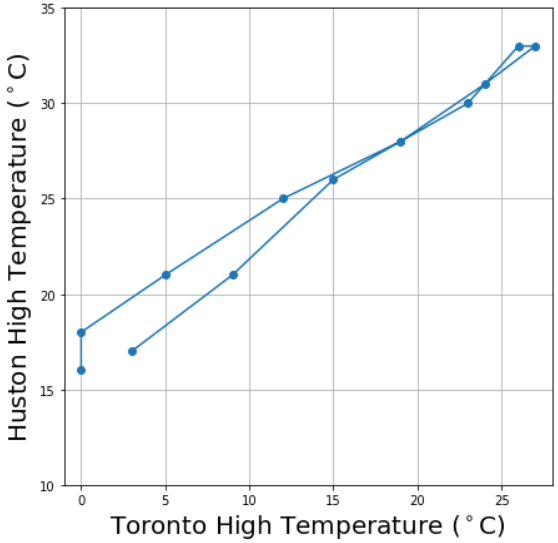
Example: two axes representing two different units

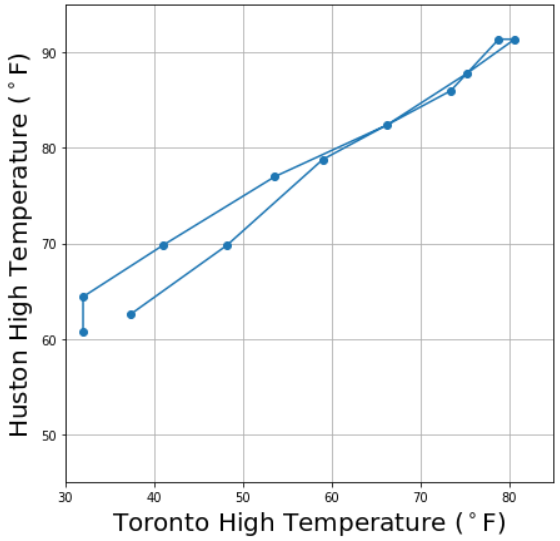






Example: Same unit & change in unit



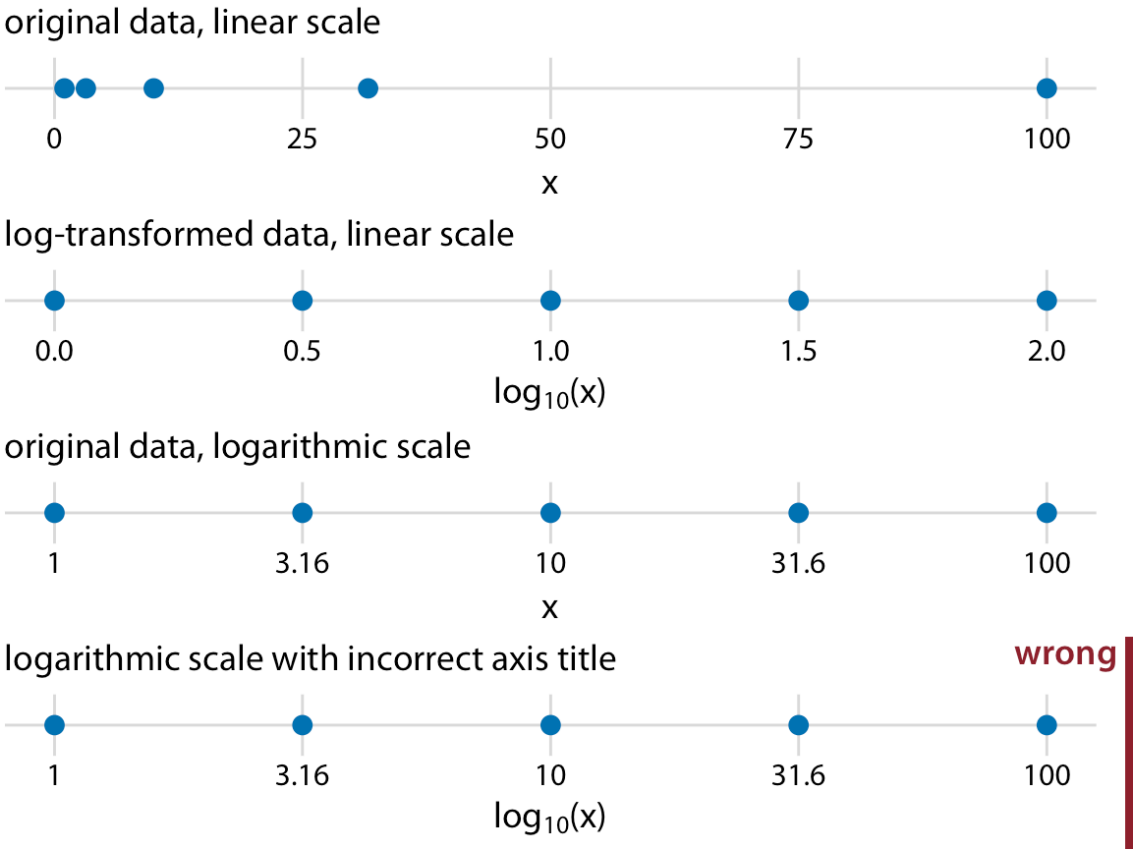


Note:

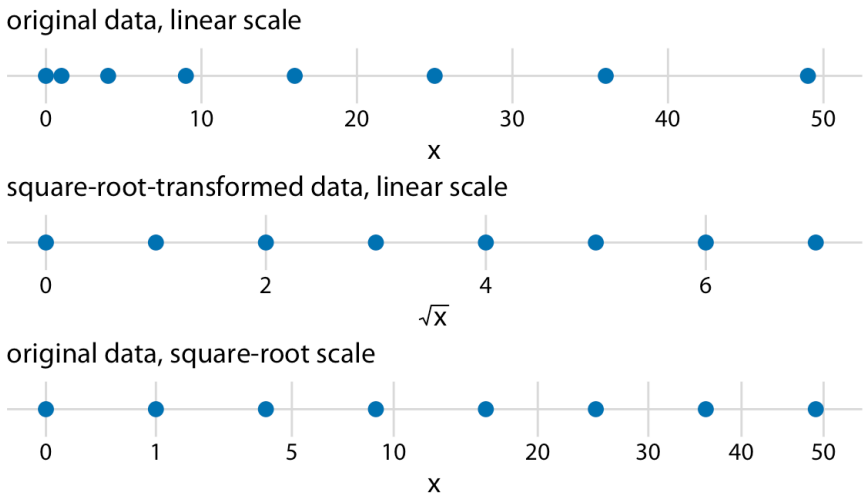
- Use equal grids for same unit
- Cartesian coordinate systems are invariant under linear transformations

- What if we want to visualize highly skewed data?
- Nonlinear axes
    - Even spacing in data units corresponds to uneven spacing in the visualization

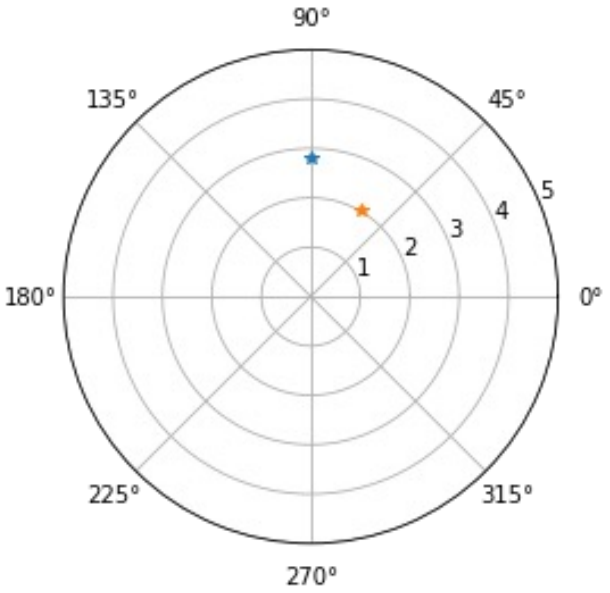
- *log scale*


original data, linear scale
0
25
50
75
100
x
log-transformed data, linear scale
0.0
0.5
1.0
1.5
2.0
log10(x)
original data, logarithmic scale
1
3.16
10
31.6
100
x
logarithmic scale with incorrect axis title
wrong
1
3.16
10
31.6
100
log10(x)

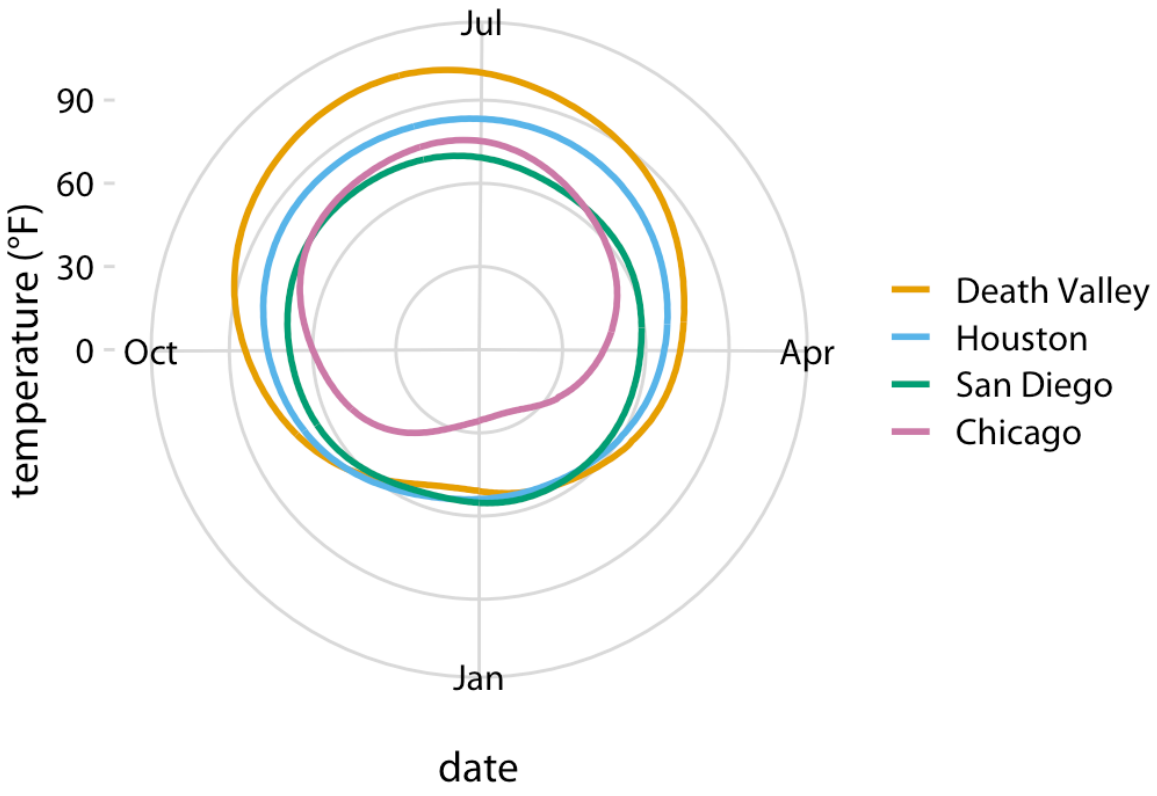
- square-root scale:
  - Also compresses larger numbers into a smaller range
  - Unlike a log scale, it allows for the presence of 0



- Coordinate systems with curved axes
  - *polar* coordinate
    - Pole
    - Radius
    - Polar angle



- Example


Jul
90
60
30
0
Oct
Apr
Jan
temperature (°F)
date
Death Valley
Houston
San Diego
Chicago

- Curved axes: geospatial data
  - Maps. Locations on the globe are specified by their longitude and latitude