Data Distortion

CMPSC 105 – Data Exploration

ALLEGHENY COLLEGE

General Principles: Be Honest

A graphic tells the truth when the visual representation of the data is consistent with the numerical representation.

— Edward Tufte

Why Data Distortion Matters

Visualizations strongly influence human perception

- Small design choices can:
  - Exaggerate trends
  - Hide variation
  - Mislead decision-making
- Distortion can be:
  - Intentional (persuasion, marketing, politics)
  - Unintentional (poor defaults, lack of awareness)

A graph can be *technically correct* and still be visually dishonest

Visual Honesty (Tufte, 1998)

A graph is honest when:

- Visual change = numerical change
- Scales are consistent and interpretable
- The design does not add meaning that isn't in the data

Common distortion mechanisms

1. Aspect ratio manipulation
2. Scale manipulation
3. Excessive scale range

Distortion Type 1: Aspect Ratio

Adjusting the Aspect Ratio

- Aspect ratio = height/width of the plotting area
- Changing only the height
  - Make trends look steeper
  - Make changes look dramatic or trivial
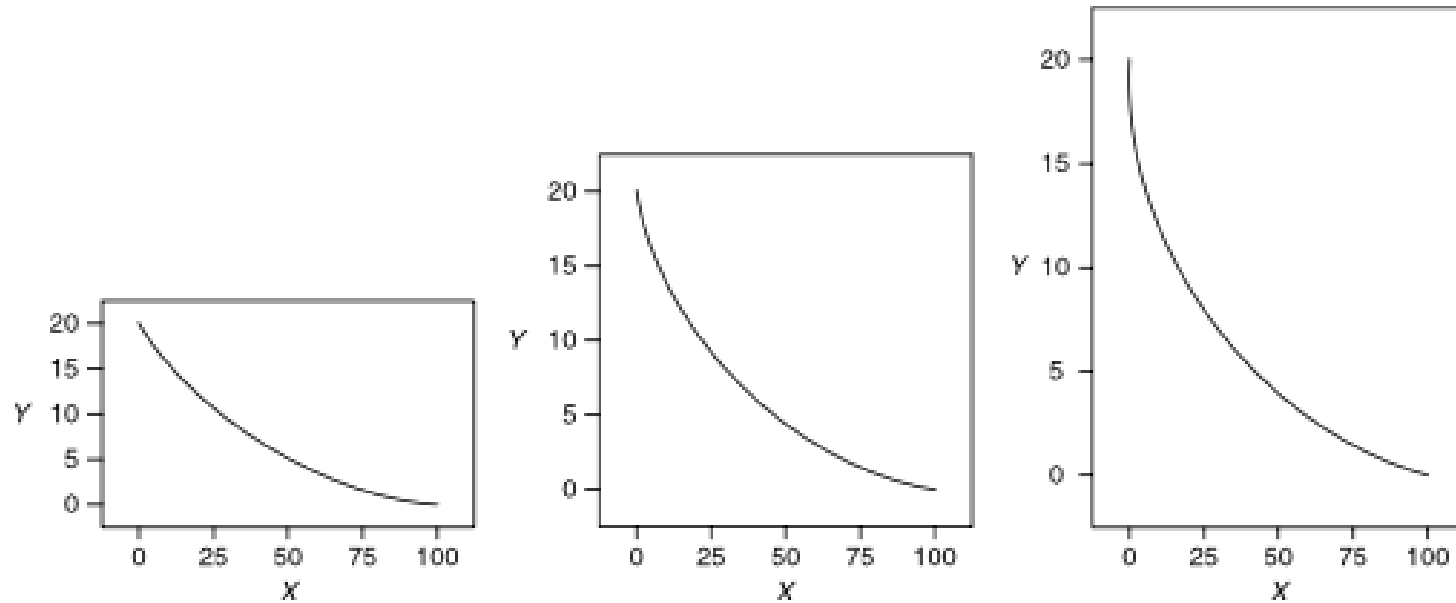
Distortion Type 1: Aspect Ratio

Figure 2.4   Making a more dramatic statement by adjusting only the aspect ratio

Aspect Ratio: What's Happening?
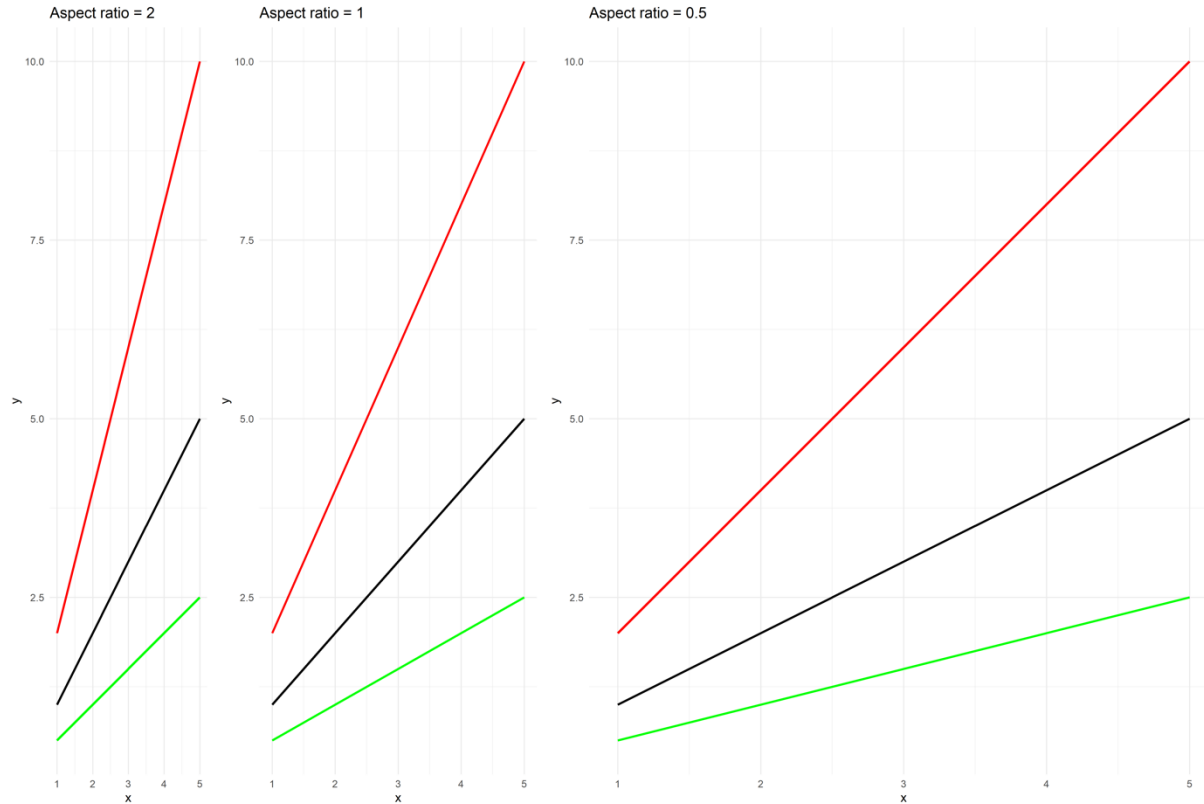
Same data → different perception

- Left: Flatter → "slow change"
- Right: Steeper → "rapid growth"
- No numbers changed, only geometry

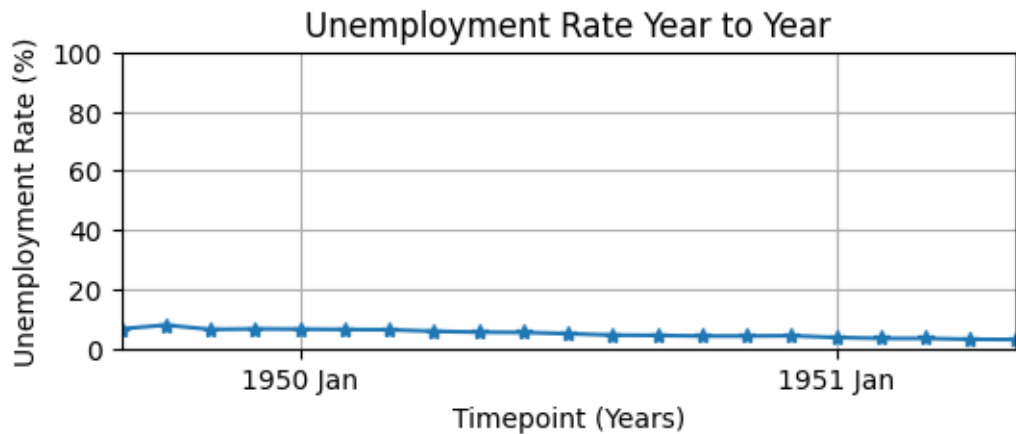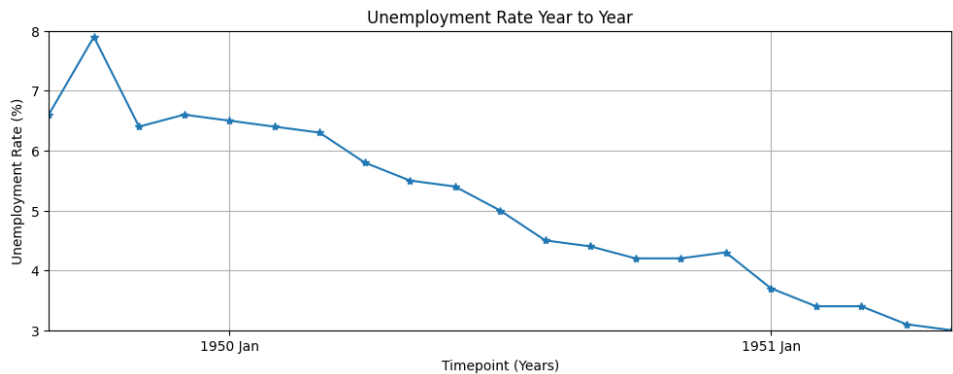➤ Aspect ratio should reflect data meaning, not storytelling goals
➤ Avoid:
  o Extremely tall or compressed plots
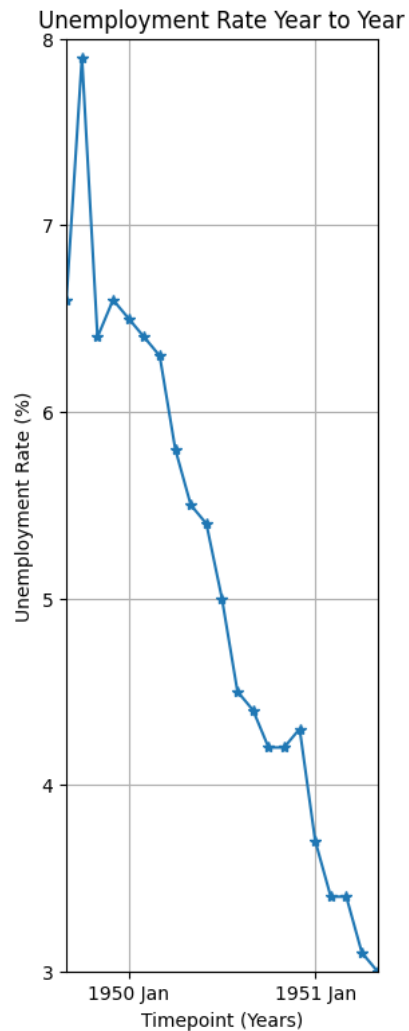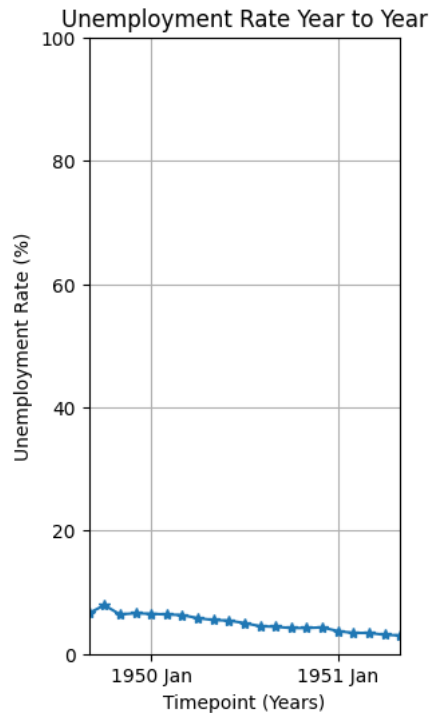 o "Chart stretching" to create drama

Distortion Type 1: Aspect Ratio

Distortion Type 1: Aspect Ratio

Distortion Type 1: Aspect Ratio

Unemployment Rate Year to Year

Distortion Type 2: Irregular Scales

Manipulating the Scale

- Occurs when:
  - Axis intervals are uneven
  - Equal visual spacing ≠ equal numerical spacing

Why this is dangerous

- Humans assume: Equal distance = equal quantity

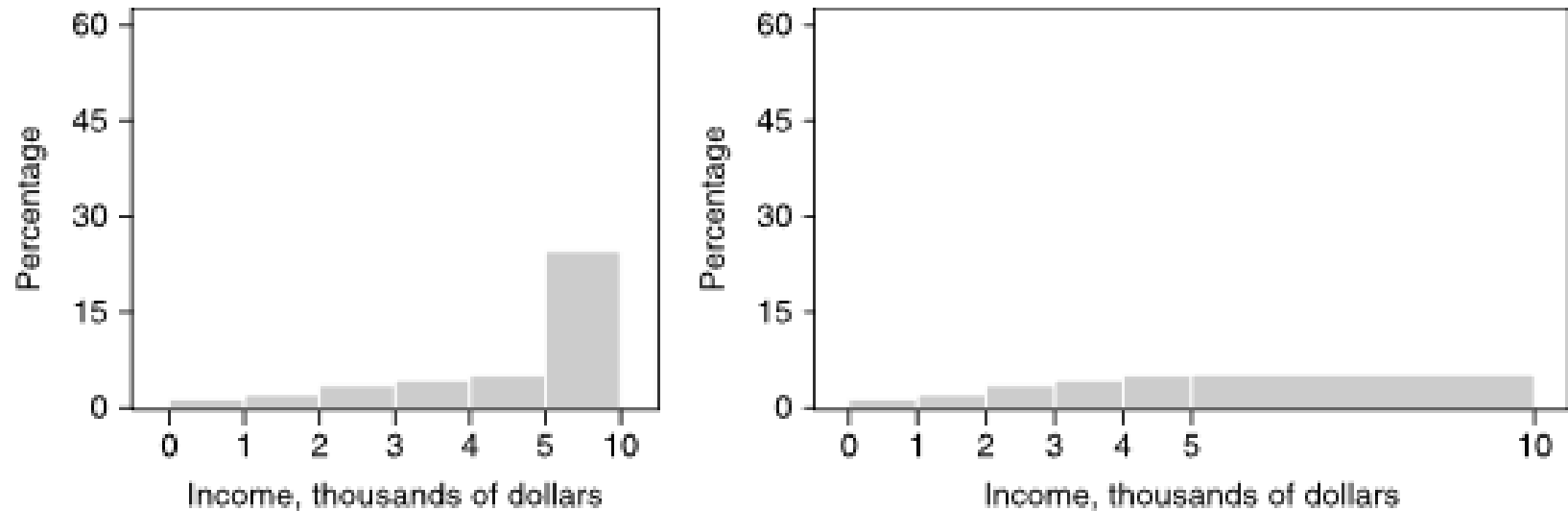# Distortion Type 2: Irregular Scales

Figure 2.5   An irregular scale distorts magnitudes of unit values

Income Histogram

What goes wrong?

- One x-axis "unit" represents:
  - $1,000 in some places
  - $5,000 in the last
- The bar for $5k–$10k: is not comparable to other bars

Corrected version

- Uses consistent bin widths
- Percentages should reflect fair comparisons

Distortion Type 3: Excessive Scale Range (Zooming out)

Increasing the Scale Range
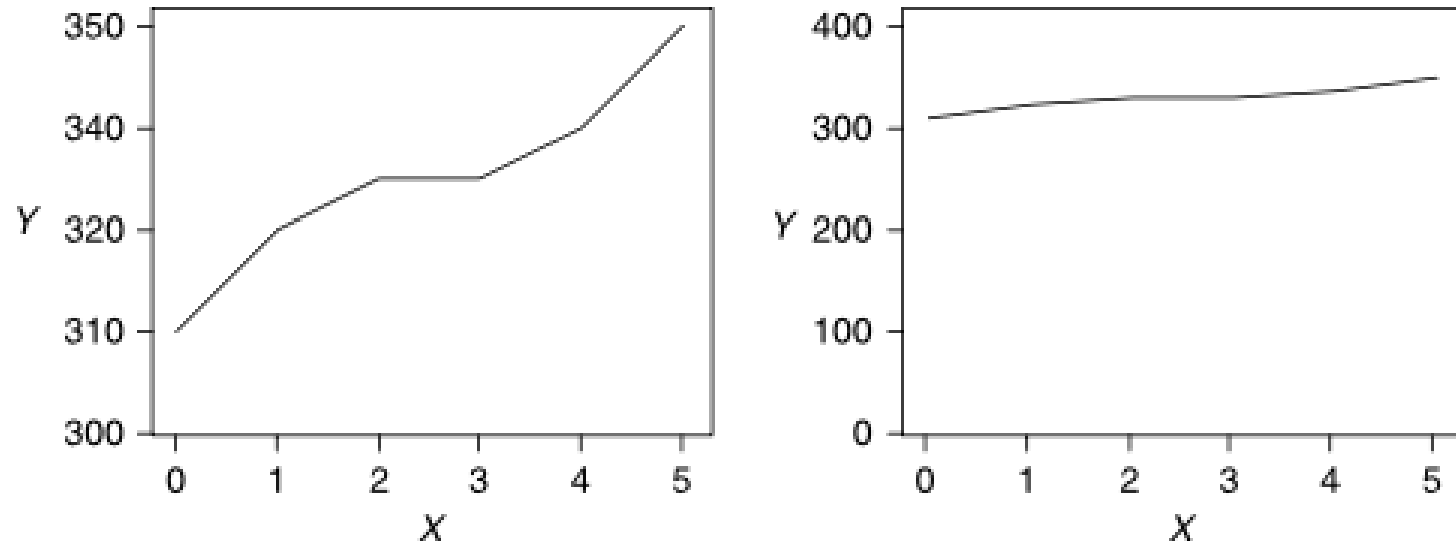
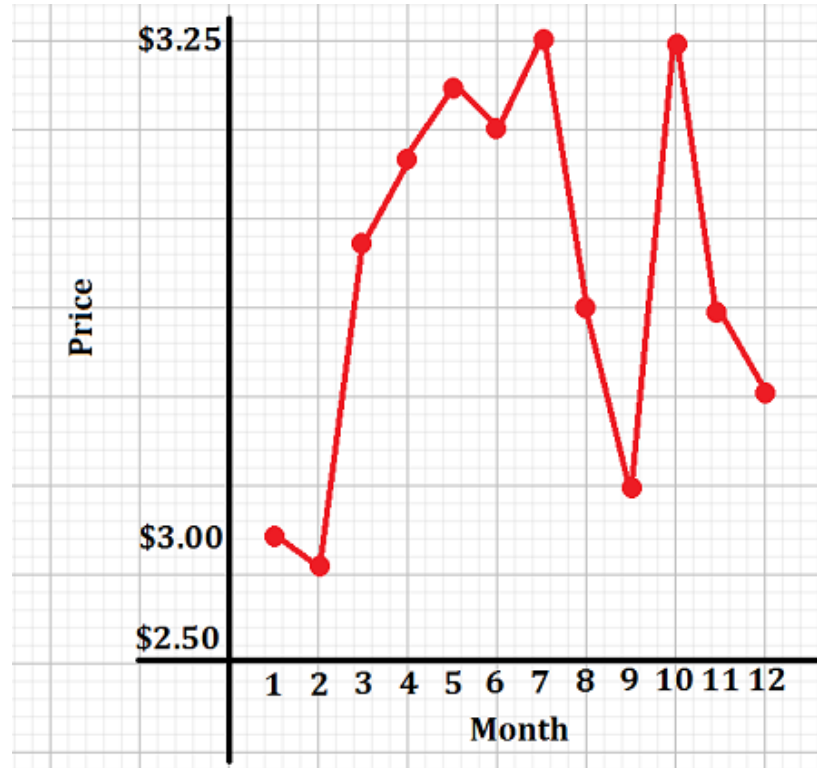- Large y-axis range: Compresses variation

Figure 2.6   Increasing the range of a scale may hide important detail

Example: Hidden Detail (Fig. 2.6)

What's the issue?

- Left plot:
  - Reveals small but important changes
- Right plot:
  - Same data
  - Variation appears negligible

Masks meaningful changes (Zooming in)

Summary: Common Data Distortions

Distortion Type	What Changes	What's Misleading
Aspect ratio	Plot geometry	Trend magnitude
Irregular scale	Axis intervals	Relative size
Large scale range	Axis limits	Variability

How to Avoid Distortion

- Keep axis scales even and labeled

- Meaningful Choose aspect ratios that match the data's natural variation

- "Would a different scale change the story?"

Think–Pair–Share

Which distortion is most common in:

- News media?

- Marketing?

- Academic papers?

Is distortion always unethical?

- Or sometimes unavoidable?