

# Redundant Coding and Best Practices in Data Visualization

# Introduction

- Redundant coding refers to encoding data using multiple aesthetic dimensions to enhance comprehension.
- Helps mitigate challenges associated with color perception and small visual elements.

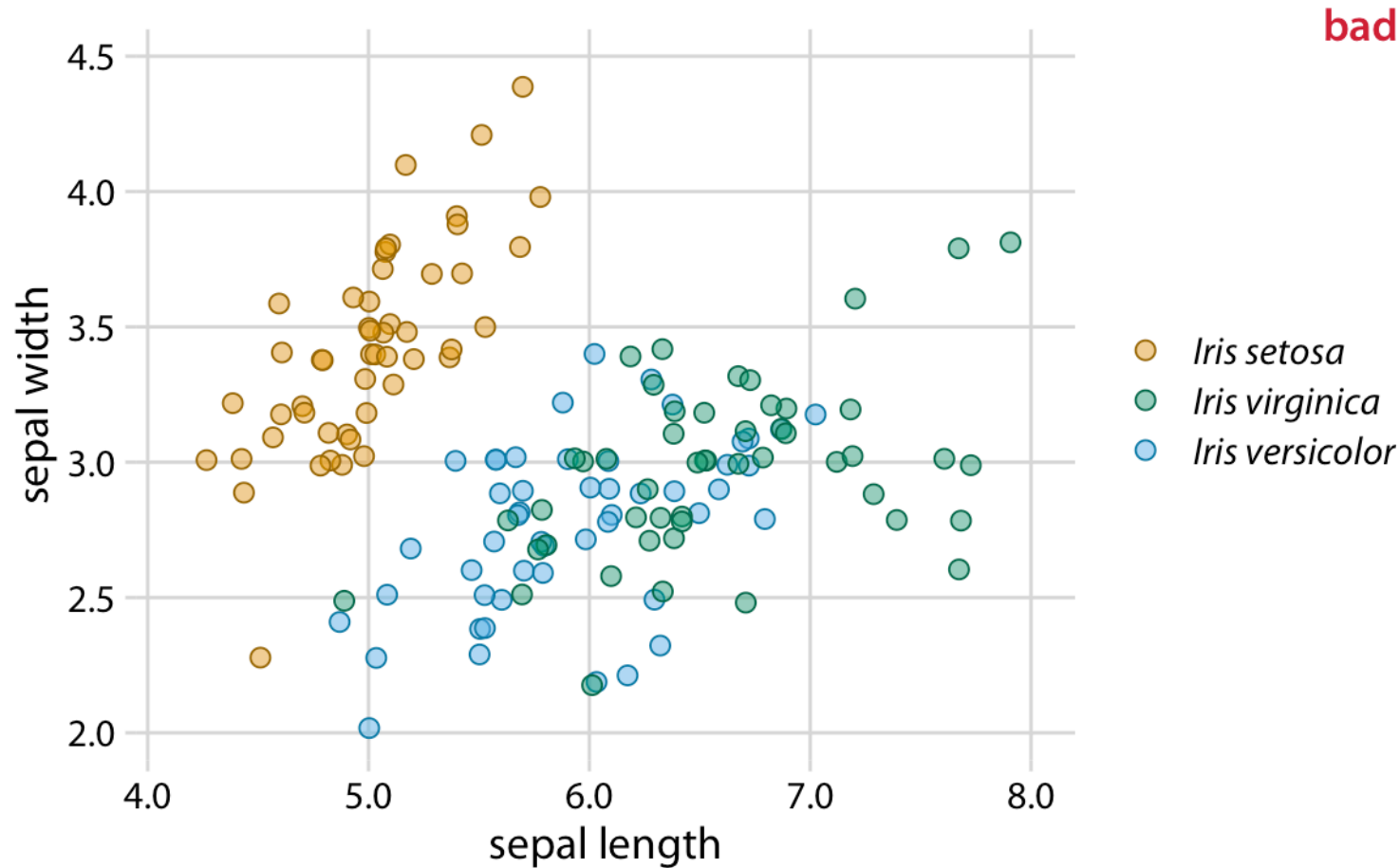
# Why Redundant Coding Matters

- **Color Limitations:** Color alone may not be sufficient for differentiating data points.
- **Accessibility:** Color-vision deficiency (CVD) affects the perception of data.
- **Small Elements:** Small visual elements can be hard to distinguish by color alone.

# Designing Legends with Redundant Coding

- **Problem:** Legends relying solely on color can be difficult to interpret.
- **Solution:**
  - Use multiple aesthetics (e.g., color + shape in scatter plots).
  - Swap colors to avoid confusion in CVD cases.
  - Ensure legibility under grayscale printing.

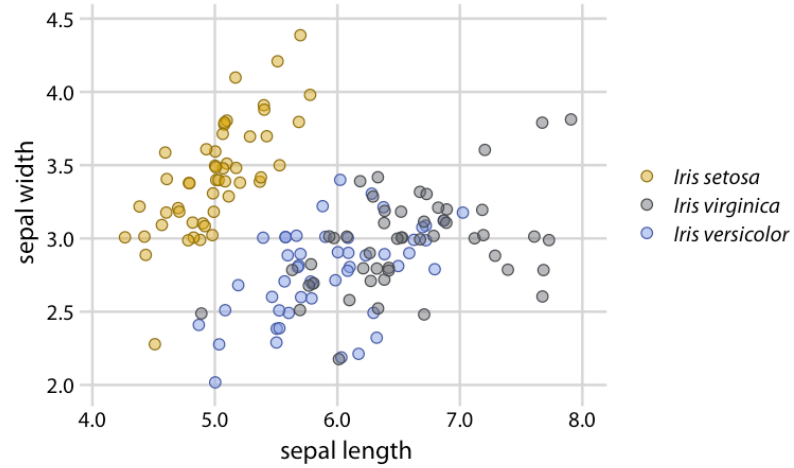
# Iris Sepal Comparison Issues



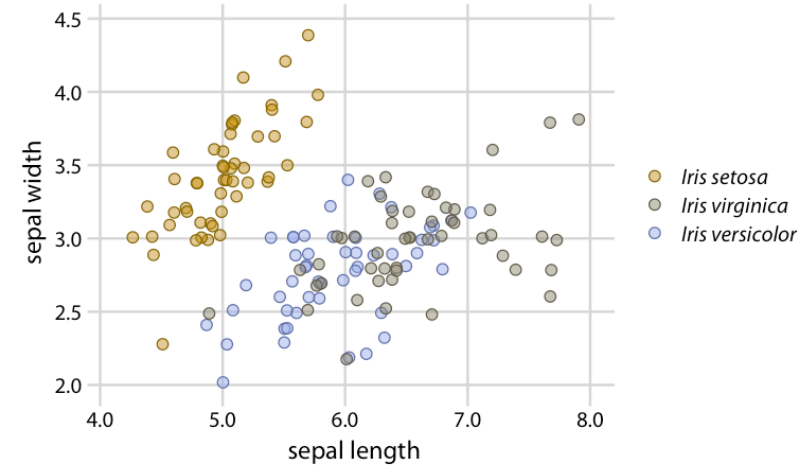
Sepal width vs. sepal length for three iris species (setosa, virginica, versicolor). Each point represents one plant sample with jitter applied to prevent overplotting. The figure is labeled “bad” as virginica (green) and versicolor (blue) are hard to distinguish.

# Color-vision-deficiency simulation

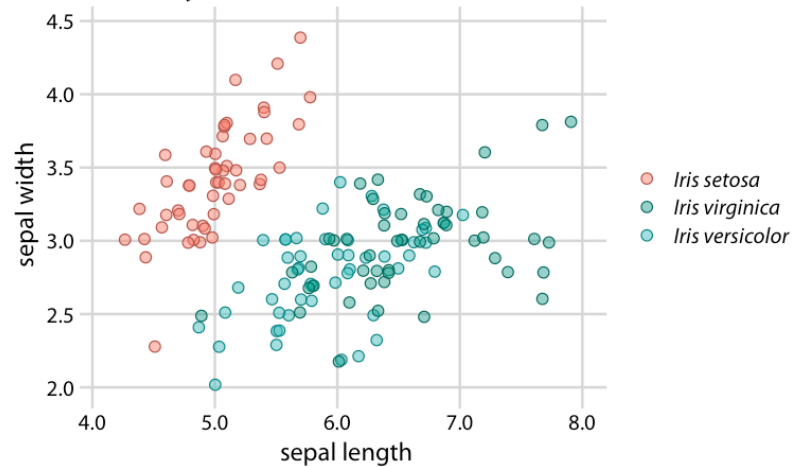
deuteranomaly



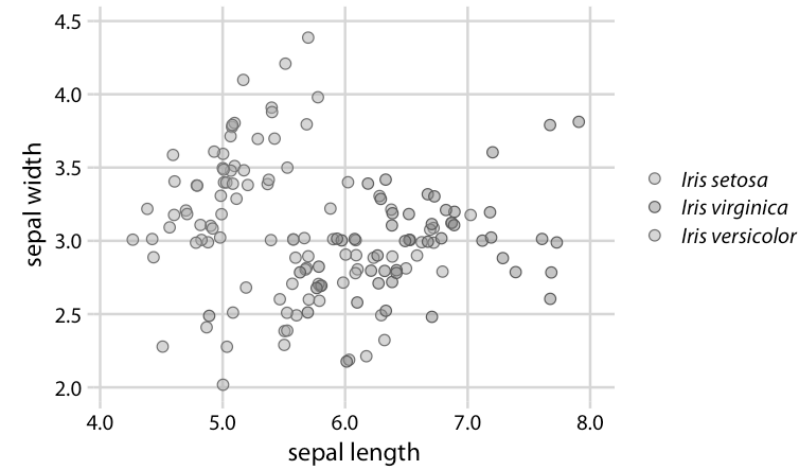
protanomaly



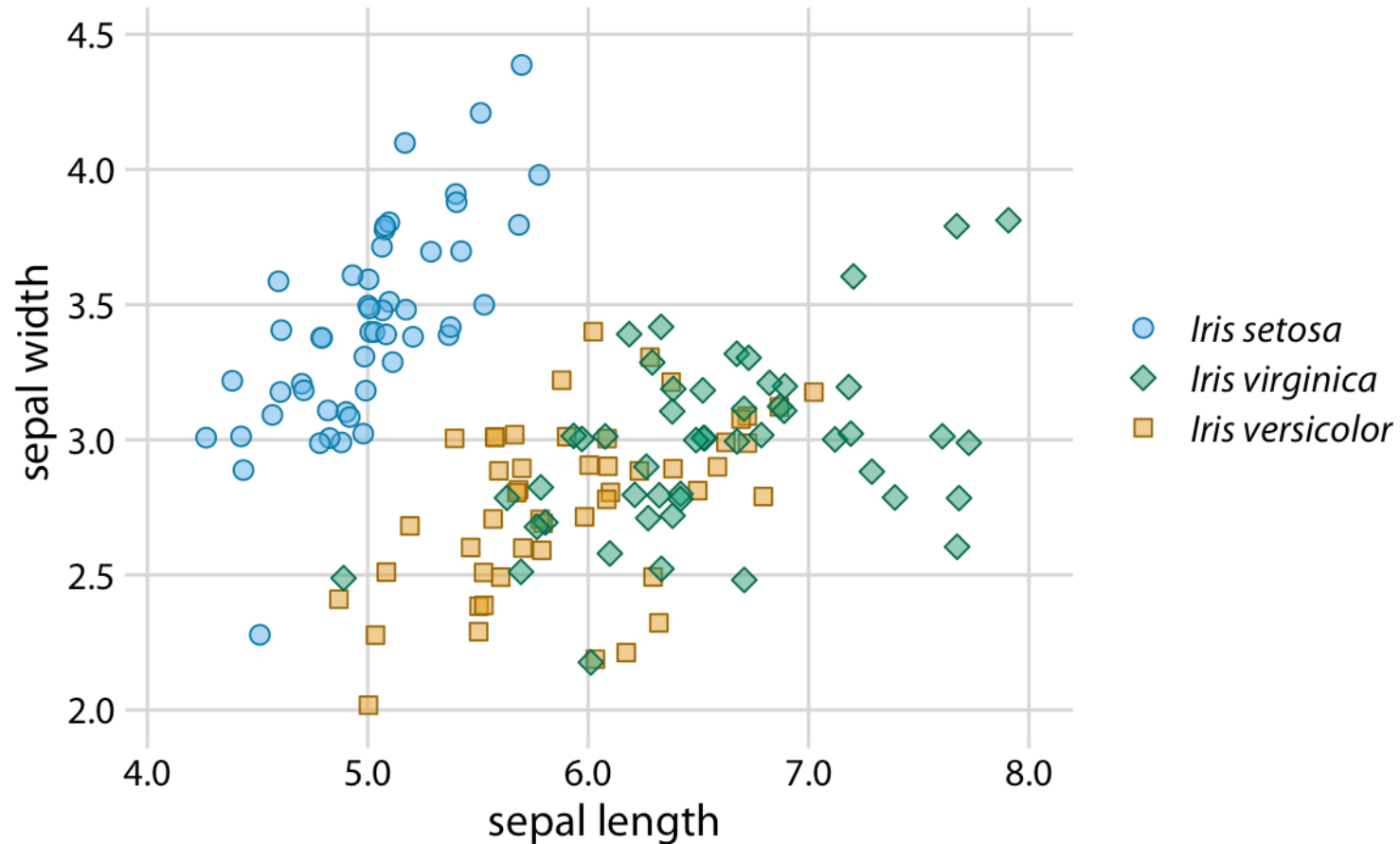
tritanomaly



desaturated

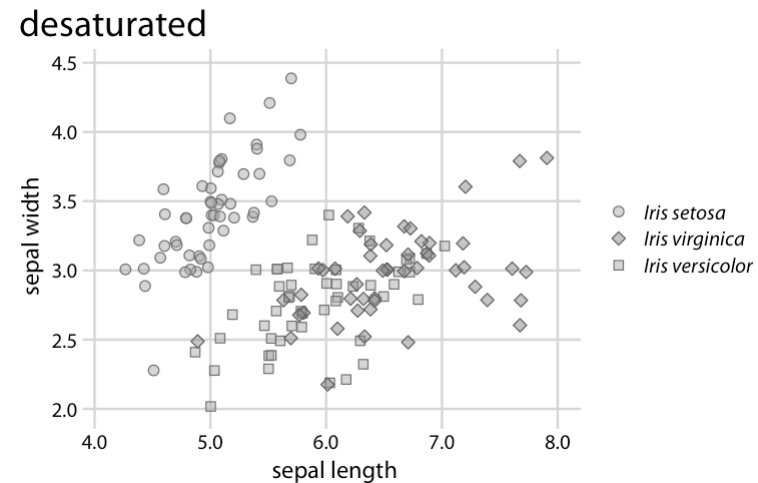
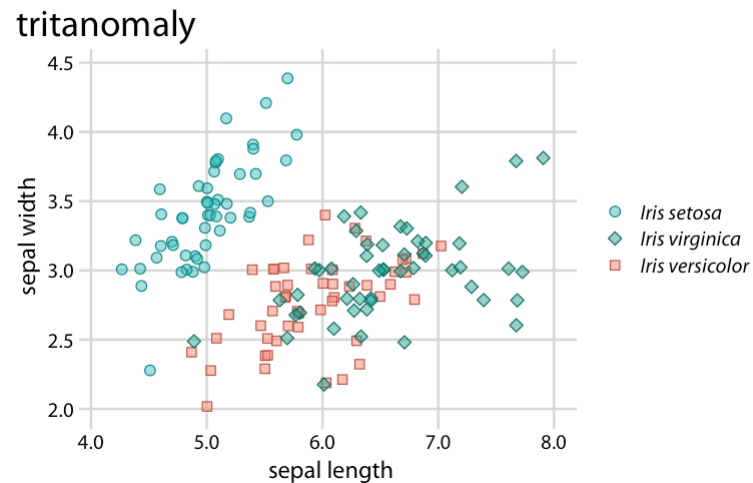
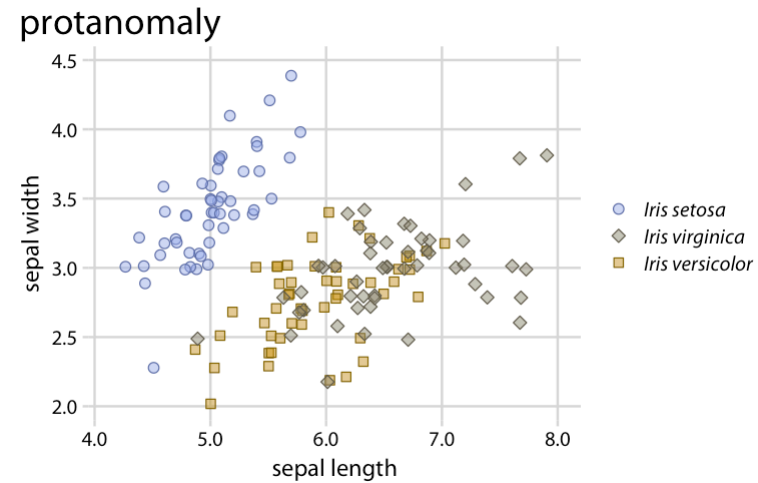
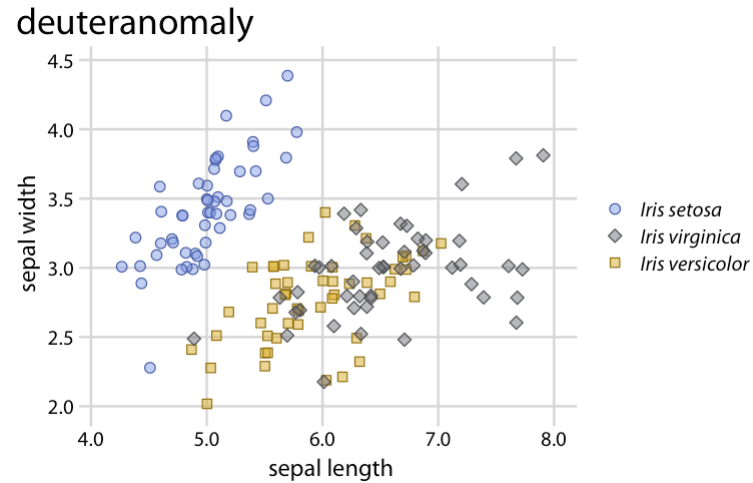


# Iris Sepal Comparison



Sepal width vs. sepal length for three iris species, with swapped colors for *Iris setosa* and *Iris versicolor* and distinct point shapes for each species.

# CVD-Friendly Visualization



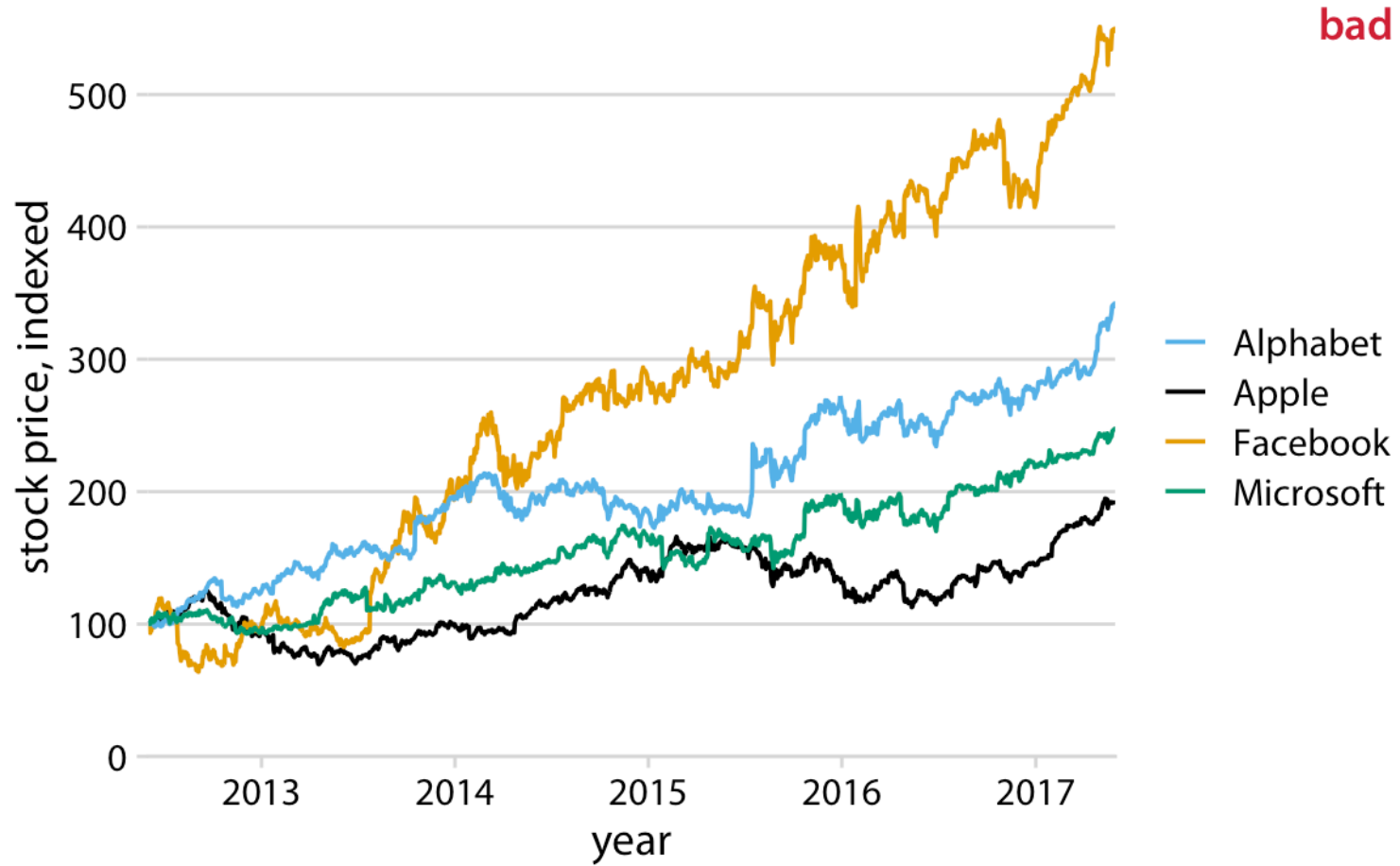
Color-vision-deficiency simulation. Because of the use of different point shapes, even the fully desaturated gray-scale version of the figure is legible.



# Improving Line Graphs

- **Problem:** Line graphs with color-only differentiation require extra effort to match legend entries.
- **Solution:**
  - Order legend items to match the visual order in the graph.
  - Use different line styles (solid, dashed, dotted) where appropriate.

# Tech Stock Trends

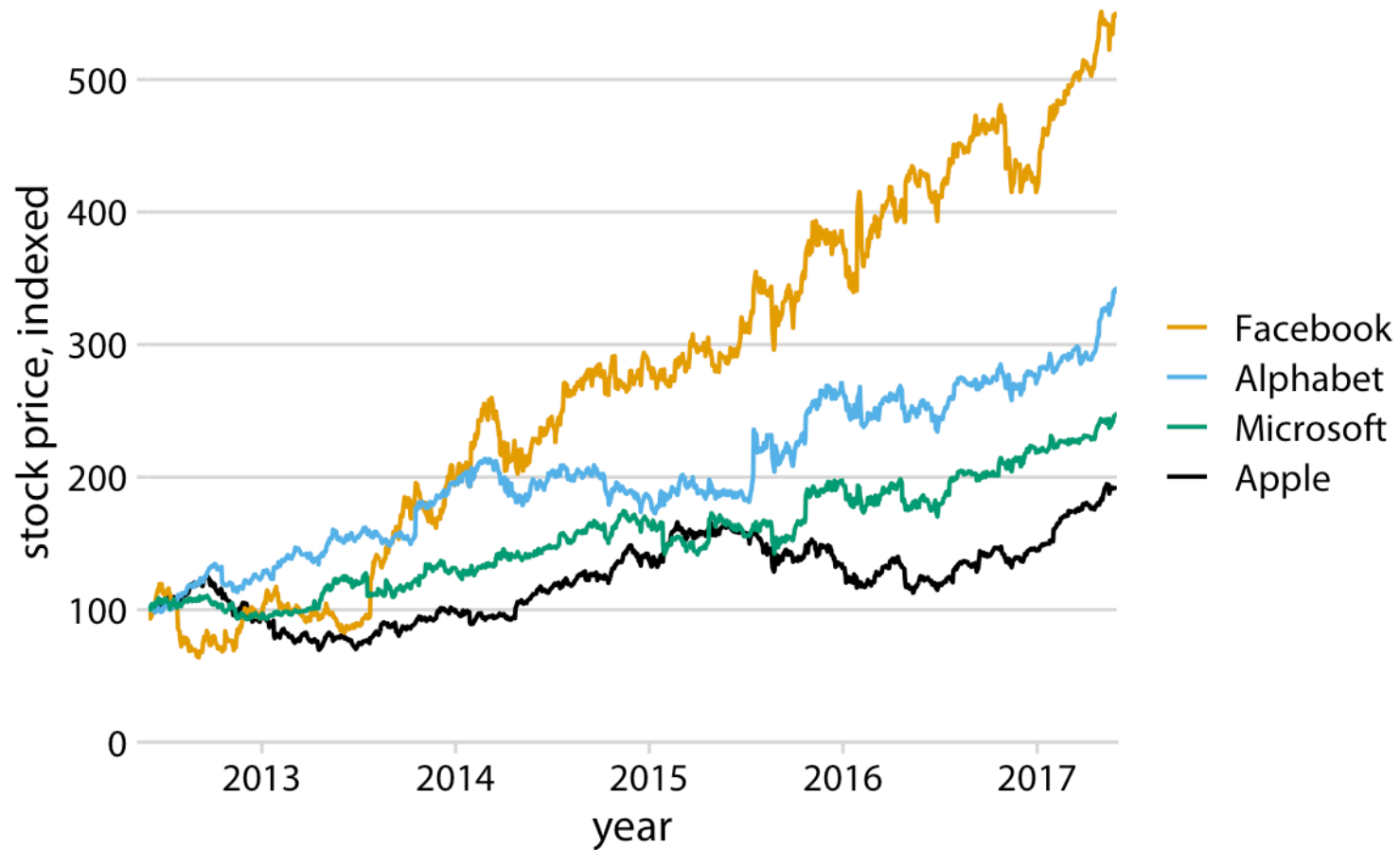


Stock price trends for four major tech companies, normalized to 100 in June 2012. Matching legend entries to data curves requires effort.

# Designing Figures Without Legends

- **Problem:** Legends increase cognitive load by requiring viewers to match colors/symbols to data.
- **Solution:**
  - Use **direct labeling** by placing labels next to data points.
  - Apply color-coding to text labels for better association.
  - Utilize density plots as legend replacements in scatter plots.

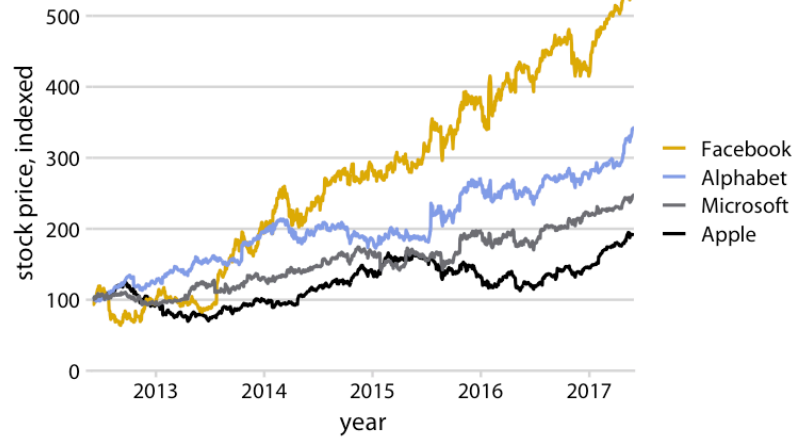
# Tech Stock Trends Over Time



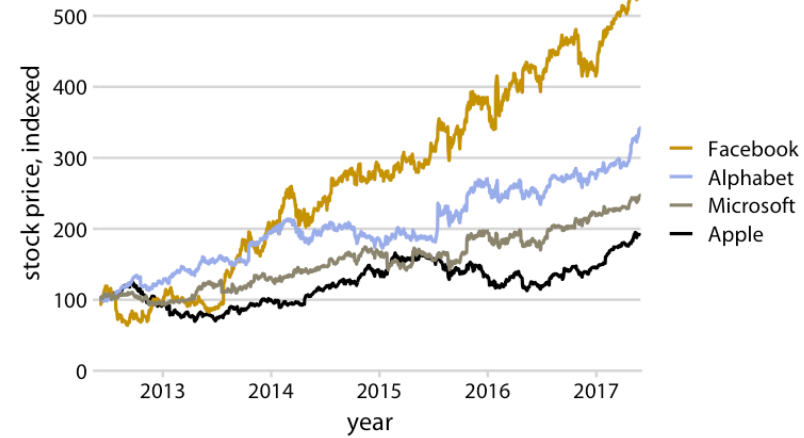
Stock price over time for four major tech companies. The stock price for each company has been normalized to equal 100 in June 2012.

# Color-vision-deficiency simulation

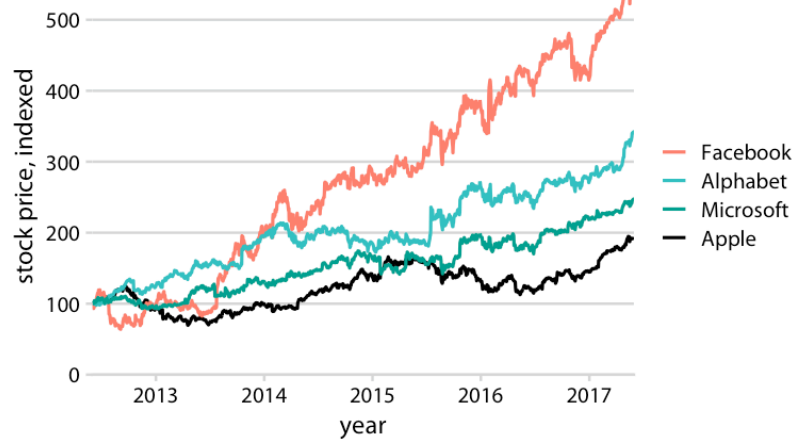
deuteranomaly



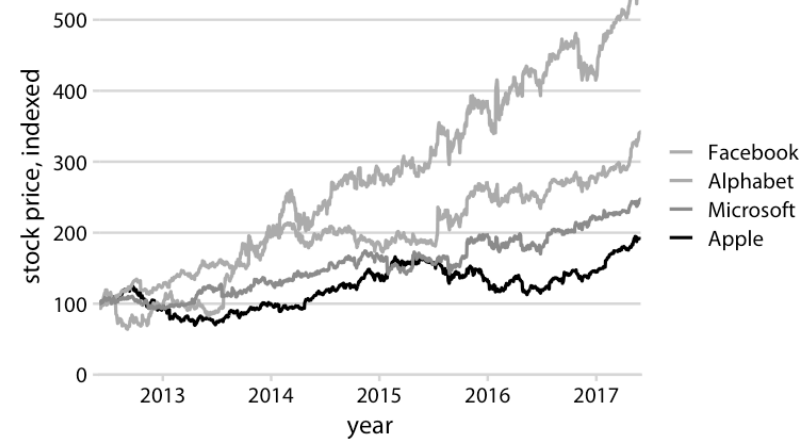
protanomaly



tritanomaly



desaturated



# Integrating Legends into the Data Axis

- **Problem:** Separate legends can make reading graphs more difficult.
- **Solution:**
  - Embed legends within the axis (e.g., color bar running along the x-axis).
  - Align legend information with the primary visualization element.

# Case Studies & Examples

- **Scatter Plots**

- Without redundant coding: Data points are differentiated only by color.
- With redundant coding: Different shapes + color used for distinction.

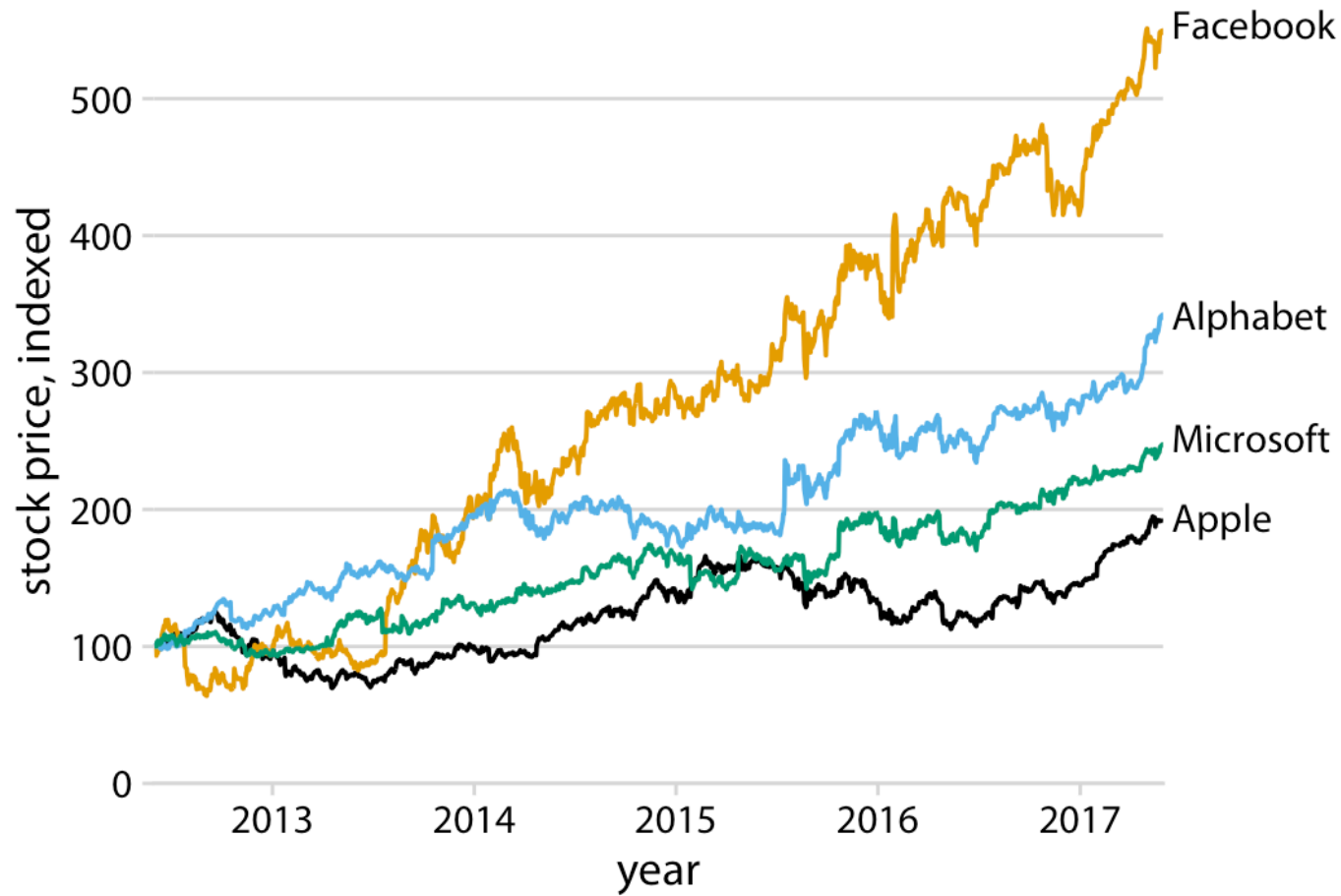
- **Stock Price Trends**

- Without redundant coding: Legend does not align with the perceived order of data.
- With redundant coding: Legend order matches the visual trend.

- **Direct Labeling in Density Plots**

- Instead of a separate legend, labels are placed directly within the plot.

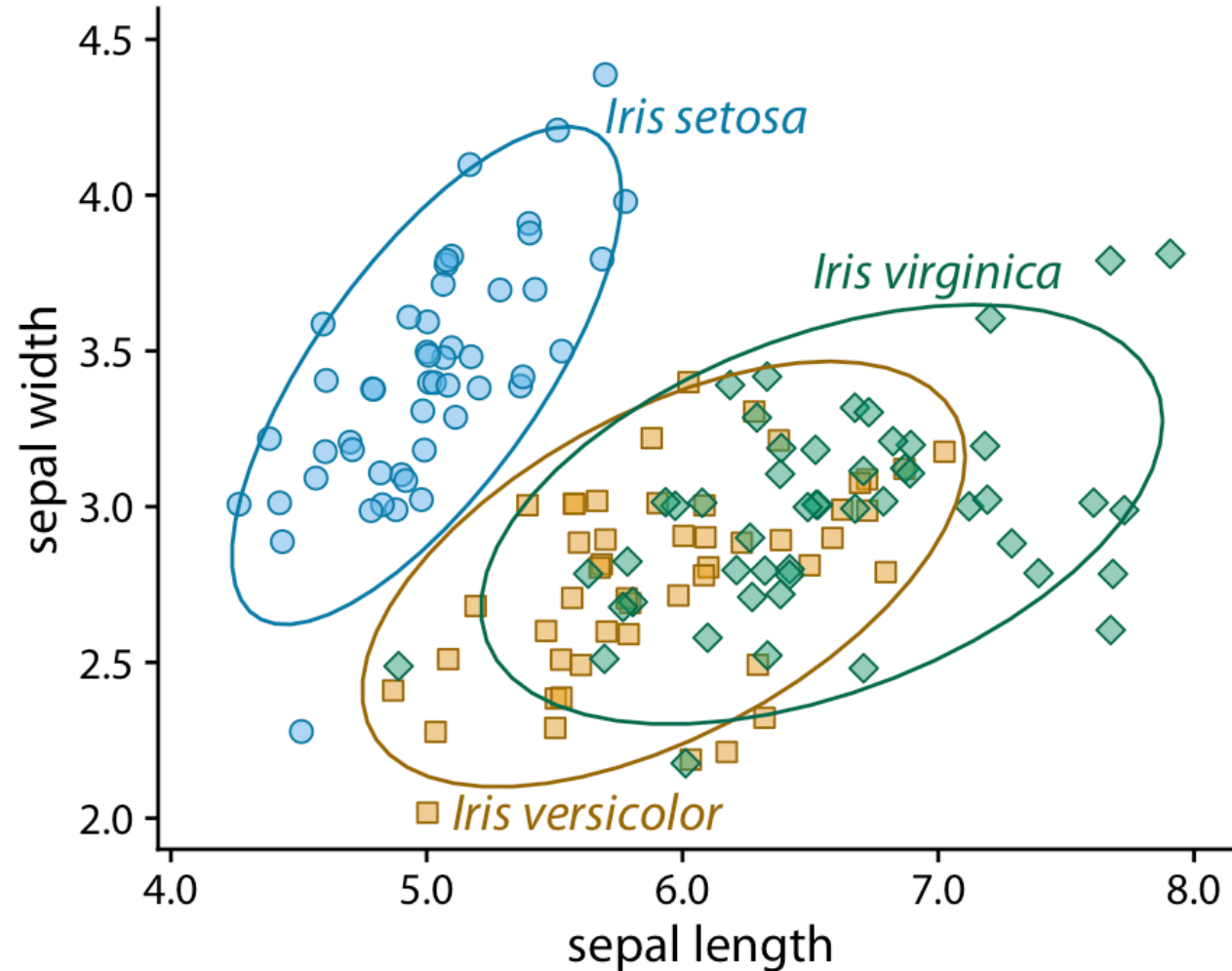
# Tech Stock Trends Without Legends



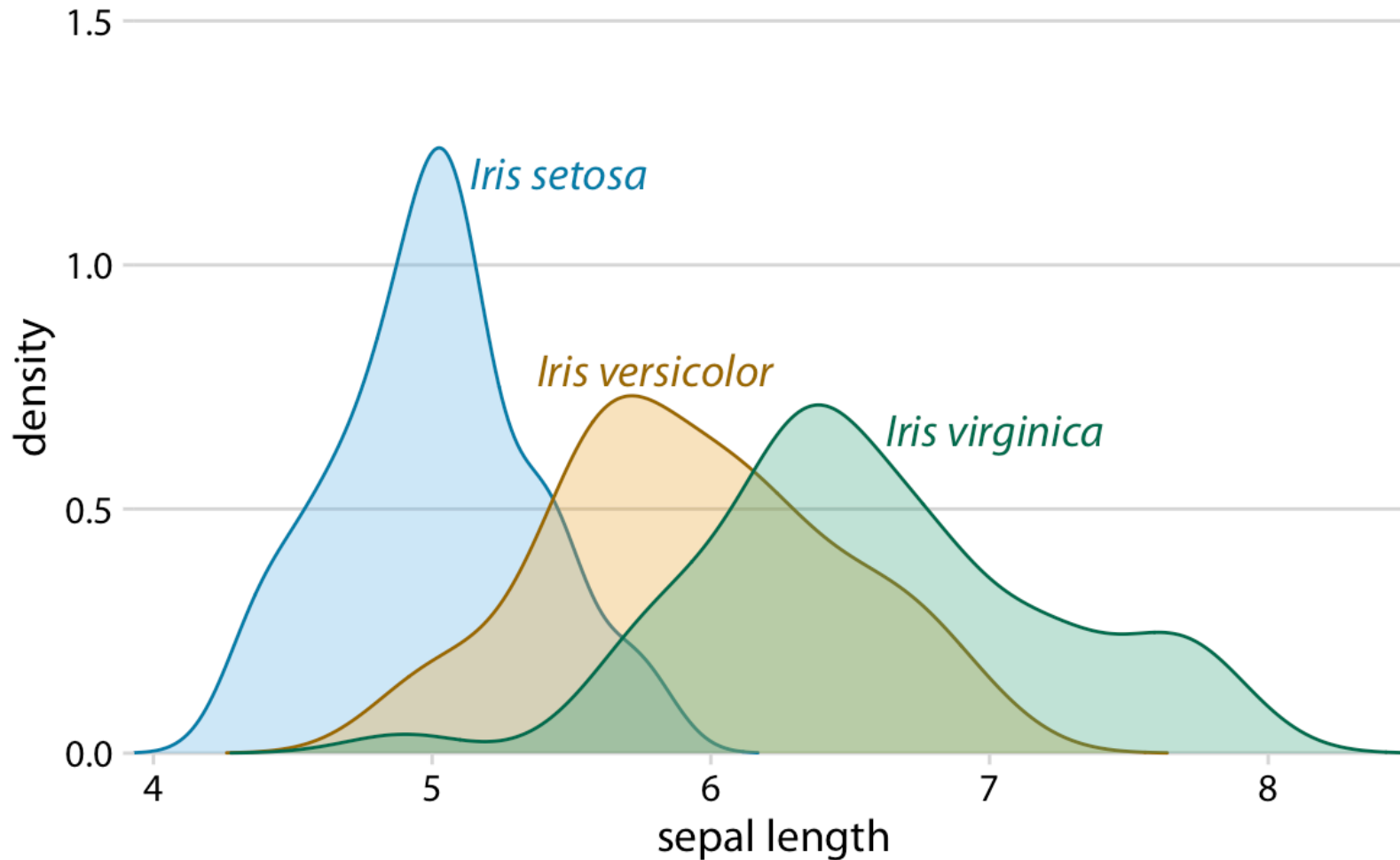
Stock price over time for four major tech companies. The stock price for each company has been normalized to equal 100 in June 2012.



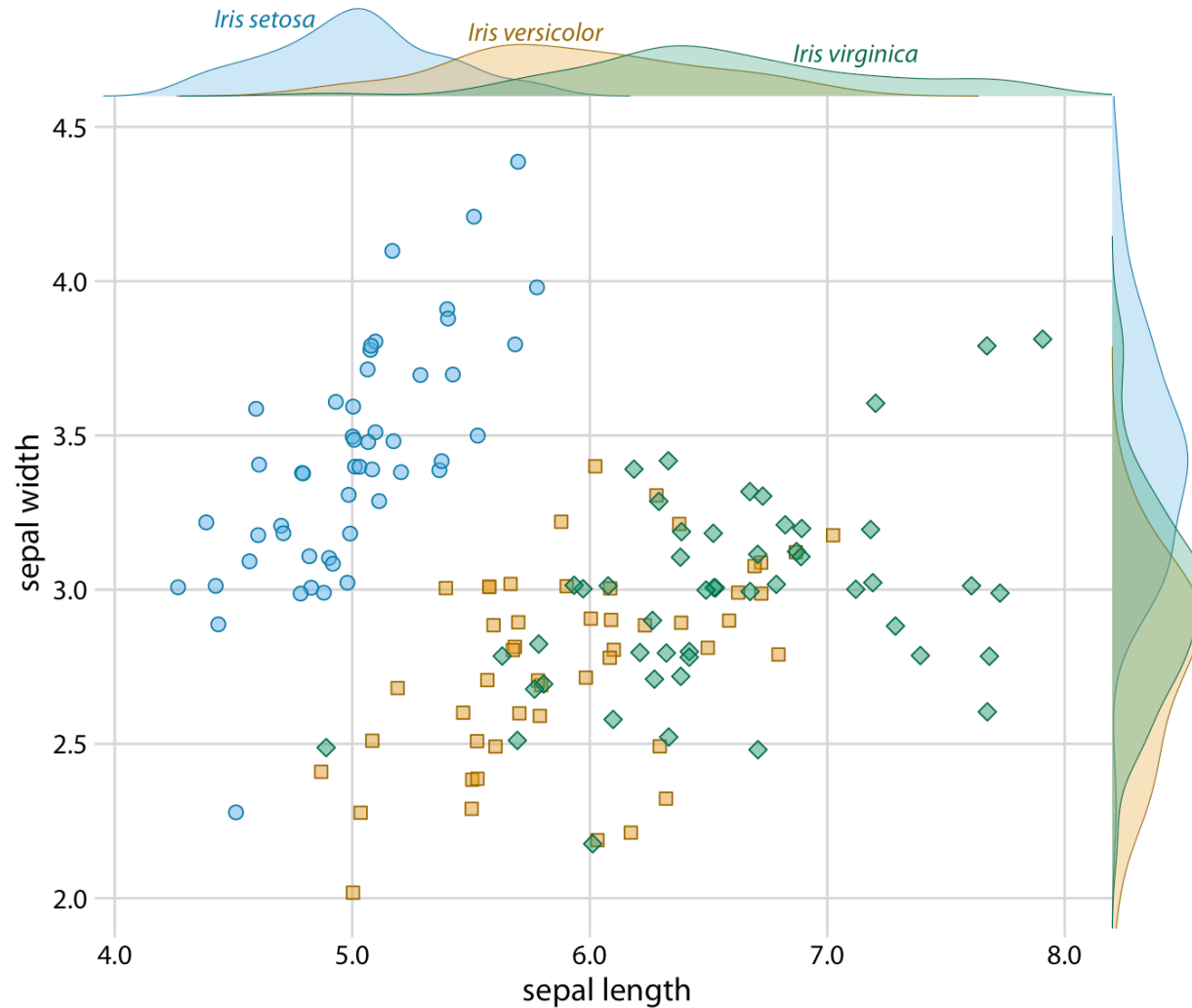
# Sepal Width vs. Length for Iris Species, Grid Removed



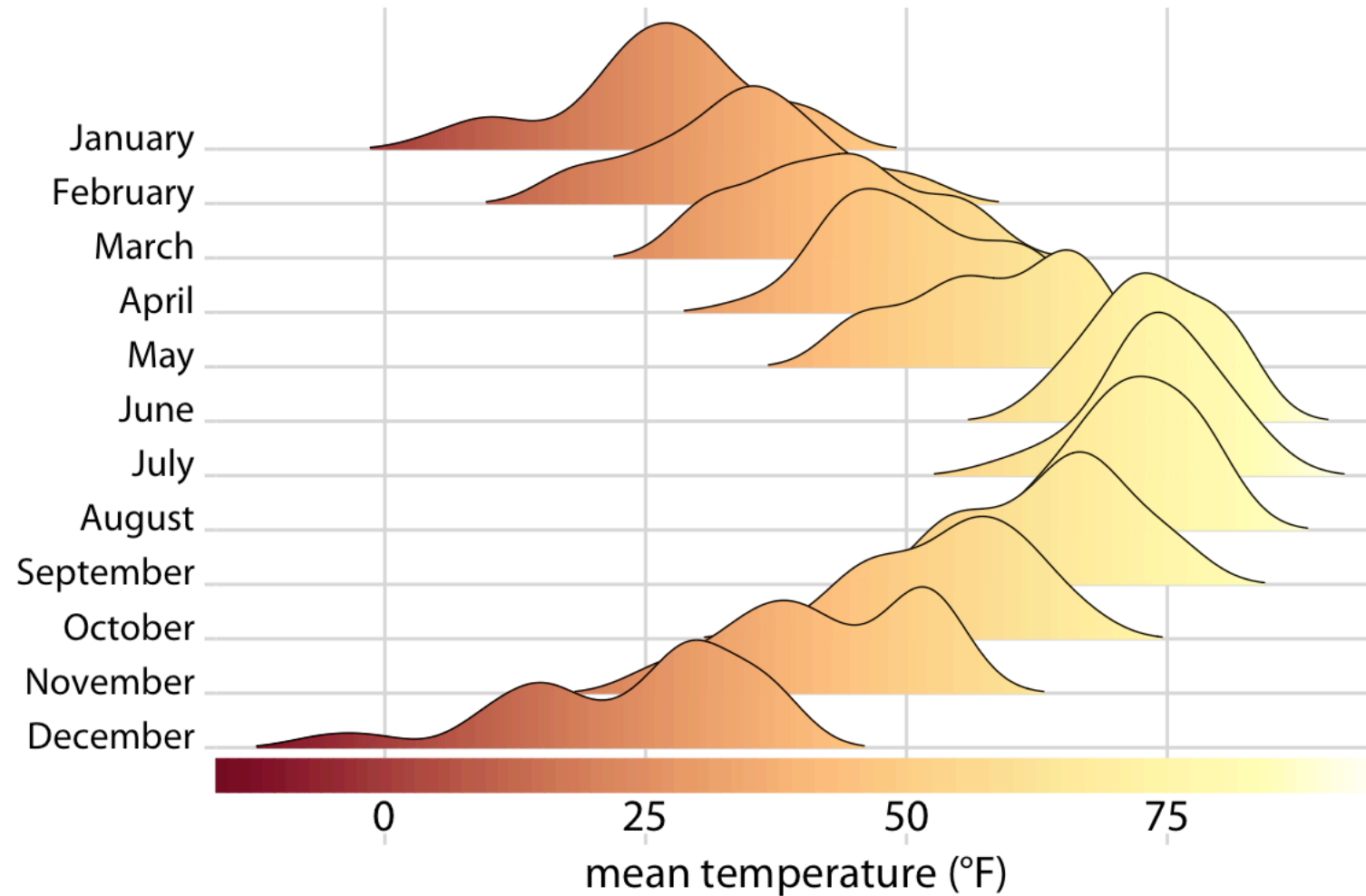
# Density Estimates of Sepal Lengths with Direct Labels



# Sepal Width vs. Sepal Length with Marginal Density Estimates



# Lincoln Temperatures with Color Scale



Temperatures in Lincoln, Nebraska (2016) shown by x-axis position and color, with a color bar for reference.

# Conclusion

- Redundant coding enhances clarity and accessibility in data visualization.
- Direct labeling, ordering legends logically, and using multiple aesthetics improve interpretation.
- Integrating legends into axes reduces cognitive load for viewers.
- **Key Takeaway:** Always consider how visual elements interact and ensure data remains distinguishable for all audiences, including those with CVD.