Unit V

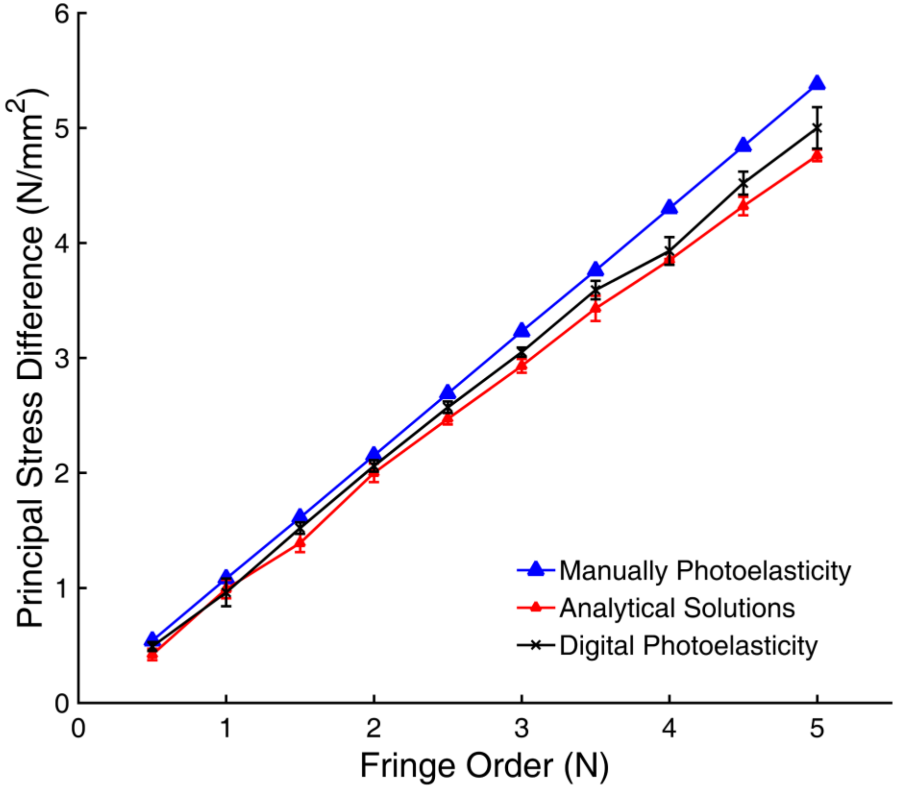
The Principle of Proportional Ink

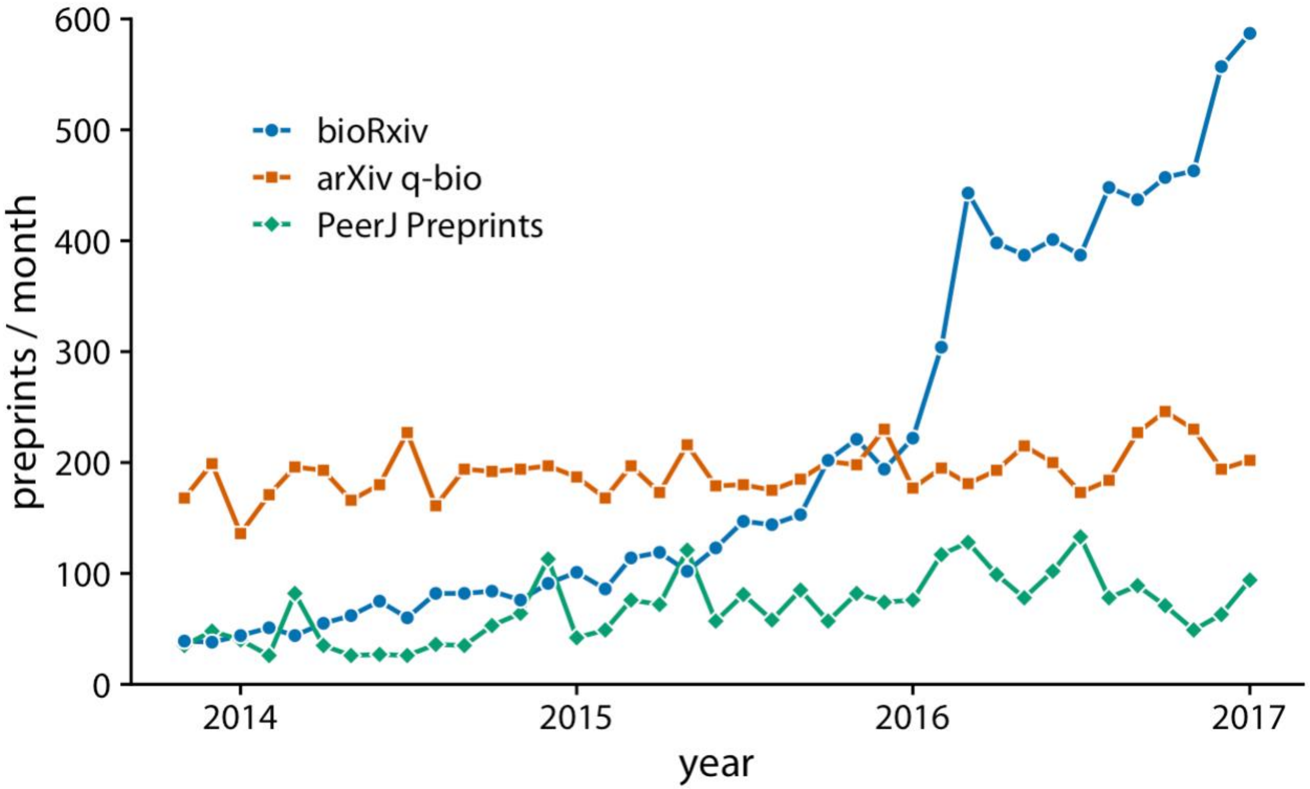
Principles of Figure Design

- What are the different does and don'ts while designing a figure/graph/images.
  - Clarity: The primary principle of figure design is to ensure clarity in conveying information.
  - Hierarchy and Emphasis: Use visual cues such as size, color, and font style to establish a hierarchy of information within the figure.
  - Consistency: Use consistent formatting, color schemes, and labeling.
  - Labelling and Annotations: Clearly label all axes, data points, and relevant features in the figure.
  - Balance, proportion and scaling: Maintain a sense of balance and proportion in the figure design.
  - Aesthetics: Visually appealing figures that engage the reader and present the information in an aesthetically pleasing manner.

The Principle of Proportional Ink

- When a shaded region is used to represent a numerical value, the area of that shaded region should be directly proportional to the corresponding value.
- Ink in this case refer to: lines, points, shared areas, and text.





Visualizations Along Linear Axes

- Visualizations along linear axes refer to the use of one or more linear axes to represent data in a visual form. This approach is commonly used in data visualization to display numerical or quantitative data along a continuous axis.
- Different charts that can be used:
  - Line Charts: Line charts are one of the most common types of visualizations along linear axes. They use a continuous horizontal or vertical axis to represent time, quantity, or any other numerical variable. Line charts are effective for showing trends, patterns, and changes in data over time or across different categories.
  - Area Charts: Area charts are similar to line charts but with the area below the line filled in with color or pattern. They are useful for comparing the cumulative values of different variables over time or for highlighting the distribution of data within a category.
  - Bar Charts: Bar charts use rectangular bars to represent data values along a linear axis. The length or height of each bar corresponds to the value of the data point it represents. Bar charts are ideal for comparing and contrasting categorical or discrete data.

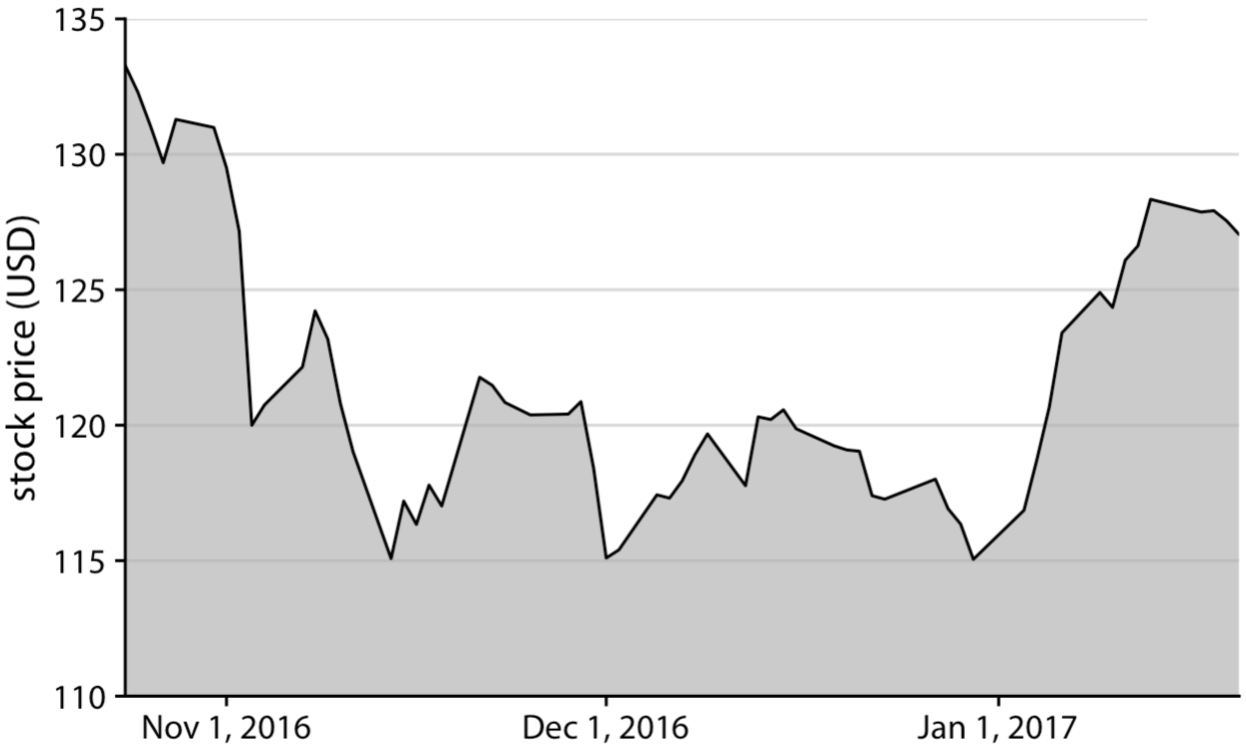
Visualizations Along Linear Axes

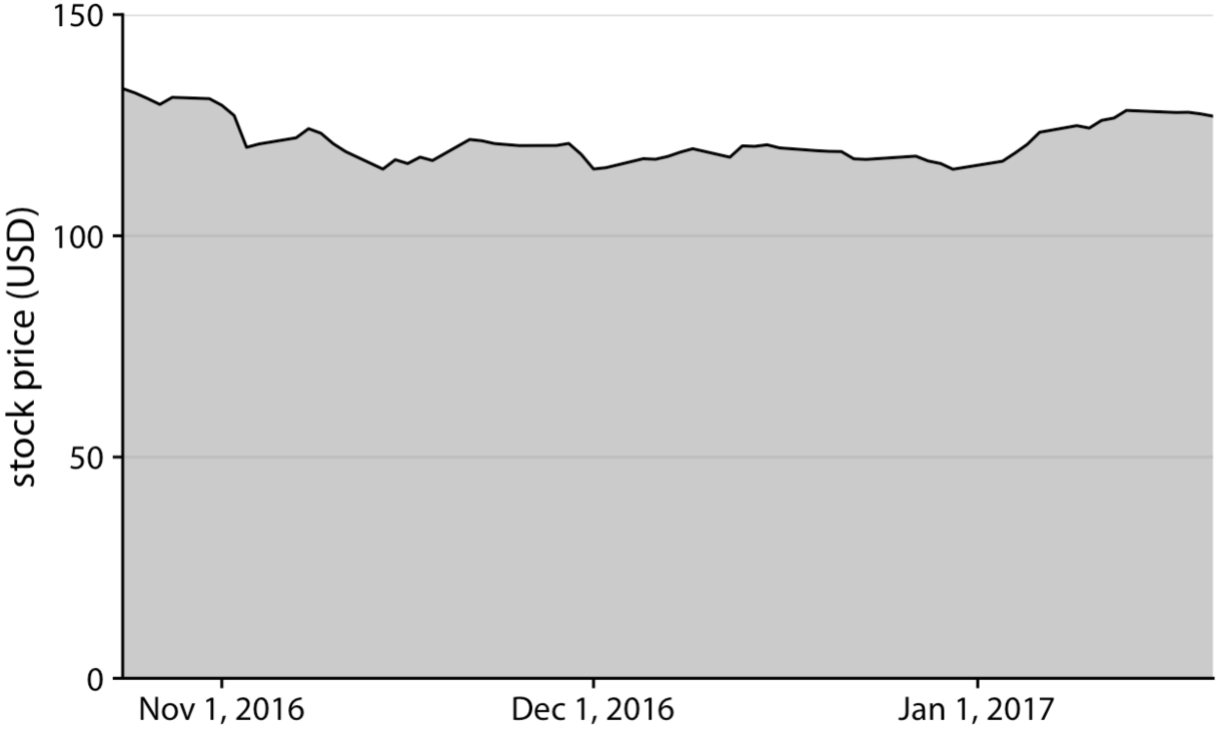
- Histograms: Histograms are used to visualize the distribution of continuous data along a linear axis. They divide the data range into bins or intervals and display the frequency or count of data points within each bin. Histograms provide insights into the shape, central tendency, and spread of the data.
- Scatter Plots: Scatter plots utilize linear axes to represent two continuous variables. Each data point is plotted as a point on the chart, with one variable mapped to the x-axis and the other variable mapped to the y-axis. Scatter plots are helpful for identifying relationships, correlations, or clusters in the data.
- Box Plots: Box plots, also known as box-and-whisker plots, provide a summary of the distribution of a continuous variable along a linear axis. They display key statistics such as the median, quartiles, and outliers. Box plots are useful for comparing distributions or identifying potential anomalies in the data.
- Heatmaps: Heatmaps use a grid-like layout with linear axes to represent data values as colors. They are effective for visualizing large matrices or tables of data, where each cell's color intensity corresponds to the magnitude or significance of the data point.

Visualizations Along Linear Axes

- Scaling/proportions/balance along axis

Stock price of Facebook (FB)





Visualizations Along Logarithmic Axes

- Visualizations along logarithmic axes involve using logarithmic scaling on one or more axes to represent data in a visual form. Logarithmic scaling is particularly useful when dealing with data that spans a wide range of values, allowing for a more balanced representation and better visualization of both small and large values.
    - Logarithmic Line Charts: Line charts with logarithmic scaling on one or both axes are commonly used when displaying data with exponential growth or decay. Logarithmic scaling compresses the data at higher values, making it easier to observe and compare trends across different magnitudes of data points.
    - Logarithmic Bar Charts: Similar to line charts, bar charts can also utilize logarithmic scaling on one or both axes. This is useful when dealing with data that exhibits a large disparity in values, enabling a clearer comparison between different categories or data points.

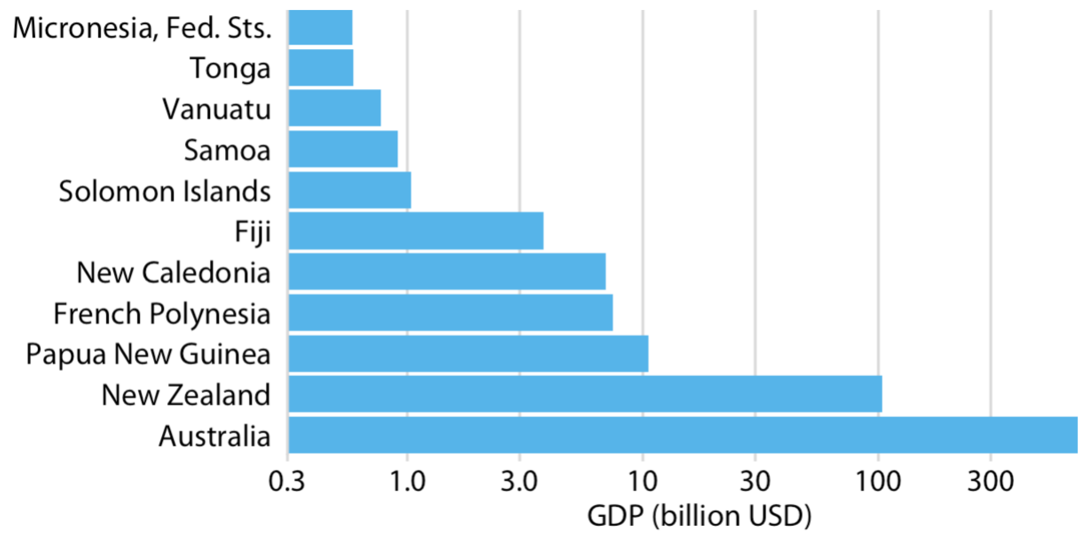
Visualizations Along Logarithmic Axes

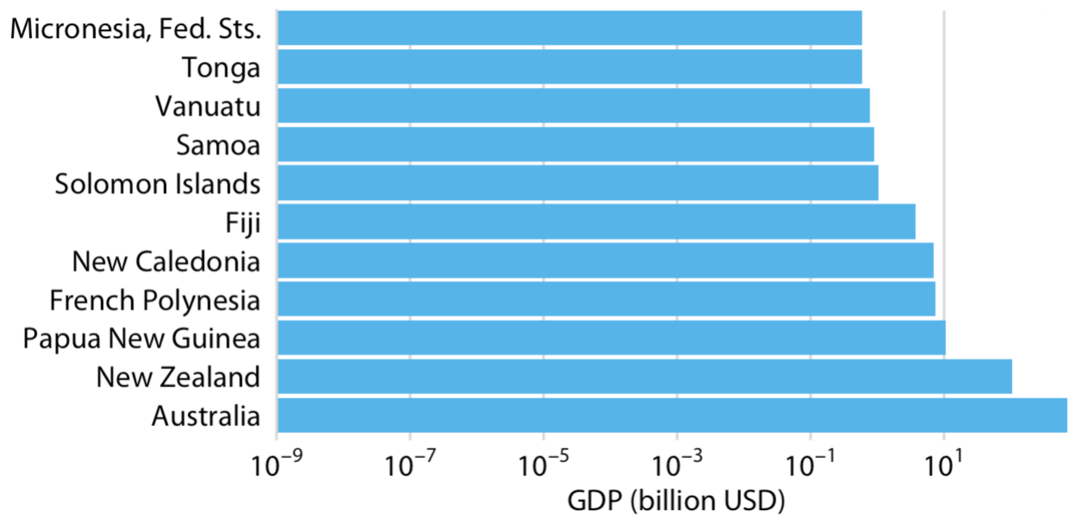
- Logarithmic Scatter Plots: Scatter plots can benefit from logarithmic scaling when visualizing data that spans several orders of magnitude. By applying logarithmic scaling to the axes, the plot can reveal relationships or patterns that may not be apparent with linear scaling, particularly when dealing with data that includes outliers or extreme values.

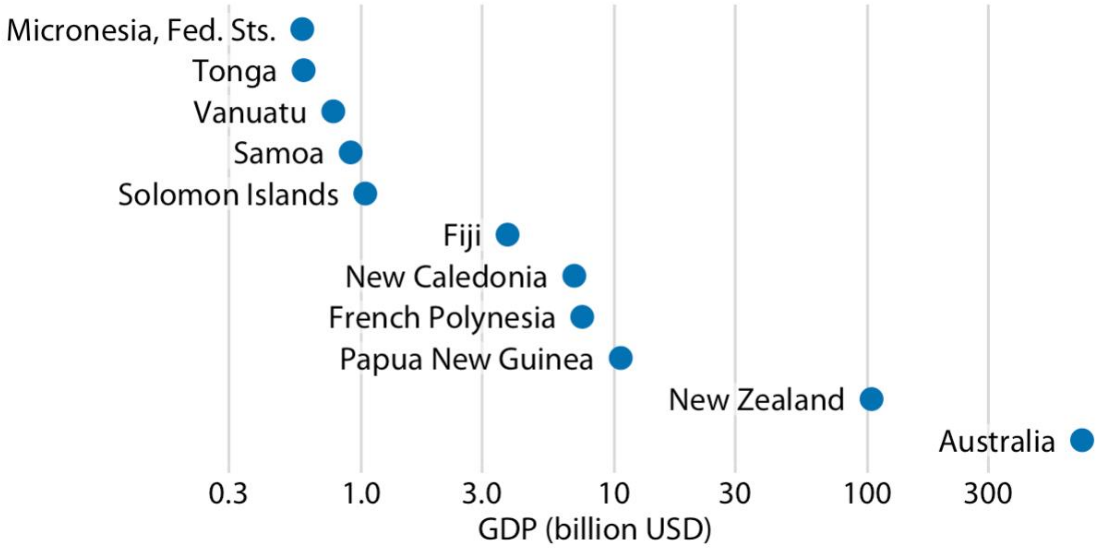
- Logarithmic Histograms: Histograms can be transformed into logarithmic histograms by applying logarithmic scaling to the frequency or count axis. This helps visualize the distribution of data with a wide range of values, emphasizing the distribution in lower frequency regions while maintaining visibility for higher frequency regions.

- Logarithmic Heatmaps: Heatmaps can also utilize logarithmic scaling when representing data values. This is particularly useful when visualizing data with highly skewed distributions or when trying to emphasize smaller values that might otherwise be overshadowed by larger values.

Visualizations Along Logarithmic Axes

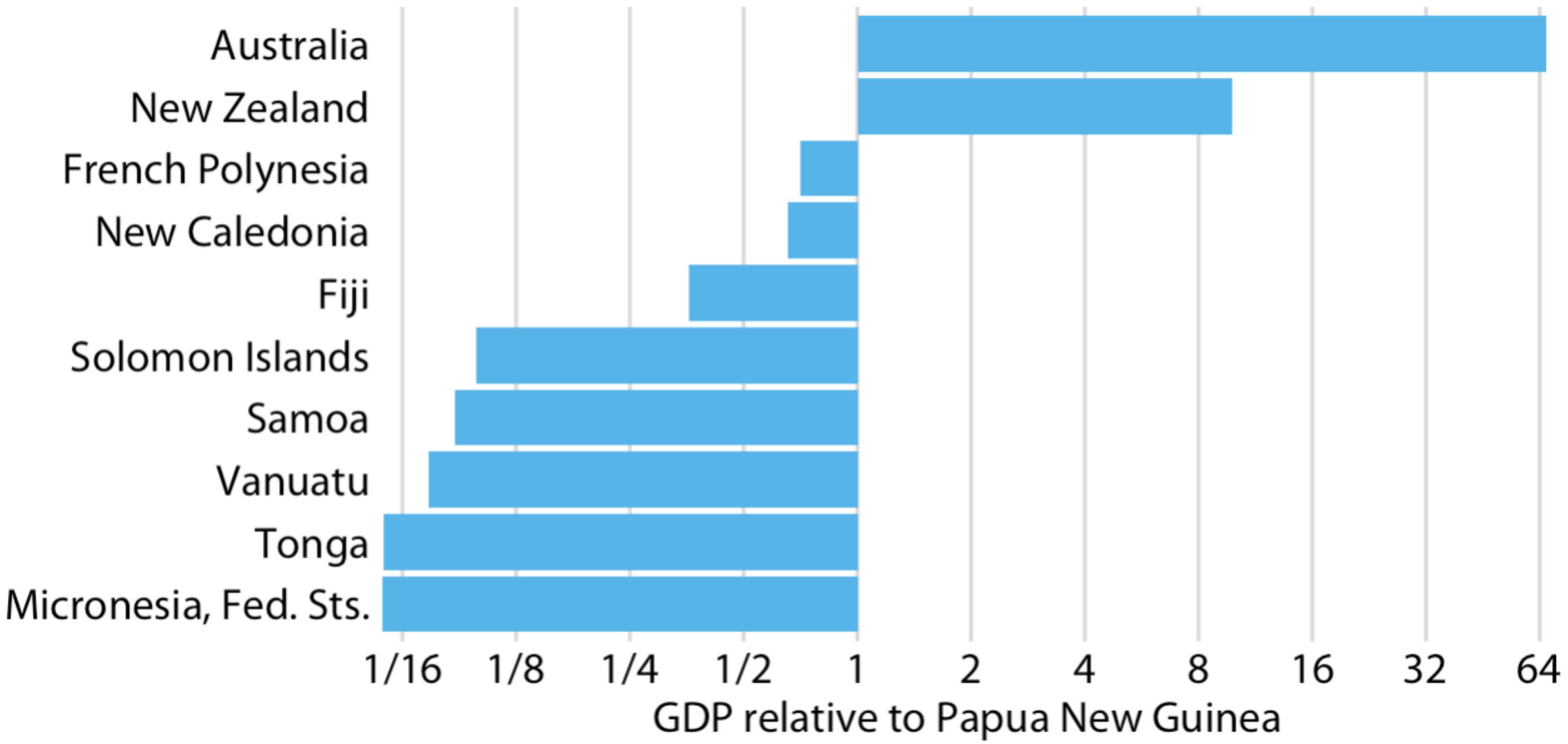






Data representation as proportions (wrt a data)

- GDP in 2007 of countries in Oceania, relative to the GDP of Papua New Guinea.



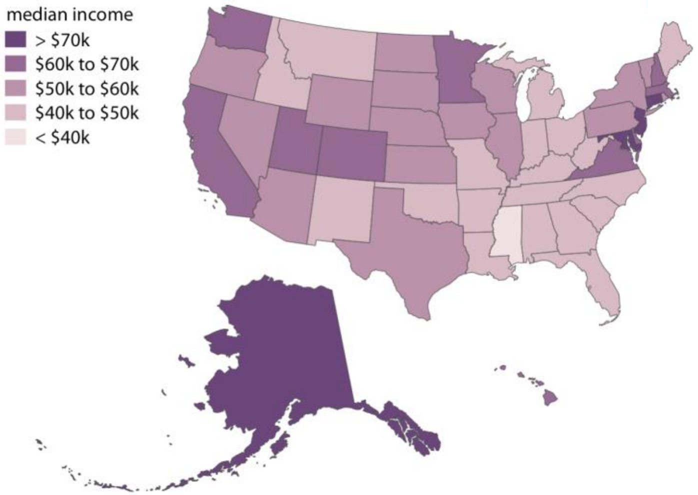
Direct Area Visualizations

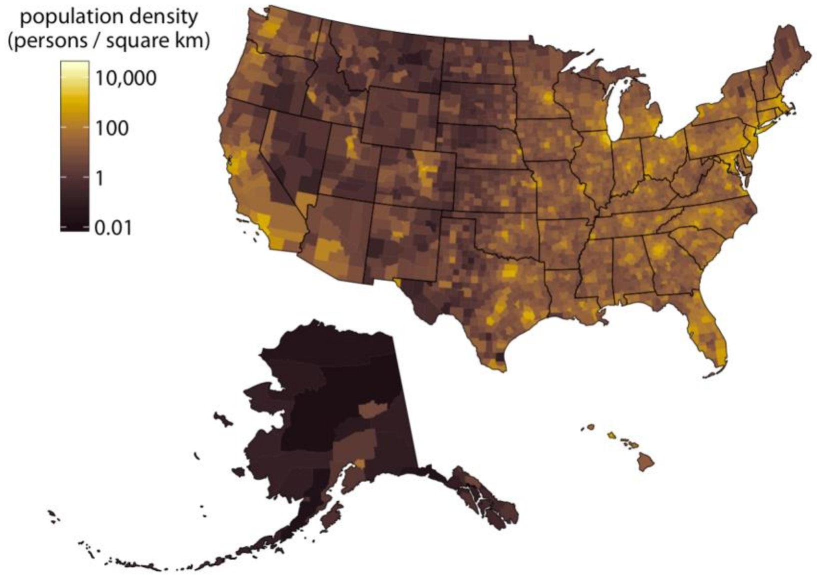
- Direct area visualizations are a type of data visualization technique that represent data using the area of geometric shapes. By mapping data values to the size of the corresponding shapes, direct area visualizations provide an intuitive way to compare and understand data patterns based on the visual perception of area.
  - Bubble Charts: Bubble charts are a common form of direct area visualization where data points are represented as circles or bubbles. The size of each bubble corresponds to a specific data value, such as magnitude or frequency, while the x and y coordinates can represent other variables. Bubble charts are effective for visualizing three-dimensional data, as they combine the use of position and size to convey information.
  - Pie Charts: Although pie charts primarily rely on angle or arc length to represent data, the area of each pie slice can also be used to convey information. The size of the pie slice can indicate the proportion or percentage of a specific category or variable. While pie charts are widely used, it's important to note that they are often criticized for potential misinterpretation and difficulty in accurately comparing slice sizes.

Direct Area Visualizations

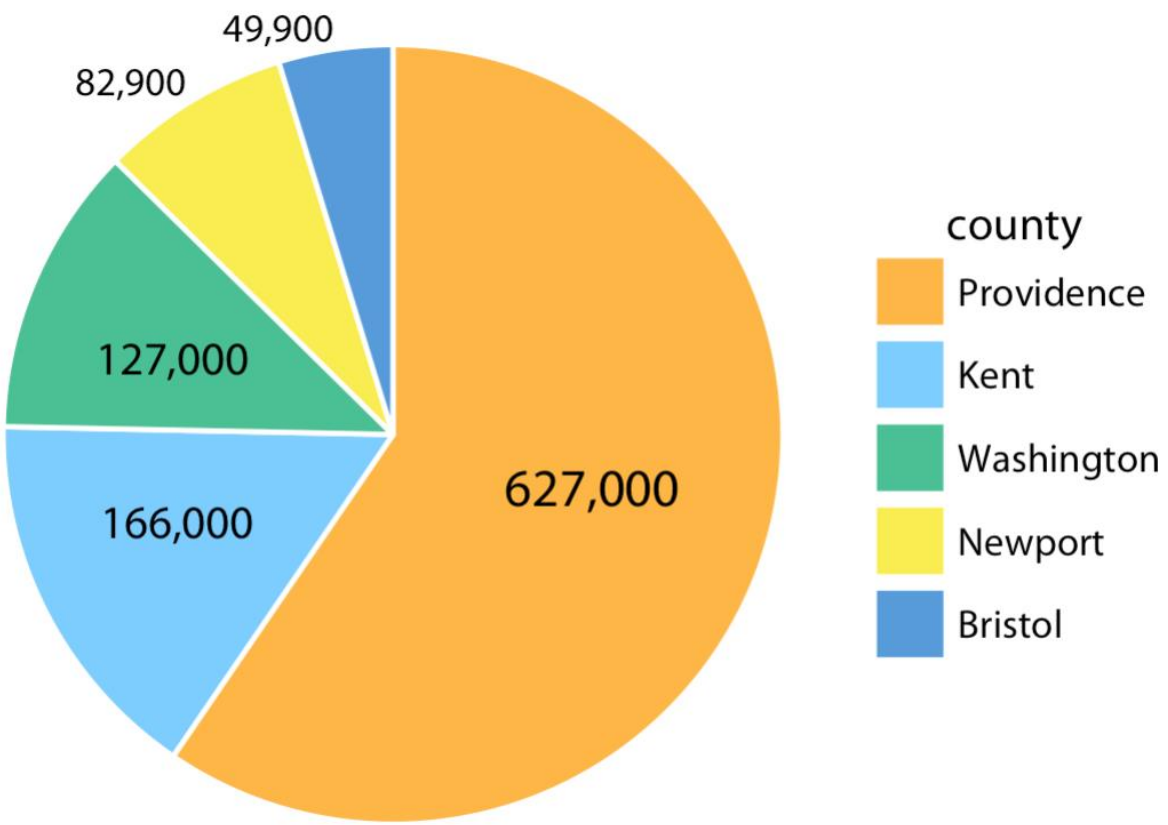
- Heat Maps: Heat maps are a type of direct area visualization that use color gradients to represent data values within a two-dimensional grid or matrix. The intensity of the color corresponds to the magnitude or density of the data, allowing for easy identification of patterns and variations across the grid. Heat maps are commonly used in various fields, including data analysis, geographic information systems (GIS), and biological research.
- Choropleth Maps: Choropleth maps use different shades or colors to represent data values for specific geographic regions, such as countries, states, or zip codes. The color intensity or pattern indicates the magnitude or density of the data within each region. Choropleth maps are widely used for visualizing spatial data, such as population density, election results, or economic indicators.

Choropleth Maps:





Direct Area Visualizations



Handling Overlapping Points

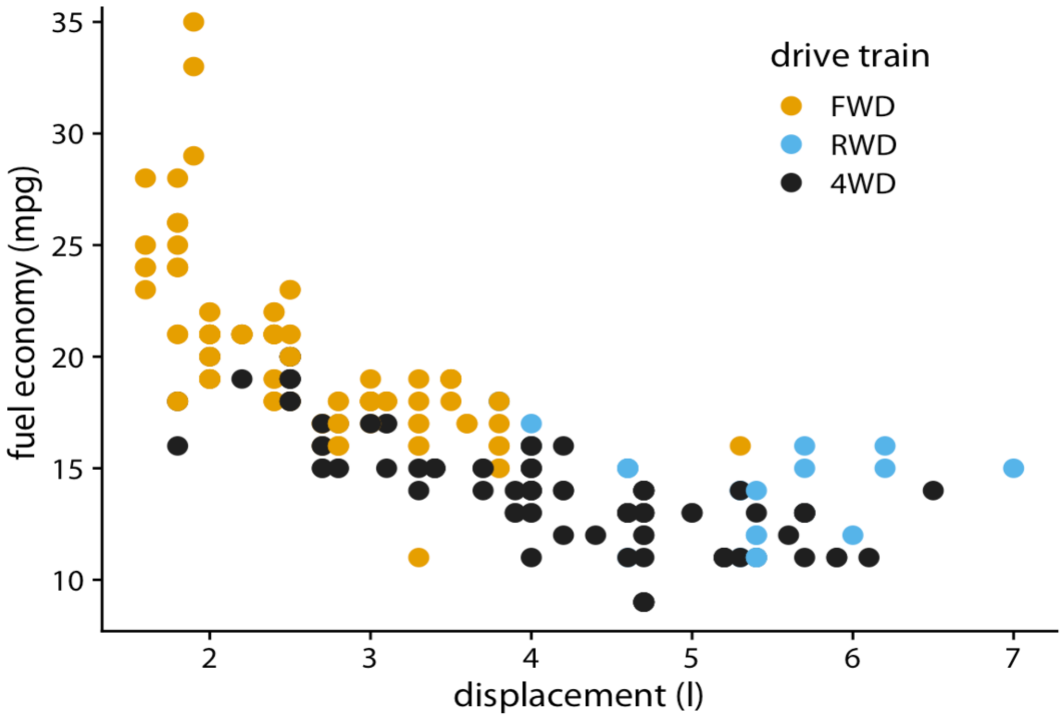
Handling Overlapping Points

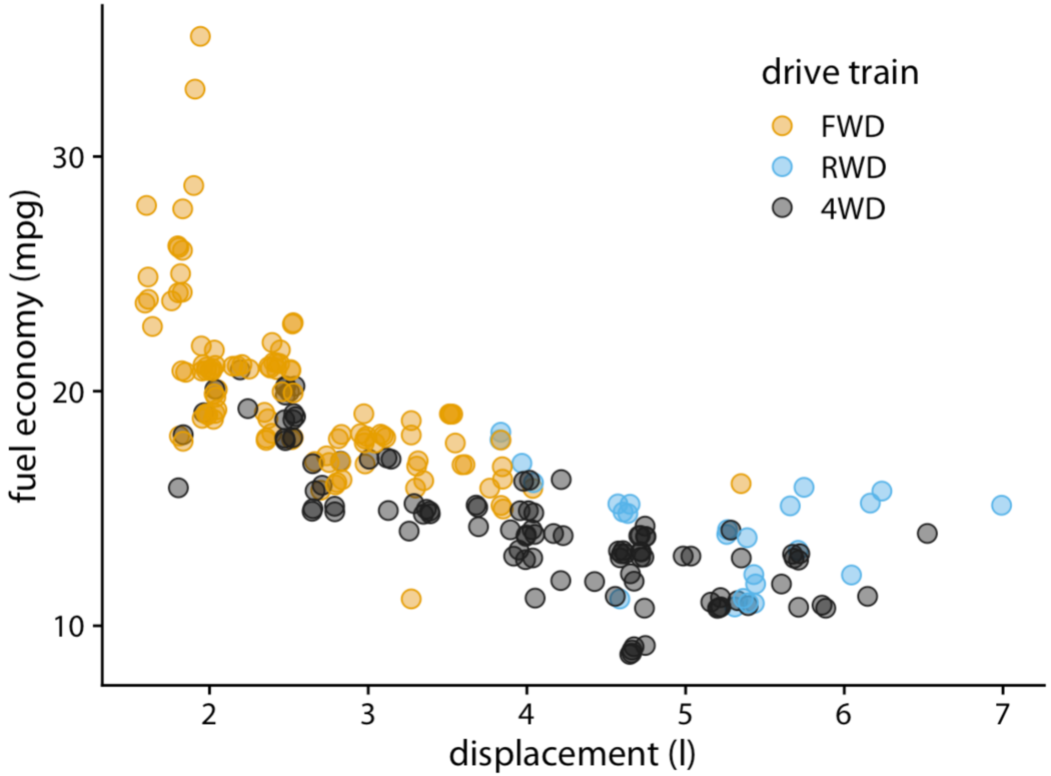
- In data visualization, overlapping points can occur when multiple data points have the same or very similar coordinates on a plot. This can pose challenges in effectively visualizing and interpreting the data. Here are some considerations and techniques for handling overlapping points.
  - Transparency and Alpha Blending: One way to address overlapping points is by adjusting their transparency or opacity. By reducing the opacity of the points, overlapping areas become more visible, allowing viewers to discern patterns and density. This technique is particularly useful when dealing with large datasets where individual points may overlap.
  - Point Jittering: Point jittering involves adding a small amount of random noise to the position of each point. By introducing slight variations in the placement of overlapping points, they become spread out, reducing the amount of overlap. This technique is commonly used when visualizing categorical data or discrete values.
  - Density-Based Techniques: Density-based techniques aim to represent the density of points in a particular area rather than individual data points. Heat maps, contour plots, or kernel density estimation can be used to visualize the density of overlapping points, providing a smoother representation of the data.

Handling Overlapping Points

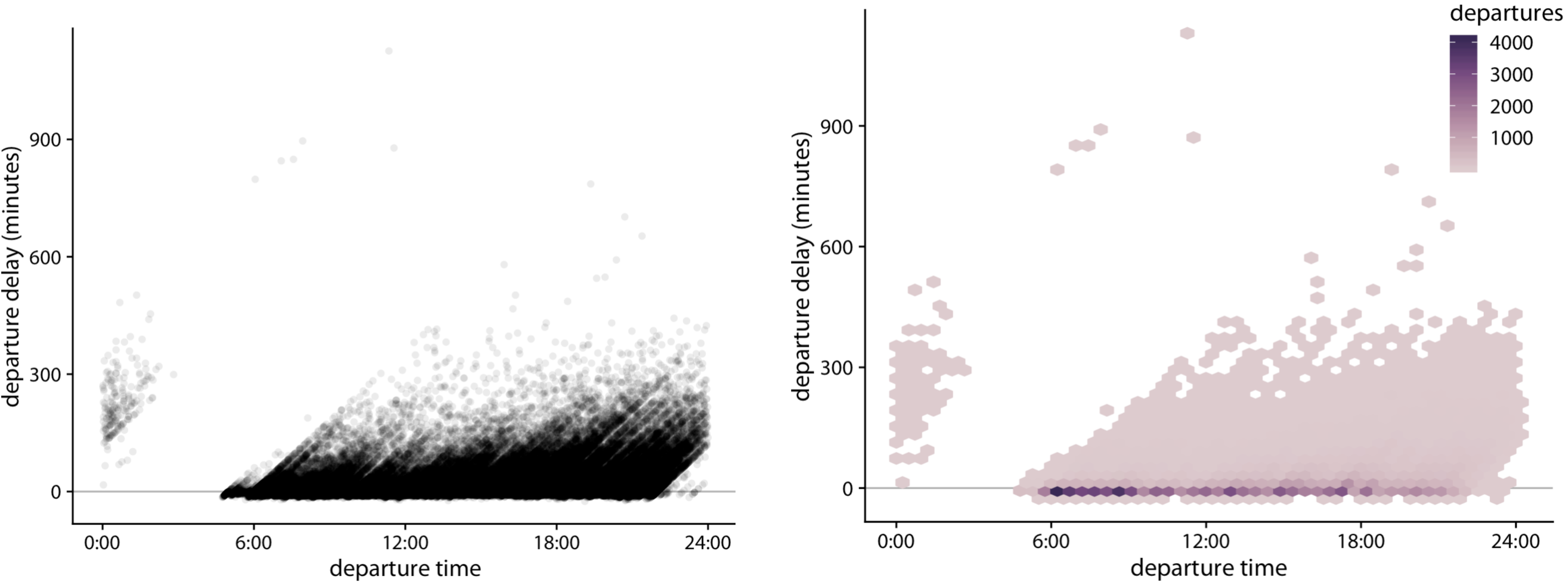
- Size and Shape Variation: Varying the size or shape of points based on an additional variable can help distinguish overlapping points. By mapping a relevant variable to point size or shape, overlapping points can be differentiated, providing additional insights into the data.
- Use of Hierarchical or Clustered Visualizations: In cases where overlapping points are particularly dense, hierarchical or clustered visualizations can be employed. Techniques such as dendrograms, treemaps, or clustering algorithms can group similar points together, reducing visual clutter and enabling a clearer representation of the data.

Partial Transparency and Jittering





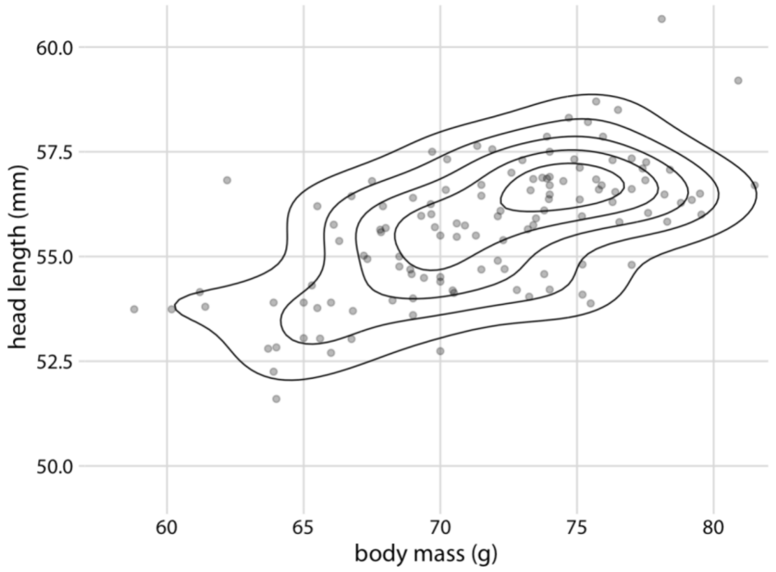
2D Histograms

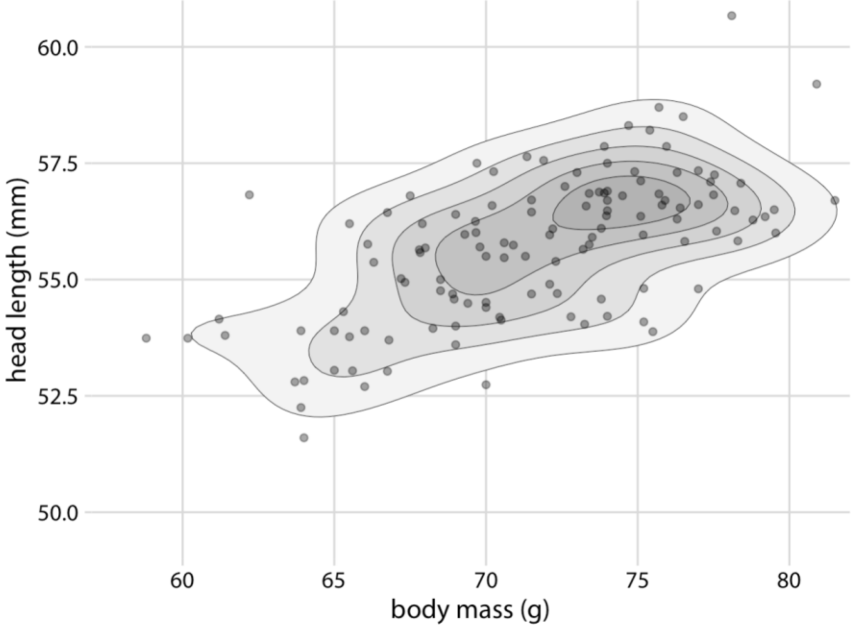


Departure delay in minutes versus the flight departure time. Each colored hexagon represents all flights departing at that time with that departure delay. Coloring represents the number of flights represented by that hexagon.

Contour Lines

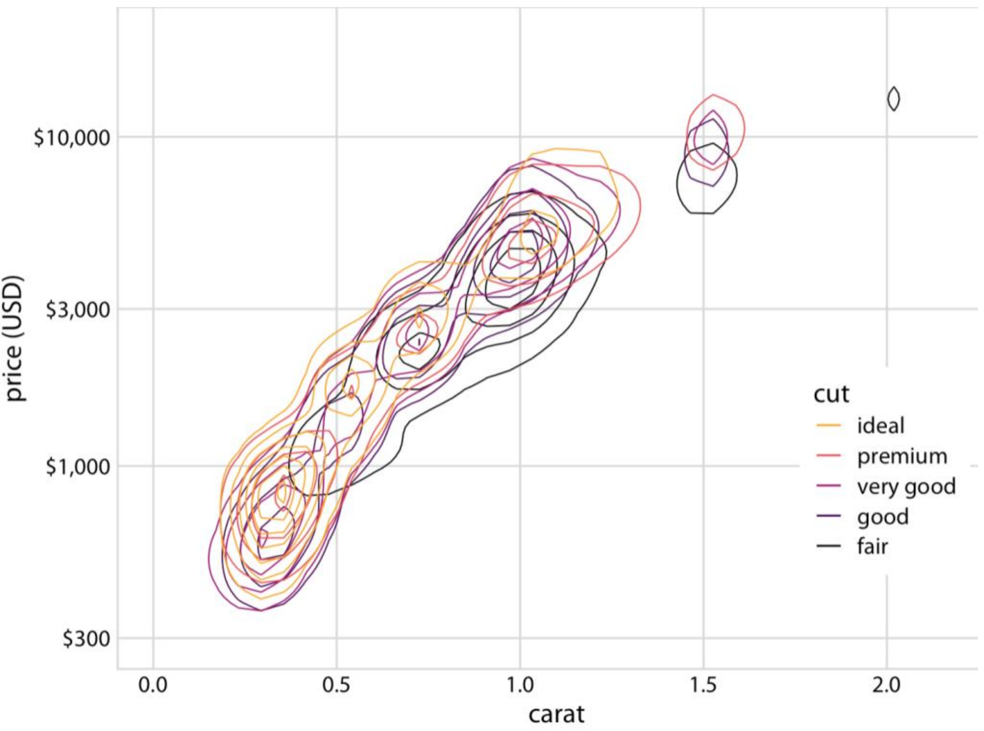
- Instead of binning data points into rectangles or hexagons, we can also estimate the point density across the plot area and indicate regions of different point densities with contour lines. This technique works well when the point density changes slowly across both the x and the y dimensions.



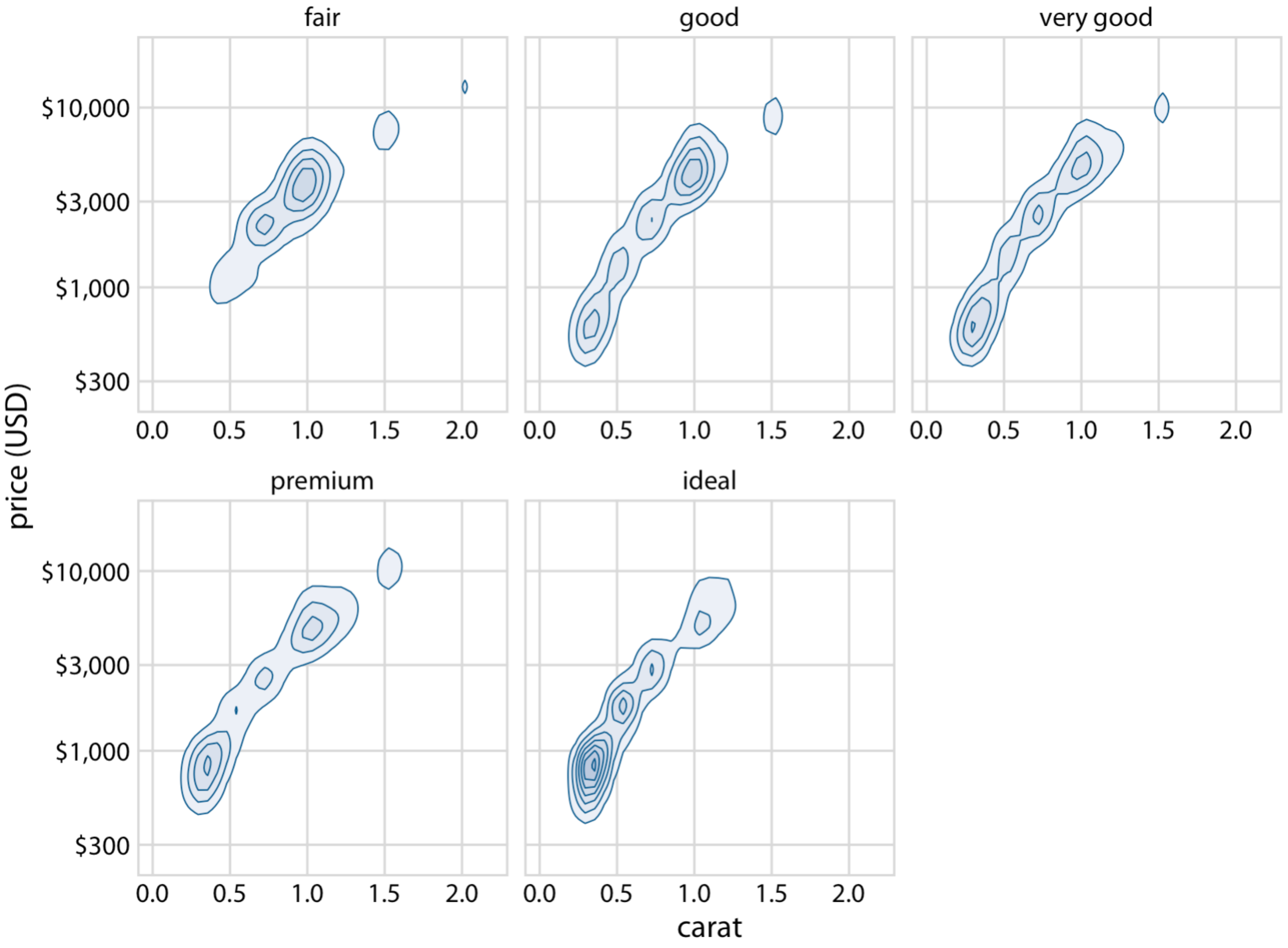


Areas enclosed by the contour lines are shaded with increasingly darker shades of gray. This shading creates a stronger visual impression of increasing point density toward the center of the point cloud.

Contour Lines



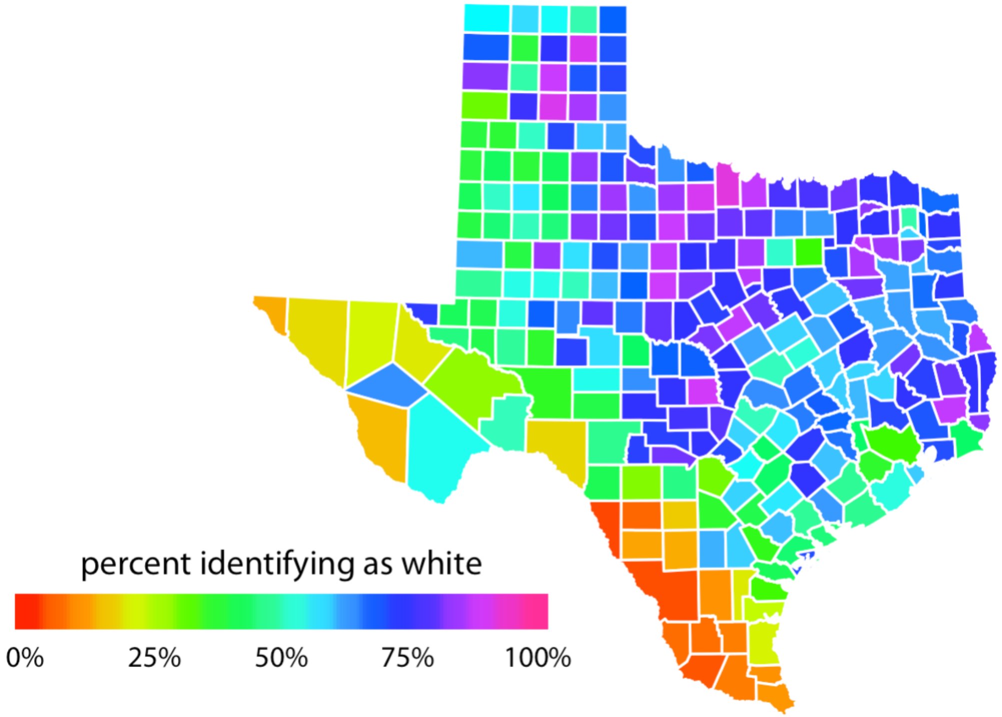
Price of diamonds versus their carat value.



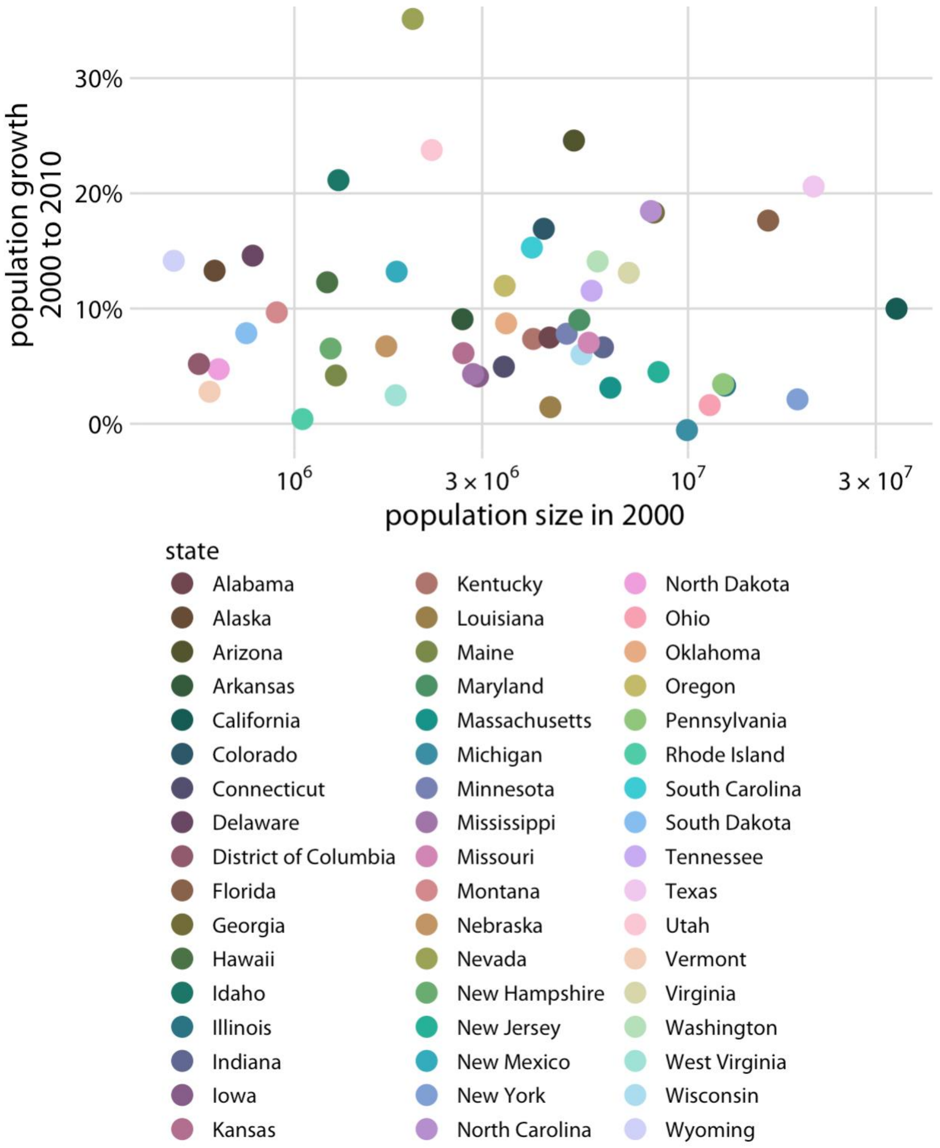
Common Pitfalls of Color Use

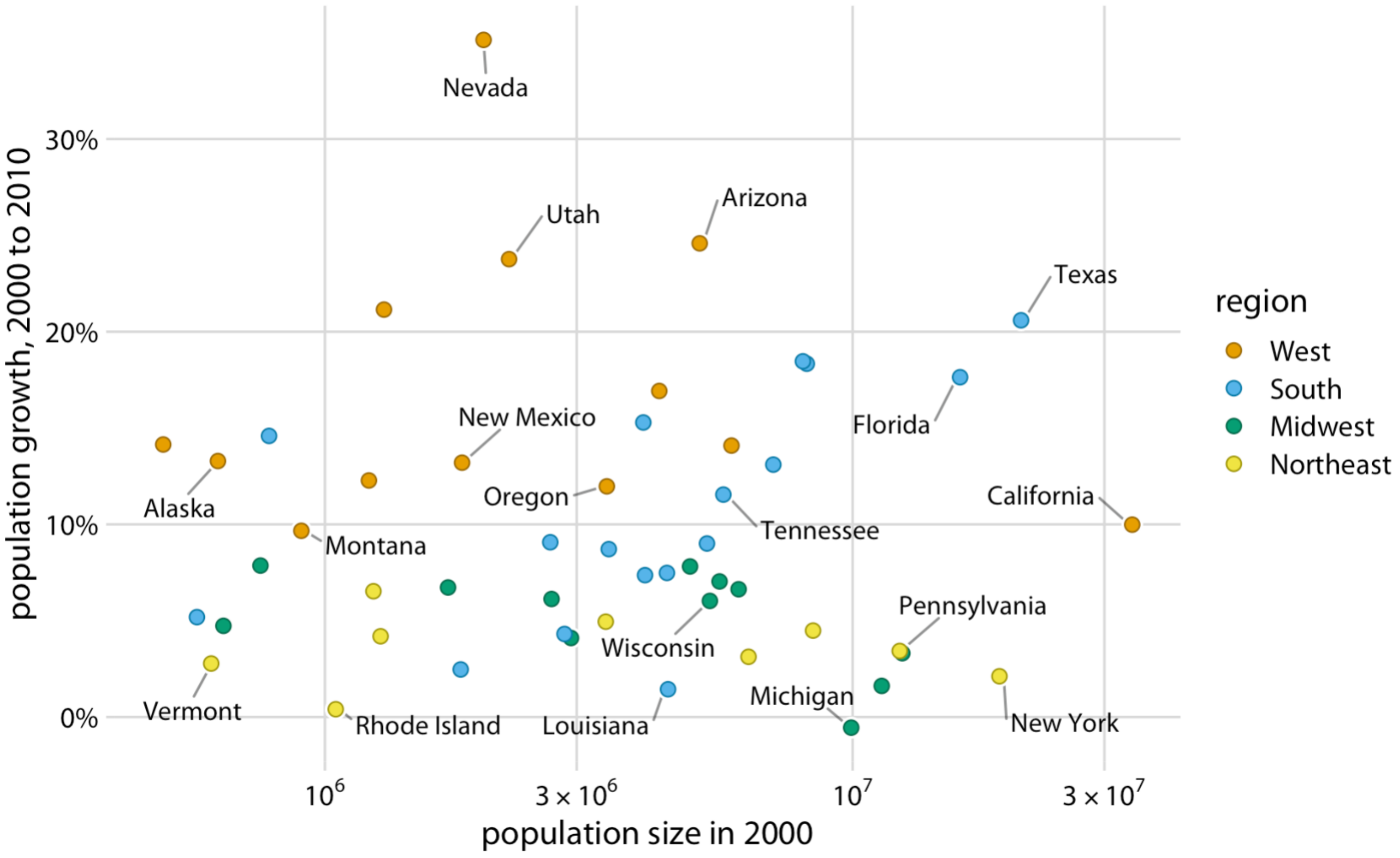
Common Pitfalls of Color Use

- Few bad choices for colouring a figure
  - Bad colour combinations: many a times closely related colours cannot be distinguished.
  - Bad choice such as brick red or florescent green can distract audience.
  - Sometimes, most commonly occurring colour blindness can mislead audience.

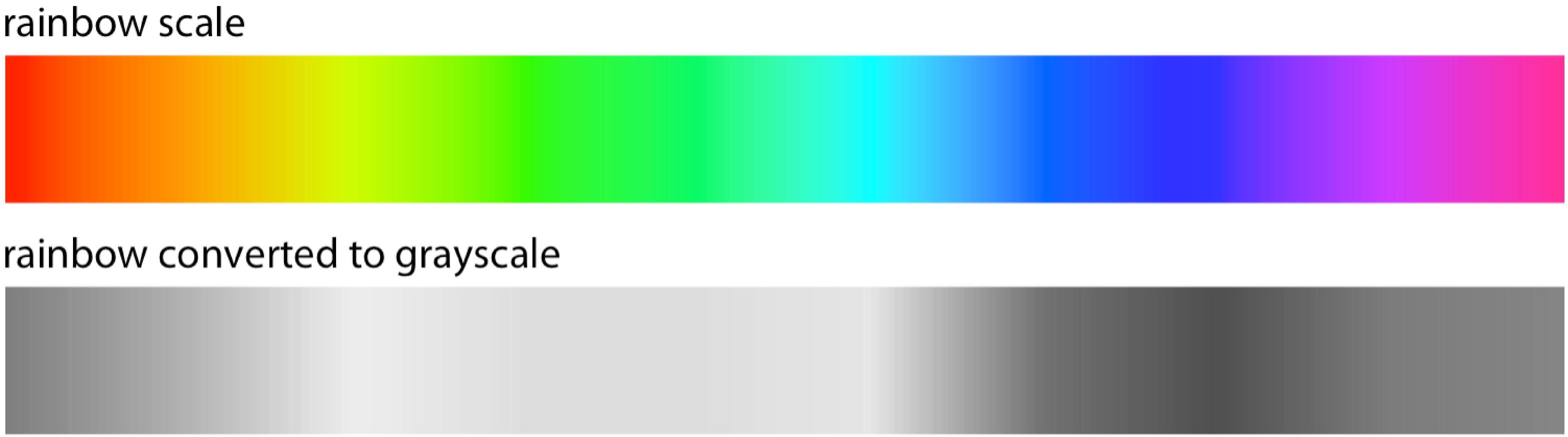


Encoding Too Much or Irrelevant Information



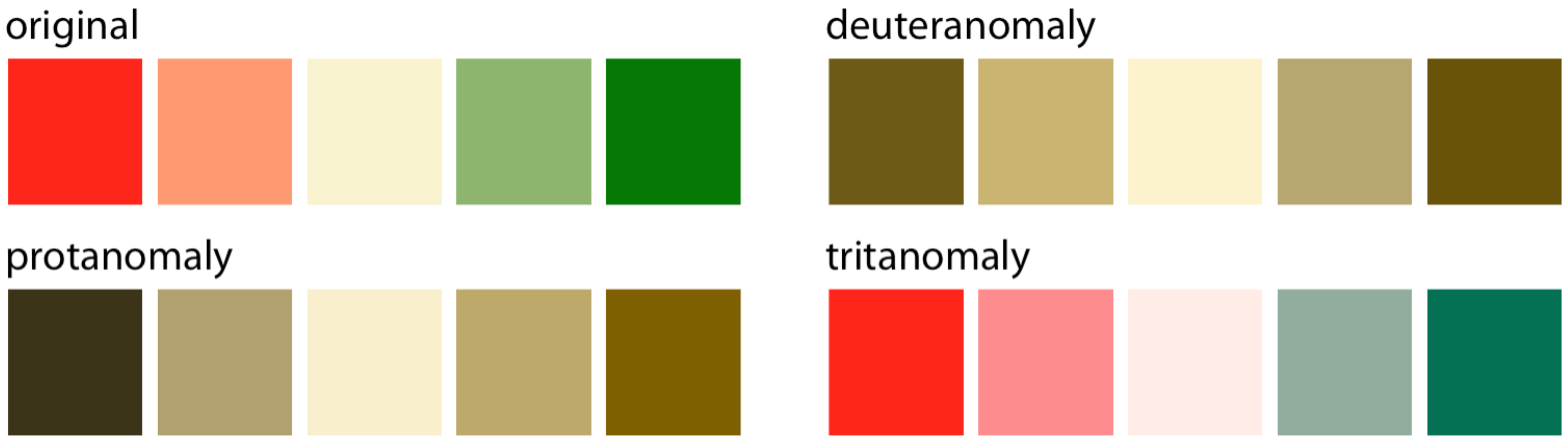


Using Nonmonotonic Color Scales to Encode Data Values



Not Designing for Color-Vision Deficiency

- These readers may not be able to distinguish colors that look clearly different to most other people.
- Red–green variant – Problem with red and green colour.
- Blue–yellow variant – Problem with blue and yellow colour.



Colourblind-friendly colour scale

Name	Hex code	Hue	C, M, Y, K (%)	R, G, B (0–255)	R, G, B (%)
Orange	#E69F00	41°	0, 50, 100, 0	230, 159, 0	90, 60, 0
Sky blue	#56B4E9	202°	80, 0, 0, 0	86, 180, 233	35, 70, 90
Bluish green	#009E73	164°	97, 0, 75, 0	0, 158, 115	0, 60, 50
Yellow	#F0E442	56°	10, 5, 90, 0	240, 228, 66	95, 90, 25
Blue	#0072B2	202°	100, 50, 0, 0	0, 114, 178	0, 45, 70
Vermilion	#D55E00	27°	0, 80, 100, 0	213, 94, 0	80, 40, 0
Reddish purple	#CC79A7	326°	10, 70, 0, 0	204, 121, 167	80, 60, 70
Black	#000000	N/A	0, 0, 0, 100	0, 0, 0	0, 0, 0