Introduction

Data visualization refers to the visual representation of data using charts, graphs, maps, and other visual elements. The purpose of data visualization is to help people understand and make sense of complex data by presenting it in a clear, concise, and visually appealing way.

Properties of data visualization

- Simplify complex data: Data visualization can simplify complex data, making it easier to understand patterns and relationships between variables.
- Provide insights: By highlighting important trends and patterns, data visualization can provide insights that may not be apparent in raw data.
- Support decision-making: Data visualization can support decision-making by making it easier to identify trends, patterns, and outliers that may require action.
- Engage audiences: By presenting data in an engaging and visually appealing way, data visualization can capture the attention of audiences and help them connect with the data on an emotional level.

Good, Ugly, Bad, and Wrong Figures

Good figure: A figure or graph which is aesthetically right and clearly display the information to the audience

Ugly figure: data is fine but aesthetically wrong and is not eye friendly.

Bad figure: A figure that has problems related to perception; it may be unclear, confusing, overly complicated, or deceiving

Wrong figure: A figure that has problems related to mathematics; it is objectively incorrect.
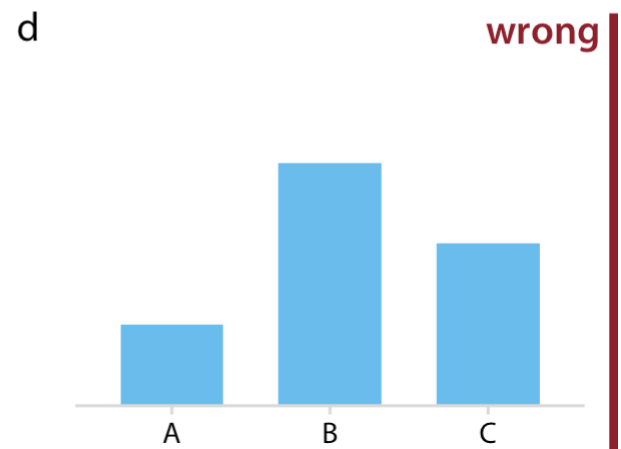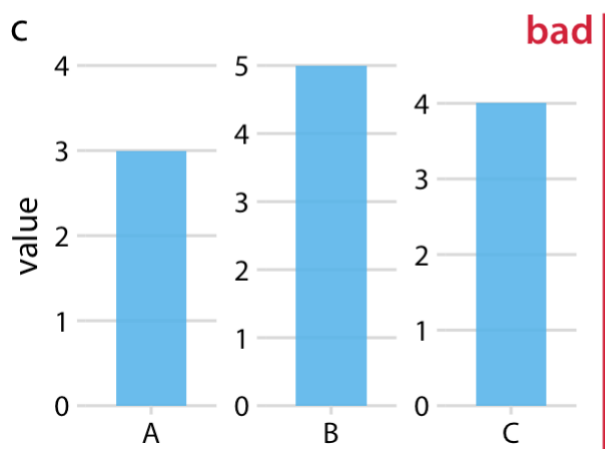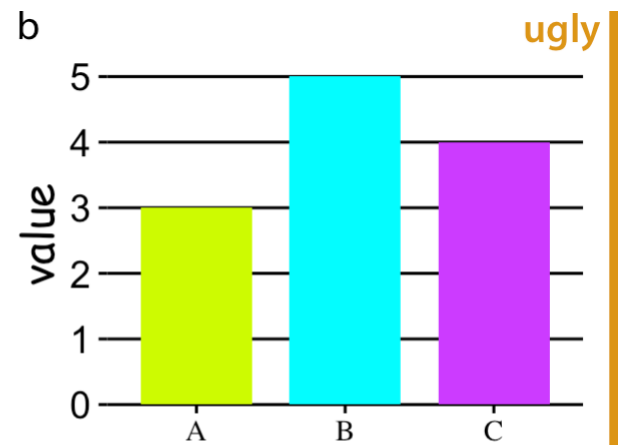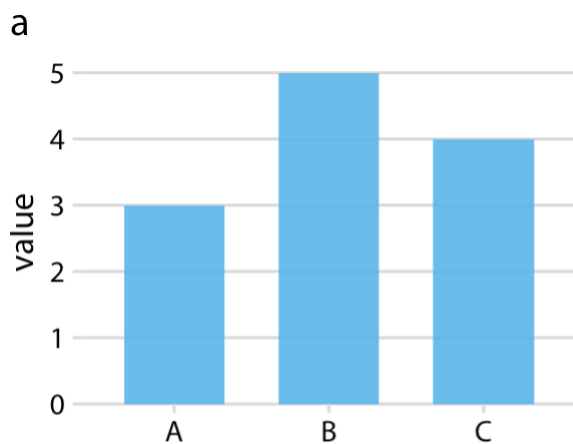
Aesthetics and Types of Data

Data visualization is an essential tool for communicating complex data to a diverse audience. It is the process of converting data into visual representations, such as graphs, charts, maps, and infographics, to help people understand and analyze the data better. In data visualization, aesthetics and the types of data play a crucial role in communicating the message effectively. In this article, we will discuss the aesthetics and types of data in data visualization in detail.

➢ Different components of Aesthetics in data visualization

Aesthetics refers to the visual design elements used in data visualization, such as colour, typography, layout, and shapes. These design elements are used to create an appealing and effective visualization that communicates the intended message to the audience.

Colour

Colour is one of the most important design elements in data visualization. It can be used to highlight important information, distinguish between different data points, and create visual contrast. However, the wrong use of colour can also create confusion and distort the intended

message. When using colour in data visualization, it is essential to choose colours that are easy to distinguish, consistent across the visualization, and accessible to the intended audience.

Typography

Typography refers to the use of fonts and text in data visualization. It can be used to emphasize key information, create hierarchy, and improve readability. However, using too many different fonts or font sizes can make the visualization cluttered and difficult to read. Therefore, it is important to use typography that is legible, consistent, and appropriate for the intended audience.

Layout

Layout refers to the arrangement of visual elements in data visualization. It can be used to guide the audience's eye through the visualization, create visual hierarchy, and improve readability. However, an overly complex or cluttered layout can make the visualization difficult to understand. Therefore, it is important to use a simple and clear layout that emphasizes the most important information.

Shapes

Shapes refer to the use of geometric shapes and icons in data visualization. They can be used to represent different data points, create visual interest, and improve readability. However, using too many different shapes or icons can create confusion and distract from the intended message. Therefore, it is important to use shapes and icons that are simple, consistent, and appropriate for the intended audience.
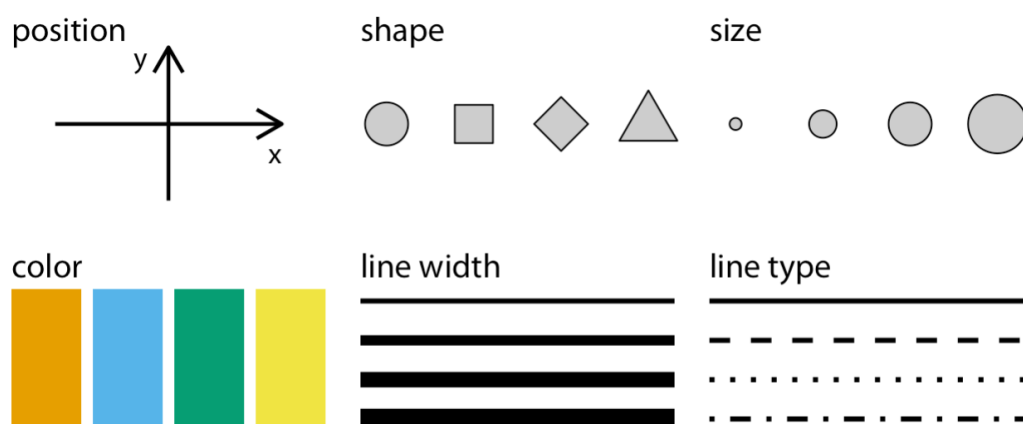
Figure: Commonly used aesthetics in data visualization: position, shape, size, colour, line width, line type. Some of these aesthetics can represent both continuous and discrete data

(position, size, line width, colour), while others can usually only represent discrete data (shape, line type).

➢ Types of Data in Data Visualization

There are different types of data that can be visualized in data visualization. Understanding the types of data is important because it determines the appropriate type of visualization to use.

Quantitative Data

Quantitative data is numerical data that can be measured and analyzed. It includes data such as age, weight, income, and temperature. Quantitative data can be visualized using charts, graphs, and histograms. Some of the most commonly used visualizations for quantitative data include bar charts, line charts, and scatter plots.

Categorical Data

Categorical data is data that falls into specific categories or groups. It includes data such as gender, race, and type of car. Categorical data can be visualized using bar charts, pie charts, and stacked bar charts. These visualizations help to compare the frequency of data points in each category.

Temporal Data

Temporal data is data that is related to time. It includes data such as sales over time, weather over time, and stock prices over time. Temporal data can be visualized using line charts, area charts, and stacked area charts. These visualizations help to show trends and patterns over time.

Spatial Data

Spatial data is data that is related to geographic locations. It includes data such as population density, crime rates, and weather patterns. Spatial data can be visualized using maps, choropleth maps, and heat maps. These visualizations help to show the distribution of data points across different geographic locations.
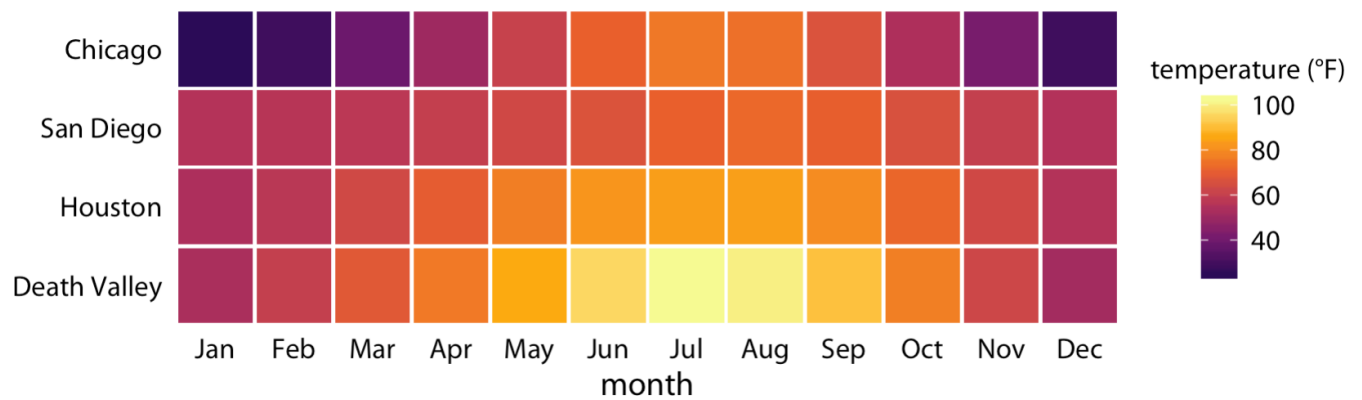
Figure: Monthly normal mean temperatures for four locations in the US.

Coordinate Systems and Axes

Coordinate systems and axes are fundamental components of data visualization. They provide the framework for plotting and presenting data in a clear and understandable way. In this article, we will discuss coordinate systems and axes in data visualization in detail.

Coordinate Systems

A coordinate system is a mathematical framework used to plot and locate points in two or three-dimensional space. In data visualization, coordinate systems are used to plot data points in a graphical representation. There are two main types of coordinate systems used in data visualization: Cartesian coordinate systems and polar coordinate systems.

Cartesian Coordinate Systems

A Cartesian coordinate system is a two-dimensional plane that is divided into four quadrants. The x-axis represents the horizontal axis, and the y-axis represents the vertical axis. The intersection of the x and y-axis is called the origin. Cartesian coordinate systems are commonly used in data visualization to plot quantitative data such as line charts, scatter plots, and bar charts.
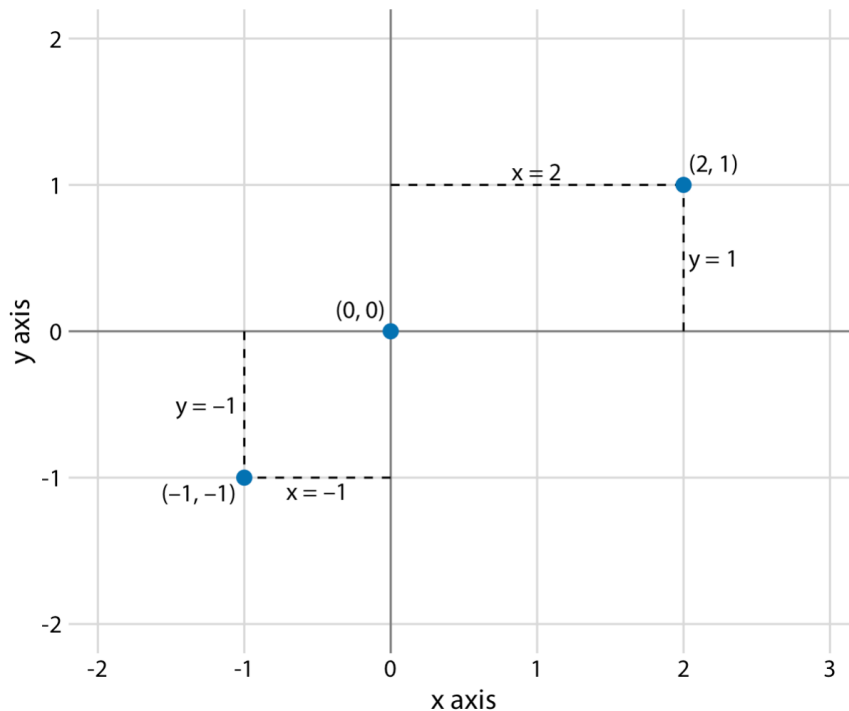
Figure: Standard Cartesian coordinate system. The horizontal axis is conventionally called x and the vertical axis y. The two axes form a grid with equidistant spacing. Here, both the x and the y grid lines are separated by units of one. The point (2, 1) is located two x units to the right and one y unit above the origin (0, 0). The point (−1, −1) is located one x unit to the left and one y unit below the origin.

Coordinate Systems with Curved Axes

In data visualization, coordinate systems with curved axes are used to plot data points in a curved or circular space. This type of coordinate system is used to represent data that has a circular or radial pattern, such as data on a clock face or a radar chart.

One common example of a coordinate system with curved axes is the polar coordinate system. In a polar coordinate system, data points are located by their distance from the center point and their angle of rotation. The distance from the center point is represented by the radius of the circle, and the angle of rotation is represented by the position of the data point along the circumference of the circle.

Another example of a coordinate system with curved axes is the logarithmic coordinate system. In a logarithmic coordinate system, the axes are curved and represent logarithmic scales instead of linear scales. This type of coordinate system is often used to represent data that has a large range of values, such as data on a seismic scale or a decibel scale.

Coordinate systems with curved axes can be effective in presenting data that has a circular or radial pattern, or data that has a large range of values. However, they can be more challenging to interpret than traditional Cartesian coordinate systems with linear axes. When using a coordinate system with curved axes, it is important to label the axes accurately and provide clear explanations of the scale being used to avoid confusion for the viewer.

Polar Coordinate Systems

A polar coordinate system is a two-dimensional plane that is based on a central point or origin. In a polar coordinate system, points are located by their distance from the origin and their angle of rotation. Polar coordinate systems are commonly used in data visualization to plot cyclic data such as circular graphs and pie charts.
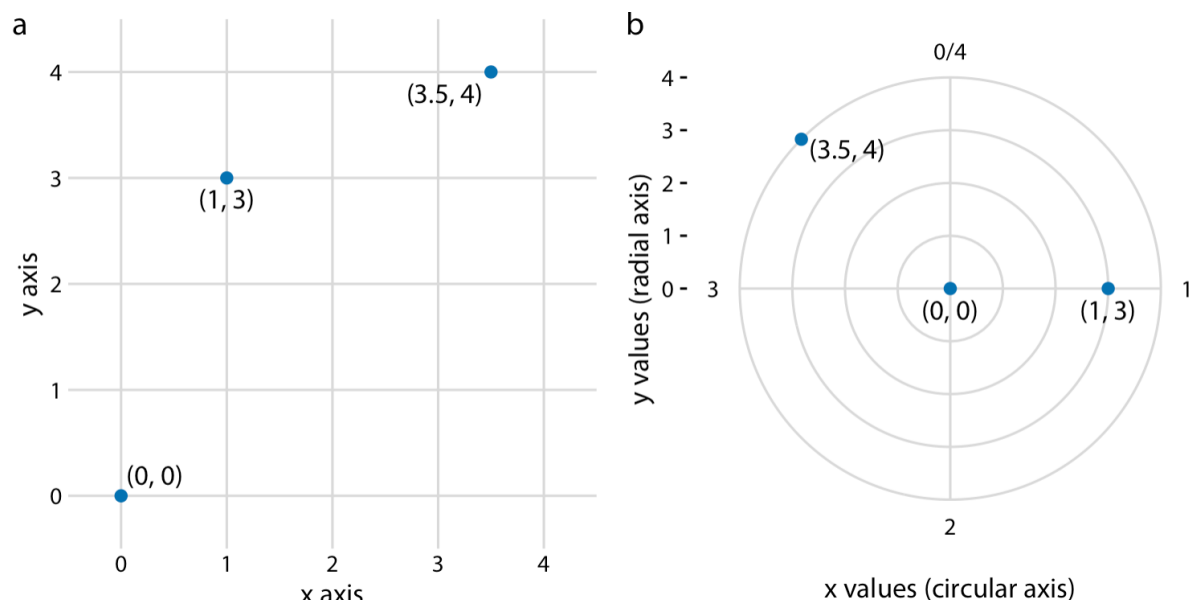

Figure: Relationship between Cartesian and polar coordinates. (a) Three data points shown in a Cartesian coordinate system. (b) The same three data points shown in a polar coordinate system. We have taken the x coordinates from part (a) and used them as angular coordinates and the y coordinates from part (a) and used them as radial coordinates. The circular axis runs from 0 to 4 in this example, and therefore x = 0 and x = 4 are the same locations in this coordinate system.

Properties of coordinate system:

Axes:
An axis is a line used to represent a quantitative variable in a coordinate system. There are two types of axes: horizontal axes and vertical axes.
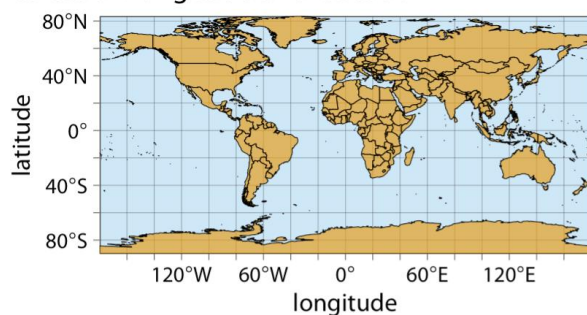
Horizontal Axes

Horizontal axes, also known as x-axes, are used to represent the independent variable in a data visualization. The independent variable is the variable that is being manipulated or controlled in the experiment or study. The horizontal axis is usually located at the bottom of the graph, and the labels on the axis indicate the values of the independent variable.
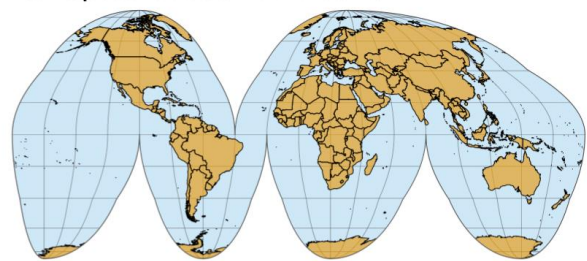
Vertical Axes

Vertical axes, also known as y-axes, are used to represent the dependent variable in a data visualization. The dependent variable is the variable that is being measured or observed in the experiment or study. The vertical axis is usually located on the left side of the graph, and the labels on the axis indicate the values of the dependent variable.

In data visualization, the axes play an important role in providing context and meaning to the plotted data. They help to identify the range of values for each variable and provide a reference point for interpreting the data. The labeling of axes is critical for clear communication of the information being presented. They should be labeled accurately and descriptively with units of measurement if applicable.
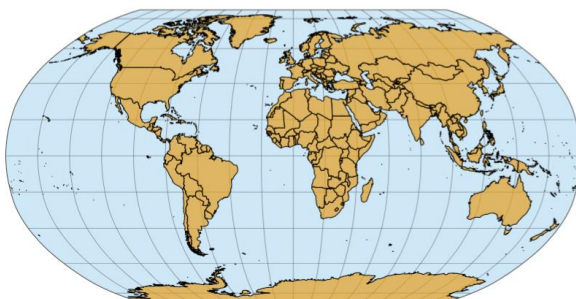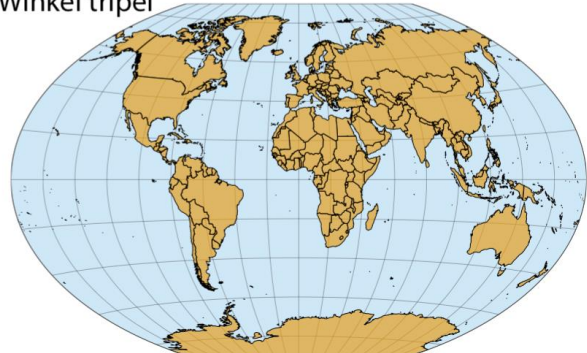

Figure: Map of the world, shown in four different projections. The Cartesian longitude and latitude system maps the longitude and latitude of each location onto a regular Cartesian coordinate system. This mapping causes substantial distortions in both areas and angles

relative to their true values on the 3D globe. The interrupted Goode homolosine projection perfectly represents true surface areas, at the cost of dividing some land masses into separate pieces, most notably Greenland and Antarctica. The Robinson projection and the Winkel tripel projection both strike a balance between angular and area distortions, and they are commonly used for maps of the entire globe.

Colour Scales:
Colour scales are a fundamental component of data visualization. They are used to represent and communicate data in a visual way. A colour scale is a range of colours that are used to represent values on a numeric scale. Each colour on the scale represents a different value, with darker or brighter colours indicating higher or lower values.

Colour scales can be qualitative or quantitative. Qualitative colour scales are used to represent categorical data, where each colour on the scale represents a different category. Quantitative colour scales are used to represent numerical data, where each colour on the scale represents a range of values on a numeric scale.

One popular type of quantitative colour scale is the gradient scale. Gradient scales use a range of colours that transition smoothly from one colour to the next, representing a range of values. For example, a gradient scale might start with light blue for low values, transition to dark blue for moderate values, and end with red for high values.

When using colour scales, it is important to choose colours that are easy to distinguish and that are suitable for the type of data being presented. For example, using red and green together can be difficult for people with colour vision deficiencies. In addition, colour scales should be used consistently throughout a visualization to avoid confusion.

Colour as a tool to distinguish:
Color can be used to distinguish between different categories of data, such as different types of products or different regions of the world. This can be accomplished by using a different color for each category, or by using a color scale that assigns a range of colors to different values or levels of a variable.

Color can also be used to distinguish between different values or levels of a variable. For example, a color scale could be used to represent different levels of income, with darker shades of green representing higher income levels and lighter shades representing lower income levels.

In addition, color can be used to distinguish between different data points on a chart or graph. For example, different data points could be represented by different colors on a scatter plot, with each color representing a different category or value.

When using color to distinguish between different categories or values, it is important to choose colors that are easily distinguishable and that do not cause confusion for the viewer. Color combinations that are difficult to distinguish, such as red and green, should be avoided. In addition, color should be used consistently throughout a visualization to ensure clarity and accuracy in the presentation of data.
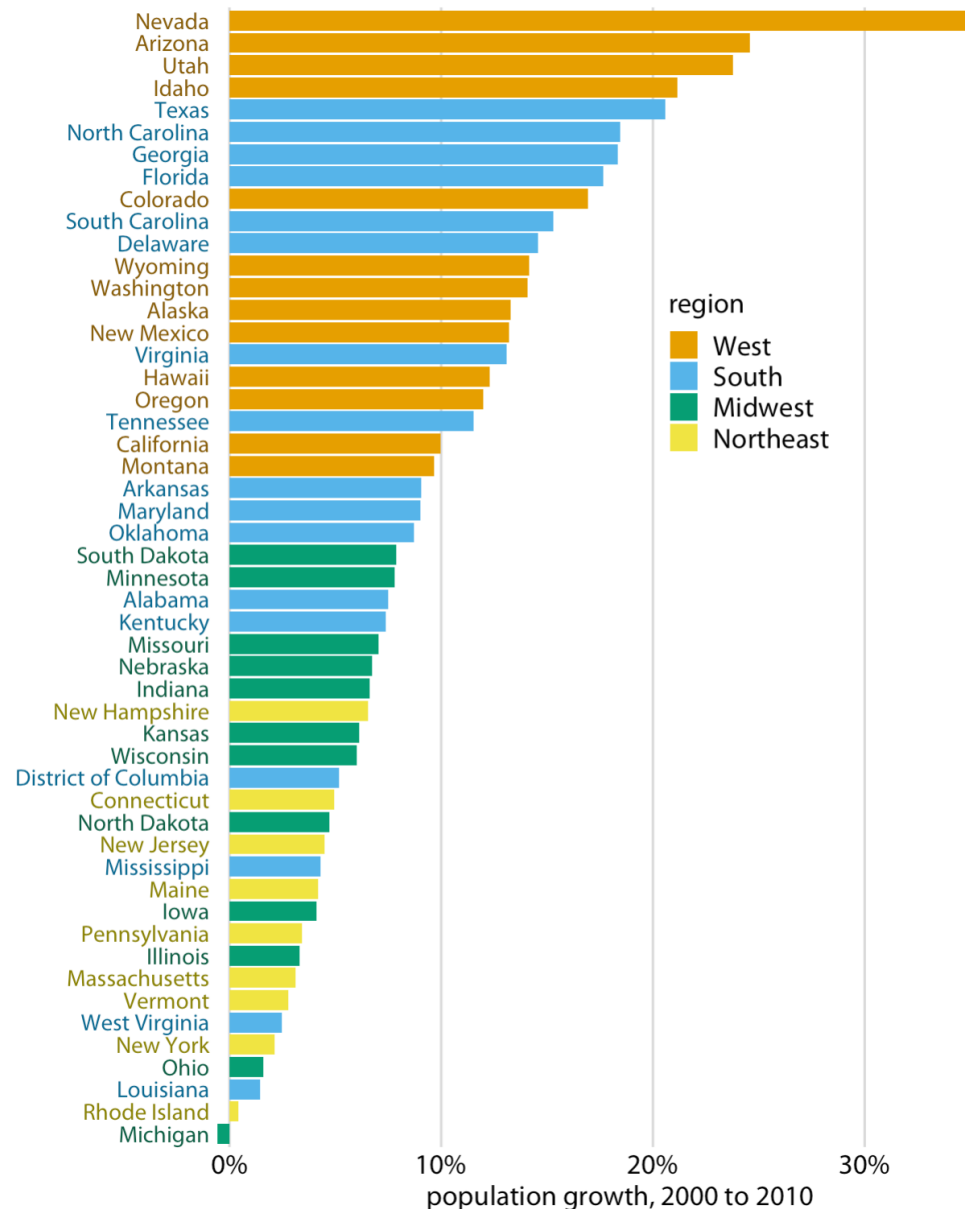

Figure: Population growth in the US from 2000 to 2010. States in the West and South have seen the largest increases, whereas states in the Midwest and Northeast have seen much smaller increases (or even, in the case of Michigan, a decrease). Data source: US Census Bureau.
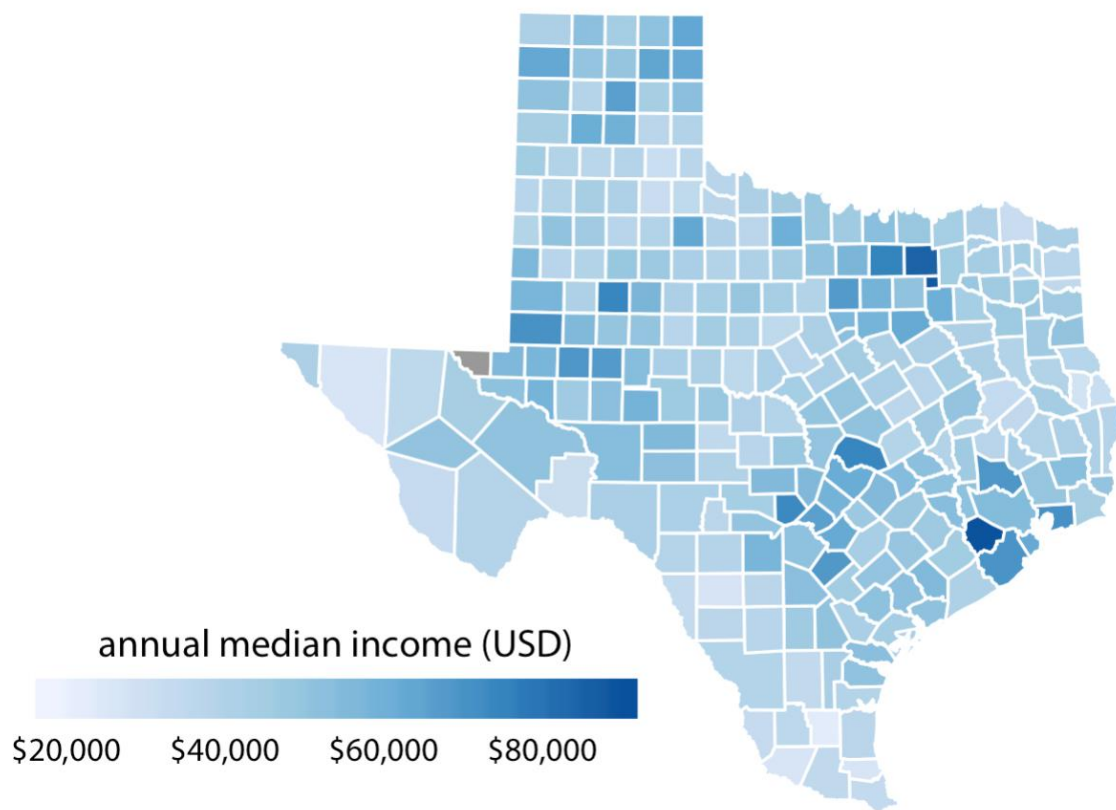
Figure: Median annual income in Texas counties. The highest median incomes are seen in major Texas metropolitan areas, in particular near Houston and Dallas. No median income estimate is available for Loving County in West Texas, and therefore that county is shown in gray. Data source: 2015 Five-Year American Community Survey.

Color as a Tool to Highlight: