CS5346

INFORMATION VISUALIZATION

AXES and SCALES

Reference : Fundamentals of Data Visualization by Claus Wilke
clauswilke.com/dataviz/

Assoc Prof. Bimlesh Wadhwa | School of Computing | NUS | bimlesh@nus.edu.sg

Lecture outline

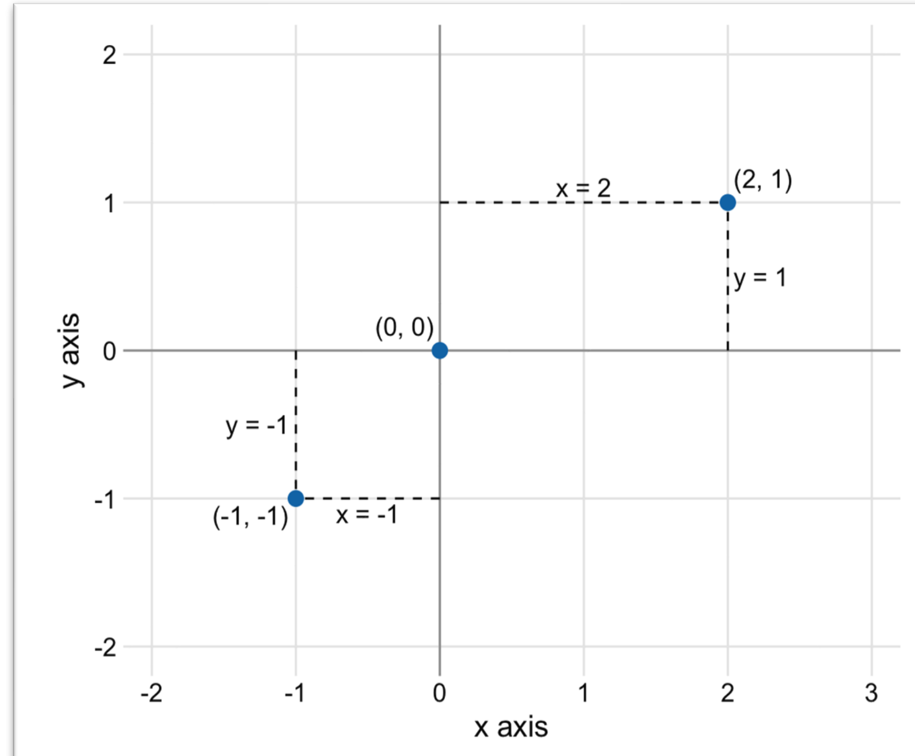
- Cartesian coordinates, Axes and Scale - Quantitative and Categorical scales
- Dual-scaled axes – best practices
- Nonlinear axes – Logarithmic and sqrt scales
- Curved axes
- Color scales – Qualitative, Sequential, Divergent, Accent

Cartesian coordinates, Axes and Scale
Quantitative and Categorical scales

Cartesian Coordinates, Axes and Scale

2D Cartesian coordinate system : each location is uniquely specified by an x and a y value.

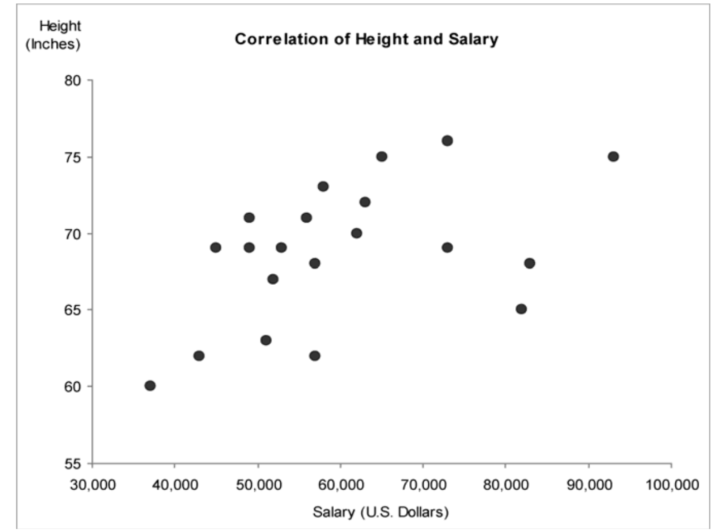
The x-axis and y-axis run orthogonally to each other, and data values are placed in an even spacing along both axes.

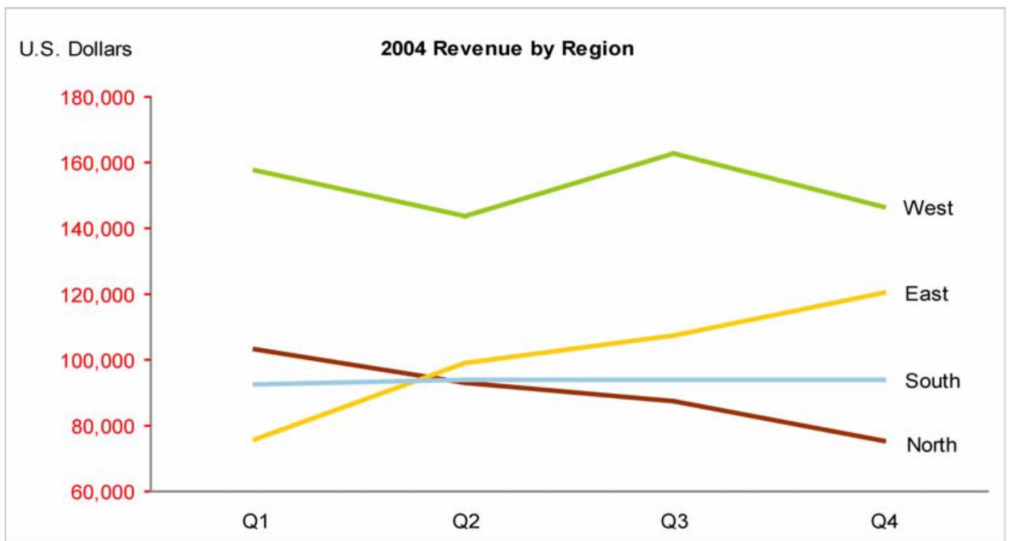


The scale of an axis is **the units into which the axis is divided**.
The units are marked by ticks, labels, and grid lines.

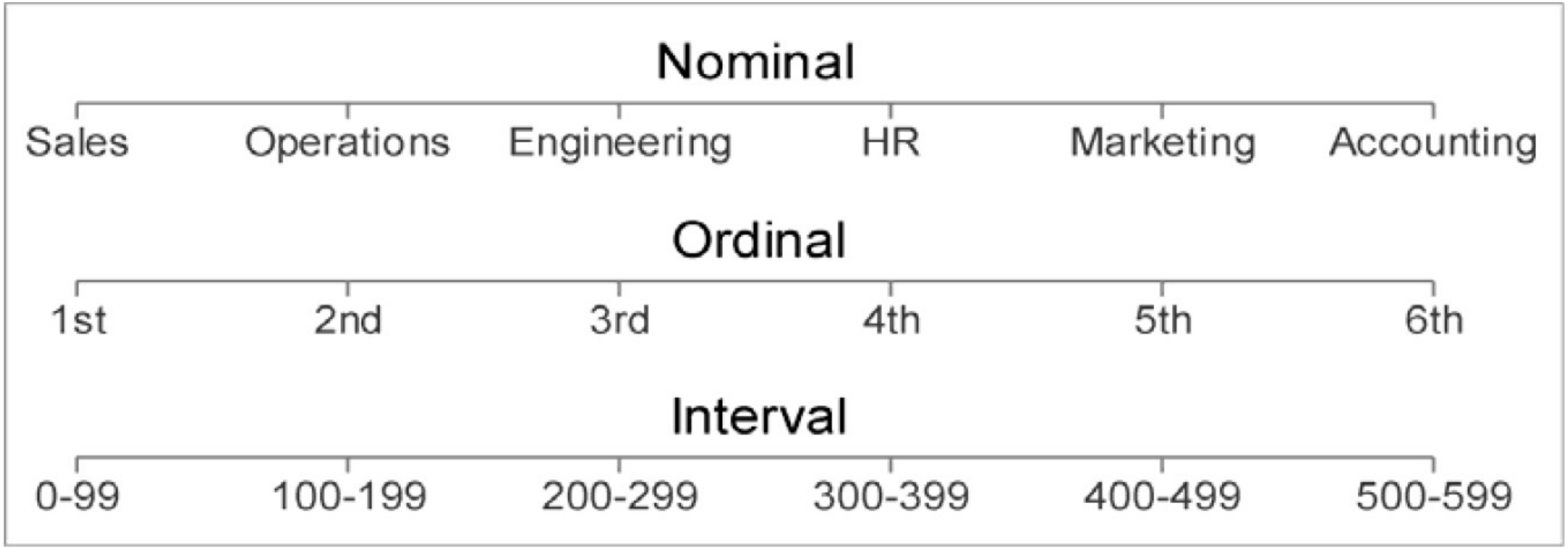
Categorical and Quantitative scales

Most two-dimensional graphs consist of one **categorical scale** and one **quantitative scale** (in the example below), OR
quantitative scales along both axes (in the example on the right)





More about Categorical Scales



More about Categorical Scales

Nominal scales:

Discrete categories

No relation to one another;
No particular order

They don't represent quantitative values.

Ordinal scales:

intrinsic order

But cannot quantify the difference between two values

Interval Scale: “space in between,”

intrinsic order; difference between the value is known(~equal intervals)

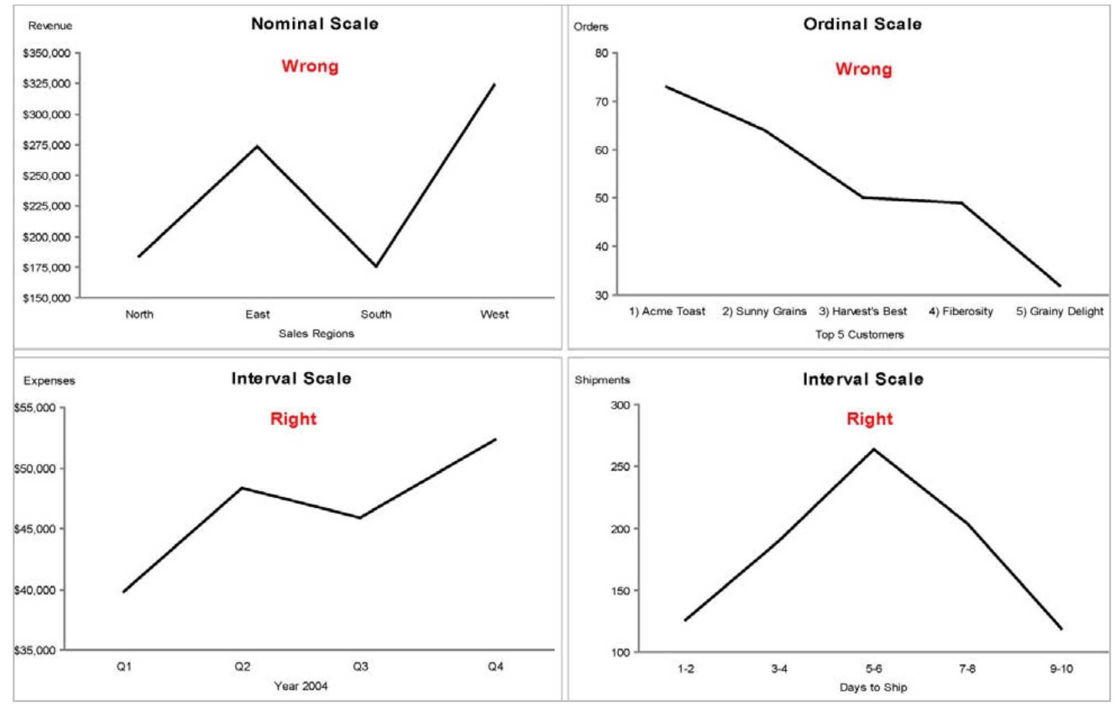
An interval scale ***starts out as a quantitative scale*** *that is* ***then converted into a categorical scale*** *by subdividing the range of values into a sequential series of smaller ranges of equal size and by giving each range a label.*

Best Practices

- In nominal and ordinal scales, use bars instead of lines

 -- the individual items are not related closely enough to be linked with lines

- Lines (as in a line graph) are used to encode data along an interval scale.



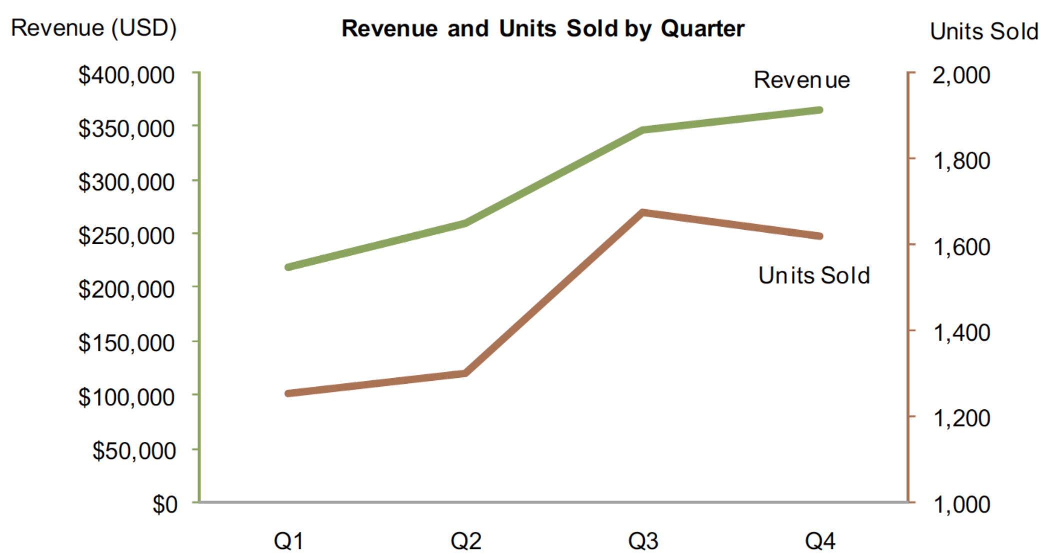
With interval scales, you are <u>not</u> forced to use lines; you can use bars as well.
If you want to emphasize changes from one item to the next, line works best.
If, you want to emphasize individual items, such as individual months, or to support discrete comparisons of multiple values at the same location along the interval scale, such as revenues and expenses for individual months, then bars work best.

Dual-scaled Axes – best practices

Dual-scaled Axes

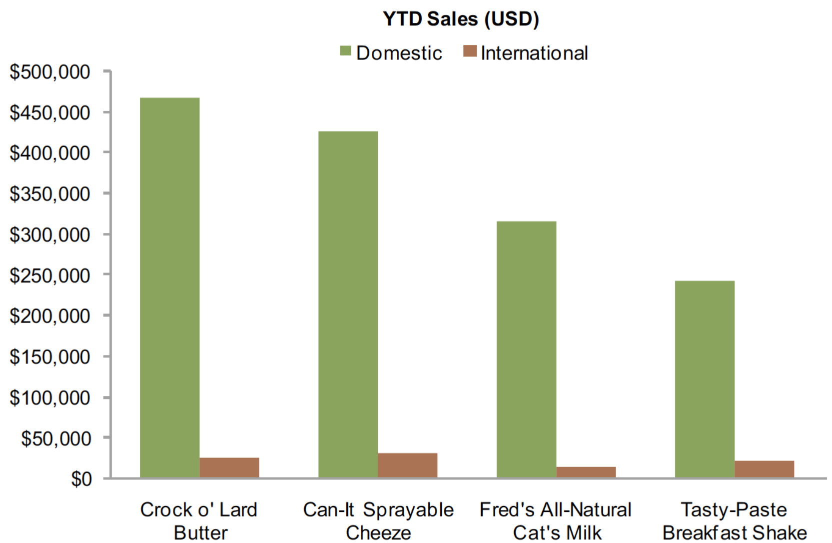
Example Source: https://www.perceptualedge.com/articles/visual_business_intelligence/dual-scaled_axes.pdf

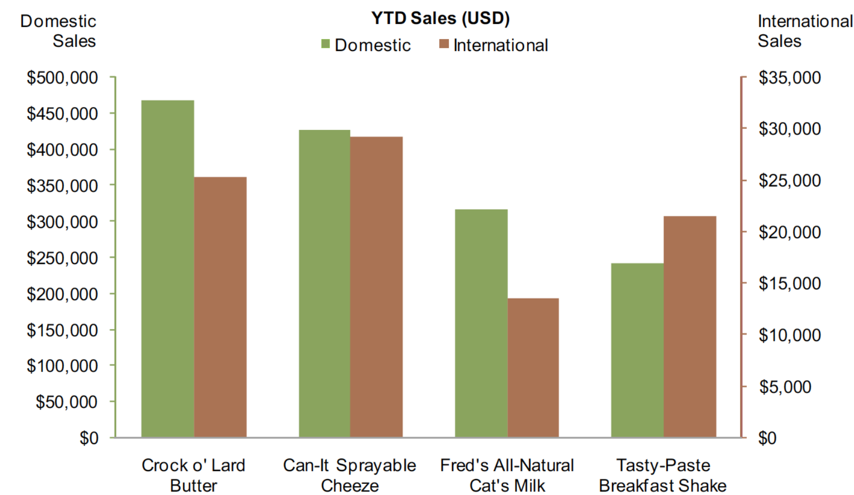
- Two quantitative scales on a single axis
- When: to combine two different units of measure in a single graph to support meaningful comparisons



Dual-scaled Axes

Best Practice: A graph should only include a dual-scaled axis when needed to compare data sets that have different units of measure.





Dual-scaled Axes

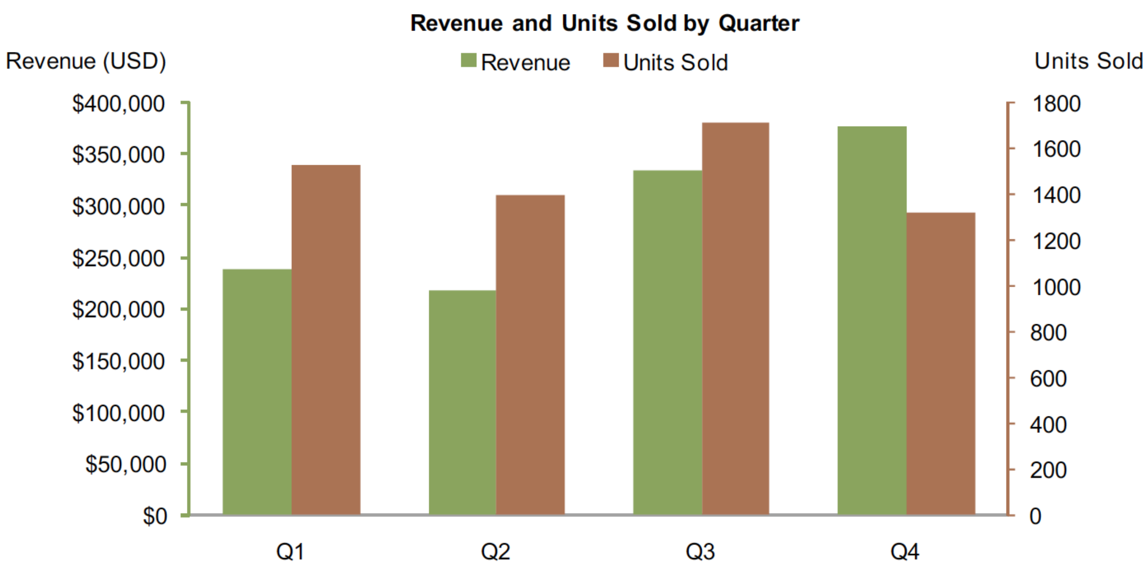
Best Practice :
Magnitude comparisons between values with different units of measure and scales are not appropriate, and should therefore be discouraged.

Because **bar graphs are designed for magnitude comparisons**, a graph with a dual-scaled axis should never exclusively encode values as bars.

Example :

Left axis : product revenue

Right axis: the number of units sold

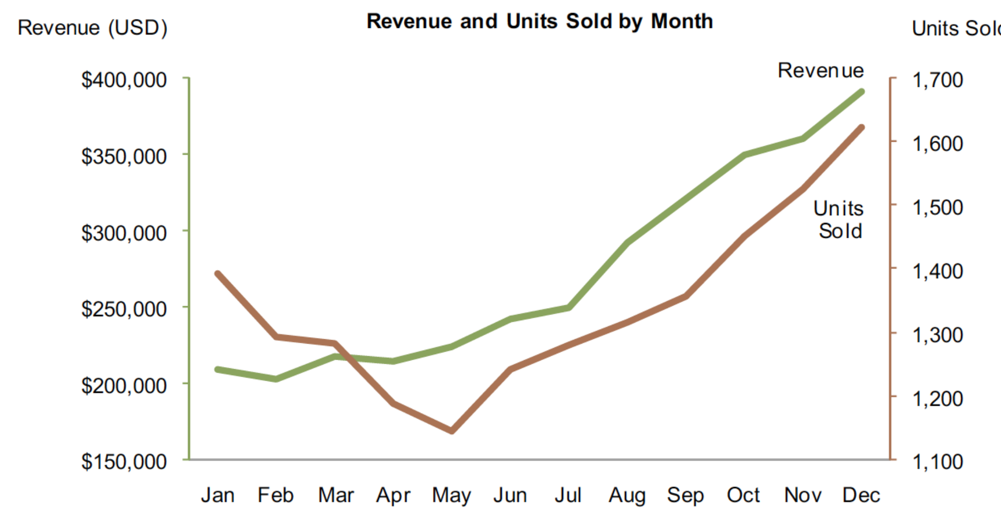


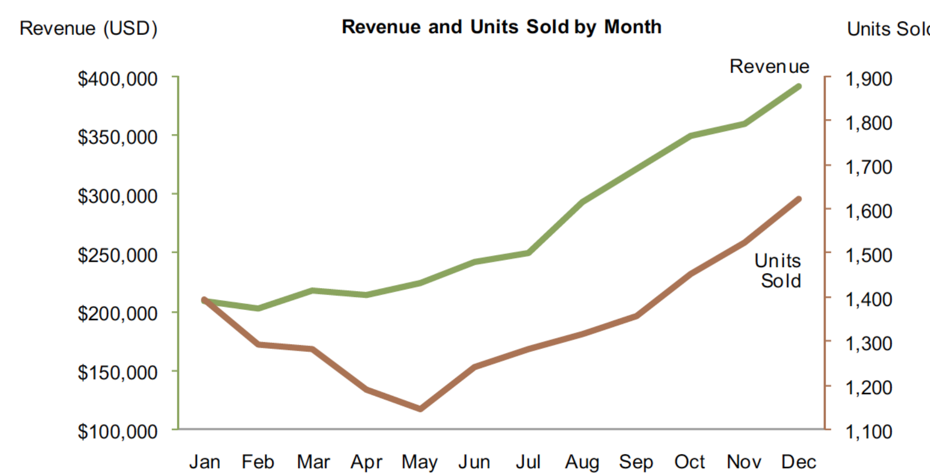
The two sets of bars have completely independent units of measure and scales.

The magnitude comparisons between them are completely meaningless.

Dual-scaled Axes

Lines should be used to display values in graphs with dual-scaled axes, because only lines focus attention on the overall pattern formed by the values and comparisons of those patterns.

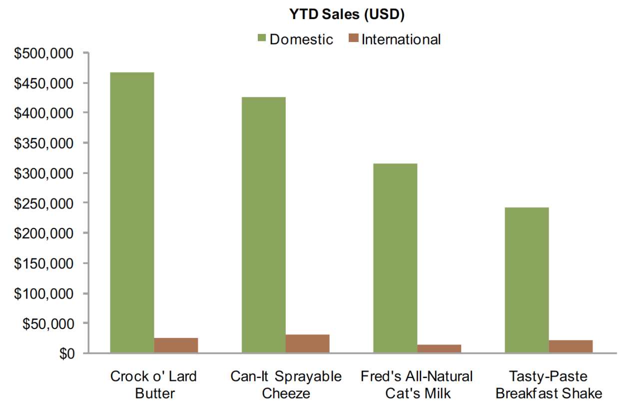


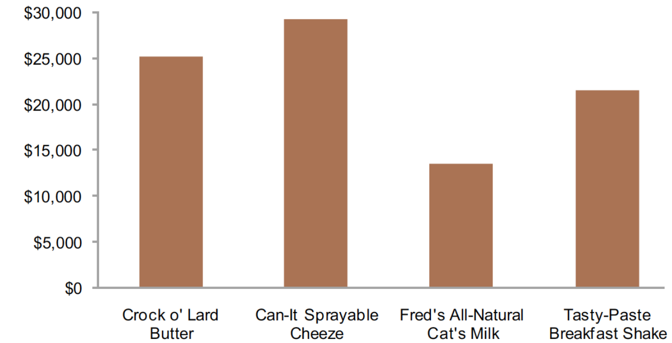


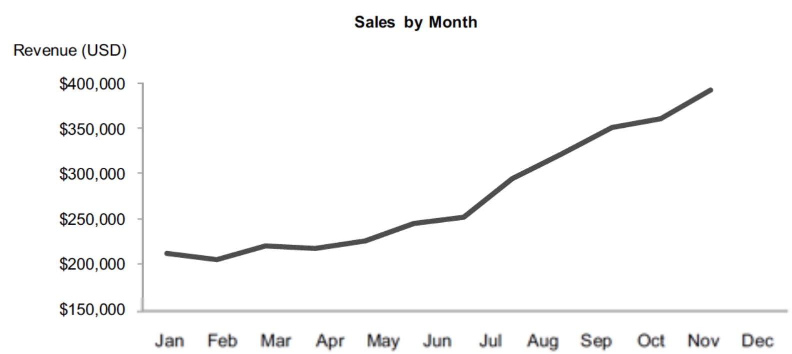
Dual-scaled Axes

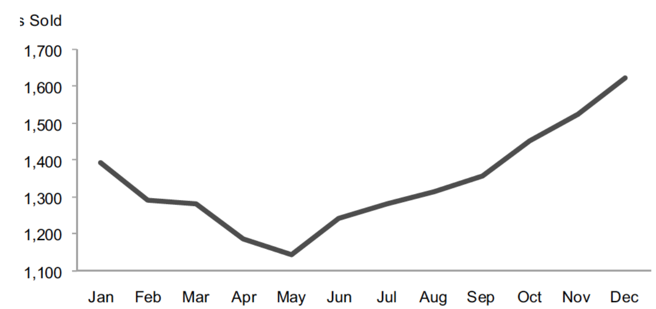
Best Practice

Avoid using more than one quantitative scale on a single axis, because, to some degree, this encourages users to (meaninglessly)compare magnitudes of values between them.









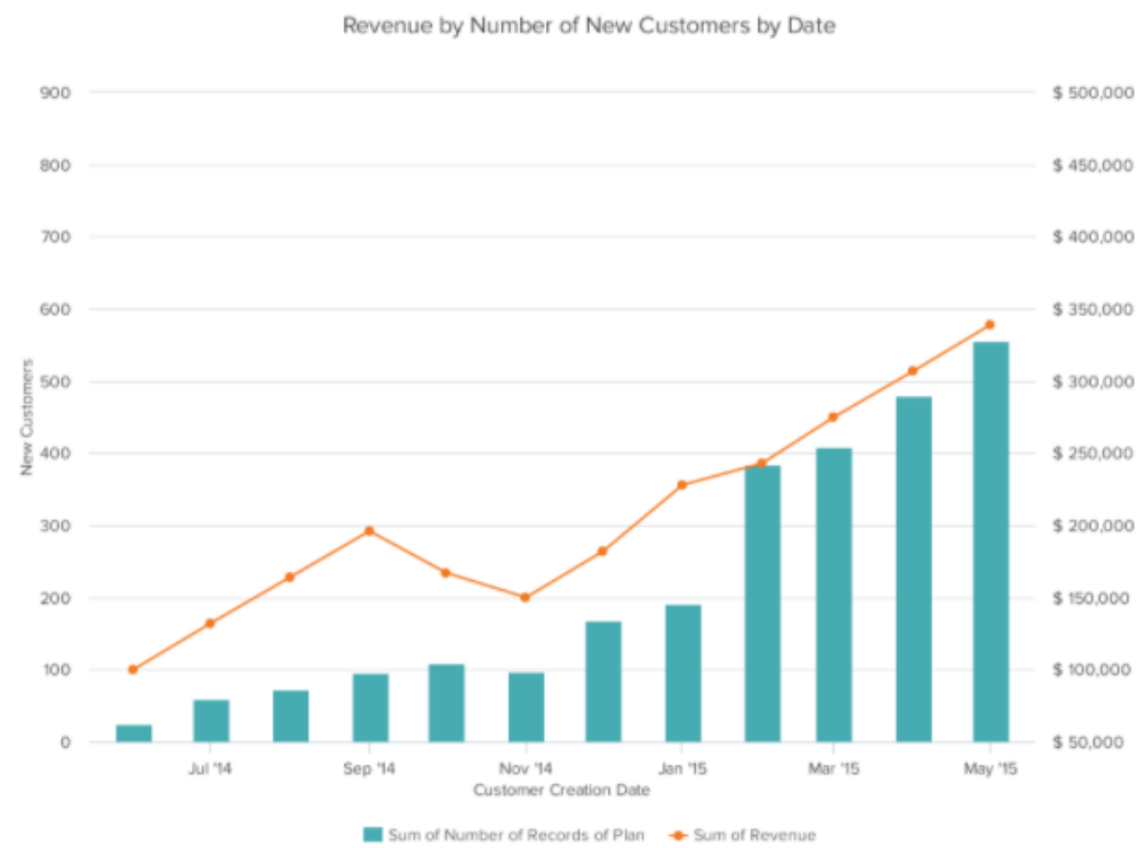
Dual-scaled Axes

Best Practice

Use the y-axis on the left side for the primary variable

Use different graphing styles to illustrate the two data sets

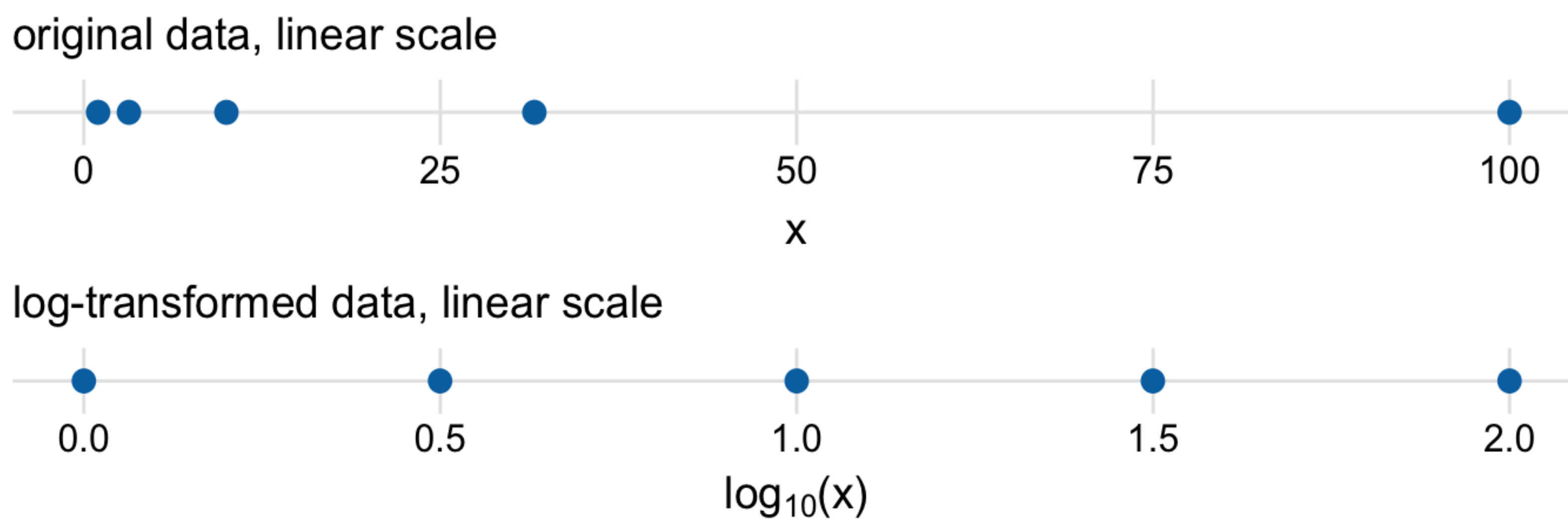
Choose distinguishing colors for the two data sets.

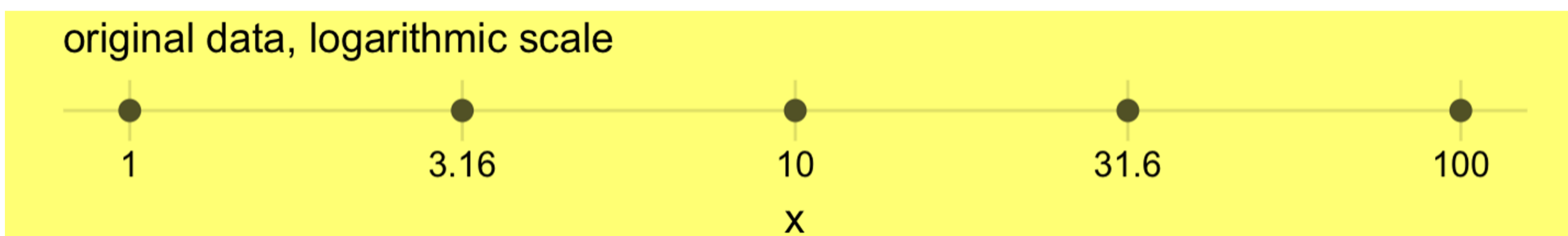


Source: https://blog.hubspot.com/marketing/types-of-graphs-for-data-visualization

Nonlinear axes – Logarithmic and sqrt scales

Nonlinear axes – Log scale





Log scale

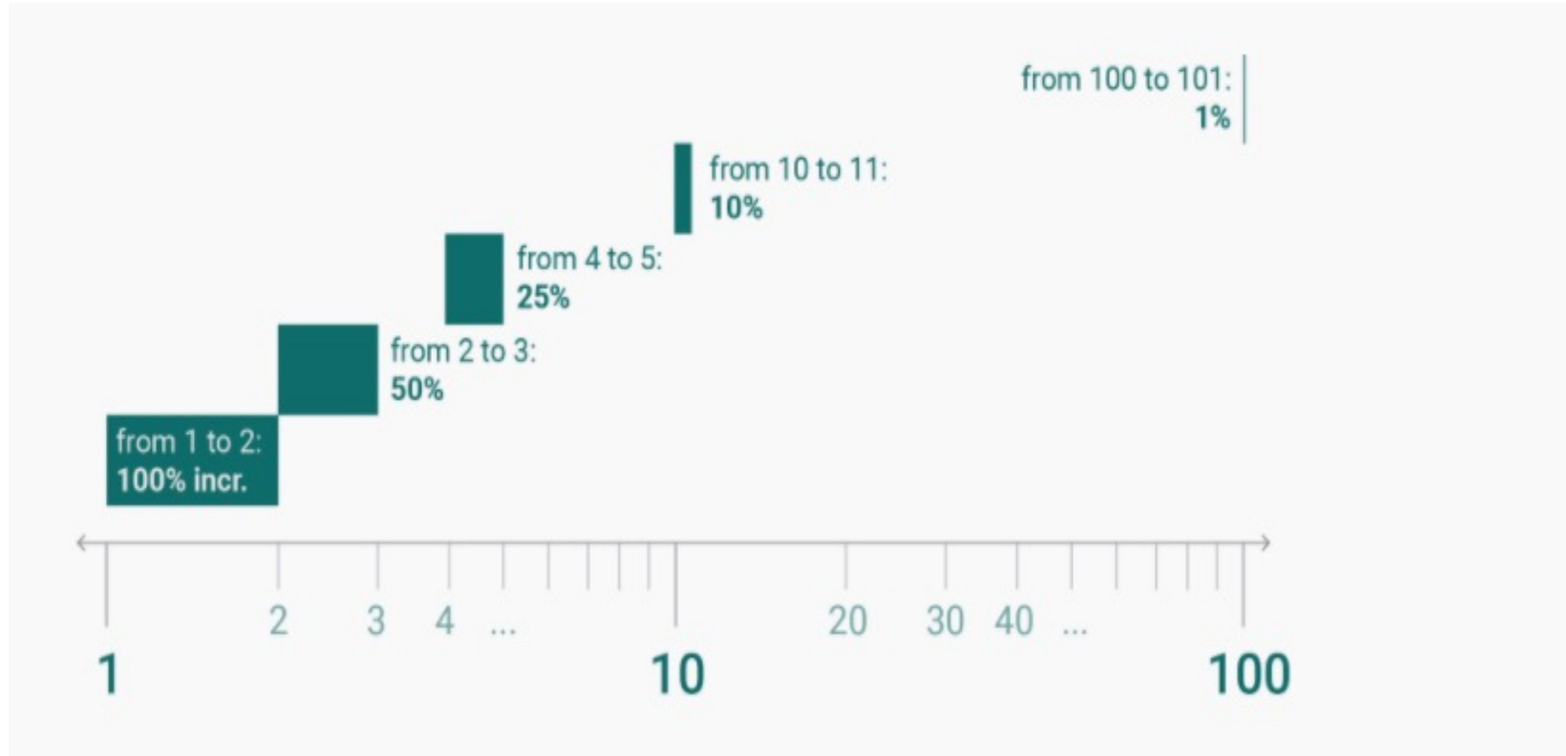
Log scales show relative values instead of absolute ones.

Log scale is concerned with percentages:

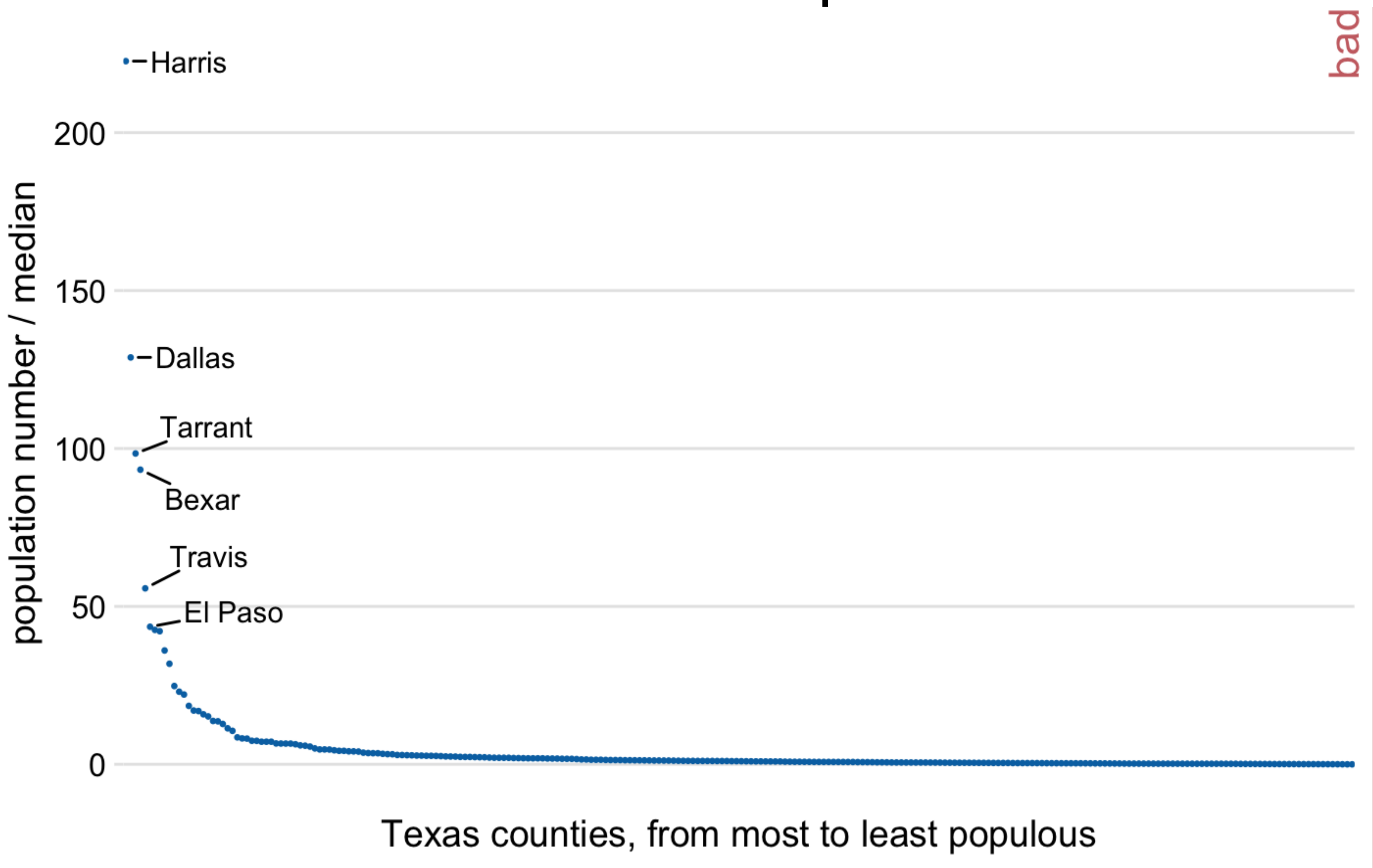
the difference between 100 and 101 is a 1% increase,

while the difference between 1 and 2 is a 100% increase.

So on a log scale, the distance between 100 and 101 is roughly 1% of the distance between 1 and 2.



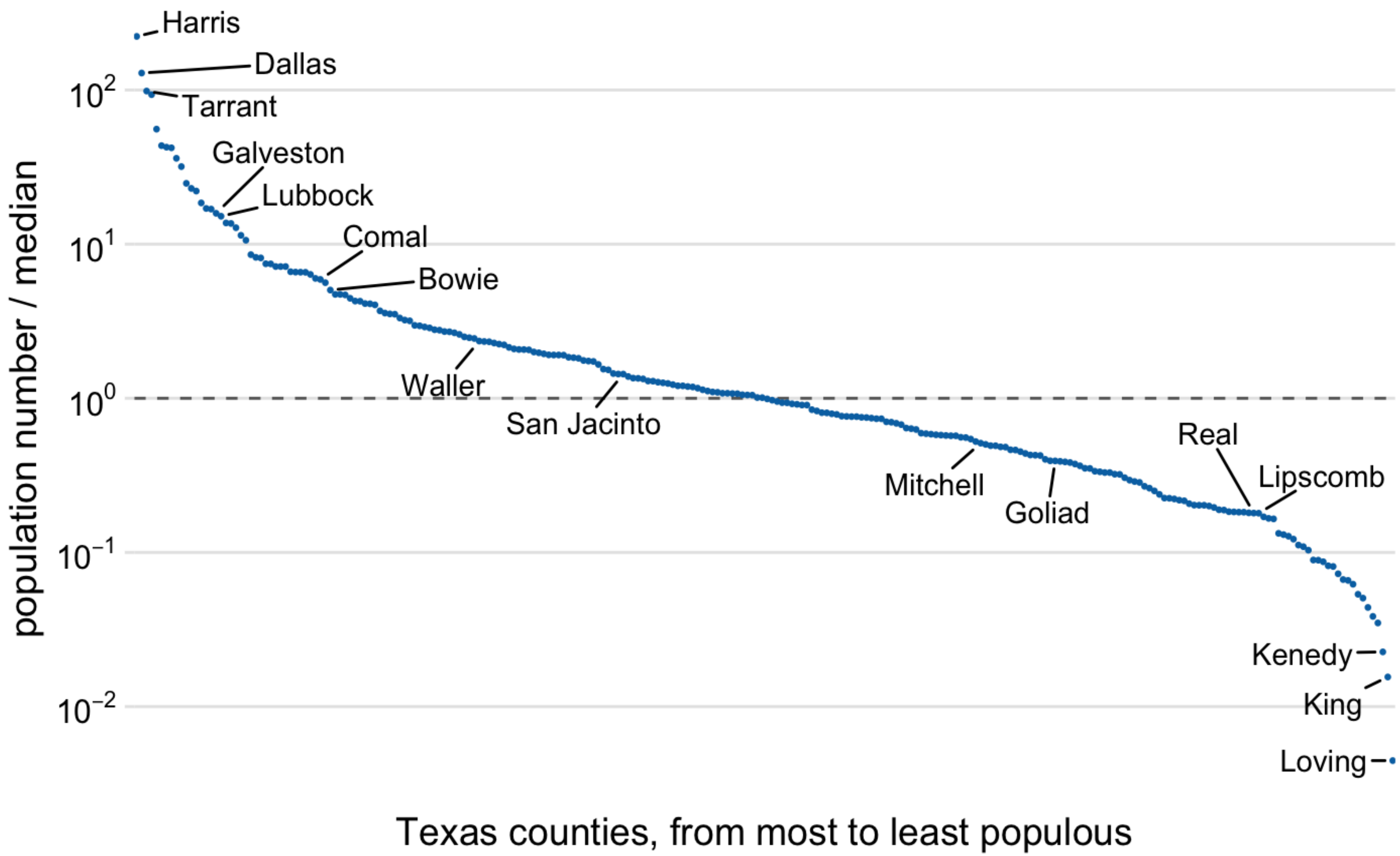
Example



Population sizes of Texas counties relative to their median value. By displaying a ratio on a linear scale, we have overemphasized ratios > 1 and have obscured ratios < 1.

As a general rule, ratios should not be displayed on a linear scale.

Example

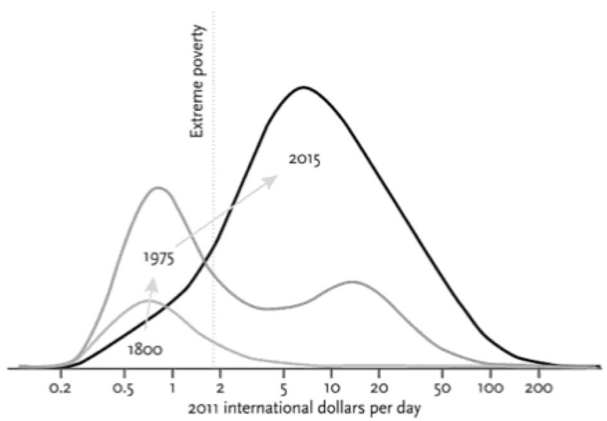


Population numbers of Texas counties relative to their median value. Select counties are highlighted by name. The dashed line indicates a ratio of 1, corresponding to a county with median population number. The **most populous** counties have approximately **100 times more** inhabitants than the median county, and the **least populous** counties have approximately **100 times fewer** inhabitants than the median county.

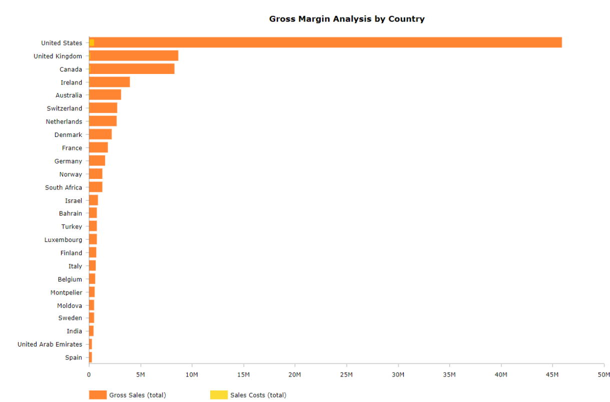
Example

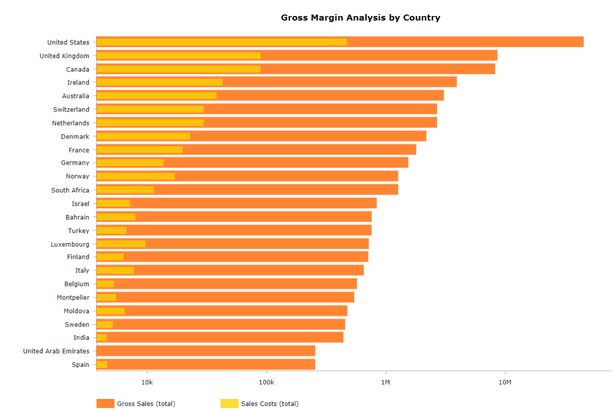
- Linear scales : more natural to users
- Non-Linear eg Logarithmic scale : useful for relatively rare cases for users who do(or are trained to) think in logarithms.

- When is it used : e.g. when there is a large value range in your dataset.
  - i.e. the values are based on order of magnitude.
  - Each mark on an axis represents a value that is a multiple of the previous mark on the axis.

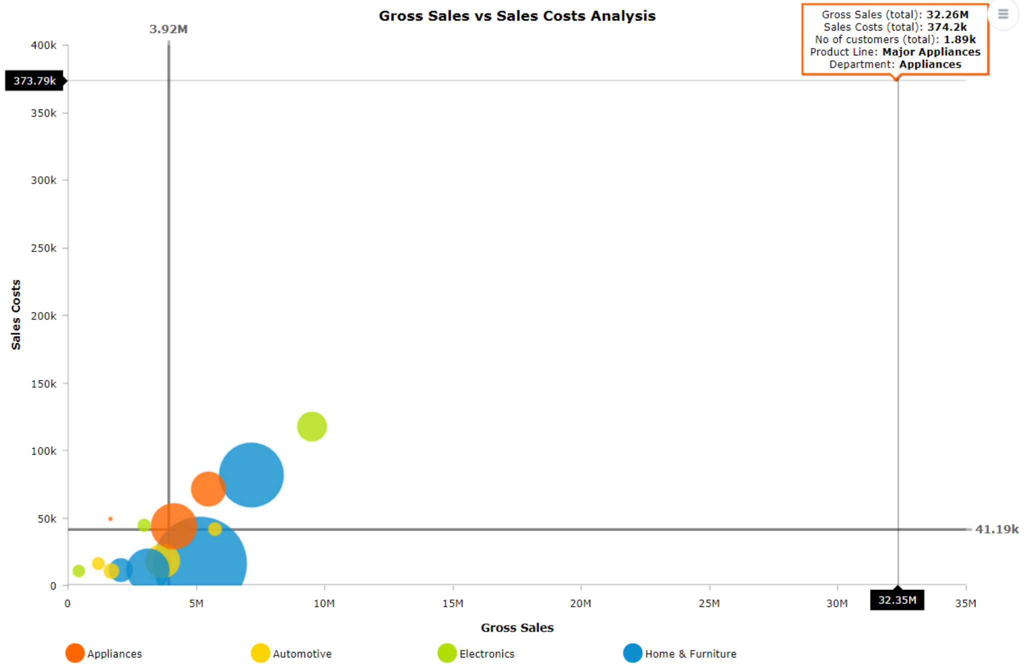


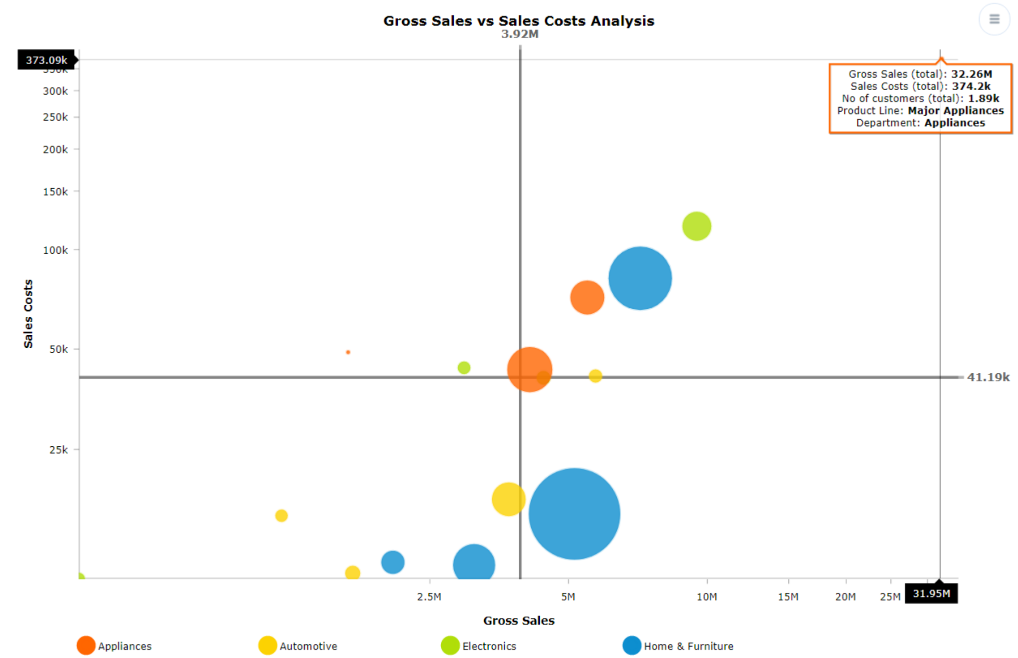
Had a linear scale been used, the income distribution would appear significantly skewed with a peak nearer to the lower end and a long declining tail extending to the right.

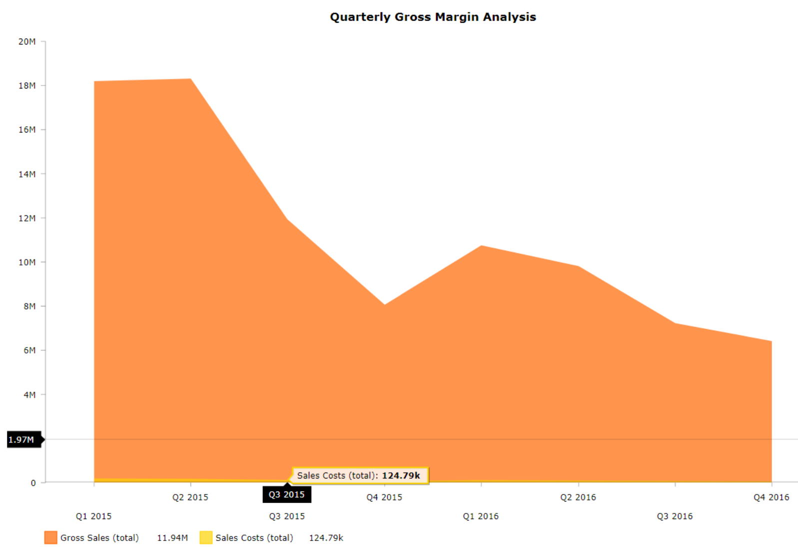


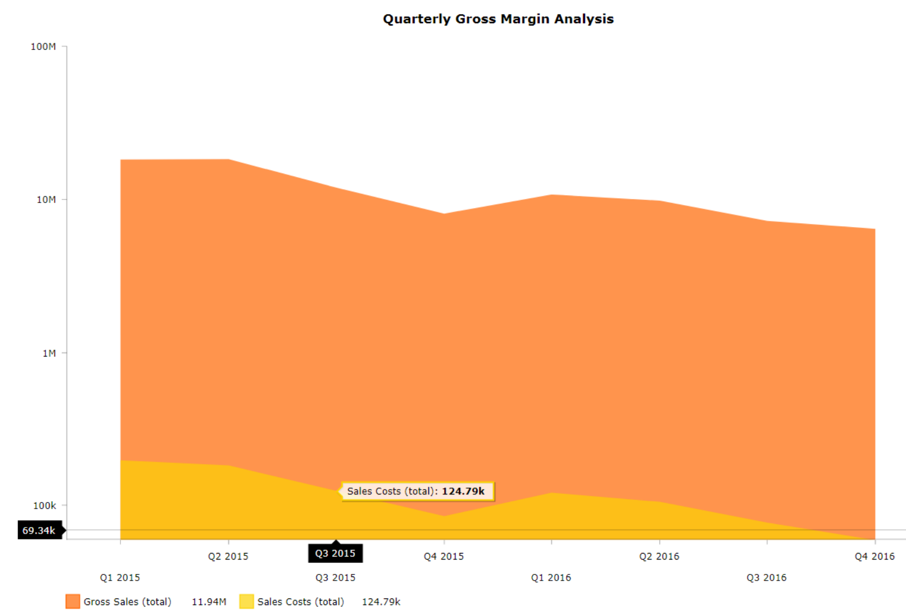


Source: https://www.dataclaritycorp.com/the-power-of-logarithmic-scale/





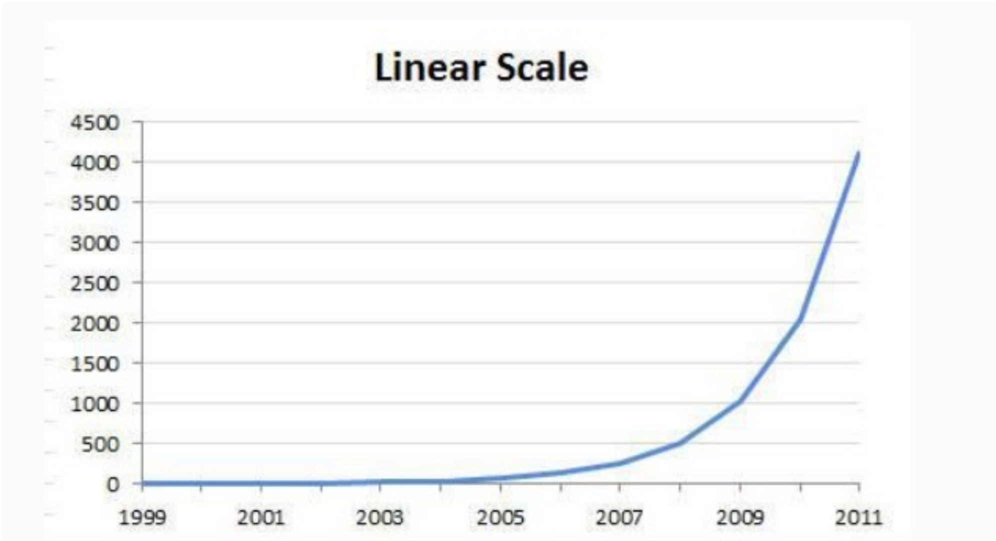


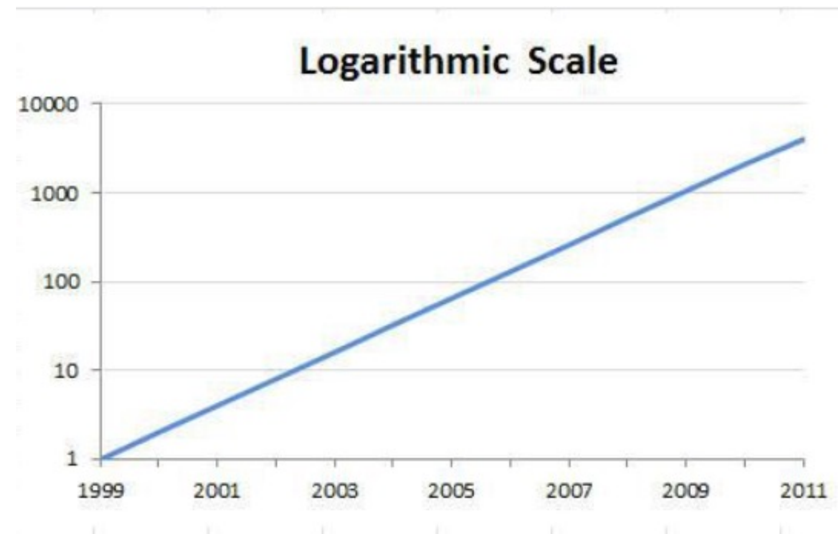


Source: https://www.dataclaritycorp.com/the-power-of-logarithmic-scale/

Example

- When to use : e.g. to visualize rate of change.





Source: https://www.forbes.com/sites/naomirobbins/2012/01/19/when-should-i-use-logarithmic-scales-in-my-charts-and-graphs/?sh=4f08a6305e67

Same distances on log scale show the same percentage growth or decline.

Year 0- > **1** car.

Year 1-> 2 cars

Year 2 -> 4 cars

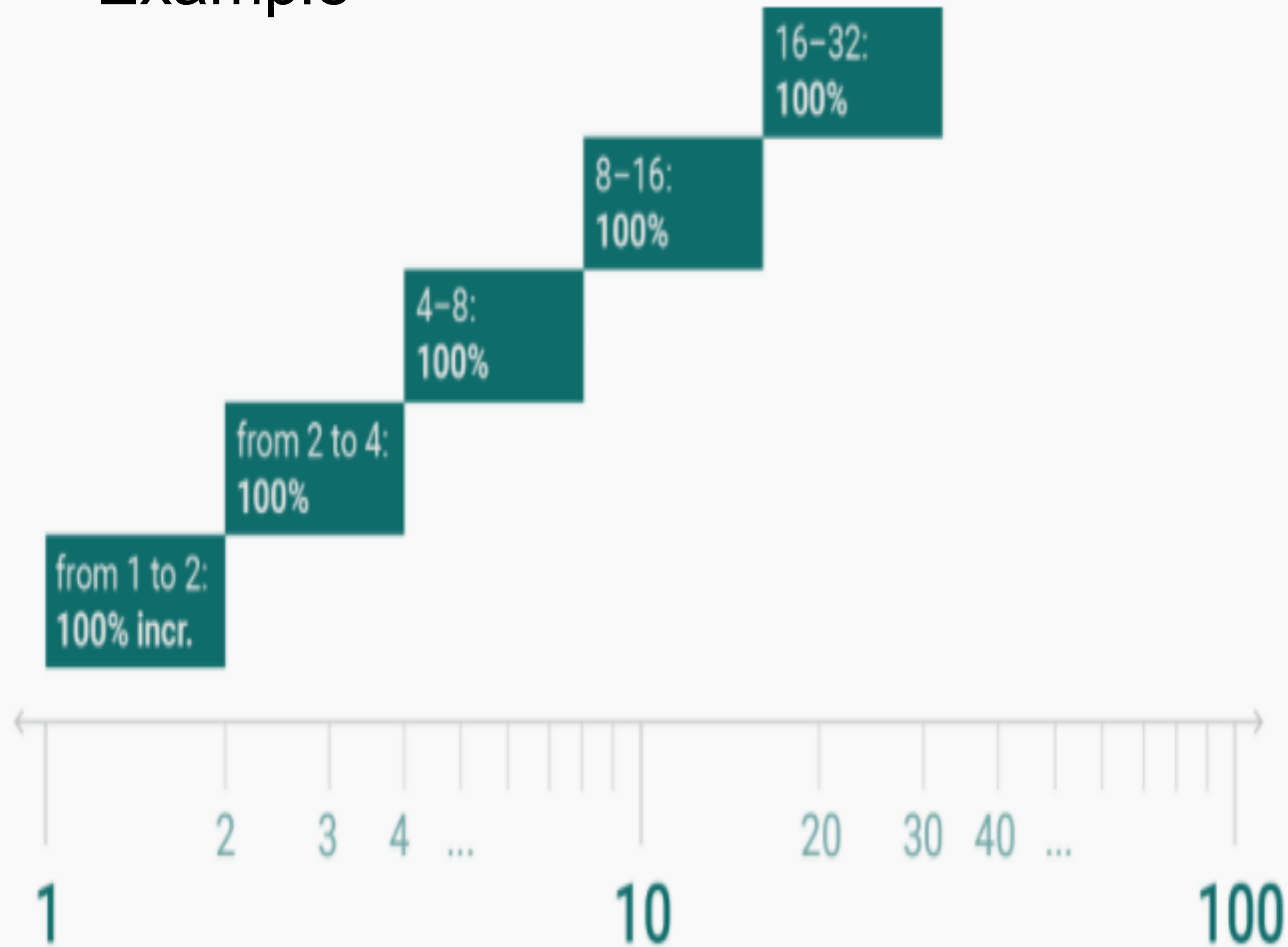
----Year 25 -> 30+ **million** cars.

On a linear scale,--> a line chart would go insanely high

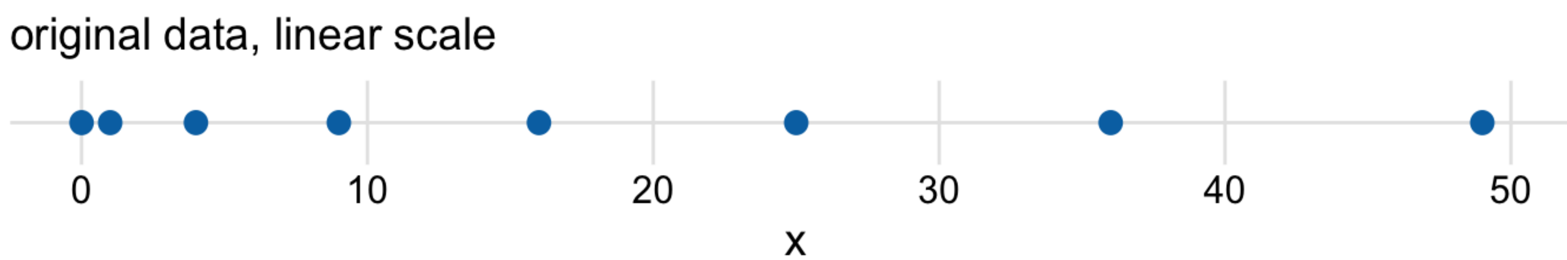
On a log scale, each amount would take up the same space,

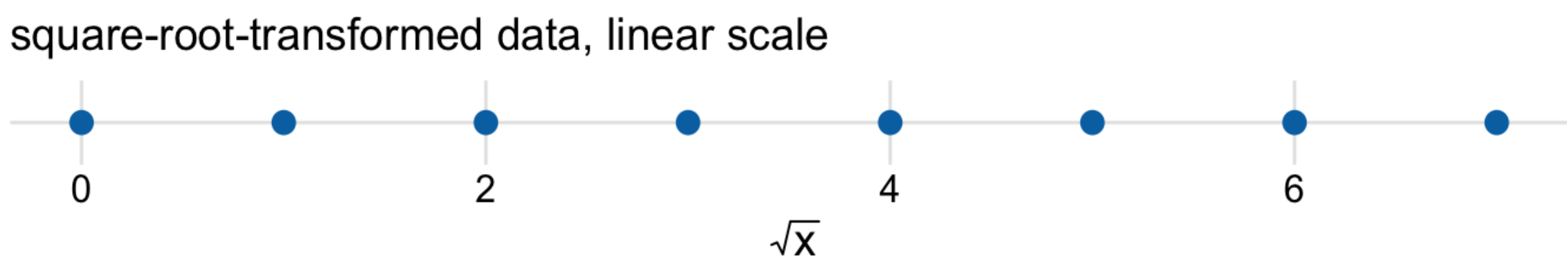
because, the number of cars **grows at a fixed rate (100%)**.

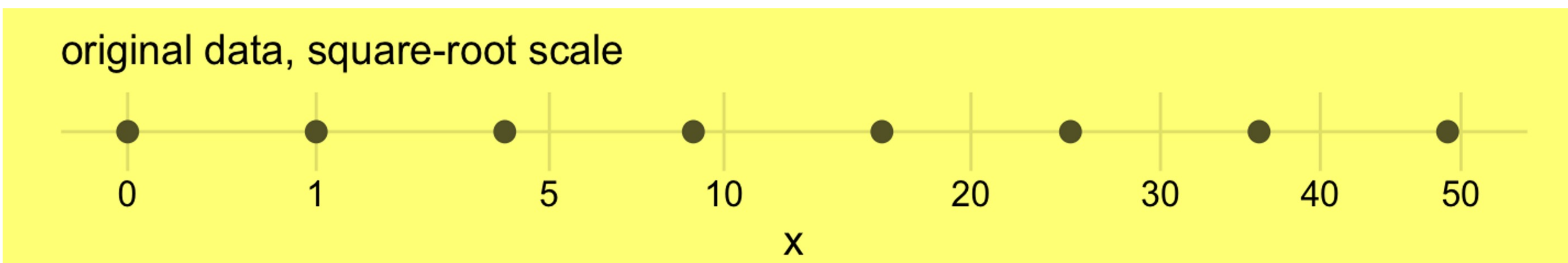
Example



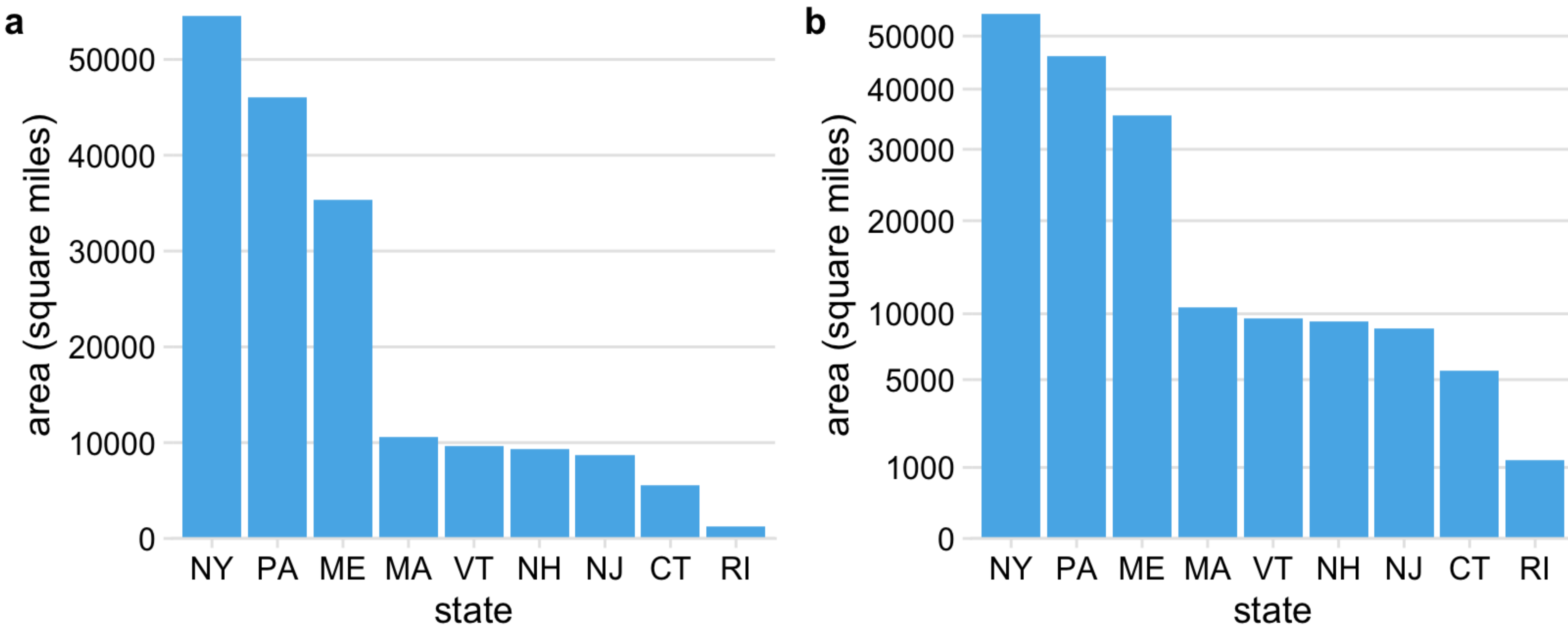
Other Nonlinear axes e.g. sqrt scale







Example

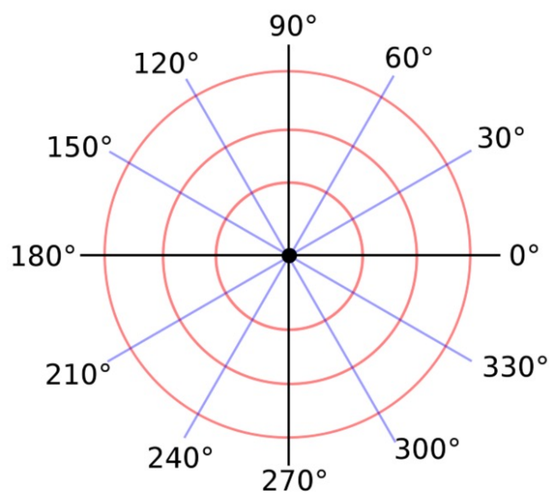


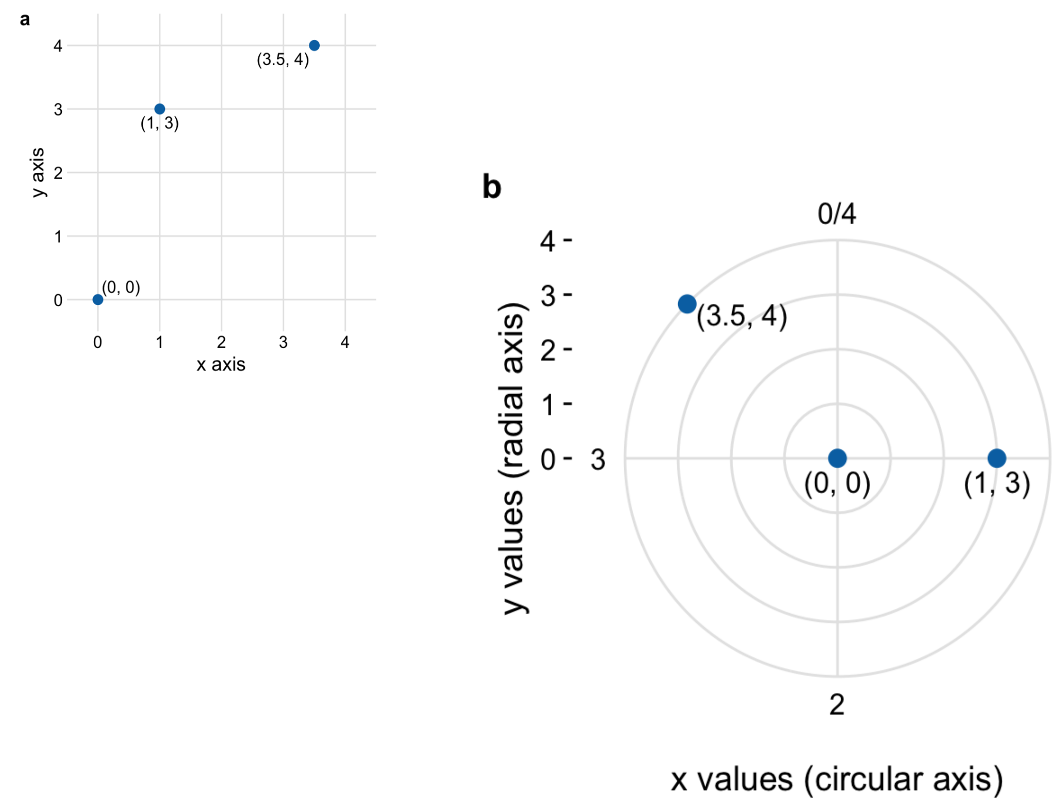
Areas of Northeastern U.S. states. (a) Areas shown on a linear scale. (b) Areas shown on a square-root scale. Data source: Google.

Curved axes

Curved Axes

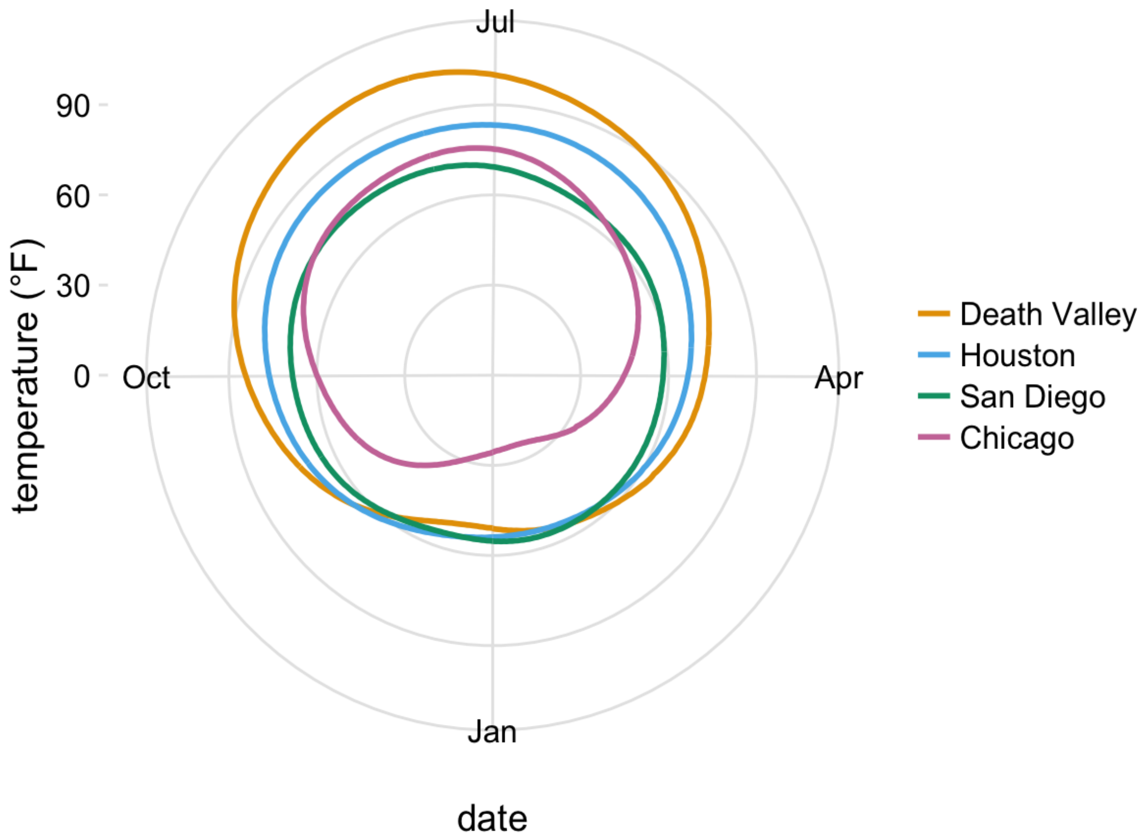
In the *polar* coordinate system, we specify positions via an angle and a radial distance from the origin, and therefore the angle axis is circular.

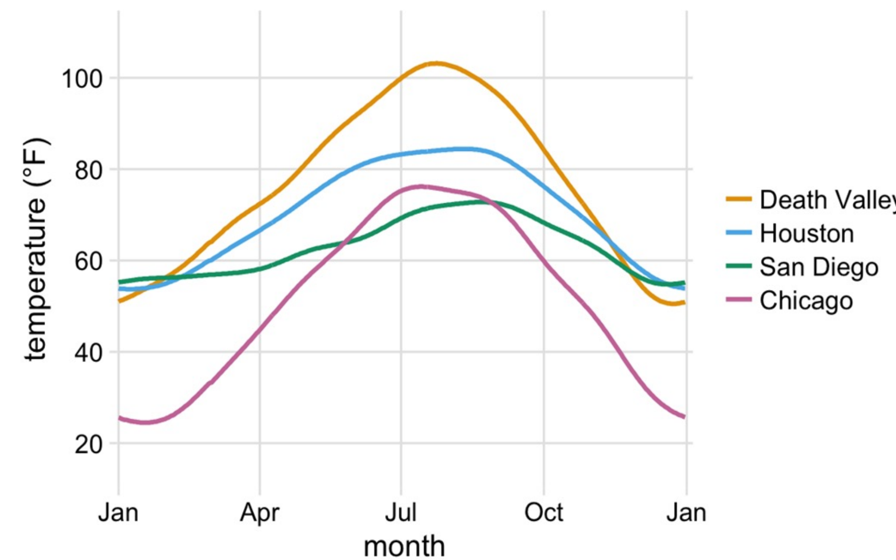




Polar coordinates are most appropriate in any context where the phenomenon being considered is inherently tied to direction and length from a center point in a plane, such as spirals.

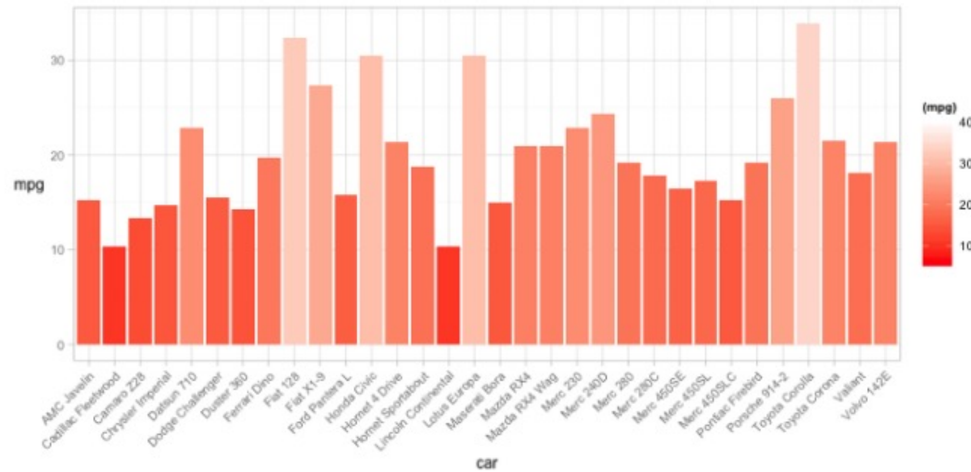
Example



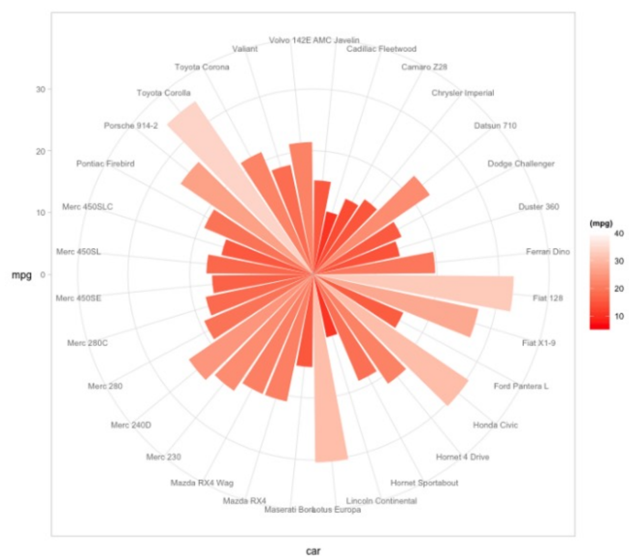


The radial distance from the center point indicates the daily temperature in Farenheit, and the days of the year are arranged counter-clockwise starting with Jan.

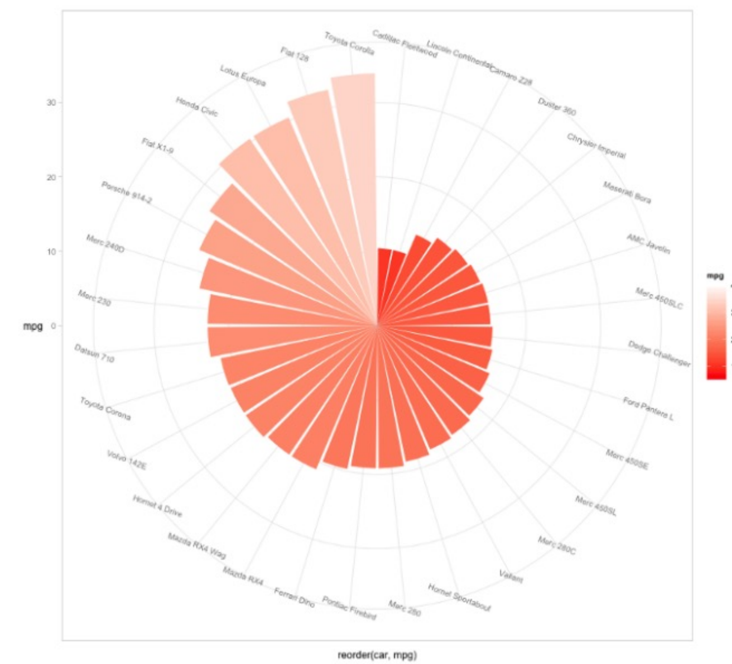
Example



Normal bar plot



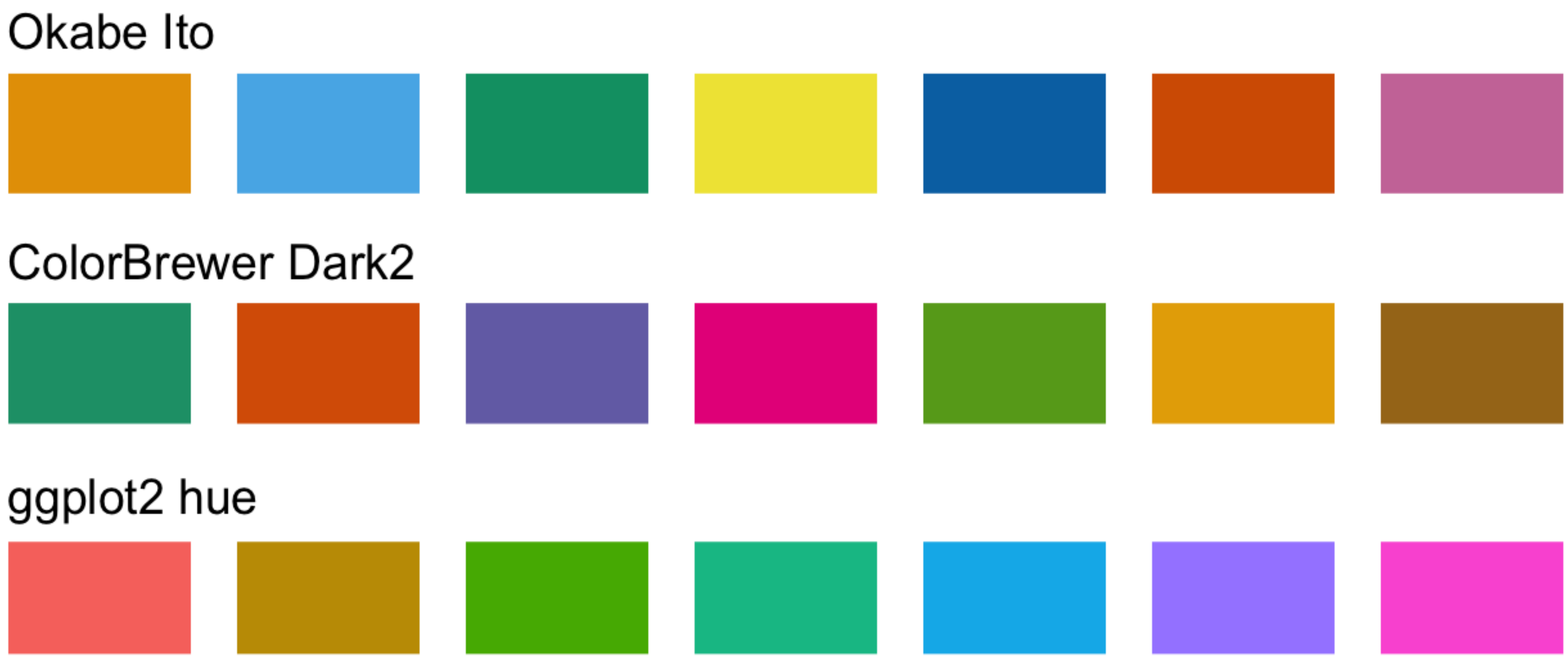
Bars in polar plot



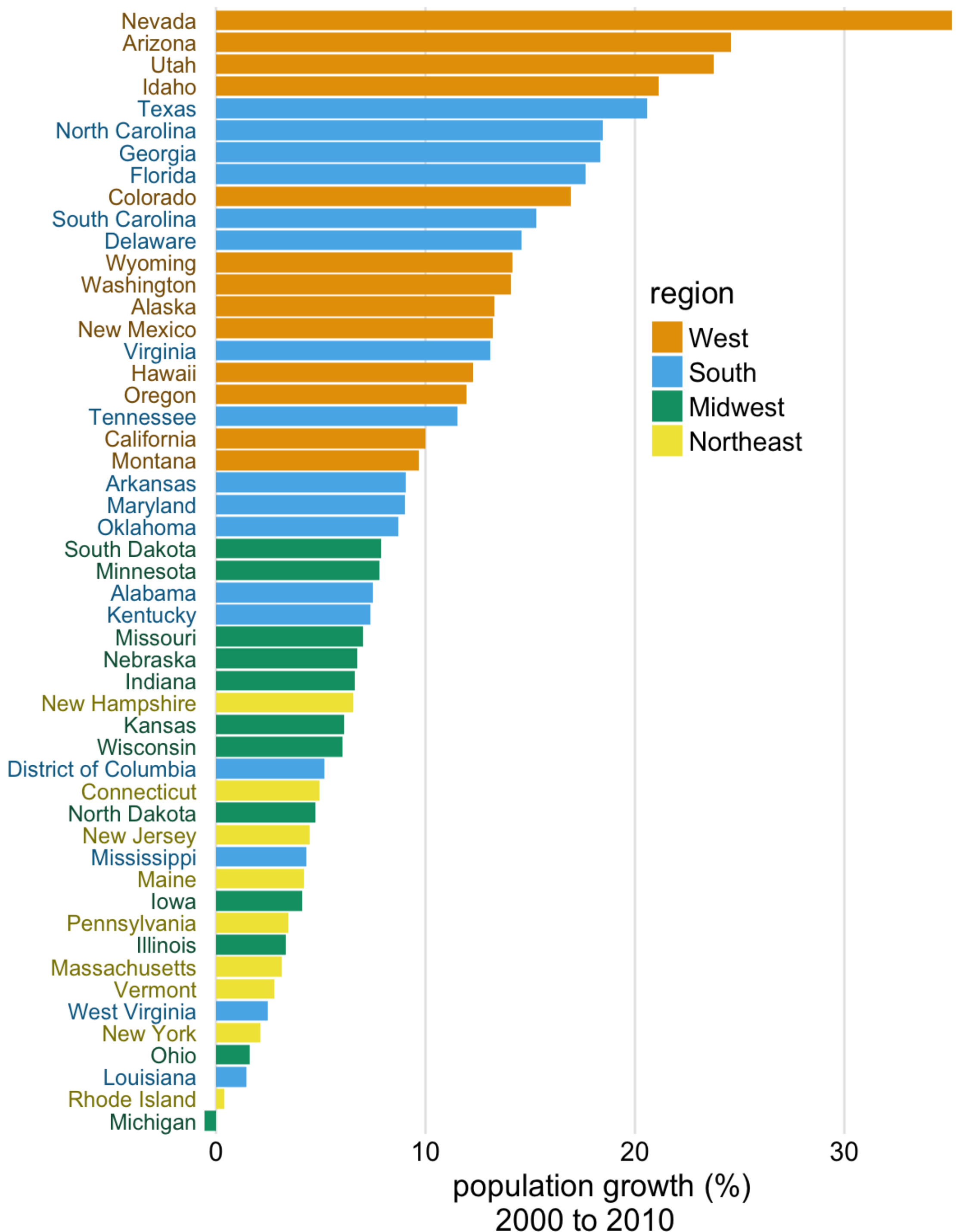
Color scales – Qualitative, Sequential, Divergent, Accent

Color scales

Qualitative color scales: colors to distinguish data values



(Okabe and Ito 2008) (ColorBrewer project -Brewer 2017) (ggplot2)



Example

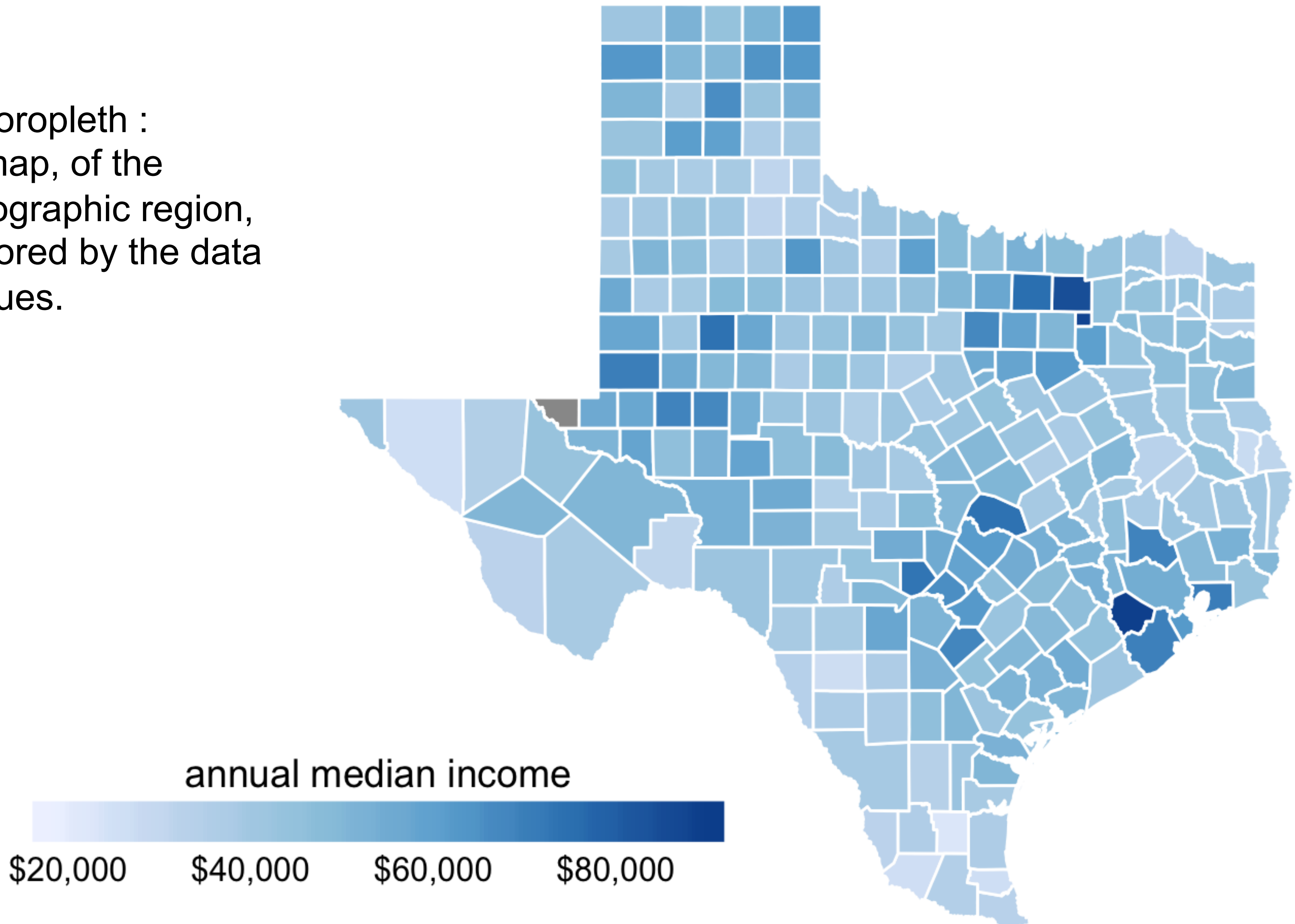
Population growth in the U.S. from 2000 to 2010. States in the West and South have seen the largest increases, whereas states in the Midwest and Northeast have seen much smaller increases or even, in the case of Michigan, a decrease.

Color scales

Sequential color scale : (i) ordering (ii) distance between two specific values.

Example

Choropleth :
a map, of the geographic region, colored by the data values.

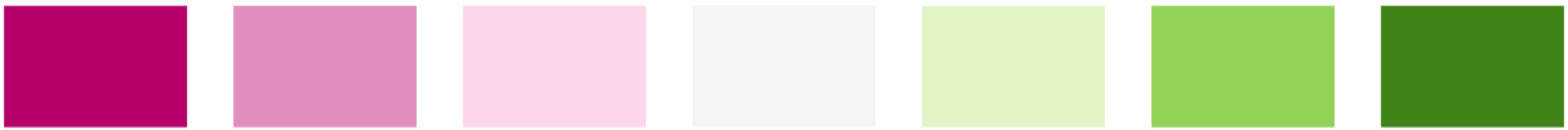


Color scales

Diverging color scale : deviation of data values relative to a neutral midpoint.

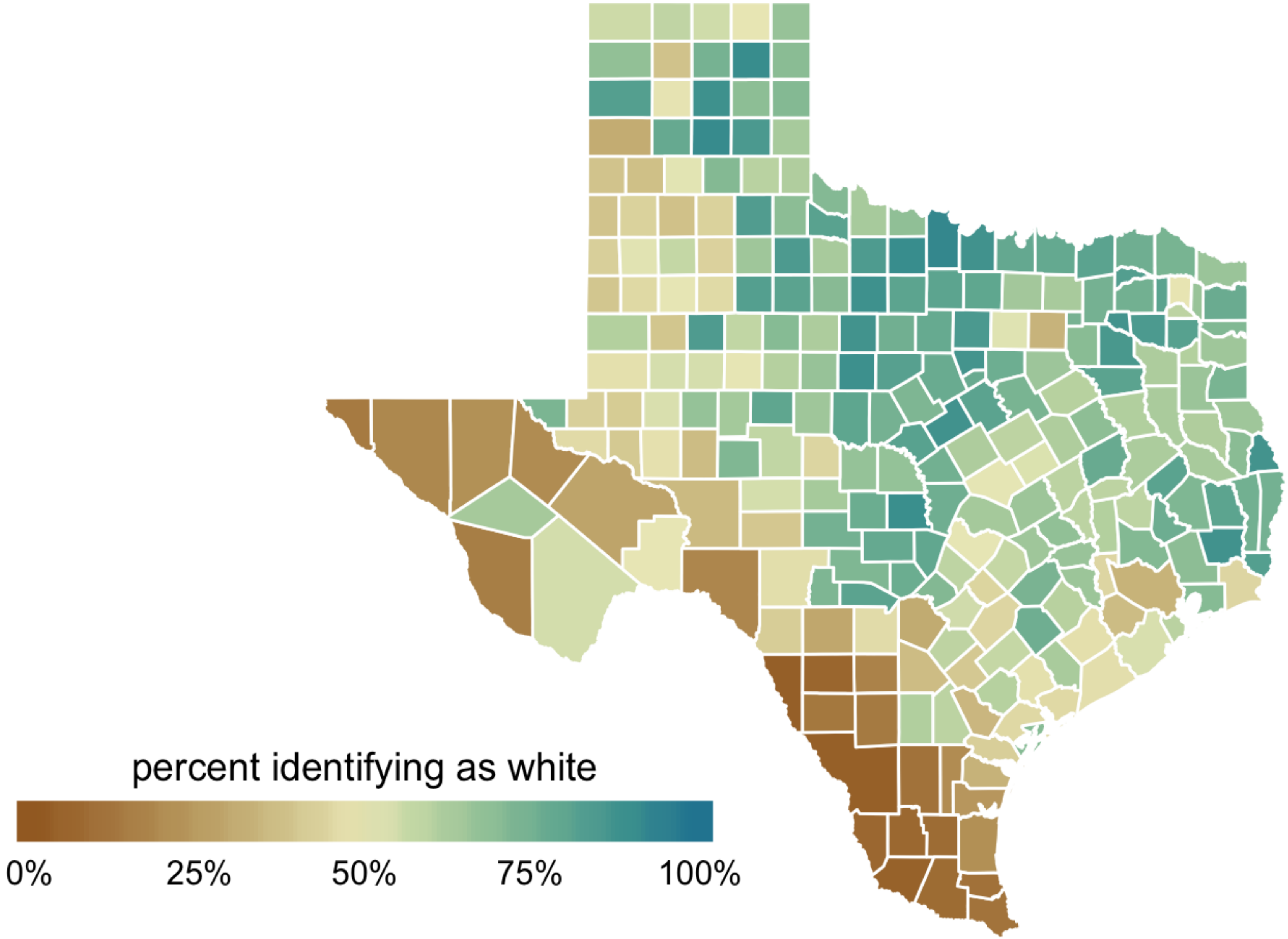
CARTO Earth

ColorBrewer PiYG



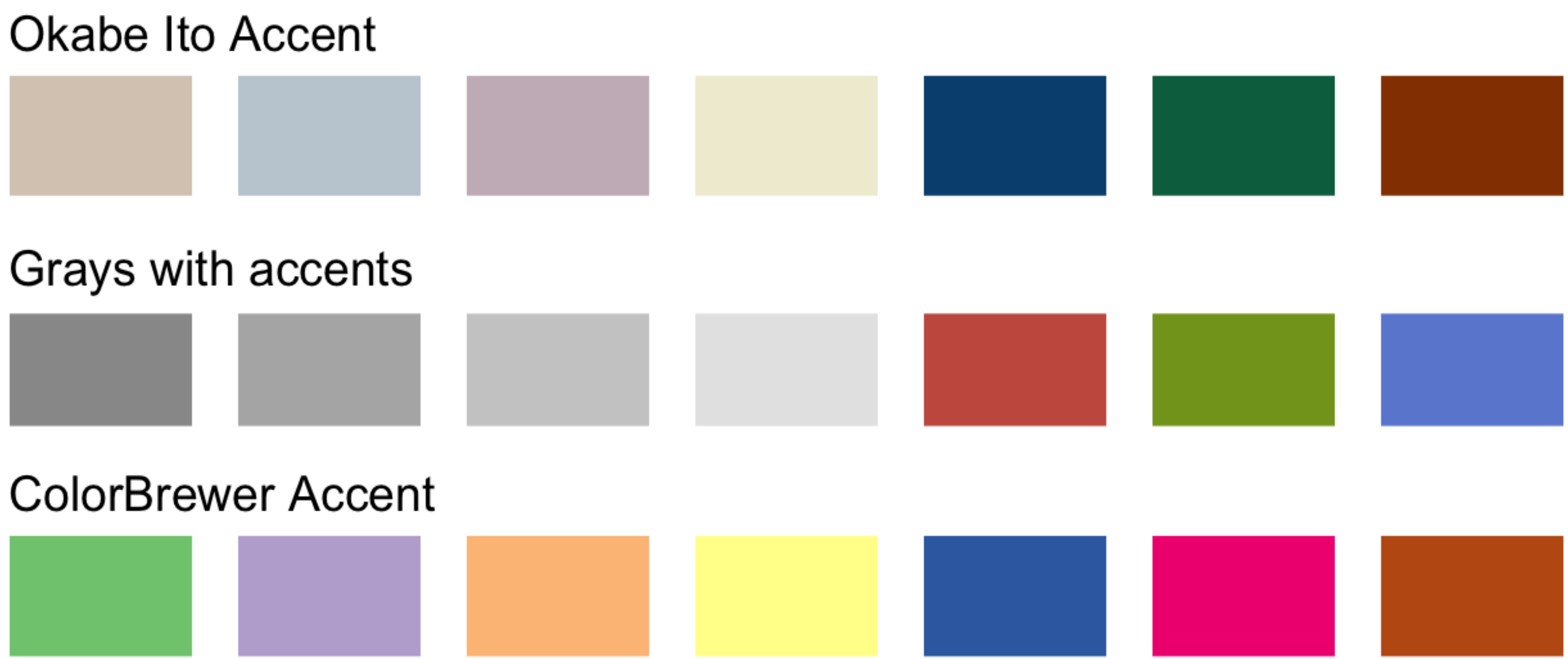
Blue-Red

Example

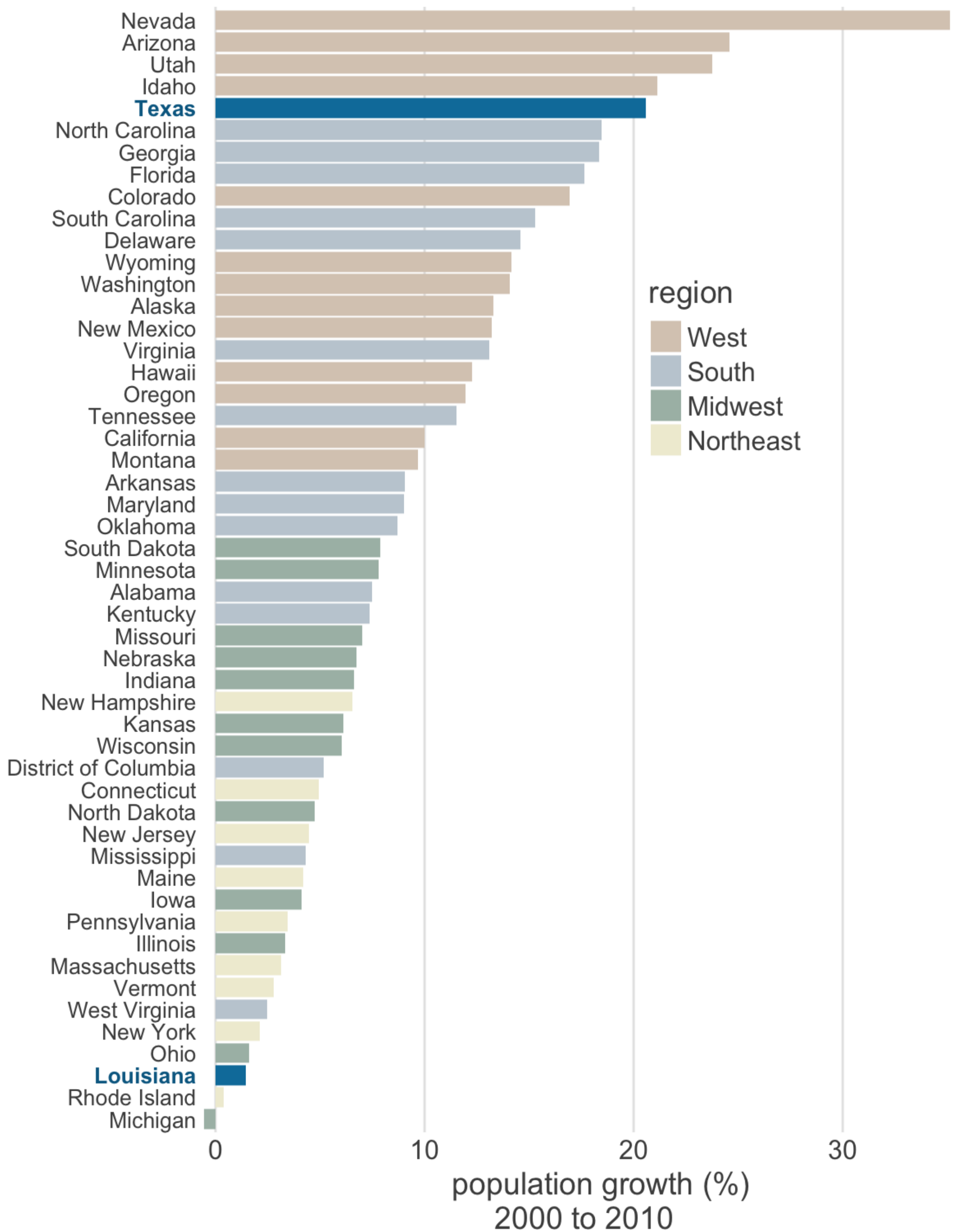


Color scales

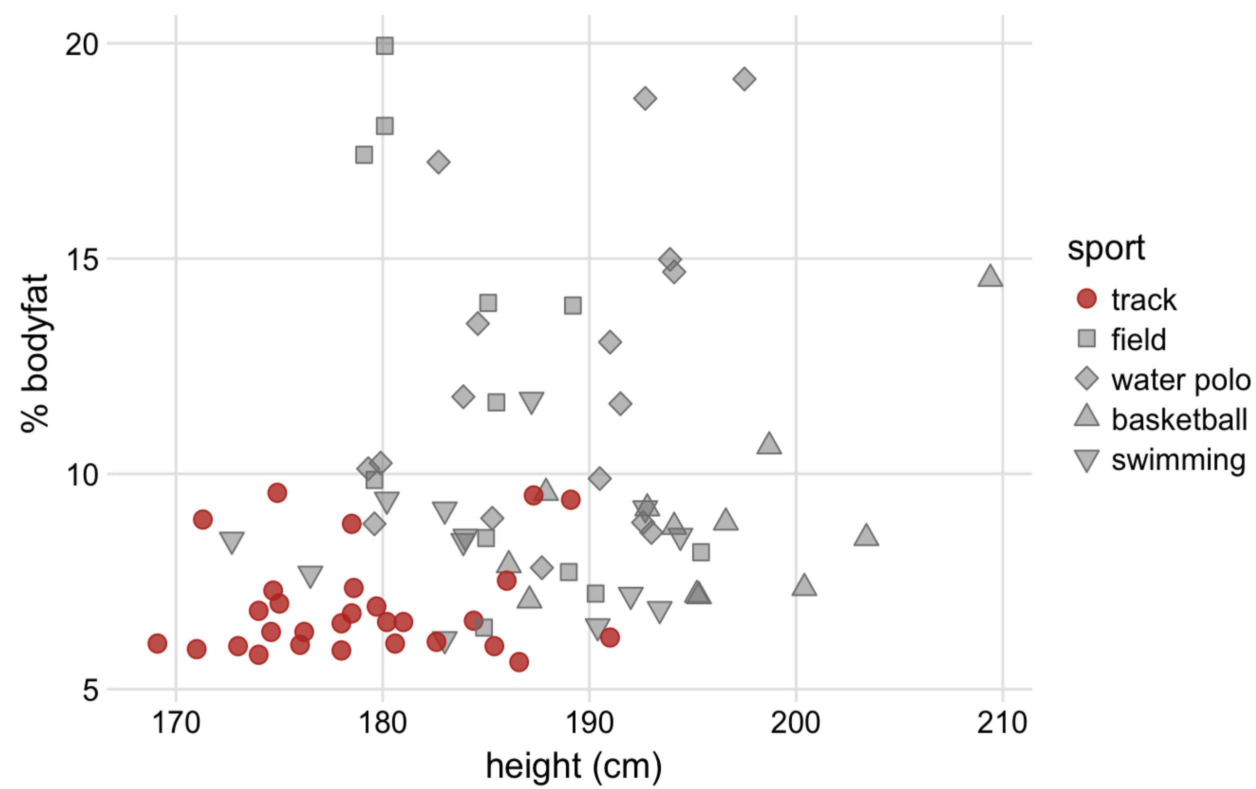
Accent color scales : contain both a set of subdued colors and a matching set of stronger, darker, and/or more saturated colors; eg to show popout



Example



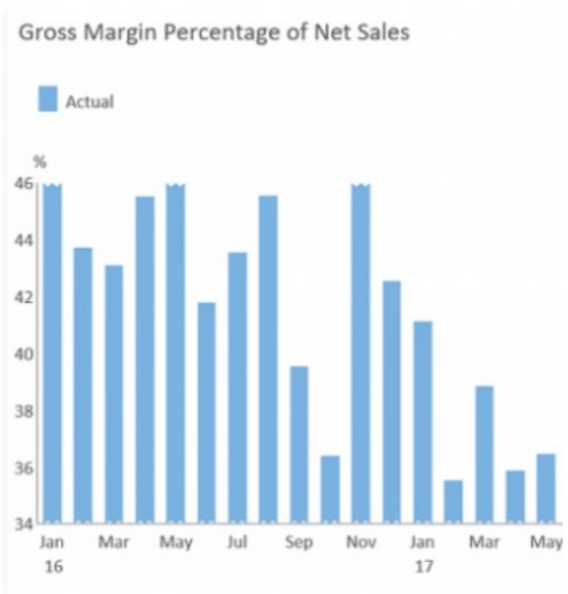
Nevada
Arizona
Utah
Idaho
Texas
North Carolina
Georgia
Florida
Colorado
South Carolina
Delaware
Wyoming
Washington
Alaska
New Mexico
Virginia
Hawaii
Oregon
Tennessee
California
Montana
Arkansas
Maryland
Oklahoma
South Dakota
Minnesota
Alabama
Kentucky
Missouri
Nebraska
Indiana
New Hampshire
Kansas
Wisconsin
District of Columbia
Connecticut
North Dakota
New Jersey
Mississippi
Maine
Iowa
Pennsylvania
Illinois
Massachusetts
Vermont
West Virginia
New York
Ohio
Louisiana
Rhode Island
Michigan
0
10
20
30
population growth (%)
2000 to 2010
region
West
South
Midwest
Northeast

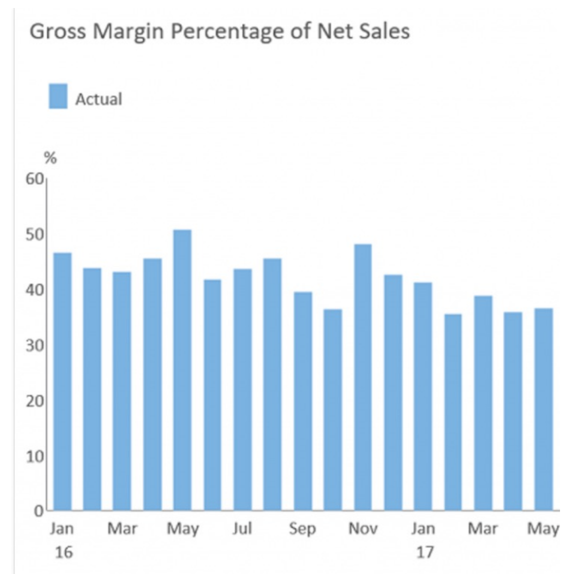


% bodyfat
20
15
10
5
170
180
190
200
210
height (cm)
sport
track
field
water polo
basketball
swimming

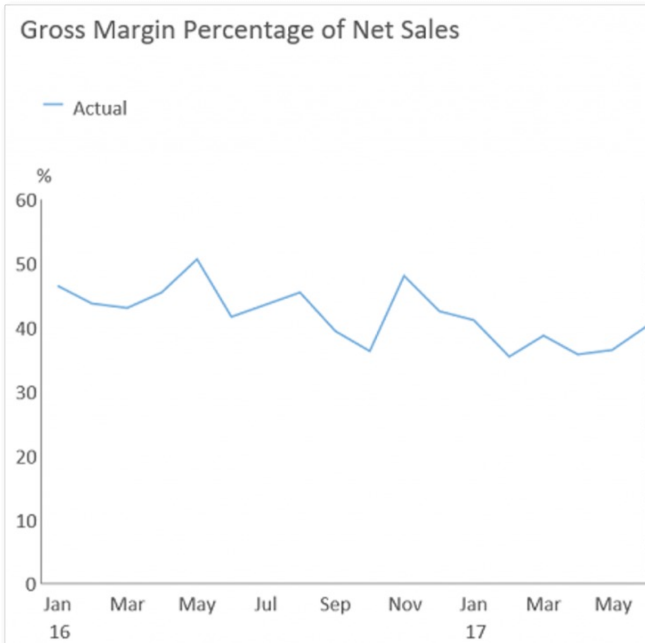
Key points

- Choosing the scale for a graph
    - Common design constraints or choices to make
        - -Zero(or any other particular value) to be included in a scale
        - -The range of the tick marks on each scale
        - -Dual axes





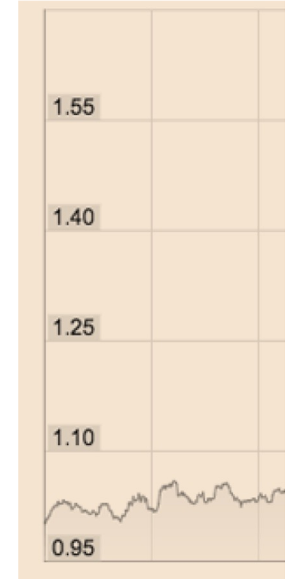
Reminder : Best practices

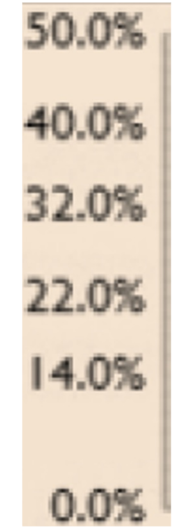


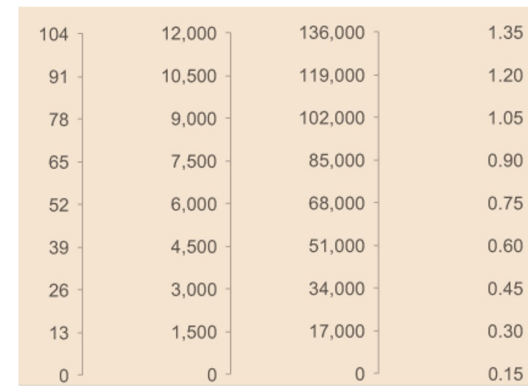
Remidner: Best practices

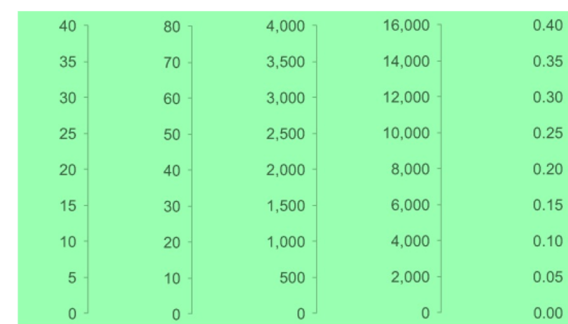
- All intervals on the scale should be equal
  - Avoid scales that make perceiving values in the graph can be harder than it should be.

- Good scale intervals : a power of 10 or a power of 10 multiplied by 2 or 5
  - Exceptions to above rule : Months (3 or 12), seconds (60), RAM in Gigabytes (4 or 16) etc.
    - A scale of months of 0, 5, 10, 15, 20 would be less cognitively fluent than 0, 3, 6, 9, 12, for months/year data









References

https://dfrieds.com/data-visualizations/when-use-log-scale.html

https://www.dataclaritycorp.com/the-power-of-logarithmic-scale/

http://colorbrewer2.org
Matthew Ward et al: Interactive Data Visualization

http://urbaninstitute.github.io/graphics-styleguide/
This site contains guidelines that are in line with data visualization best practices and proven design principles.