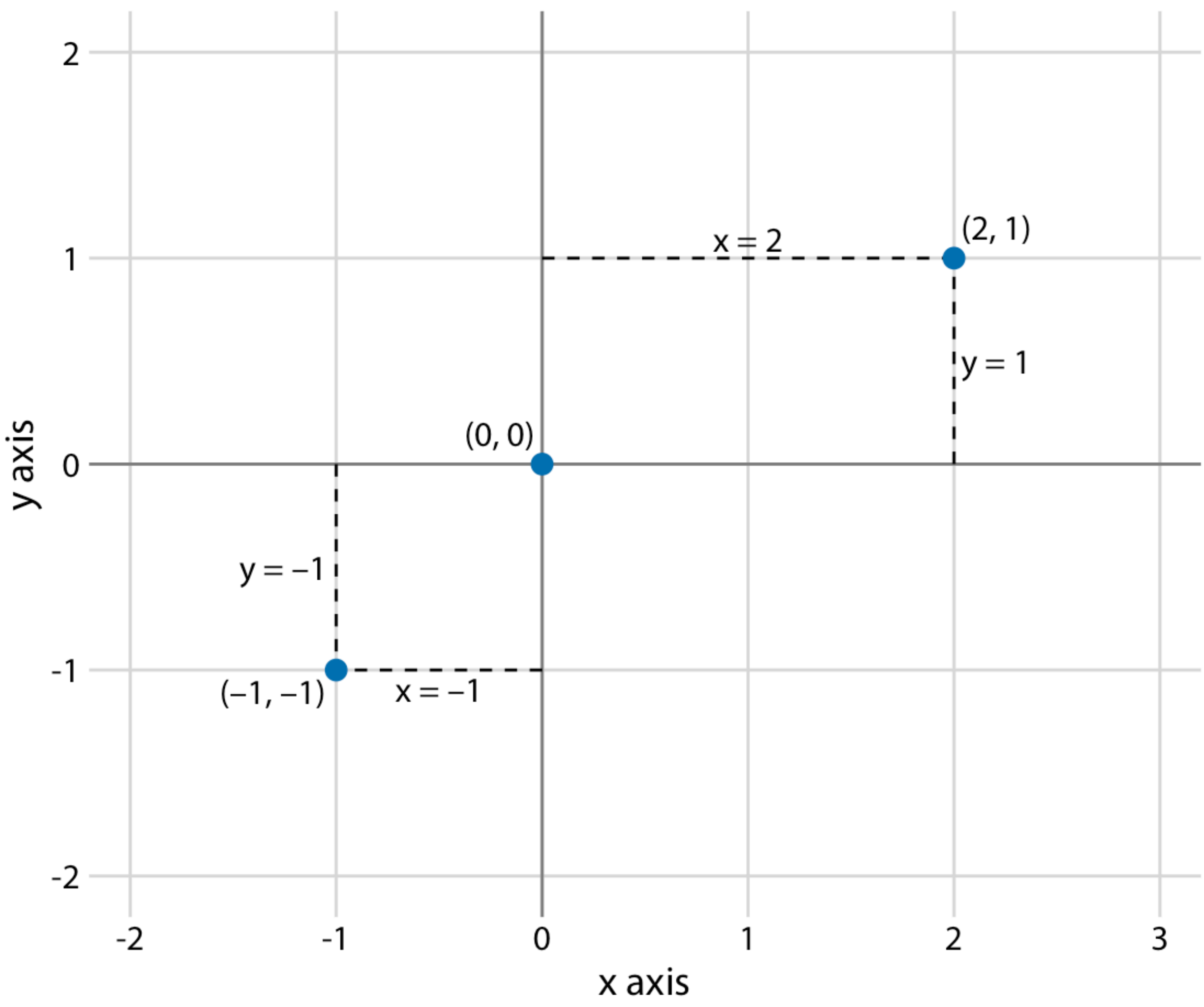
Coordinate systems and axes -

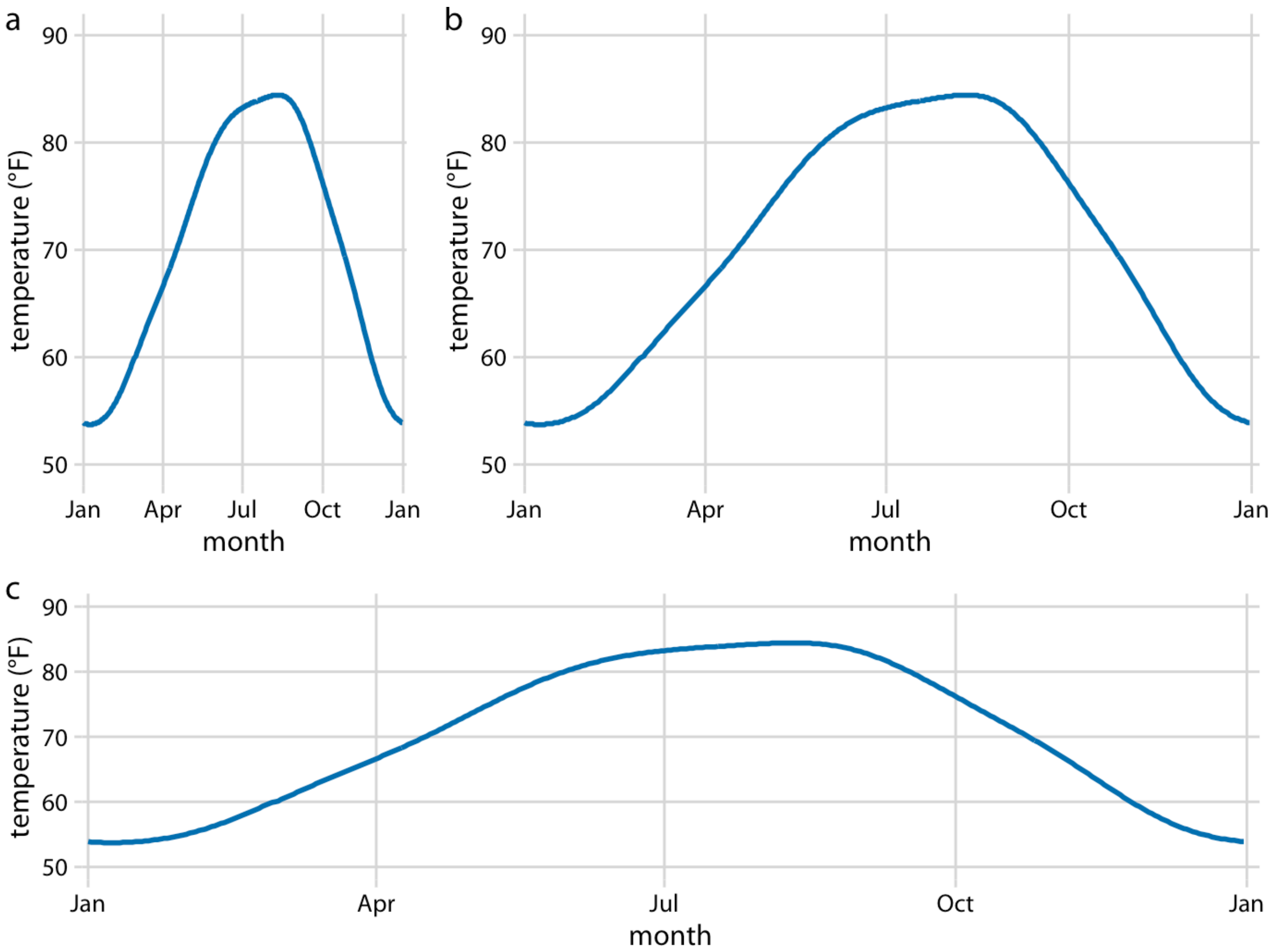
to make any sort of data visualization, we need to define position scale, which determine where in a graphic different data values are located. the combination of a set of position scales and their relative arrangement is called co-ordinate system. in 2d two numbers are required to uniquely specific point these two scales are usually but not necessarily the x and y axis. we also have specify relative geometric arrangement of these scales. Conventionally, the x-axis runs horizontally and the y-axis vertically, but we could choose other arrangments. for example we could have the y-axis at an acute angle relative to the x-axis.

Cartesian Coordinates -

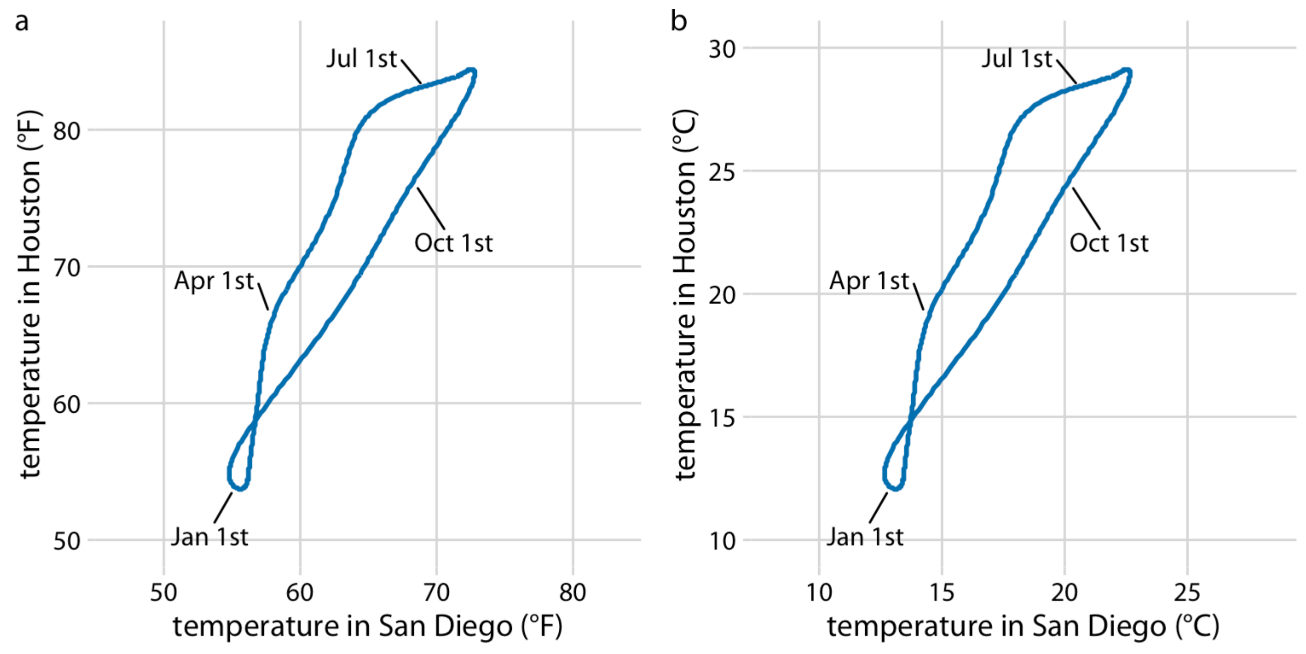
the most widely used coordinate system for data visualization is the 2d Cartesian coordinate system where each value uniquely specified by an x and a y value. the axes are continuous position scale and they can represent both positive and negative real numbers.



we plotted temperature vs. days of the year. the y-axis of the figure is measured in degrees Fahrenheit(a grid line equal 20 degrees), and the x axis is measured in months(a grid line is equal to every third month). whenever the two axes are measured in different units we can stretch or compress one relative to the other and maintain a valid visualization of data.



on the other hand, if the x and the y axis are measured in the same units, then the grid spacing for two axes should be equal, such that the same distance along x or y axis corresponds to the same number of data units. for example we can plot temperature in houston, Texas against the temperature in san Diego, CA for every day of the year since the same quantity is plotted along both axes, we need to make sure that the grid lines forms perfect square.



Nonlinear axes -

1. logarithmic scale or log scale

log scales are linear in multiplication, such that a unit on the scale corresponds to multiplication with a fixed value. multiplication on a log scale looks like addition on a linear scale.

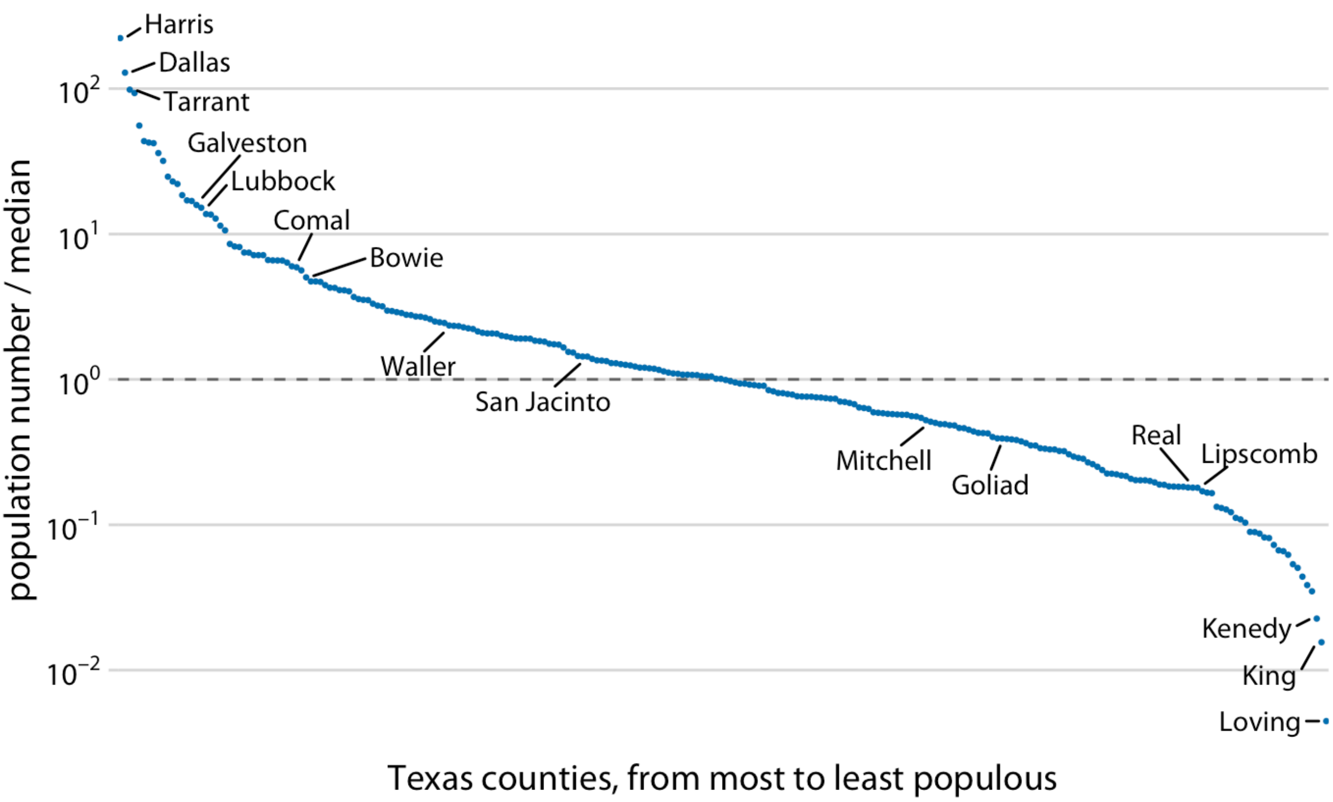
To create a log scale, we need to log-transform the data values while exponentiating the numbers that are shown along the axis grid lines.
if we are converting data values in log then we need to title the axes with log(data), if we transforming the scale of axis then we need not to title the axis with log(data) because we can see it like we are giving log spacing in axis.

mathematically there is no difference between plotting the log-transformed data on linear scale or the original data on a logarithmic scale. the only difference lies between labelling for the individual axes.

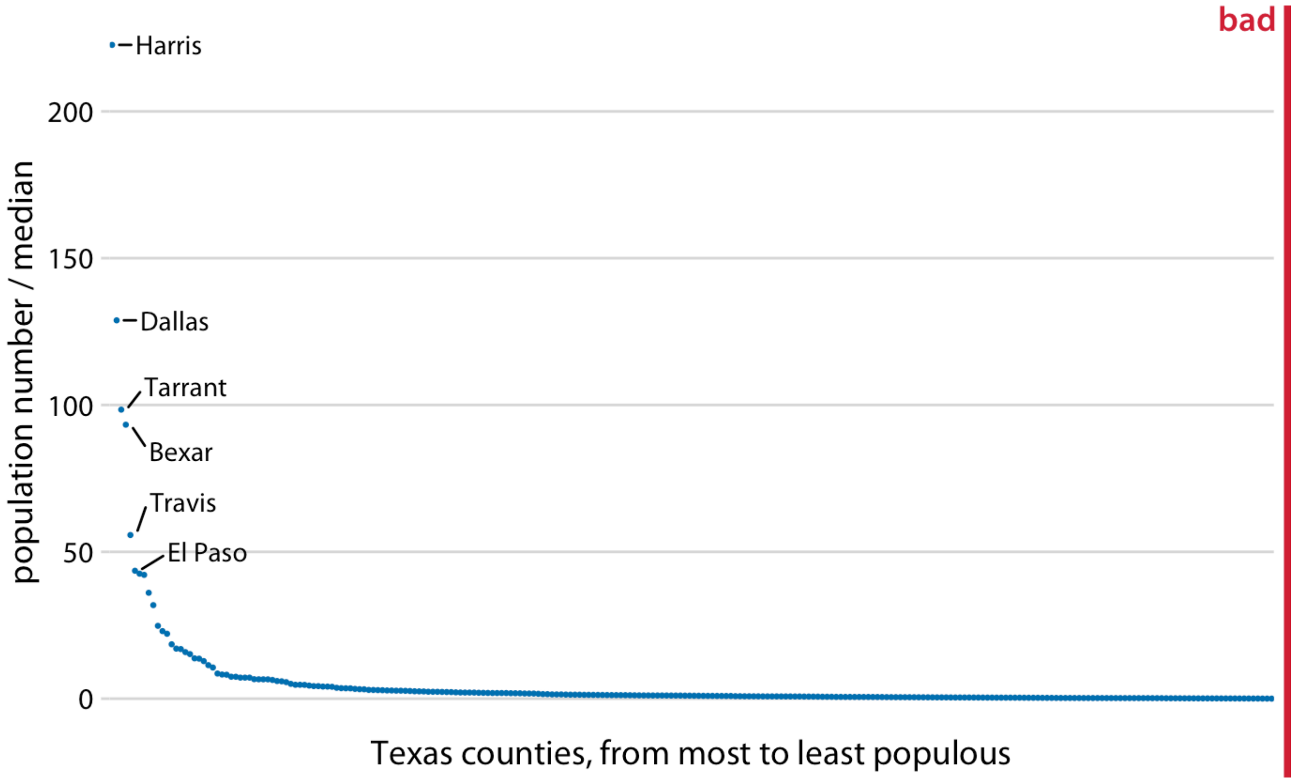
note to improve visualization-
in most cases, the labelling for a logarithmic scale is preferable, because it places less mental burden on the reader to interpret the numbers. when working with log-transformed data, we can get confused about

whether the data were transformed using the natural logarithm or the logarithm to base 10. "log(x)" doesn't specify a base at all. i recommend that you always verify the base with log-transformed data, always specify the base in the labelling of the axis. log scale are the natural choice for any data that have been obtained by multiplication or division.

eg. we have taken the number of inhabitants in each county in Texas and divided it by the median number of inhabitants. the resulting ratio is a number that can be larger or smaller then 1. A ratio of exactly 1 implies that the corresponding county has the median number of inhabitants. when visualizing these ratio on log scale we can see that population in Texas counties are symmetrically distributed around the median, and the most populous counties have over 100 times more inhabitants then median, and least populous have over 100 times fewer inhabitants. this visualization clearly show that log-transformed data looks like multiplication on linear scale. log10(1) = 0 so median is on mid point.



important note -
log scale are always start with 1 because negative number and 0 is not defined in log().
log scales are frequently used when the data set contains numbers of very different magnitudes or difference between data values is very large.

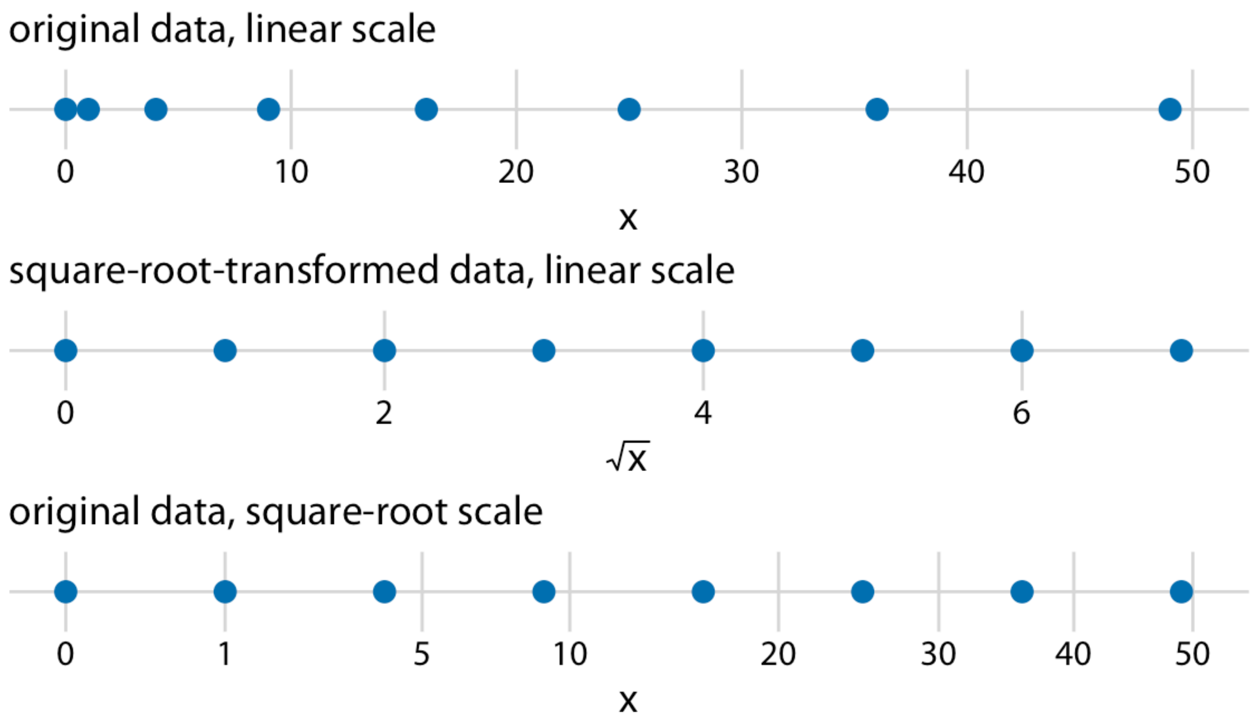


on a log scale, the value 1 is the natural midpoint, similar to the value 0 in linear scale.
by displaying population/median ratio on a linear scale we have overemphasised ratio > 1 and have obscured ratio < 1 and by plotting this ratio on a linear scale we can see that we are not able to make distinguish in between least populous counties they are making a single line. As a general rule, ratios should not be displayed on a linear scale.

2. Square-root scale -

what if we wants to see a country with 0 inhabitants? this county will not be see log scale because its log value is infinite for overcome this condition square-root scale is use. this uses square-root transformation just like a log transformation.

A square-root scale compresses larger number into a smaller range but not like log scale. it allows the presence of 0.

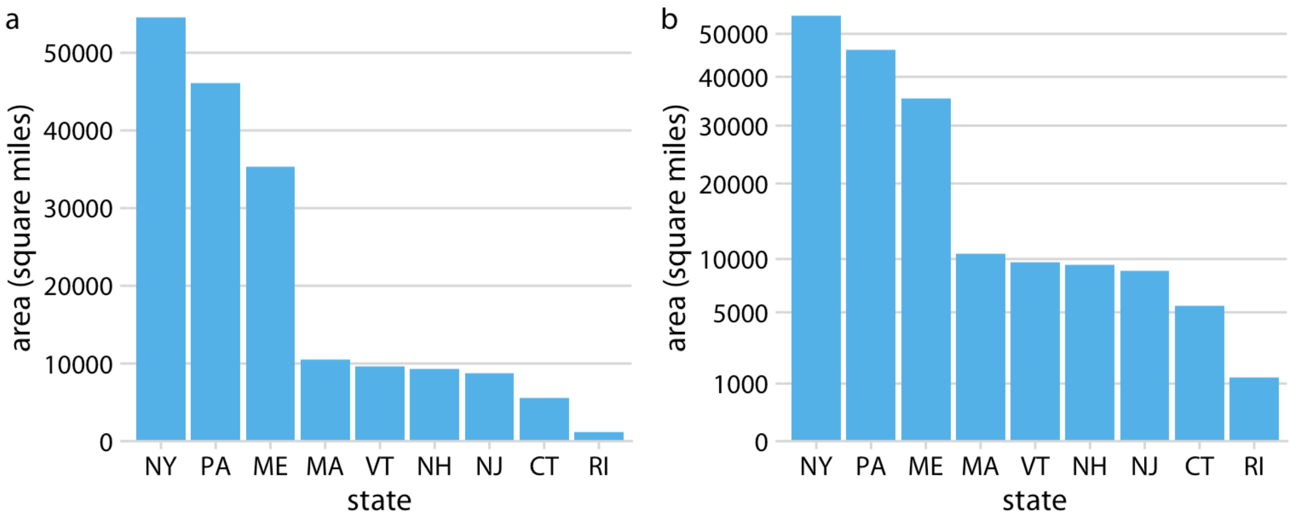


problem with root scale -

1. On a linear scale one unit step corresponds to addition or subtraction of and a log scale corresponds to multiplication or division by a constant value, no such rule exist for a square-root scale. The meaning of a one unit square-root scale depends on the scale value at which we are starting.
2. it is unclear how to best place axis ticks on a square-root scale. for example positions 0, 4, 9, 25, 36, 49, 81 would be highly unintuitive. alternative we could place them at linear intervals 10, 20, 30, 40, 50, 60, 70, 80, 90 but this would result in either too few ticks near the low end of the scale or two many near the high end for overcome this condition you have to manually specify the ticks.

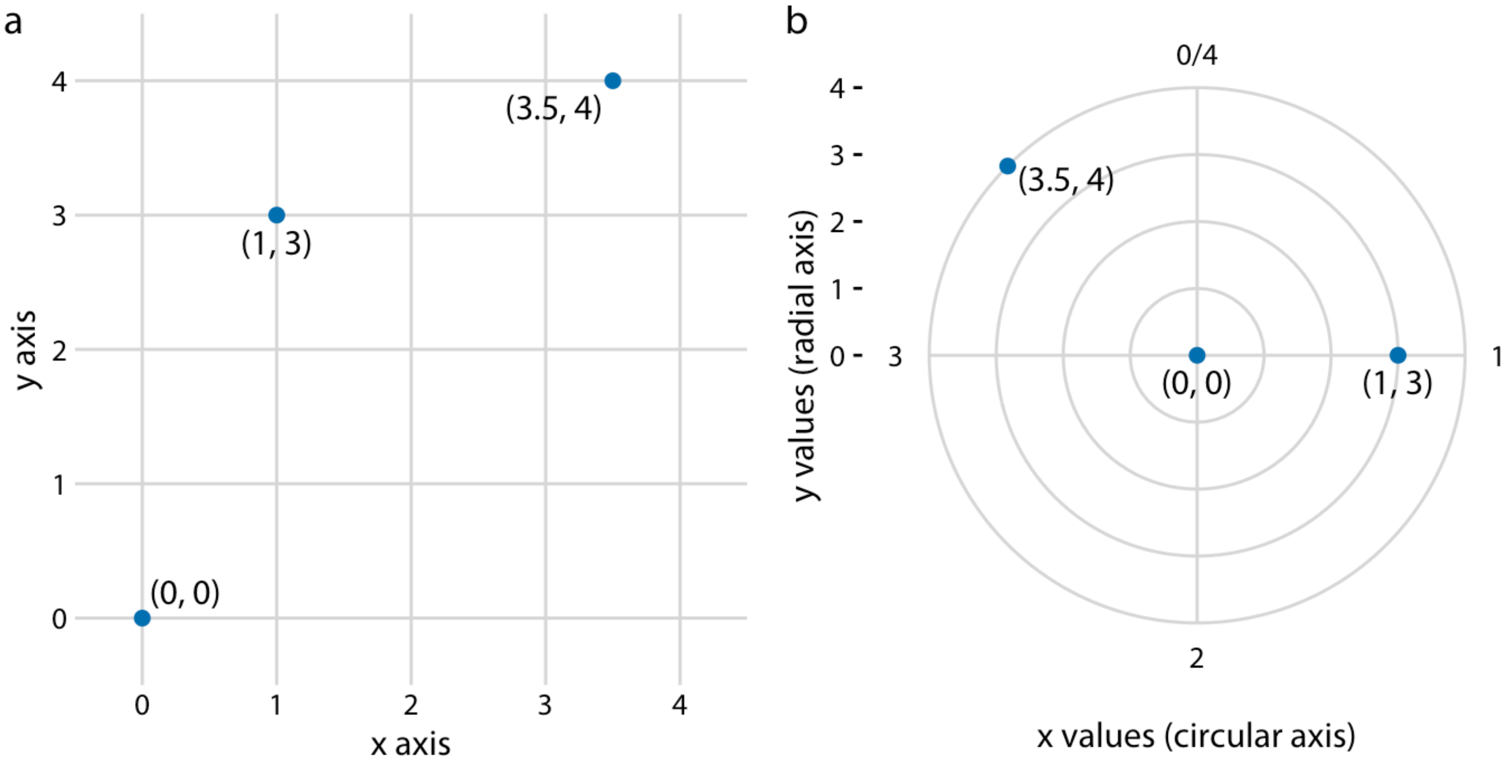
log scale is the natural scale for ratios, and the square-root scale is the natural scale for data that come in squares.

eg. in one scenario in which data are naturally squares is geographic regions. if we show the ares of geographic regions on a square-root scale, we are highlighting the regions. figure shown below shows the areas of states in the U.S. Northeast on both a linear and a square-root scale. even though the areas of these states are quite different, the time it will take to drive across each time will more closely resemble the figure on the square-root scale.

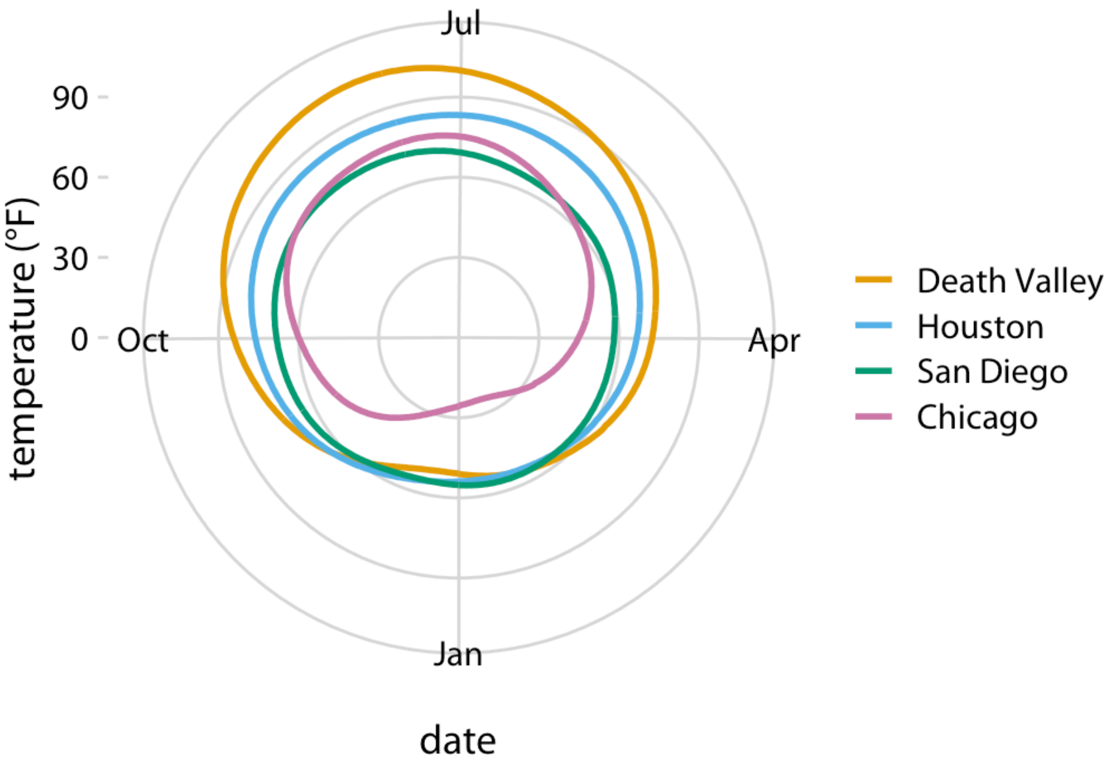


Coordinate system with Curved axis -

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

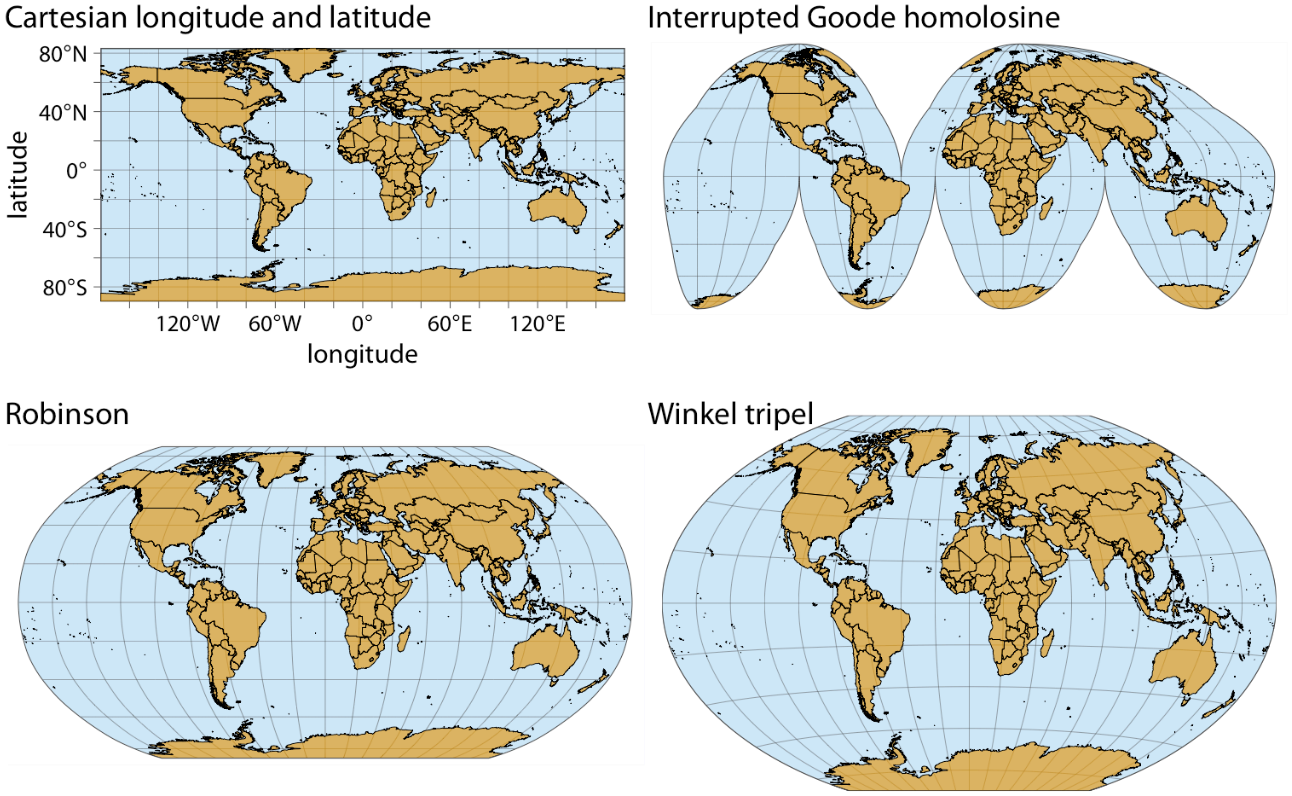


Polar coordinates can be useful for data of a periodic nature. for example consider the days in a year. December 31st is the last day of the year, but it is also one day before the first day of the year. if we want to show how some quantity varies over the year, it can be appropriate to use polar coordinates with the angle coordinate specifying each day and temperature on radial axis. By plotting the temperature normals in polar coordinate system, we emphasise the cyclical property they have.the polar version of the graph highlights how similar the temperature are in Death vally, Houston as San Diego from late fall to early spring. in the Cartesian coordinate system, this fact is obscured because December and January are in opposite direction.



In the figure he radial distance from the centre is representing the temperature and the days of the year are arranged on angle axes.

A second condition we can use curved axes in context of geospatial data i.e. maps. Location on the globe are specified by their longitude and latitude, because earth is sphere drawing latitude and longitude as Cartesian is misleading.


Cartesian longitude and latitude
80°N
40°N
0°
40°S
80°S
latitude
120°W
60°W
0°
60°E
120°E
longitude
Interrupted Goode homolosine
Robinson
Winkel tripel