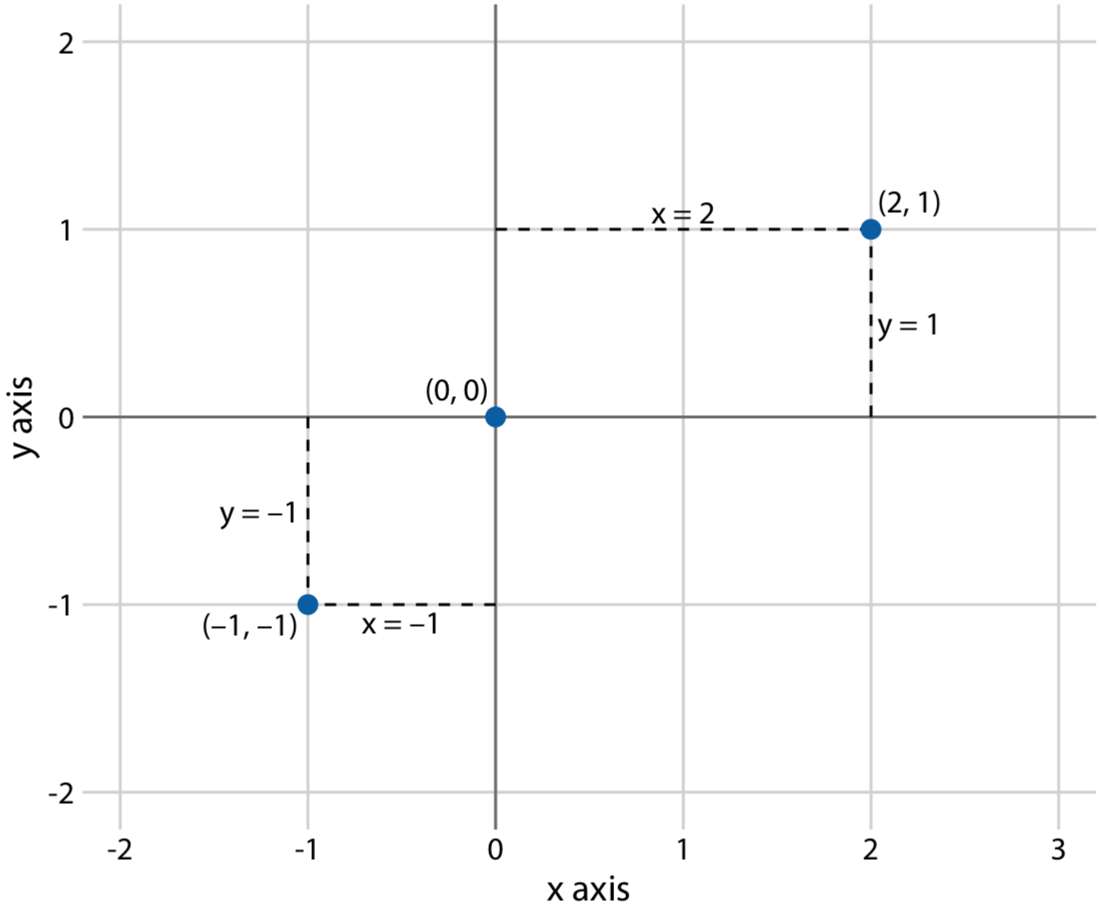
SESSION 3

Coordinate systems and axes

To make any sort of data visualization, we need to define position scales, which determine where in a graphic different data values are located. We cannot visualize data without placing different data points at different locations, even if we just arrange them next to each other along a line. For regular 2d visualizations, two numbers are required to uniquely specify a point, and therefore we need two position scales. These two scales are usually but not necessarily the *x* and *y* axis of the plot. We also have to specify the relative geometric arrangement of these scales. Conventionally, the *x* axis runs horizontally and the *y* axis vertically, but we could choose other arrangements. For example, we could have the *y* axis run at an acute angle relative to the *x* axis, or we could have one axis run in a circle and the other run radially. The combination of a set of position scales and their relative geometric arrangement is called a *coordinate system.*

Cartesian coordinates

The most widely used coordinate system for data visualization is the 2d *Cartesian coordinate system*, where each location is uniquely specified by an *x* and a *y* value. The *x* and *y* axes run orthogonally to each other, and data values are placed in an even spacing along both axes (Figure 3.1). The two axes are continuous position scales, and they can represent both positive and negative real numbers. To fully specify the coordinate system, we need to specify the range of numbers each axis covers. In Figure 3.1, the *x* axis runs from -2.2 to 3.2 and the *y* axis runs from -2.2 to 2.2. Any data values between these axis limits are placed at the respective location in the plot. Any data values outside the axis limits are discarded.



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.

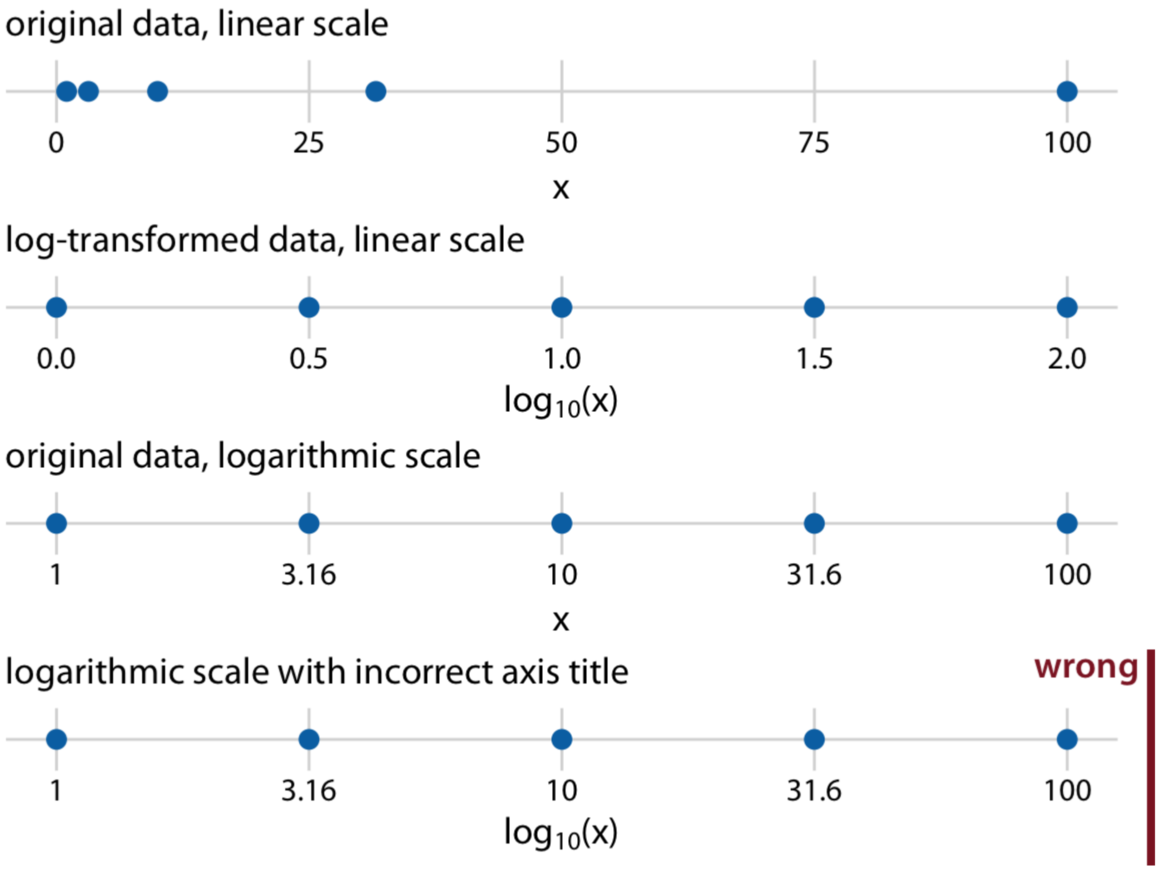
Nonlinear axes

In a Cartesian coordinate system, the grid lines along an axis are spaced evenly both in data units and in the resulting visualization. We refer to the position scales in these coordinate systems as *linear*. While linear scales generally provide an accurate representation of the data, there are scenarios where nonlinear scales are preferred. In a nonlinear scale, even spacing in data units corresponds to uneven spacing in the visualization, or conversely even spacing in the visualization corresponds to uneven spacing in data units.

The most commonly used nonlinear scale is the *logarithmic scale* or *log scale* for short. Log scales are linear in multiplication, such that a unit step on the scale corresponds to multiplication with a fixed value. 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. This process is demonstrated in Figure, which shows the numbers 1, 3.16, 10, 31.6, and 100 placed on linear and log scales. The numbers 3.16 and 31.6 may seem a strange choice, but they were chosen because they are exactly half-way between 1 and 10 and between 10 and 100 on a log scale. We can see this by observing that $10^{0.5}=\sqrt{\boxed{10}}\approx 3.16100.5=10\approx 3.16$ and

equivalently $3.16\times 3.16\approx 103.16\times 3.16\approx 10$.

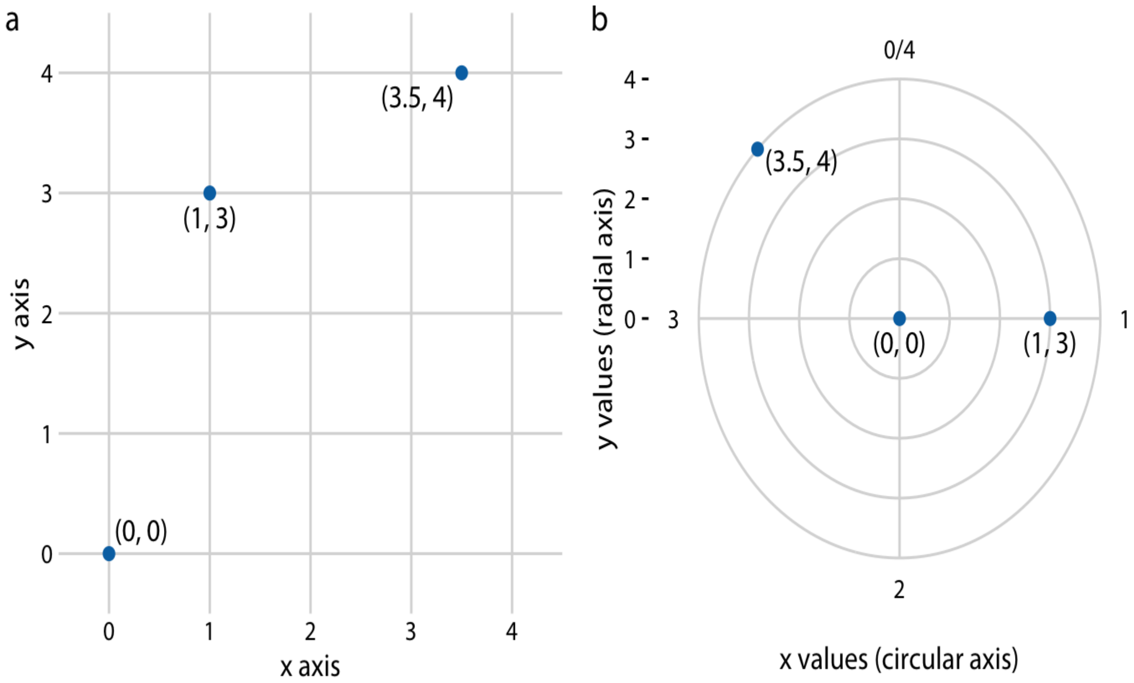
Similarly, $10^{1.5}=10\times 10^{0.5}\approx 31.6101.5=10\times 100.5\approx 31.6$.



Relationship between linear and logarithmic scales. The dots correspond to data values 1, 3.16, 10, 31.6, 100, which are evenly-spaced numbers on a logarithmic scale. We can display these data points on a linear scale, we can log-transform them and then show on a linear scale, or we can show them on a logarithmic scale. Importantly, the correct axis title for a logarithmic scale is the name of the variable shown, not the logarithm of that variable.

Coordinate systems with curved axes

All coordinate systems we have encountered so far used two straight axes positioned at a right angle to each other, even if the axes themselves established a non-linear mapping from data values to positions. There are other coordinate systems, however, where the axes themselves are curved. In particular, 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.



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.