3.1 Scale *(Jonathan W. Chipman, December 2020)*

Definition and expression of scale

Scale is among the most important concepts in the spatial sciences, and an understanding of scale is necessary for the design and use of maps and other geovisualizations. Formally, **scale** is defined as *the ratio of the distance on a map to the corresponding distance on the ground*, as shown in equation 3.1.1:

$$s = \frac{d}{D} \quad (3.1.1)$$

where

s = scale

d = map distance

D = ground distance

The scale of other spatial representations, such as remote sensing images or 3D models, can be described similarly, by substituting "image" or "model" in place of the word "map" in the above definition. Likewise, the rest of this section will discuss scale in terms of maps, but much of this material applies to other forms of geovisualization as well.

Scale may be expressed in text as a ratio or **representative fraction (RF)**, e.g., "1:25,000" or "1/25,000", both of which are pronounced as "1 to 25,000", or in **unit equivalents** ("1 cm = 250 m"). All three of these examples mean the same thing: one unit of distance on a map is equal to 25,000 of the same units of distance on the ground. Scale may also be expressed on a map by use of graphical indicators such as a **scale bar** (Figure 3.1.1). When a map is enlarged or shrunk, its scale changes, and digital maps may change their scales depending on the viewing device used. In such cases, a graphical scale such as the bars in Figure 3.1.1 is more reliable than the text scales described above, because the graphical scale will expand or shrink along with the map itself. Design and placement considerations for scale bars are discussed in Chapter 8.

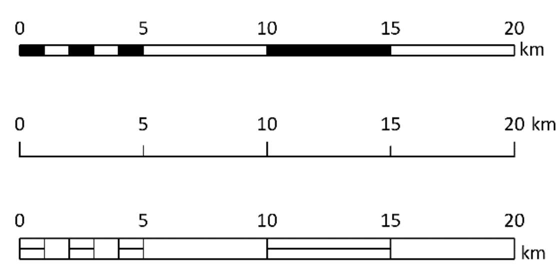

Figure 3.1.1 Examples of scale bars.

Note that the terms "large scale" and "small scale" can be confusing. Because the fraction 1/100 is much larger than the fraction 1/1,000,000, cartographers would describe 1/100 as large scale and 1/1,000,000 as small scale. On a large-scale map, individual features will appear large in size, and the map will thus usually cover a small area. In contrast, a small-scale map would show the same features as very small, and would be able to cover a much larger area. This is illustrated in Figure 3.1.2; a large-scale map of Boston (*[a]*, original scale = 1:4000) covers a small area but shows details down to the level of individual buildings, while a small-scale map (*[b]*, original scale = 1:260,000) covers a larger area with much less detail. This is the opposite of the common usage in most other fields, where "scale" is used as a synonym for "extent", and a "large-scale study" means one that would cover a large area.

Even among cartographers and other spatial scientists, there is no consistent division of the ranges referred to as **large scale**, **medium scale**, and **small scale**. As shown in Table 3.1.1, different authors have used very different thresholds to define these categories. In general, the thresholds for categories of map scales tend to be smaller than the equivalent categories for scales of aerial photographs or other remote sensing images. As methods for the collection of spatially detailed geographic data have proliferated – including the global positioning system (GPS), autonomous or remotely-piloted "drones", and high-resolution satellite imaging – geovisualization at large spatial scales is becoming more common, and the concepts of "small" and "large" scale are thus shifting. In this work, rather than adopt a new or existing set of definitions for these categories, we will simply recognize that different applications and purposes require very different scales, and what is considered "small" or "large" in one context may be labeled differently in another. It is more important to understand the relative concepts of *smaller* and *larger* scale.

Table 3.1.1 Examples of definitions of small, medium, and large scale.

Source	Small scale	Medium scale	Large scale	Subject
Raisz (1962)	1:1 million and smaller	1:1 million to 1:250,000	1:250,000 and larger	Maps
Dent et al. (2009)	1:300,000 and smaller	1:300,000 to 1:30,000	1:30,000 and larger	Maps
Peterson (2015)	1:250,000 and smaller	1:250,000 to 1:50,000	1:50,000 and larger	Maps
Field (2018)	1:500,000 and smaller	1:500,000 to 1:50,000	1:50,000 and larger	Maps
ASP (1952)	1:20,000 and smaller	1:20,000 to 1:10,000	1:10,000 and larger	Remote sensing
Lillesand et al. (2015)	1:50,000 and smaller	1:50,000 to 1:12,000	1:12,000 and smaller	Remote sensing

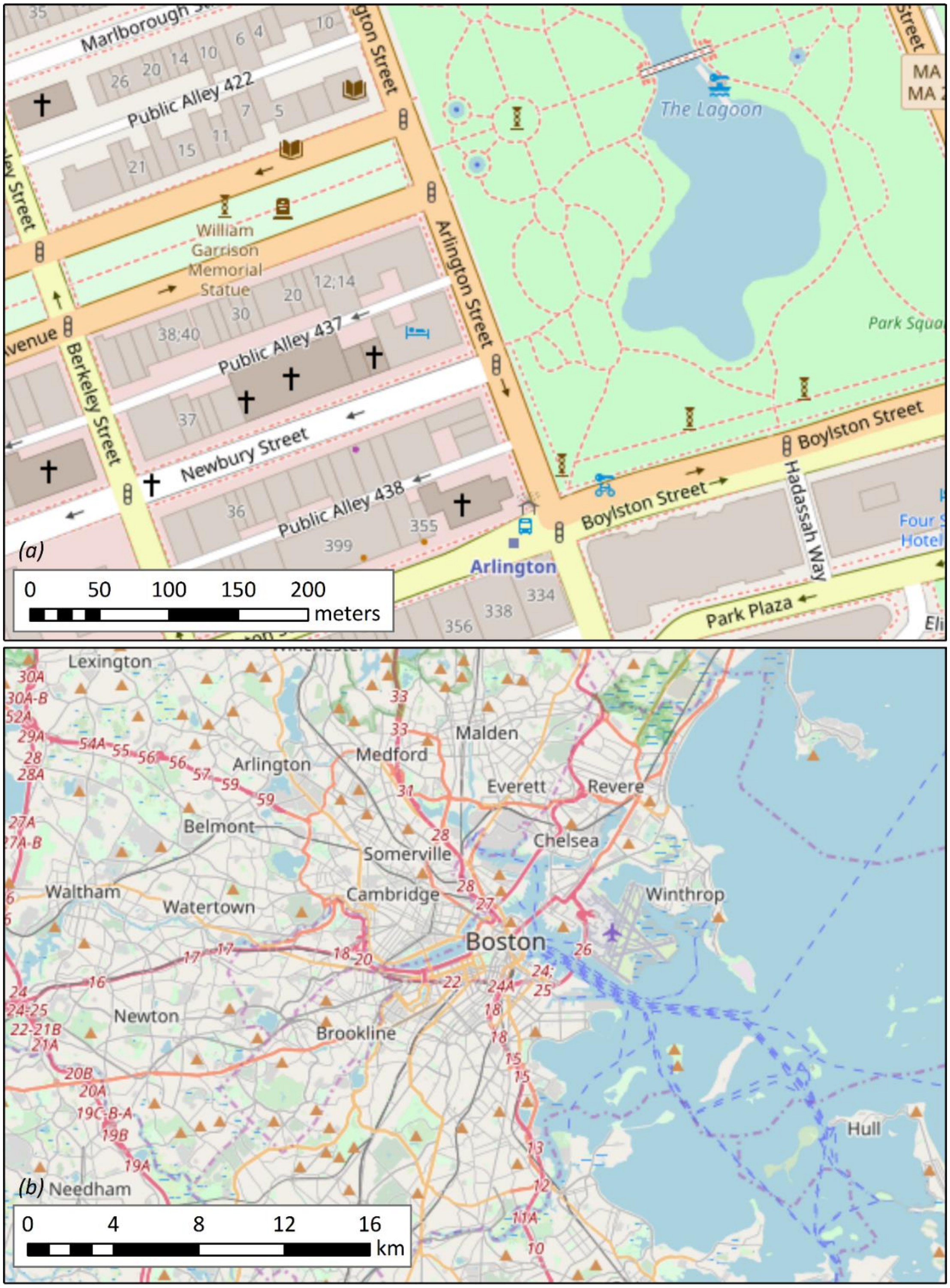

Figure 3.1.2 Maps of Boston, Massachusetts at large-scale (*[a]*, original design scale = 1:4000) and small-scale (*[b]*, original design scale = 1:260,000). Map data © OpenStreetMap contributors, licensed under Creative Commons Attribution-ShareAlike 2.0 license (CC BY-SA).

Some definitions of the term "map" require that its scale be smaller than 1:1; in other words, a map must be reduced in size from the area it depicts. While that is true of most maps, in this book we will allow maps to have any scale. For example, an ecologist working on a study of insect herbivory within a 1 m^2 sample plot of grassland could produce a map of that plot at a scale of 1:1, where 1 cm on the map equals 1 cm on the ground. Likewise, one could map tiny features of some surface observed through a microscope at a scale of 100:1, in which case the map would be much larger than the area it depicts.

Using map scale to calculate distances and areas

A map's scale can be used to calculate ground distances or areas based on measurements from the map. To determine the ground distance corresponding to a given map distance, multiply the map distance by the reciprocal of the map scale (equation 3.1.2):

$$D = d \times \frac{1}{s} \qquad (3.1.2)$$

where

D = ground distance

d = map distance

s = scale

Note that in equation 3.1.2, the ground and map distances *D* and *d* are in the same units (e.g., cm or inches). Typically, the final step in this calculation would be a unit conversion to express the ground distance *D* in more convenient units (e.g., km or miles). For example, to find the ground length of a stream segment that measures 2.6 cm on a 1:25,000-scale map, multiply its measured length (2.6 cm) by the reciprocal of the scale (25,000/1 or simply 25,000), yielding a ground length of 65,000 cm, which is then multiplied by the unit conversion factor (1 m/100 cm) to express the result as 650 m. The accuracy of this result would depend on both the accuracy of the map itself and the accuracy of the measurement process. Note also that distances calculated using map scale typically represent the *horizontal* distance between two points. In steep terrain, if those points are at different elevations, the full (three-dimensional) distance between the points will be greater than the horizontal distance.

When calculating area, the map scale must be squared, because the scale applies in both the X and Y dimensions (equation 3.1.3):

$$A = a \times \left(\frac{1}{s}\right)^2 \qquad (3.1.3)$$

where

A = ground area

a = map area

s = scale

Likewise, any linear unit conversion factor must be squared as well (e.g., there are 100 cm in one linear m, but 100^2 or 10,000 cm^2 per m^2 of area). For example, on a 1:25,000-scale map, a lake whose area is measured at 3.3 cm^2 would have a ground area of 21 hectares (ha), based on the following:

Map area (measured) = 3.3 cm^2 x (1 m^2/10,000 cm^2) = 0.00033 m^2 or **3.3 x 10^{-4} m^2**

Reciprocal of scale squared = (25,000 x 25,000) = **6.25 x 10^8**

Ground area = (**3.3 x 10^{-4} m^2**) x (**6.25 x 10^8**) = 206,250 m^2 or **2.0625 x 10^5 m^2**

Conversion to hectares: 2.0625 x 10^5 m^2 x (1 ha / 10,000 m^2) = **20.625 ha**

Rounding to two significant figures, as in initial measurement = **21 ha**

Working from an analog (printed) map, measurement of areas on maps is typically done by counting cells or dots on a grid, or by tracing the outline of an area with an instrument called a planimeter. In a GIS or other digital environment, areas can be measured using on-screen digitization.

Scale variation

These calculations assume that a map has a single, known scale. In reality, any map or image that projects the curved surface of the Earth onto a flat plane will of necessity have variation in its scale due to that curvature. This effect can usually be ignored for maps of small areas (large-scale maps) within which the curvature of the Earth is negligible, but maps of larger areas (small-scale maps) always have significant variation in scale. In some common projections used for world maps, the scale increases predictably with latitude, and more complex scale bars have been developed to indicate this (Figure 3.1.3). The projection process may also result in maps that have very different rates of scale variation in the north-south vs east-west directions. On such maps, calculating distances and areas from measurements is not straightforward, and equations 3.1.2 and 3.1.3 cannot be used directly.

Another example of scale variation occurs in perspective view (or “bird’s-eye”) maps. These maps have a scale that decreases with distance – scale is larger in the foreground and much smaller in the background. They also may have “vertical exaggeration” caused by a differing scale in the vertical (Z) dimension. These situations require care when interpreting the map, and can greatly hinder the ability to make quantitative measurements of distance or area from the map.

Figure 3.1.4 shows a map of a cross-country ski trail network in Québec, Canada, in perspective view. The map is beautifully drawn, and accurately conveys the qualitative “feel” of the landscape, but the scale bar at lower right (enlarged in Figure 3.1.5) is inappropriate and misleading, because the map’s scale varies with distance. This scale is approximately correct for the central portion of the map, but the map scale in the background (the upper portion of the map) is much smaller than indicated, and the scale in the foreground is much larger. Note that the map is also described as having a “contour interval” of 10 m (Figure 3.1.5), though there are no contour lines on this map.

Many forms of remote sensing imagery are collected in a perspective-view geometry similar to that of the map in Figure 3.1.4. In such images, scale will likewise vary as a function of distance from the sensor; points closer to the sensor will have a larger scale, while those farther away will have a smaller scale. This scale variation can be corrected through the process of **orthorectification**, in which the image is reprocessed to have approximately uniform scale throughout its area. The result is referred to as an **orthophoto** or **orthoimage**. While measurements of distances and areas on a "raw" remote sensing image will usually be incorrect due to scale variation, the corresponding measurements on an orthoimage can be used to calculate ground distances and areas using equations 3.1.2 and 3.1.3.

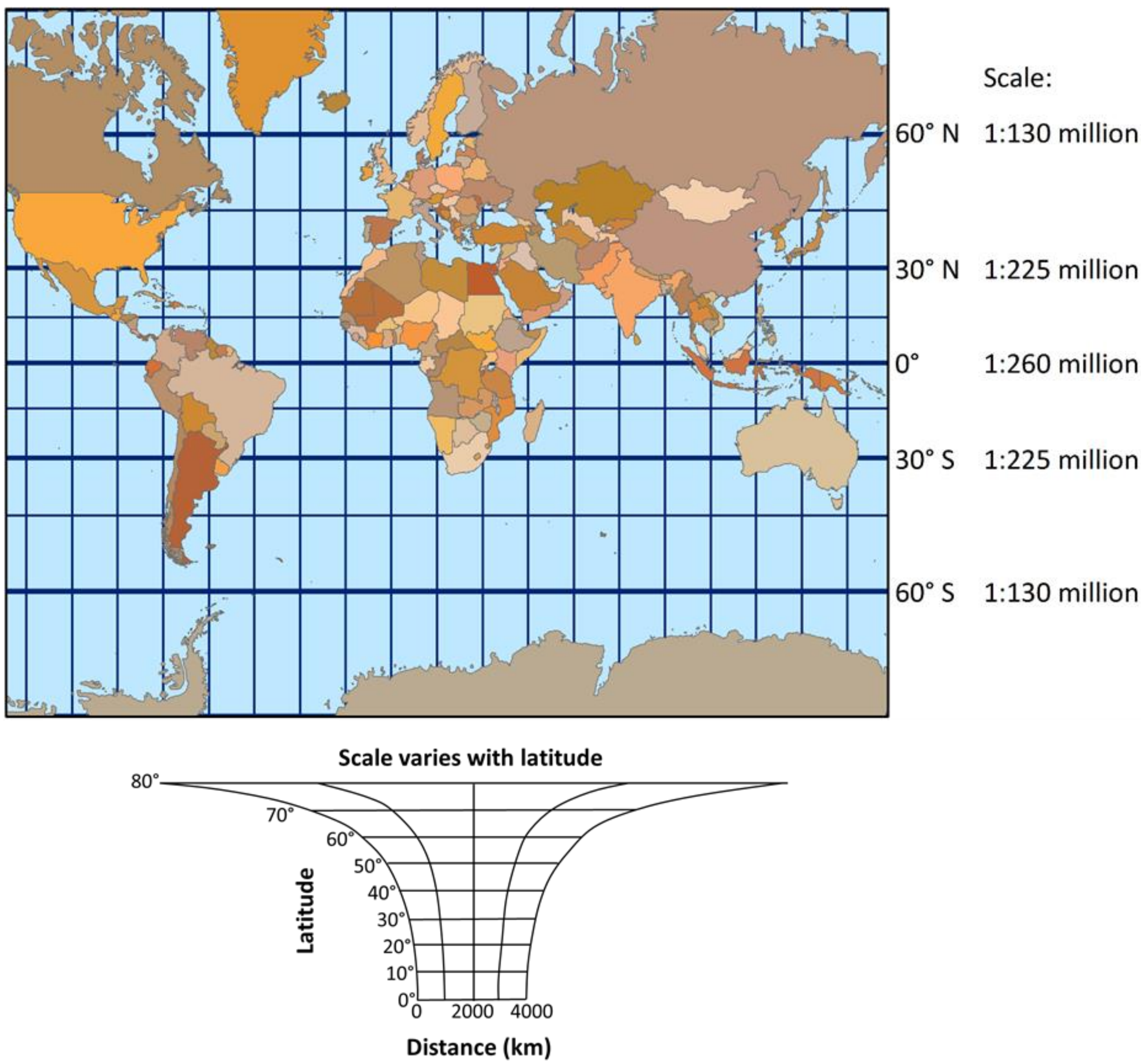

Figure 3.1.3 Scale variation as a function of latitude on world maps. Top: Map in Mercator projection, with scales listed at 30° intervals of latitude. Bottom: Two-dimensional scale bar, showing scale varying with latitude.

Figure 3.1.4 Map of a cross-country ski area, drawn in perspective view. Scale bar at lower right (enlarged in Figure 3.1.5) is inappropriate for this map's geometry. Map courtesy Centre de ski de fond Mont-Sainte-Anne, Saint-Ferréol-les-Neiges, Québec, Canada.

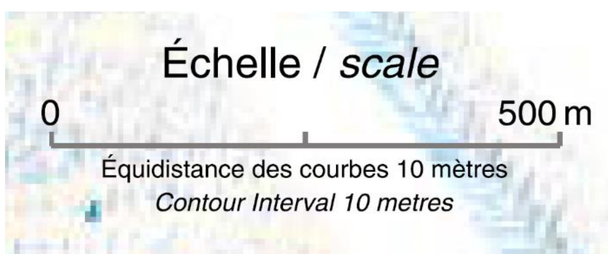

Figure 3.1.5 Enlargement of the scale bar and contour interval text from the lower right corner of the map shown in Figure 3.1.4.

Despite this non-uniformity of scale, maps and remote sensing images are often described by their **nominal (or principal) scale**. This may be an approximation of the average scale within the area depicted, or it may be the scale at an identified location, such as the Earth's equator on a world map.

In cartography it is often defined as the scale of the **generating globe**, the globe used to produce the projection for a given map. In contrast to the nominal scale, the **local scale** at any point incorporates the effects of distortion due to the map projection process, the geometry of a remote sensing system, or other factors. The distortion of scale at any point on a map or image can be quantified by its **scale factor (SF)**, the ratio of the local scale at the point to the nominal scale of the map or image (equation 3.1.4):

$$SF_p = \frac{s_p}{s_0} \qquad (3.1.4)$$

where

SF = scale factor at point *p*

s_p = local scale at point *p*

s_0 = nominal scale

For example, if the nominal scale of the map in Figure 3.1.3 is defined as the scale along the equator (1:260 million), then the scale factor for a point at 60° N (where the local scale is 1:130 million) would be given by equation 3.1.5:

$$SF = \frac{1/130{,}000{,}000}{1/260{,}000{,}000} = 2.0 \qquad (3.1.5)$$

On a globe, or a section of a globe, the SF is 1.0 at all points, because the local scale is equal to the nominal scale. In contrast, all maps on flat surfaces will have values for SF that vary spatially, with SF < 1 where the local scale is smaller than the nominal, and SF > 1 where the local scale is larger. However, as noted earlier this scale variation is often ignored on large-scale (small-area) maps where Earth curvature is small and distortion is minimal in comparison to the expected accuracy of measurement on the map.

The effects of the map projection process on map scale are discussed in more detail later in this chapter, and the geometry of remote sensing images is discussed in chapter 15.

Scale and generalization

Vector data are typically suited for use over a particular range of scales. If they are displayed at too small (coarse) a scale, the data will have too much detail for the map, causing it to look crowded, as in the southern coast of Chile shown in Figure 3.1.6*a*. At the opposite extreme, when viewed at too large (fine) a scale, the data have insufficient detail, and the map loses accuracy because the spatial positions of features are not represented adequately at that scale, as in Figure 3.1.6*c*. When preparing a map using vector GIS data, the cartographer will look for a version of the data that is suitable for the intended scale, or use *generalization* methods to reprocess finer-scale data to an appropriate level of detail for the map's scale, as in the "happy medium" example of Figure 3.1.6*b*. Meanwhile, the excessively detailed version of the data from 3.1.6*a*, where it was displayed at an original scale of 1:50 million, is more appropriate when enlarged by a factor of 10, as in 3.1.6*d* at a scale of 1:5 million.

Often, spatial data sets such as vector hydrography (showing rivers, lakes, and other water bodies) are available in different versions suitable for use at scales ranging from the continental/global down to the local scale. The relationship between map scale and generalization of mapped features is discussed in chapter 4.

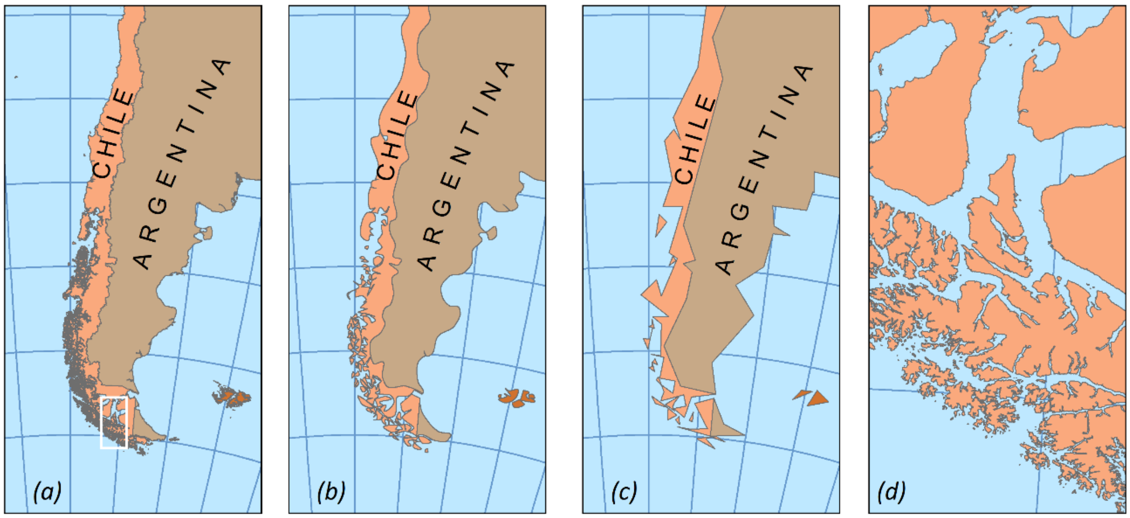

Figure 3.1.6 Vector polygons at varying degrees of generalization. *(a)* Too much detail for this scale; *(b)* Reasonable level of detail; *(c)* Too little detail for this scale; *(d)* Enlargement of (a) at appropriate scale for this data set's level of detail. Original scale 1:50 million *(a-c)* and 1:5 million *(d)*. White rectangle in *(a)* indicates extent shown in *(d)*.

Raster data have a particular *grid spacing* (or *posting*) that relates to the scale at which they are appropriate for use. Often the term *resolution* is used, though this technically refers to the smallest size feature that can be detected (or "resolved") in the data. (The concept of resolution in remote sensing is discussed further in chapter 15.) Using raster data whose grid spacing is too coarse for the viewing scale will result in a map that appears blocky or "pixelated." This depends on a third factor, the *dot (or pixel) density* of the output medium on which the image will be printed or displayed. This density factor is generally expressed in dots per inch (dpi) for analog printed output, or pixels per inch (ppi) for digital displays (metric equivalents have been defined but are not yet in common usage).

The ground dimensions covered by a single printed dot (or on-screen pixel) at a given scale can be determined using equation 3.1.2 from earlier in this section. For example, a raster image printed at 300 dpi (where 1 dot = 1/300th of an inch) and a scale of 1:100,000 would have dot dimensions of approximately 8.5 x 8.5 m on the ground:

$$D = d \times \frac{1}{s} = \left(\frac{1\text{ in}}{300}\right) \times \left(\frac{100{,}000}{1}\right) = 333\text{ in} \times \frac{1\text{ m}}{39.37\text{ in}} = 8.5\text{ m} \quad (3.1.4)$$

meaning that each dot in the printed output is 8.5 m wide on the ground. If the original raster layer has a grid spacing that is close to this output dot dimension (or finer), it will appear visually satisfactory in the output. If the raster layer has a much coarser grid spacing, then there will be too many output dots (or pixels) within each raster grid cell. This will cause the output to look blocky, with the pixel edges visible.

This phenomenon is illustrated in Figure 3.1.7, which shows a Landsat-8 satellite image of the Kuiseb River in Namibia. In 3.1.7*a*, the river channel with its surrounding riparian forest corridor runs from lower right to upper left, with the dunes of the Namib Sand Sea to the lower left and a gray-hued stony plain on the upper right. When printed at 300 dpi at a scale of 1:100,000, each dot covers approximately 8.5 x 8.5 m on the ground, as in the preceding calculation. This is slightly finer than the 30 x 30 m grid spacing of the original Landsat image, but close to it, and the resulting image thus looks reasonably smooth and realistic (3.1.7*a*). In contrast, when displayed at 1:20,000 scale (3.1.7*b*), each 300-dpi dot covers only 1.7 x 1.7 m on the ground, meaning that each Landsat-8 grid cell will be represented by a uniform block of many small dots in the printed output. The scale of 1:20,000 is not appropriate for this data set (3.1.7*b*); the smaller (coarser) scale of 1:100,000 is better (3.1.7*a*).

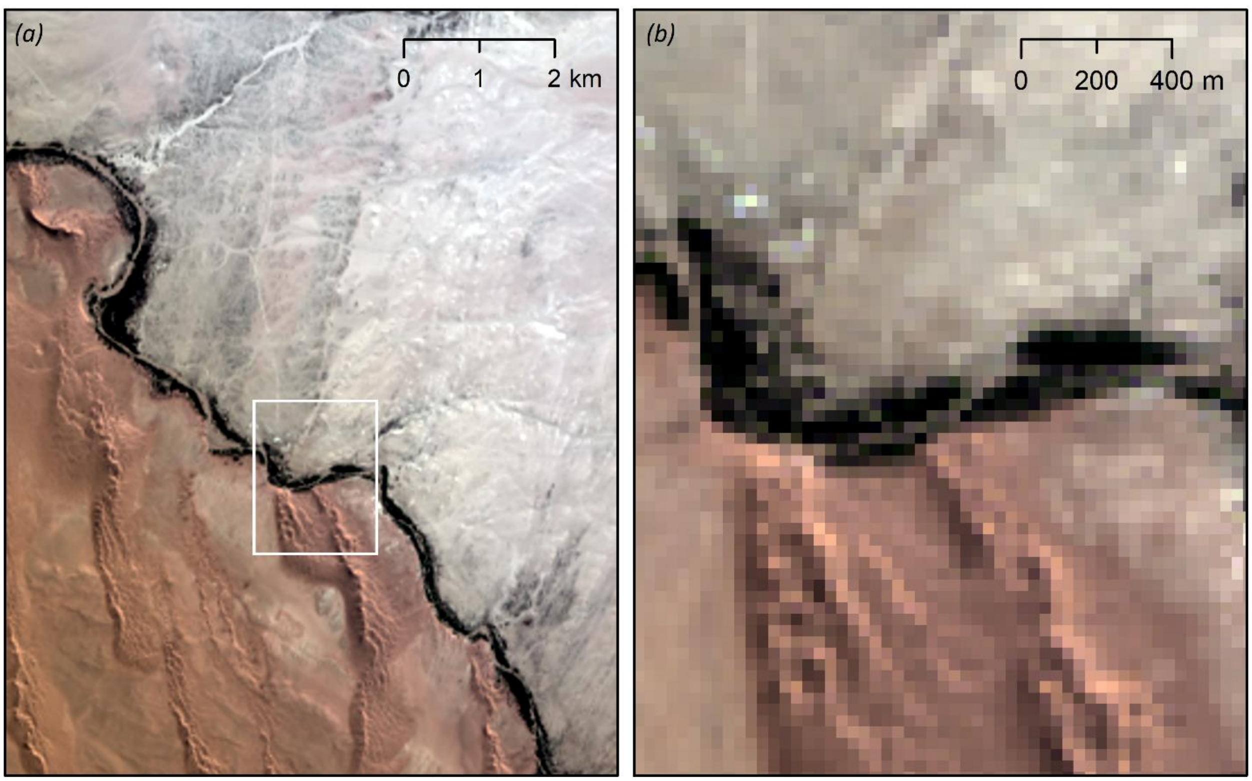

Figure 3.1.7 Landsat-8 true-color imagery of the Kuiseb River in Namibia, displayed at 1:100,000 scale *(a)* and at 1:20,000 scale *(b)*. White rectangle in *(a)* indicates extent of image in *(b)*. Image date 31 July.

References

ASP (American Society of Photogrammetry), *Manual of Photogrammetry* (2nd edition), 1952, p 504.

Dent et al., *Cartography* (6th ed.), 2009, p. 34.

Field, *Cartography*, 2018, p. 405.

Lillesand et al., *Remote Sensing and Image Interpretation*, 2015, p. 76.

Peterson, *GIS Cartography* (2nd ed), 2015, p. 9.

Raisz, *Principles of Cartography*, 1962, p. 32.