CHAPTER 3

Scale and Projection

Scale is fundamental to geography. When observing any geographical process, whether physical, biological, social, economic, political, or whatever, available evidence is always mediated by the scale at which we make our observations. This statement clarifies how wrapped up in the concept of scale are both an implicit scope or extent across which observations are made, along with some level of detail or resolution at which we note meaningful differences between one location and another. This makes clear that the notion of scale is unavoidably bound up with other key concepts such as space and location (see Chapter 2) and process (see Chapter 8). It also draws attention to the fact that scale is always in some sense socially constructed, being bound up with acts of observation at particular times and in particular places (Sheppard & McMaster, 2004).

This chapter explores some of these complexities. First from the perspective of geographical theory; in human geography, where recognition of the social construction of scale has held sway for some time; and in physical geography, where the challenges of translating across scales have long been recognized. And second from the perspective of cartography and giscience, where scale is also a key concept but one where a narrower technical definition is central. This purely technical approach, which pertains to the relationship between the physical size of phenomena in the world and the size of their cartographic representation is muddied considerably by (among other things) the emergence of seemingly scale-free, infinitely zoomable web maps, the geometry of map projections, and by cartographic generalization. Each of these topics is considered, along with a brief look at the resulting effects of scale on the analysis of spatial data.

The picture that emerges is one where, although some nuances of geographical theory concerning scale are inevitably lost in computational approaches, seriously engaging the concept demands that we recognize the force of arguments about scale's epistemological character and social construction.

SCALE IN GEOGRAPHICAL THEORY

Scale as Size or Scope

In everyday usage the concept of *scale* is more or less synonymous with the *size* of phenomena of interest. Something large-scale is big, and something small-scale is, well … small.

This leads to an apparently natural way of thinking about scales of interest, from the cosmological to the quantum. It is worth bearing in mind the truly vast range of scales on which the universe is organized and understood in this sense. The universe is estimated to have a diameter of up to 93 billion light years (Bars & Terning, 2010, p. 27), or around 10^{27}m.[1] The smallest subatomic scales observable are on the order of 10^{-18}m. Thus, the sciences collectively are concerned with phenomena whose scales range across about 45 (base 10) orders of magnitude. This puts into perspective geographers' frequent presumption to speak authoritatively about scale, when we consider that terrestrial scales extend across a much more parochial range from around 10^7 to 10^{-3}m, a mere ten orders of magnitude. The lower end of that range, arbitrarily set at one millimeter, is arguable, when geographers might be plausibly concerned with, for example, microplastic particles or microorganisms. In

---

[1] Estimates diverge widely. 93 billion light years is an upper estimate that accounts for differences between what is observable, which is limited by the speed of light and the age of the universe, and also depends on theoretical models for the expansion of the universe.

some branches of the geosciences, an argument can be made that scales range down to the molecular level, which extends the range by a half dozen or so more orders of magnitude. Even so, 15 orders of magnitude remains a narrow focus from a cosmological perspective!

Of course, even a perhaps conservative 10 orders of magnitude is unimaginably vast, requiring us to somehow hold in mind a grain of sand in relation to the whole of the Earth, an arguably futile effort.[2] Consequently, in practice, it is much more typical for researchers to restrict their scalar focus and to study entities and phenomena whose sizes encompass perhaps three or four orders of magnitude, a 1,000- or 10,000-fold range.

Scale as Hierarchy

The bracketing of phenomena into manageable size ranges leads directly to thinking about the world in terms of a series of qualitative scale ranges. Such bracketing can be justified not only for reasons of practical observation, but because processes tend to operate at particular scales and tend not to have effects at all scales. As de Boer suggests,

> [i]n principle, the form and functioning of any geomorphic system is the end product of the interaction of processes operating at all scale levels, from the smallest to the largest. Luckily, to understand a geomorphic system one does not have to consider, as a rule, every level of scale since, depending on the scale of the system and the objective of the investigation, certain levels will be dominant whereas others play a secondary role and can be ignored (1992, p. 304).

This argument is made in the specific context of geomorphology but can reasonably be applied to any subdiscipline in geography.

This leads more or less directly to notions of a hierarchically nested series of scales (see Figure 3.1). While the figure shows a set of scales that might be deployed in social, political, or economic geography, similar sequences can be considered in other domains. Biogeographers, for

---

[2] In the absence of mind-altering molecules.

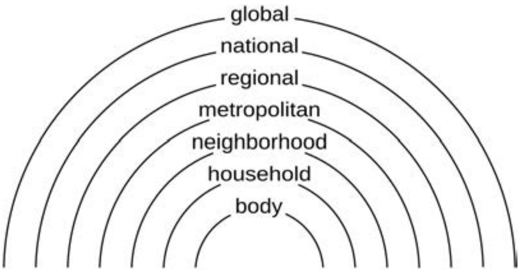

Figure 3.1. An example of how a series of qualitative scales might be considered nested one within another. See also Figure 1.2 in Herod (2011).

example, may work with a nested hierarchy (from largest to smallest) of kingdoms (or realms), regions, dominions, provinces, and districts (Morrone, 2018, p. 285). In this case there is considerable controversy about nomenclature and how such regional and scalar hierarchies can be reliably defined based, among other things, on whether flora or fauna are used to guide the determination of boundaries between areas at different scales. Similar challenges arise in any domain that is not readily encompassed by a single scale of observation and analysis.

Such controversies highlight how scalar hierarchies are an outcome of the processes relevant to understanding how the world works, from a particular perspective. As Jonathan Phillips suggests in a paper evaluating the degree to which different scales in geomorphology are related to one another and can therefore be considered linked, "even where phenomena are continuous, hierarchical structures are often imposed to make analyses tractable" (2016, p. 72). This makes clear that scale is *not* naturally given per se, but an outcome of particular sets of observational practices and analytical methods, applied in particular contexts. Of course, to the extent possible, data collection and analysis methods should be aligned with spatial (and temporal) scales relevant to particular phenomena of interest. Even where this is possible, it is clear that scale is not only dependent on the underlying nature of things, but is very much an outcome of the relationship between the phenomena under study and the apparatus and methods available for their investigation.

An important further point to make about neatly nested scale hierarchies is that the nesting is often conceptual more than it is scalar in the sense of being size-based. Taking the hierarchy proposed in Figure 3.1 as an example, there are many cases where a particular region, say California or the Midwest of the United States, is much larger than a particular nation such as Ireland or New Zealand. Again, some urban settings (Tokyo, New York, London) are larger in important respects than particular nations or regions. Further, the linkages that might exist among urban areas, regions, nations, and the global are complicated and messy, and extend both between scalar levels and within them (see Chapter 6).

Nevertheless, and in spite of the unruliness of reality relative to any conceptually clean scalar hierarchy, there are advantages to thinking in such terms. It is, for example, possible to set aside processes at scales far removed (whether higher or lower in the hierarchy) from the core scale of interest. So, in a study focused on (say) changes in the California public school system statewide, perhaps it is possible to ignore effects in individual school districts or schools on the one hand, and also to ignore effects in neighboring states, or developments in educational governance and policy at the federal or global levels. Our expectation is that larger scales (national, global) define a relatively stable context within which changes at the scalar level of interest, in this case the state or regional level, play out. Meanwhile, smaller scales such as the household level are governed day-to-day by the changes at the state or regional level that are the subject of the study.

This general framework for analysis falls under the rubric of *hierarchy theory*. Simon (1962) is among the earliest and most lucid proponents of the idea that hierarchically structured systems are common in natural and human systems because they enable levels of complexity that would otherwise fail to evolve, self-maintain, or (in the case of artificial systems) be designed and subsequently manageable. Hierarchies are *decomposable*, meaning that they consist of more or less independent subsystems that are themselves composed of more or less independent subsystems at levels below, and which together compose other subsystems at levels above

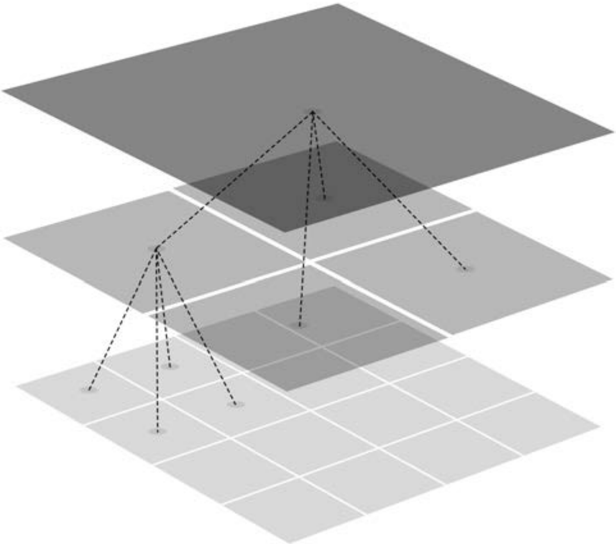

Figure 3.2. A schematic illustration of how a hierarchy of spatial elements might be organized. This hierarchy is also a tree structure.

(see Figure 3.2). Crucial to this picture is the observation that interactions *between* levels are weaker than those *within* levels between subsystems. At the same time interactions between subsystems at a particular level are weaker than those within the subsystems themselves.[3] The resulting hierarchical structure is what enables (for example) a botanist to study interactions among individual trees more or less independent of cellular-level processes on the one hand or whole ecosystem processes on the other.

---

[3] Intriguingly, elsewhere, Simon (1973, p. 23) summarizes this property thus: "Everything is connected, but some things are more connected than others." Phil Agre (2003, pp. 418–19) points to the spatial structure implied (if not required in all cases) by this perspective, which is of course reminiscent of the *first law of geography* that "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970, p. 236).

Of course, any bracketing off of levels (and scales) is always an approximation. In truth in most cases, everything matters—even if only a little—at all scales. Thus, while it may be useful to consider a nested hierarchy of scales and associated processes, it is important to recognize that such hierarchies are an external scheme imposed on the world to organize thinking from a particular perspective—economic, social, cultural, political, biological, hydrological, ecological, climatological, and so on—rather than reflective of a real set of structures in the world (see also Harvey & Reed, 1996).

The inadequacy of such schemes becomes particularly apparent when, as is often the case in geography, we want to consider the interactions across and between multiple domains or perspectives. Also, contrary to Simon's general claims, social systems, the subject matter of human geography, are often messier and less neatly decomposable than many natural systems, perhaps because social processes are less tied to specific spatial matrices through which they operate. A related argument is central to Christopher Alexander's (1965) contention that "a city is not a tree," where he argues that many of the failings of mid-20th century urban planning can be traced to reorganizing urban space into functionally distinct areas (residential, commercial, industrial) in ways that prevent such designed cities from becoming vibrant (messy!) urban places.

Scale as Socially Constructed

Emphasizing the perspective of a particular analyst sheds light on the perhaps less obvious contention prevalent in theoretical human geography that scale is best understood as *socially contested and produced*. This position is commonly traced back to an article by Neil Smith, the subtitle of which invokes "the production of geographical scale" (1992). As Jones III et al. (2017) argue, the thinking underlying this idea is found in Smith's earlier work on *uneven development* (Smith, 1984) and its relationship to gentrification (Smith, 1982). There, Smith discusses distinct scales at which capitalist processes of uneven development unfold, such that, “[w]hile gentrification represents the leading edge of spatial restructuring at the urban scale, the process is also occurring at the regional and international scales” (Smith, 1982, p. 151). There is not much suggestion here that scale itself is a product of these processes. Rather the urban, regional, and international scales are pre-given levels at which processes of investment and disinvestment operate. Elsewhere in the same paper, however, Smith hints at the idea that a particular scale is an outcome of a particular set of social processes, when he argues that “[t]he urban scale as a distinct spatial scale is defined in practice in terms of the reproduction of labor power and the journey to work” (1982, p. 146).

Although this shift is more implied than directly stated, it is nevertheless significant relative to contemporary treatments of scale that relied more on pre-given nested hierarchies. Exemplary of this is Peter Taylor’s “Geographical scales within the world-economy approach” (1981), which defines global, national, and urban scales as, respectively, scales of reality, of ideology, and of everyday life. While Taylor identifies each scalar level with a particular set of processes, he is also clear about the primacy of the global scale. In the context of then unfolding rapid changes in the industrial structure of developed economies he argues that “the state and its ideology stand between the experience of people we wish to radicalize and the reality of the world-economy which exploits and destroys them” (Taylor, 1981, p. 10), thus making clear that scale, or more accurately the way in which processes at different scales operate, has material effects on outcomes. This framework is systematically set out in a later paper (Taylor, 1982) with a political geographic emphasis.

As recounted by Jones III et al. (2017), the idea that scalar levels themselves are an outcome of processes comes into focus more in later work by Smith and Dennis (1987), who suggest that a key aspect of the same industrial transformations that concern Taylor is a redefinition of the scale—a *rescaling*—of economic regions, from relatively localized ones centered on single metropolitan areas, to more extensive polycentric ones. In a U.S. context, such economic regions may remain subnational, but elsewhere they might cross national boundaries as the communications and logistical networks that underpin economic activity expand their reach. This is not specific to the United States, but was common to other deindustrializing regions in the early 1980s (Massey, 1995). The details of the economic geography arguments are unimportant for our purposes. What matters is the idea that the scales of analysis are themselves produced as an outcome of processes under study.

A nice formulation of this idea is spelled out in a later paper by Smith, where he suggests that "[i]t is possible to conceive of scale as the geographical resolution of contradictory social processes of competition and cooperation" (1992, p. 64). While Smith is discussing political and economic processes, this formulation could equally be applied to ecological or geomorphological processes, provided that we think broadly about notions of competition and cooperation. The concept is readily applicable to ecology in an obvious way, where competition and mutualism between species will play out in the scale of habitats, home ranges, ecotopes, and so on. In geomorphology, processes of uplift and erosion might be considered analogous. It is notable that Smith, jumping off from a consideration of those evicted in New York's 1990s battles over gentrification, goes on to consider "a sequence of specific scales: body, home, community, urban, region, nation, global" (1992, p. 66), thus significantly extending the reach of the concept, beyond the narrowly political and economic emphasized in the earlier debates.

Nevertheless, given its origins in economic and political geography, it is unsurprising that further theoretical development through the 1990s of theories of scale as socially constructed tended to emphasize economic and political processes under capitalism. Among many contributions to these themes was work by Smith (1992), Herod (1991), Brenner (1997), and Swyngedouw (1997). Taylor (1999) widened the scope in recognizing the significance of the domestic sphere and particularly the idea of "home," although paradoxically he is concerned with the idea of home at scales from the domestic to the global, and in concluding is strongly focused on the whole-Earth or global scale.

A necessary corrective to this tendency was administered by Sallie Marston, who argues for much closer attention to the scale of the household, which is foregrounded by consideration of social reproduction:

> [p]reoccupied with questions of capitalist production, contemporary writing about scale in human geography has failed to comprehend the real complexity behind the social construction of scale and therefore tells only part of a much more complex story (2000, p. 233).

This intervention led to some concern expressed by Brenner that there was a danger in expanding the remit of the concept of scale, that it would be confused with other key concepts such as "space, place, locale, location, territoriality, distanciation, network formation and so forth" (2001, p. 597). He suggests in particular that the idea of scale is in danger of being confused with the notion of place; when researchers discuss, for example, the scale of the home, they are really concerned with the home as a place, rather than with a set of scalar relations among levels in a hierarchy. This is a plea for a return to a more consistently hierarchical (as Brenner puts it, "plural") perspective on scale. In response, Marston and Smith (2001) forcefully assert the importance of continuing to develop theories of scale, and suggest that Brenner, in his appeals to the work of Lefebvre, is guilty of confusing not place and scale, but space and scale. It is debatable how much light such debates shed on the matter at hand, and whatever the merits of the argument Brenner (2001) is certainly guilty of taking a line that under current social relations minimizes feminist geographies (see Monk & Hanson, 1982). Others commenting on the scale literature suggest that much of the debate is confused because different notions of scale-as-extent, scale-as-level, and scale-as-(social)-process are deployed interchangeably, muddying the debate (see, e.g., Sayre, 2005).

The End of Scale?

Rather surprisingly, given this robust defense of the concept of scale, only a few years later, Marston et al. (2005) argue for a human geography "without scale." The paper's title is slightly misleading, as it is really hierarchical notions of scale, rather than scale per se, that the authors object to:

> we argue that attempts to refine or augment the hierarchical approach cannot escape a set of inherent problems [. . .] in place of the hierarchical [. . .] we offer a flat alternative, one that does not rely on the concept of scale (2005, p. 417).

A flat perspective is an ontological position that asserts that everything that exists whether material or immaterial, and whatever its size or duration, exists on an equal footing such that none inherently determines, controls, or contains the others (see also §Related Strands in Geographical Thought, Chapter 8). In this perspective, a hierarchy of nested scales makes no sense. Events at some notional national scale do not predetermine or dictate events regionally. Researchers should not approach their objects of study with preconceived ideas about scale, but should instead be open to seeing things for what they are:

> a flat ontology must be rich to the extent that it is capable of accounting for socio-spatiality as it occurs throughout the Earth without requiring prior, static conceptual categories (2005, p. 425).

They argue that for all the vigorous, sometimes contentious debates, scale as a concept never really escaped Taylor's early, hierarchical conceptualization. While much ink had been spilled on the subject, with mixed results, it still seems extreme to seek "to expurgate scale from the geographic vocabulary" (2005, p. 422).[4]

This position is, I think, best understood as a political argument, not a theoretical one. Hierarchically nested, vertical scale thinking has consequences for how we understand events:

> Invariably, social practice takes a lower rung on the hierarchy, while 'broader forces', such as the juggernaut of globalization, are assigned a greater degree of social and territorial significance. Such globe talk plays into the hands of neoliberal commentators [. . .] the standard trope [. . .] is to shift blame 'up there' and somewhere else (the 'global economy'), rather than on to the corporate managers who

---

[4] Perhaps, after well over a decade of it, Marston et al. had had enough of the scale debates and just wanted it all to end; informal conversations with colleagues at the time (and since) suggest that if so, then they were not alone.

> sign pink slips. In this fashion 'the global' [. . .] can underwrite situations in which victims of outsourcing have no one to blame, a situation possibly worse than blaming oneself (2005, p. 427).

Paying greater attention to specific interactions among people and things, regardless of any prior notion of their scale, will enable a clearer sense of why and how things happen as they do. In place of the distraction of scale, where "states of affairs rarely fit neatly into scalar (or local–global) [levels]" (Woodward et al., 2010, p. 274), the *site* is proposed as an alternative approach to framing research. What this means in practice is clearer only with regard to more specific examples, such as the suggestion in response to Prytherch (2007) that a productive site for shedding light on the operations of Walmart is the shipping container (Woodward et al., 2008, p. 82). The shipping container is a complex driver and outcome across scales of material, economic, regulatory, legal, labor, and other relations (Levinson, 2006). Tracing these effects can help us understand how Walmart works as a complex *assemblage* (see §Related Strands in Geographical Thought, Chapter 8).

Overall, these post-scale objections to scale oppose pre-supposing a fixed hierarchy of globally applicable scales, rather than scale and scalar effects per se. MacKinnon (2011) argues persuasively for a more nuanced position, acknowledging a tendency to reify specific scalar levels (local, urban, national), but also an unsustainable denial of scalar aspects in flat ontologies. Complex assemblages of human and nonhuman objects exert their effects at some scales and not others. There is also a tension such that emphasizing constant flux in flat ontologies denies the persistence of some scales—such as the national or urban levels. At the same time, the persistence of these levels can cause more structuralist perspectives to assume that these levels are permanent and not continuously made and remade through particular practices. In Chapter 8 we will see how process philosophies loosen some of these tensions.

For now, MacKinnon's (2011) argument that scalar levels have meaningful real existence, but that these depend heavily on ongoing practices and narratives to maintain them, is a constructive attempt to resolve the difficulty of continuing to work with scale while recognizing its contingent character (see also Manson, 2008). Crucially this position recognizes the scalar thinking implicit in recognizing the messy complexity of phenomena. It is hard to coherently maintain a scale-free perspective on things and argue for a flat ontology, when this puts (for example) individuals, trade unions, corporations, shipping containers, shipping lanes, and national law into relation with one another, yet many of those phenomena have intricate (often hierarchically scaled) internal structure.[5] While these details may not matter in some contexts, so that immediate interactions are not scalar, but are concrete grounded events, this surely does not mean that such structure is irrelevant. Taking sites seriously surely involves pursuing the intricately scaled nature of the world.

SCALE IN GISCIENCE

The political nuances and ramifications of scale find no direct analog in giscience and cartographic representations, in spite of the inherently political nature of GIS and mapping, which are so often central to the social construction of actually existing scales (see Chapters 4 and 5). Even so, as discussed in relation to physical geographic perspectives on scale, it is widely understood that scale mediates how all geographical phenomena are perceived, represented, and experienced.

This is directly recognized in a framework for formally thinking about scale presented by Dan Montello at the first meeting of a series that became the Conference on Spatial Information Theory (COSIT). Montello begins by reiterating the central importance of scale:

> As a problem for geography [. . .] scale has always been a concern of cartographic coding and decoding. But once the scale of the cartographic representation is fixed, all the decisions made with the map become largely scale-independent. A clustered pattern is a clustered pattern. It is this scale-independence of maps, of course, that gives them their great power and utility (1993, p. 312).

---

[5] Broadly this argument is made by Prytherch (2008) in a brief rejoinder to Woodward et al. (2008).

The disconnect between this statement and then-contemporary discussions of scale in human geography is striking. As will become apparent, Montello's assertion of the scale-independence of maps is also arguable (see §The Arbitrariness of Boundaries, Chapter 5) from a giscience perspective and indeed, he goes on to argue that from the perspective of human perception, space is anything but scale-independent. Further, since it is potentially desirable to embed human understandings of space in computational models in giscience, it is important to consider how human perception of space changes with scale. Montello briefly considers earlier frameworks (e.g., Ittelson, 1973; Gärling & Golledge, 1989), noting in these the importance attached to multiple perspectives for human understanding, and the resulting distinction between large-scale (i.e., extensive) spaces that require movement through them for proper understanding, by contrast with small-scale spaces which can be made sense of from a single perspective without such movement. Montello then proposes a more detailed qualitative classification of scales into a hierarchy from *figural*, to *vista*, through *environmental*, to *geographical* (see Figure 3.3).

Figural space is the realm where an object or image of an object can be comprehended in a single view because it is "projectively smaller than the body" (Montello, 1993, p. 315). The perceptual emphasis of Montello's framework is clear when we consider that this scale encompasses both the Moon and a coffee cup! Vista space is as large or larger than the body but can be comprehended from a single viewpoint (hence the name). The qualitative and fuzzy nature of the scheme is again clear when Montello cites as examples of such spaces "single rooms, town squares, small valleys, and horizons" (1993, p. 315). Environmental spaces are larger than the body and can only be comprehended by moving through them, often over long periods of time. The most familiar instance of such space is the experience we have of getting to know a city, slowly stitching together a mental map of how its neighborhoods, streets, and landmarks are spatially related to one another. Geographical spaces are larger still, indeed so large that they cannot be meaningfully understood by moving through them, but instead must be reduced down to a figural space in the form of maps or other representations. This points to

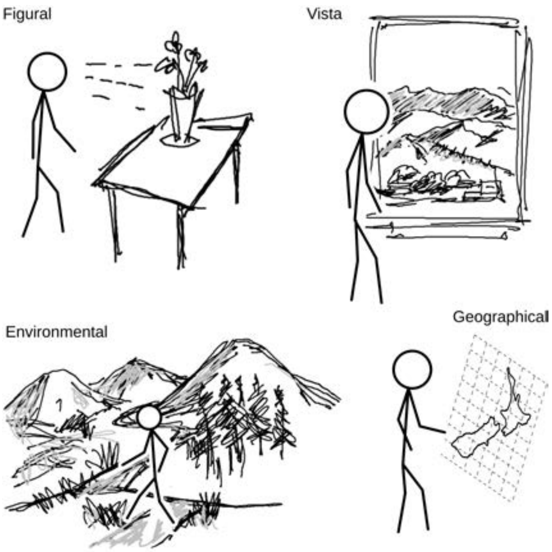

Figure 3.3. Montello's (1993) qualitative classification of spatial scales.

a deep connection between maps and geographical spaces, which can only be apprehended cartographically. It is interesting that this framework echoes work in human geography centered on the scale of the body, albeit out of a rather different motivation. A similar framework drawing on an unpublished presentation by David Zubin (for a summary of this work, see Mark et al., 1989, pp. 13–17) is presented by Helen Couclelis (1992), but she emphasizes the size of objects not their projective size.

Such relatively rich, qualitative frameworks for thinking about scale are rare in giscience, but have attracted more interest in environmental psychology than in geography. In giscience, scale is more often considered from a technical perspective as the relationship between the world and its representation in data or maps—or in Montello's terms the relationship between the figural space of the map and the geographical space of the world.

On traditional paper maps, this relationship is expressed as a *representative fraction* that explicitly tells us the ratio of a distance on the printed map to the distance in the world between the corresponding locations. Typical topographic map scales expressed in this way are 1:50,000 and 1:63,360, respectively equivalent to 1 cm representing 500 m, and 1 inch representing 1 mile. These would generally be considered medium scales, while more detailed large-scale maps have smaller ratios of (say) 1:2,500 or 1:10,000. Small-scale maps, on the other hand, are used to show whole countries at scales such as 1:1,000,000 (1 cm to 10 km) up to perhaps 1:100,000,000 (1 cm to 1,000 km).[6]

Scale and the Web Map

For experienced map users, particular numerical representative fractions become second nature. For many others, a scale bar which directly shows what distance in map space is equivalent to some stated distance in the real world is more useful. On contemporary web maps, scale bars are essential. The reason is mundane but nevertheless an important pointer to how different contemporary digital mapping is from paper-based precursors. Considering this in a little more detail also leads directly to a consideration of map projections.

Cartographers no longer have much control over the final physical form in which maps are presented to users. Screen size and resolution (i.e., the size of individual pixels in a display) are dependent on individual map users' devices and can vary widely. The same tile-based web map delivered on a 30-in, 1920×1080 monitor might appear almost seven times larger than on a smartphone with a 5·8-in, 1440×2560 screen. In the former case a 256×256 pixel tile image would measure almost 89×89 mm but only 13×13 mm in the latter. A scale bar representation will change size with the screen display and is self-correcting, but the representative fraction for each of these would be wildly different and is not

---

[6] It is often observed that small numerical ratios correspond to highly detailed large-scale mapping while large ratios correspond to less detailed small-scale maps, leading to terminological confusions, when large-scale studies may be using the word scale to refer to scope or extent, in a manner more aligned with our earlier discussion of scale as size.

reliably knowable to the map provider at the moment when the map is served to a requesting web client (the computer or phone in this case).

Although platforms are constantly changing, what we might call the classic web map is based on a nested hierarchy of scales, or *zoom levels* of detail. Level 0 is a single square map tile (tile 0/0/0) that encompasses the whole Earth—or at any rate, most of it—see Figure 3.4. This tile is repeatedly subdivided into quarters, with each subdivision increasing the scale or *zoom level,* resulting in tiles that provide more detail but across a smaller region of Earth's surface. This nested hierarchy is how the ability to smoothly zoom in and out from planetary scale to the high level of detail associated with local street directions is supported. At any particular zoom level, a web map image is composed of a set of tiles at the appropriate level laid out to produce the appearance of a continuous map. This enables websites and applications to progressively load the map information in small pieces (the tiles) rather than having to load a very large map image of the whole Earth at any particular scale.

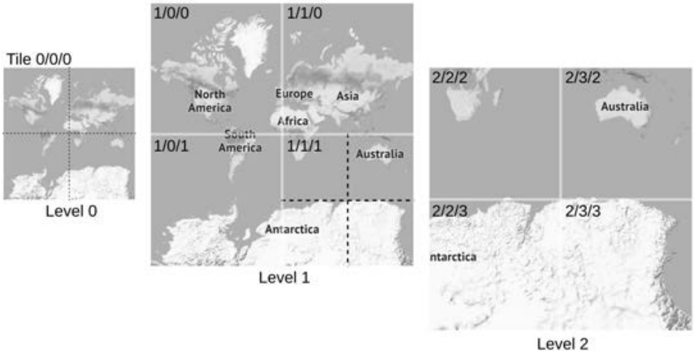

Figure 3.4. Tiles from levels 0, 1, and 2 of the web map hierarchy. Each level is obtained from the level above by progressive subdivision into quarters. Map tiles in the *terrain* design by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL.

When this approach was first introduced, map tiles were preprocessed image files (as are those in Figure 3.4). More recently, the implementation details have changed such that the nested tile hierarchy provides a reference framework, similar to a quadtree (see Samet, 1990), within which the map elements (roads, buildings, and so on) needed to produce a useable map at a desired zoom level and extent can be retrieved and rendered in the desired style. This approach is referred to as a *vector tiling* although the tiles are now a hierarchical reference frame that controls the rendering of geometries from a database rather than permanently stored already rendered images. This is desirable for reasons of storage efficiency and flexibility. Typical current whole Earth tile stores extend from zoom level 0 to level 21. At zoom level 21 there are 2^{42} or around 4 trillion tiles. Even allowing that around 70% of tiles are water—and could be stored as a single uniform blue tile in many cases—the storage requirements for a tiled image map at this level of detail are formidable: somewhere on the order of 100 petabytes. It is easier to store geospatial data than pre-rendered images when faced with numbers like these.

Apart from the obvious way in which this scheme relates to theoretical conceptualizations of scale (compare Figures 3.1 and 3.4) what is the scale of these map tiles? Because of the lack of control over the final size at which tiles are presented to a user, we cannot express scale as a representative fraction, but need to think instead in terms of the distance per image pixel. For a level 20 tile (say) a square tile has edge length $40{,}075/2^{20}$ km or around 38.2 m. Tiles are usually rendered as 256×256 images, so each pixel at this level is around 15 cm across *at the equator*. At a different latitude the scale is different, however. For example in Wellington, New Zealand, at latitude 41°17′20″ south the parallels are not 40,075 km around as at the equator but only 30,112 km, meaning that a single pixel at level 20 at this latitude represents only 11·2 cm.

Scale and Map Projection

The scale disparities among web map tiles at the same level in the hierarchy, but at different locations on Earth's surface, draw attention to a central technical challenge of mapping that significantly complicates the otherwise fairly trivial concept of map scale. Zoomed in close, the scale disparities just noted do not matter much: A level 20 tile at the equator and one at 40° latitude will never appear in the same map view, given that they are separated by several hundred thousand intervening tiles. But in tiles at or near the top of the hierarchy, the problem is severe, as demonstrated in Figure 3.5. This apparently square map represents substantially different distances from east to west than from north to south, and also has very different scales in the east–west direction at the center of the tile, than at the extreme north or south edge.

The problem, of course, is that while a map (or computer screen) is usually a flat two-dimensional surface, the Earth's surface is curved in the third dimension, forming a globe. Traditional maps—and most, if not all digital, on-screen maps—must resolve this fundamental mismatch between the flat surface of the map and the curved surface of the Earth. A *map projection* specifies the relationship between a location on the curved surface of the Earth and the spatial representation of that location on a map. This involves defining the relationship between Earth-centered latitude–longitude (ϕ, λ) coordinates and map-centered (x, y)

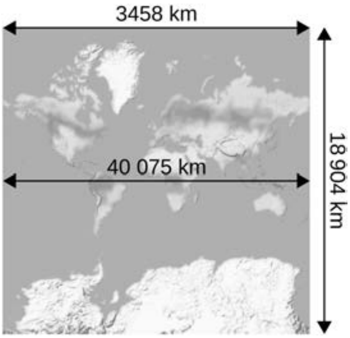

Figure 3.5. Some key approximate distances on tile 0/0/0, illustrating the impossibility of assigning a single scale to this particular projection. Map tile in the `terrain` design by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL

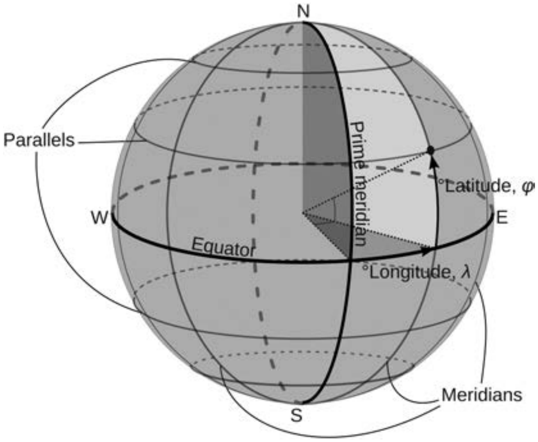

Figure 3.6. The definition of latitude and longitude relative to the Earth's surface.

coordinates. Latitude and longitude are defined relative to the Earth understood as an approximate sphere rotating around an axis through its poles (see Figure 3.6). Lines of latitude (or parallels) are circles parallel to the equator defined by their angular offset north or south of the equator measured at the center of the Earth. Lines of longitude (or meridians) are great circles passing through the poles. A prime meridian is defined to pass through Greenwich, London, and meridians are measured by their angular offset at the center of the Earth from this datum. Any point on the Earth's surface can be located using the geocentric coordinate system. Corresponding points on a map will have a pair of (x, y) coordinates in the map space, and the map's projection is defined by the two relationships

$$x = f_1(\phi, \lambda)$$
$$y = f_2(\phi, \lambda)$$

which specify how to project any particular location on the Earth's surface onto the map space. Most projections in common use can only

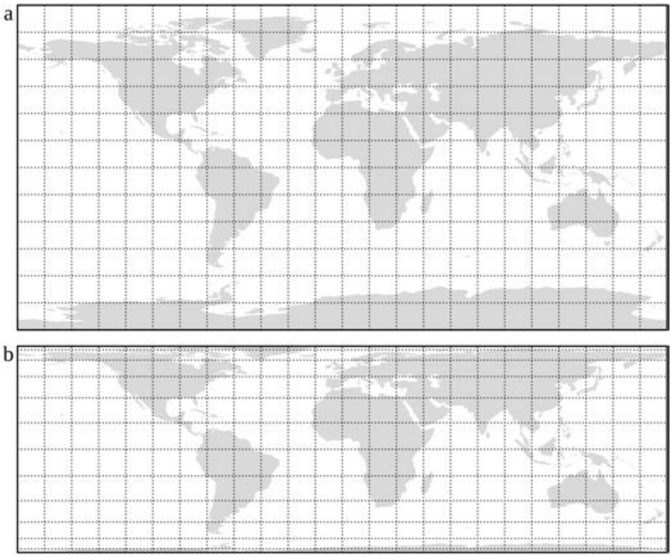

Figure 3.7. Two simple world projections. (a) Platte-Carrée, and (b) Lambert's Cylindrical Equal Area.

be described using complicated mathematical formulae, which are not particularly illuminating to the present discussion.[7]

A simple projection is the null projection where $x = k\lambda$ and $y = k\phi$, sometimes known as Platte Carrée, which treats longitude and latitude, respectively, as the x and y plotting coordinates, scaled so that the nominal map scale is correct at the equator (see Figure 3.7a). Variations on this projection where the scale is correct at the $\pm\phi_0$ parallels are obtained when we set $x = k\lambda \cos\phi_0$. Another simple example is Lambert's Equal Area Cylindrical projection where $x = k\lambda$ and $y = k \sin\phi$ (see Figure 3.7b). The rescaling of the north–south dimension in this case has the effect of making this an equal-area projection such that regions of equal area on the Earth's surface are mapped onto equal-area regions

---

[7] See, for example, Snyder's *Map Projections* for details (1987).

in the map. These are examples of *cylindrical* projections often characterized by the meridians being portrayed as parallel north–south lines, and the poles by the north and south edges of a rectangular map. The latter is an obvious distortion in such projections, since on Earth's surface the meridians converge to a point at the poles.

The cylindrical projection of tile 0/0/0 shown in Figure 3.5 is a variation on one of the oldest known map projections, that of Gerardus Mercator. This projection was designed to have the useful property that straight lines on the map are lines of constant bearing, which makes navigation by compass much simpler (see Figure 3.8). A side effect of this property is that the projection is *conformal*, meaning that the shapes of areas on the Earth's surface are preserved. However, a further side effect of preserving the shapes of regions is that relative areas of regions are grossly exaggerated the farther they are from the equator.

A little thought makes it clear why this occurs. Meridians continuously converge from the equator where they are at maximum spacing toward the poles where they meet. A degree of longitude at the equator

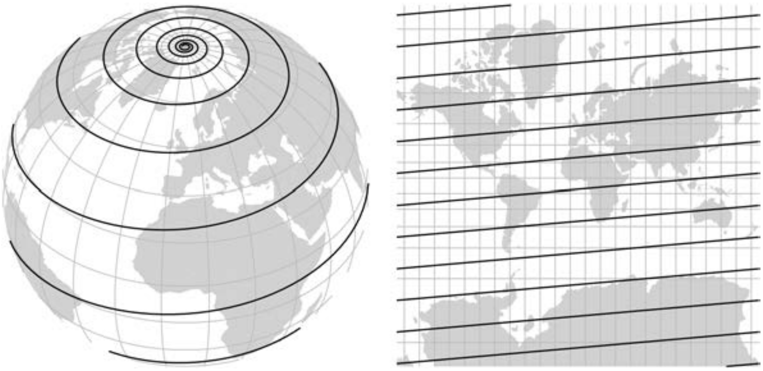

Figure 3.8. A loxodrome on the sphere and in the Mercator projection. This line of constant bearing spirals toward the pole without ever reaching it, but resolves as a straight line in the Mercator projection, greatly simplifying navigation. Note how exaggerated the Mercator projected view is at high latitudes.

is approximately 111 km but this decreases with increasing latitude, by a factor cos ϕ. Since the Mercator projection is cylindrical with meridians remaining equally spaced regardless of distance from the equator, there is an exaggeration of scale in the east–west direction by a factor $1/\cos\phi$ at latitude ϕ. To preserve the shape of regions, the north–south scale must be exaggerated by the same factor. This means that, for example at ±60°, where the scale in both directions is exaggerated by a factor of 2 relative to the equator, areas are exaggerated by a factor of 4. In fact, the exaggeration of scale in the Mercator projection is so severe that the poles cannot be mapped, because they are infinitely far from the equator in the projected coordinate space. Earlier, it was noted that tile 0/0/0 of the web map reference system does not include the whole Earth, and this is why. To give a square tile the maximum latitudes included in tile 0/0/0 are at around $\pm 85°3'4.06''$.

There is no particular reason for choosing this projection as the basis for interactive zoomable web maps, although repeated division into quarters of square tiles is computationally convenient. Indeed, long before tiled web maps (or for that matter the web!) a system along these lines was proposed by Tobler and Chen (1986) using a cylindrical *equal area* projection with standard parallels ±55.654° (see Figure 3.9). The dominance in recent years of Web Mercator has reopened arguments about the political impact of map projections in reinforcing dominant political and cultural ideas concerning the relationship between the global north and south. Fortunately, more recent developments seem likely to result in an interactive 2.5D globe being presented to users at whole Earth scales in many cases, which may slowly correct misconceptions embedded over time by the prevalence of the Mercator projection (see Crampton, 1994).

While the political impacts of various map projections (intended or not) are important, the key issue here is not so much that any given map projection distorts, as that it is well suited for some purposes but not for others. The Mercator projection is *really useful* for navigation, but terrible (in almost all cases) for small-scale thematic mapping. The deeper point, as already implied by different perspectives on space discussed in Chapter 2 (see §Absolute Space in GIS, Chapter 2), is that there

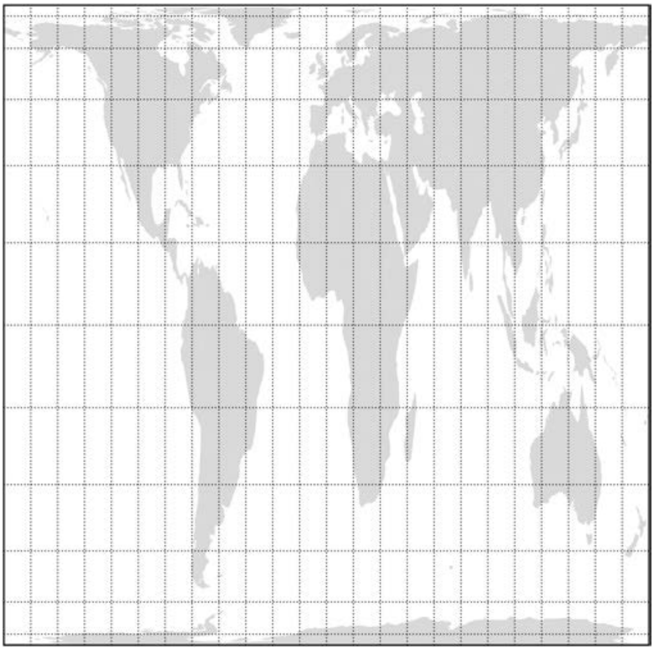

Figure 3.9. Tobler and Chen's equal area cylindrical projection in a square, which was proposed as a potential basis for a recursively subdivided "scheme for the storage of geographic information on a global basis" (1986, p. 370).

is no particular reason to prioritize narrow geometric concerns such as distance, area, and direction in computational representations. The cartograms in Figure 2.2 can be considered examples where the scale is one relating population to area according to some representative fraction in exactly the same way that a more conventional map scale does.

Map projection is a wide-ranging topic central to any serious consideration of cartographic representation. The issue for the present discussion is how a chosen projection can subvert any particular choice of scale in complex ways. Even taken on its own terms, there are no maps—not even topographic maps—that preserve the same representative fraction map scale across their full extent.[8] This can also be understood as reflecting the observer-dependent nature of scale discussed in the theoretical geography literature. Another aspect of the topic to which we return is how different ways of seeing the world imply different map projections (see §Graph Drawings as (Possible) Projections, Chapter 6). Once we understand projections in this way, the potential to expand the concept beyond a narrow focus on geometric fidelity is clear and opens up a wider range of representations in geography to being understood as map projection.

Scale-Dependencies: Resolution and Generalization

At the outset of my discussion of scale I considered the range of spatial extents across which the work of geographers might range, settling on something like 10 orders of magnitude, from (say) grains of sand to (say) continents. Attending to the extremes of this range and meaningfully relating them to one another is close to impossible, necessitating in most cases a narrower focus, so that researchers concerned with particle grain sizes are more likely to do their work at catchment scales or smaller, for example. This points to an important alternative emphasis in thinking about scale—that of *grain* in relation to *extent*. This approach emphasizes the relationship between the finest details or smallest phenomena of interest relative to the overall extent of the areas across which we are studying them. The wider this range, the more difficult (or even impossible) scaling up from processes at one scale to processes at other scales will generally be. Logistical concerns often limit our ability to zoom in arbitrarily closely (i.e., to reduce the grain size) given a particular study extent. These considerations are nicely summarized by Wiens:

---

[8] This claim is technically correct, but in practical terms an exaggeration. For example, assuming that the nominal map scale is correct at the center of a 1:50,000 scale, generously proportioned one meter square map sheet, then at the corners the nominal scale would be off by on the order of only 0.005%, assuming a widely used transverse Mercator projection. The world really is not flat, but locally, flat is a pretty good approximation!

> [e]xtent and grain define the upper and lower limits of resolution of a study; they are analogous to the overall size of a sieve and its mesh size, respectively. Any inferences about scale-dependency in a system are constrained by the extent and grain of investigation—we cannot generalize beyond the extent without [assuming scale-independence], and we cannot detect any elements of patterns below the grain. For logistical reasons, expanding the extent of a study usually also entails enlarging the grain (1989, p. 387).

This statement is made specifically with reference to field-based data collection in ecology, but is equally relevant to the more general challenges of any spatially explicit study.

From computational and geospatial data perspectives, these arguments correspond to concerns with *resolution* and *generalization*. Resolution is usually understood in relation to the smallest elements that can be resolved in imagery data. Generalization refers to the various operations applied to geospatial data to make them suitable for display at different map scales, and often has similar effects to resolution in making details available to viewers or not. Both concepts inhabit similar conceptual terrain, as expressions of the tension between the grain and extent of geospatial data of all kinds.

The resolution of sensor platforms and the data derived from them strongly affect their potential applications. Features smaller than the resolution are undetectable. Features up to twice the resolution may not be reliably detected, and only larger multipixel features can be easily distinguished one from another. Raster data can be conveniently resampled by aggregation to lower resolutions, coarsening it, by averaging pixel values in the original high-resolution layer to pixel values in the new lower-resolution layer. However, the reverse process of disaggregation by interpolation or smoothing cannot recover the original data (see Figure 3.10). Working with multiple data layers, the lowest-resolution layer may constrain the reliability of results to at or around that resolution, although this is not always so. Some phenomena, for example, mean annual temperatures may inherently only make sense at coarse grains. This example shows how spatial resolution interacts with temporal and measurement resolution (Lam & Quattrochi, 1992; Lam, 2019).

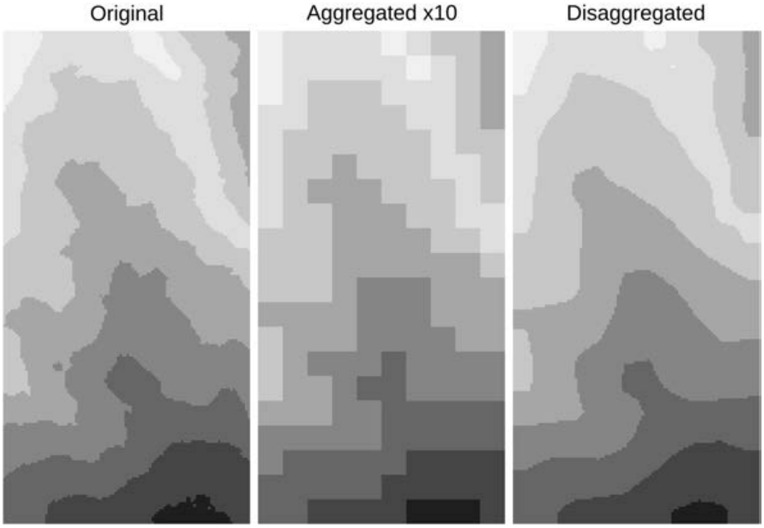

Figure 3.10. An original raster dataset aggregated by a factor of 10, then disaggregated back by interpolation, showing the data loss that results.

The same interaction between spatial, temporal, and measurement scales is true of data attributed to points or polygons. Census polygons at the "block" level (typically meaning around 100 people) will record relatively coarse information about the associated population, such as counts of persons in broad (say) 15-year age ranges. While this might (depending on the jurisdiction) be stipulated by privacy constraints, a thought experiment suffices to show that something more fundamental is going on. Given a group of 100 people, if age group counts by year were reported, there would be high variability among broadly similar census blocks. Some blocks would have zero populations reported at some ages, based solely on the census date and on the birth dates of respondents. National censuses of population happen at long time intervals of 5 or 10 years, and so, even though data will have been collected giving exact ages, reporting it to this precision is only likely to make sense at coarser spatial resolutions, for areas with populations of (say) 5,000 or more. The challenges of dealing with uncertainties in noncensus data collected on populations such as the American Community Survey further emphasize this point (Spielman et al., 2014). In this case the vagaries of sampling mean that small area (high spatial resolution) data are highly unreliable unless they are aggregated over several years of the survey.

Cartographic generalization, long practiced in the production of paper maps, also sits uncomfortably at the point where grain and extent meet. The too often quoted conceit of the one-to-one scale map[9] is intentionally absurd but nevertheless has been and remains a geospatial technological dream, most recently in the form of *digital twins* (Batty, 2018), but going back at least to "mirror worlds" (Gelernter, 1991) if not further. Such a map is impossible and absurd. The point of maps is not to mirror the world, but to represent it in specific ways for particular purposes. Including everything in small-scale maps is impossible; even including everything in a notional 1:1 scale map is impossible (that's the point of the much quoted parables). In small-scale maps, the first line of defense is selecting what to include or exclude, although that only partly addresses the challenge of simplifying the map sufficiently for it to be useful. In addition, the cartographic twins of things in the world are generalized so that they remain legible at a small scale, or elements are removed completely to avoid clutter and confusion.

Generalization is usually considered to consist of combining a variety of operations (Raposo, 2017, provides an excellent summary), among them simplification, aggregation, selection, and exaggeration (see Figure 3.11). Routine application of any one of these operations *might be* relatively straightforward (although often it is not), but combining several operations across multiple datasets to produce an overall effect in a map is extremely complex. A significant source of difficulties is anticipating the interactions among different data layers. For example, it is not a

---

[9] Variously attributable to Lewis Carroll in *Sylvie and Bruno Concluded,* Luis Borges in *On Exactitude in Science*, Umberto Eco in *On the Impossibility of Drawing a Map of the Empire on a Scale of 1 to 1*, and approvingly discussed by Jean Baudrillard (1994) in *Simulacra and Simulation.*

Figure 3.11. Map generalization showing a simplified layer after processes of selection, simplification, and aggregation—look closely!

simple matter to generalize a road layer for a particular scale of presentation on page or screen. But it becomes significantly more complicated when generalization of roads has implications for how building or parcel, or any of the other layers that interact with the road layer, should be represented. Thus, in the same way that the resolution of geospatial data is a complex interaction of the effects of spatial, temporal, and measurement scales (especially of classification), generalization reveals the subtleties of the impact of scale on giscience in practice. And indeed, early success (Douglas & Peucker, 1973) quickly gave way to recognition that automated generalization is far from trivial (Brassel & Weibel, 1988) and the emergence of multiscale maps (i.e., web maps) has intensified the design challenge (Roth et al., 2011). Even on narrowly technical grounds then, it seems likely that ongoing maintenance by different actors, for different purposes, of somewhat related, nonidentical(!) digital cousins will prevail, rather than that monolithic singular digital twins will emerge other than in specific limited domains.[10]

A further scale-dependent effect, much discussed in giscience is the modifiable areal unit problem (MAUP), which we consider in detail in §The Arbitrariness of Boundaries, Chapter 5.

---

[10] Although there can be little doubt that sweeping claims will continue to be made for the political, economic, and social efficacy of digital twins and mirror worlds!

THE SALIENCE OF SCALE

Scale is considered central by geographers of all stripes, even human geographers who hanker after a geography without scale. The salience of scale is also recognized in giscience. This differs from the case of space, where having settled on absolute space in most platforms, what is lost in that approach disappears from view. Chosen approaches to space are embedded in data formats, making their further diffusion and uptake even more likely. In this way, decisions about how to represent space are pre-encoded, and unless geographers have the capacity to develop their own formats, the limited spatial imaginary of these formats becomes a taken-for-granted default.

By contrast, scale enters the picture at every stage of working with geospatial data. Some of these stages are hidden. The selections, simplifications, and omissions in a particular dataset, arising from its scales of collection and collation, may pass unnoticed. But, how things come in or out of focus at different scales, as data are manipulated, displayed, aggregated, or combined with other data, is often highly visible. Further, these effects are explicitly considered as concerns of giscience (see, e.g., Lloyd, 2014). The impacts of scale on data collection, and subsequent processing, storage, and visualization, are too significant to escape attention as merely technical challenges demanding solutions.

How these technical challenges play out surely deserves closer attention in geographical thought more generally. Considering what is recorded and mapped (or not) at different scales might even shed light on how scales are socially constructed, and on some of the ways that scale operates (or not) politically. Interactive web maps are a recent development in this context, and it can be instructive to spend time *slowly* zooming in to such maps to see what features appear (or disappear) at successive levels of detail. As we will see in the next two chapters, what is included or not, and in what guise—generalized to a point, or rendered in more detail—is politically charged, and understanding these choices as purely technical dimensions of scale hides those politics. The scale dependencies of geospatial data are after all more than technical matters.