# 2

# *Georeferencing*

## 2.1 Needs for Metric Georeferencing

Geospatial data differ from other types of data in that data are geographically referenced. One must understand and master coordinate systems to become functionally adept in the use and application of GIS. Locations are uniquely assigned to data in reference to coordinate systems in GIS. Georeferencing enables geospatial data of disparate sources to be integrated. Without this integration of data, relationships among different variables will not be holistically explored, and high-value information based on multiple criteria (such as suitable sites) will not be created.

Consider two methods for identifying a location on the Earth's surface. The location of the White House can be expressed in a street address or latitude/longitude as follows:

- 1600 Pennsylvania Ave NW, Washington, DC 20500
- 38°53′52.6452″ N 77°2′11.6160″ W

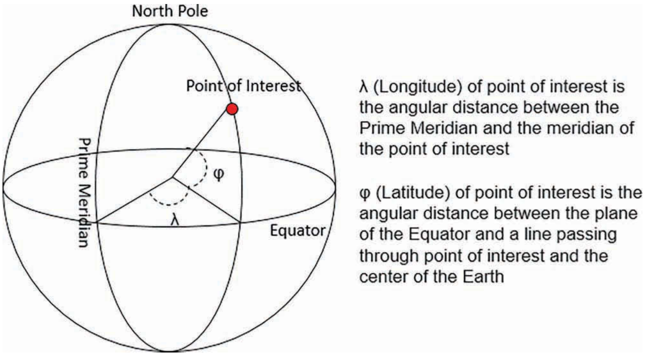
The former is a relative and nonmetric method of georeferencing whereas the latter is an absolute and metric method of georeferencing. A street address is identified relative to a street network. Thus, the location of the building would change if the street network changed. Two people in the same building cannot be distinguished if a street address is used as a method of georeferencing. In contrast to a street address, a location referenced by latitude and longitude is not measured with respect to geographic features that might change (absolute), and can be identified in an infinitely fine spatial resolution (metric). This enables any location to be identified unambiguously and precisely. One of the aspects that distinguish geospatial data from other data types is that coordinates of spatial entities are explicitly coded in a database. Such data are geographically referenced with respect to a universally recognized coordinate system in contrast to spreadsheets, photos, or computer-aided design files. To understand coordinate systems that are used in GIS, it is necessary to understand datums and map projections.

## 2.2 Understanding Datums

Just like street addresses are measured with respect to street networks, latitude and longitude are measured with respect to the surface of the Earth. Consider how your height (one-dimensional value) is measured. Height is measured with respect to the floor. You can measure the location of trees (two-dimensional value) as $(x, y)$ coordinates with respect to the origin (0, 0) in a local Cartesian coordinate system if you presume the surface is flat. The presumption is not valid any more if the goal is to measure locations in the entire Earth's curved surface. To measure locations on the Earth's surface, consider a three-dimensional earth model as a base that forms the theoretic component of a geodetic (or map) datum.

Figure 2.1 shows how latitude and longitude are determined. Latitude and longitude are measured in angular units, not linear units (such as meters) because latitude and longitude are measured in reference to a three-dimensional earth model. If you assume the Earth's surface is a sphere, latitude is the angular distance ($\varphi$) between the plane of the equator and a line passing through the point under investigation and the center of the Earth. Latitude ranges from 0° at the equator to 90° at the poles. The latitude 38°53′52.6452″ N means the location is 38°53′52.6452″ north of the equator. Longitude is the angular distance ($\lambda$) between the prime meridian and the meridian of the point under investigation. Longitude ranges from 0° to 180°. The longitude 77°2′11.6160″ W means the location is 77°2′11.6160″ west of the prime meridian.

In geodetic surveying and GIS, a sphere is not usually used as an Earth model because a sphere is not the most accurate way to characterize the Earth. After measuring the shape and size of the Earth for many years, geodesists realized that an oblate ellipsoid represents the Earth more accurately than a sphere. Accumulated centrifugal force caused by the Earth's rotation makes



λ (Longitude) of point of interest is the angular distance between the Prime Meridian and the meridian of the point of interest

φ (Latitude) of point of interest is the angular distance between the plane of the Equator and a line passing through point of interest and the center of the Earth

**FIGURE 2.1**
Determining latitude and longitude.

the equatorial axis (semimajor axis) longer than the polar axis (semiminor axis). Hence, geodetic coordinates are measured with respect to an ellipsoid. Figure 2.2 illustrates the difference between the Earth as a sphere and the Earth as an oblate ellipsoid in reference to how latitude is determined.

Although it would be ideal to have one ellipsoid for the entire world, this is not reality. Different countries have developed their own ellipsoids to best fit their local areas (Table 2.1). For instance, GRS 80, the ellipsoid used for mapping North America, models the Earth as the elliptical surface that has semimajor axis 6,378,137 m and semiminor axis 6,356,752.3141 m. Different countries have also developed a network of surveyed control points that form an empirical component of datums (Figure 2.3). Different specifications of the Earth also reflect surveying technology at the time of surveying. The United States previously used Clarke 1866 as a reference ellipsoid. This ellipsoid forms the basis of the North American Datum of 1927 (NAD27), which was phased out in the early and mid-2000s. WGS 84 is the reference system for GPS and for mapping the entire world. Latitude and longitude collected with a GPS-equipped device (such as a smartphone) is measured using WGS 84. Pseudo Mercator—widely used as a coordinate system used for web mapping—is based on WGS 84 as the web map is designed to show any location in the world unlike other projection that is usually designed to show particular areas.

Altitude is measured with respect to a geoid that approximates the mean sea level. Unlike a sphere or an oblate ellipsoid, a geoid is not a mathematical surface. The ellipsoid serves as the basis for horizontal datums, and a geoid serves as the basis of vertical datum. Coordinates reference different datums depending on what jurisdictions created the geospatial data. Figure 2.4 shows the spatial reference section of the metadata in XML format. The metadata shows that latitude and longitude are measured with respect to the North American Datum of 1983, using GRS 80 as an ellipsoid. In sum, datums consist of control points as empirical components and ellipsoids (geoids) as theoretical components for horizontal (vertical) measurement and are used in coordinate systems as reference surfaces.

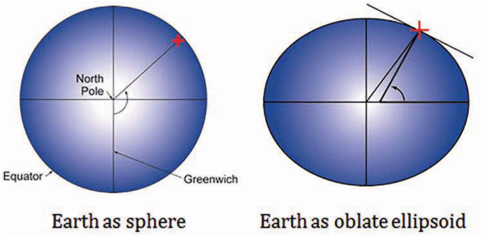
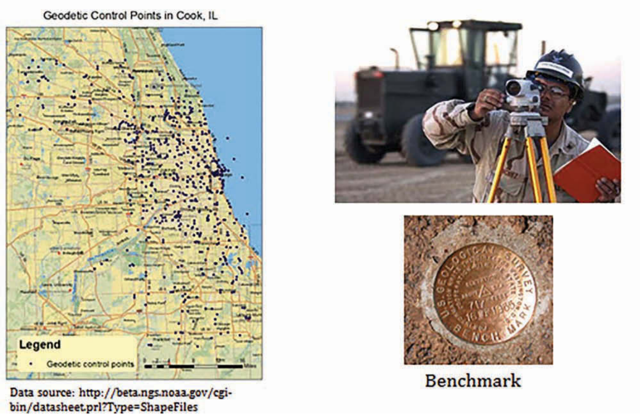


**FIGURE 2.2**
Earth as sphere vs. Earth as oblate ellipsoid. (Reprinted with permission from Paul Longley, Michael Goodchild, David Maguire, and David Rhind, *Geographic Information Systems and Science*, 2nd Edition, Wiley, 2005.)

**TABLE 2.1**

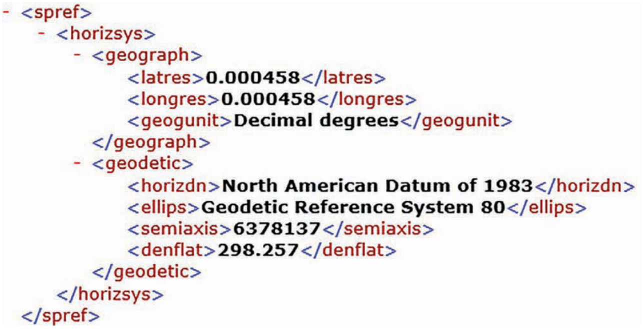
Selected common ellipsoids used for regional, national, and international mapping and GIS applications

| Ellipsoid | Semimajor Axis (m) | Semiminor Axis (m) | Use |
|---|---|---|---|
| Australian National 1966 | 6,378,160 | 6,356,774.719 | Australia |
| Clarke 1866 | 6,378,206.4 | 6,356,584.467 | North America |
| International 1924 | 6,378,388 | 6,356,911.946 | Remaining parts of the world |
| GRS80: Geodetic Reference System 1980 | 6,378,137 | 6,356,752.3141 | Adopted in North America for 1983 Earth-centered coordinate system |
| WGS84: World Geodetic System 1984 | 6,378,137 | 6,356,752.3142 | Used with GPS and by NASA |
| Normal Radius of the Earth | 6,370,997 | 6,370,997 | A perfect sphere |



Data source: http://beta.ngs.noaa.gov/cgi-bin/datasheet.prl?Type=ShapeFiles

**FIGURE 2.3**
Geodetic control points that are established through surveying.



```
- <spref>
  - <horizsys>
    - <geograph>
        <latres>0.000458</latres>
        <longres>0.000458</longres>
        <geogunit>Decimal degrees</geogunit>
      </geograph>
    - <geodetic>
        <horizdn>North American Datum of 1983</horizdn>
        <ellips>Geodetic Reference System 80</ellips>
        <semiaxis>6378137</semiaxis>
        <denflat>298.257</denflat>
      </geodetic>
    </horizsys>
  </spref>
```

**FIGURE 2.4**
Spatial reference section in the geospatial metadata.

## 2.3 Understanding Map Projection

Map projection is another key component of coordinate systems that must be understood. Map projection can be considered as the process of flattening the Earth's curved surface. Although the Earth is curved, most of media that represent the features of the Earth (such as maps and computer screens) are flat. This means that positions on the Earth's curved surface should be rendered (projected) onto a flat map surface (Figure 2.5). Map projection refers to mathematical operations that convert longitude and latitude defined on the curved Earth onto $(x, y)$ coordinates in a flat (Cartesian) coordinate system.

One way to understand how features on the curved Earth are rendered onto a flat map surface is to visualize projecting light from a lightbulb at the center of a miniature globe onto a transparent sheet of paper that wraps around it (Figure 2.6). The graticule (gridlines made of parallels and meridians) drawn on the curved Earth will be projected onto the upwrapped sheet of paper (a flat map surface). The miniature globe is called the reference globe, and the sheet of paper is called the developable surface. Mapping the contiguous United States on a flat surface can be understood as the process of drawing the contiguous United States on the reference globe, projecting light through the globe onto the developable surface, and unwrapping the developable surface.

The size and shape of the land shown in maps are artifacts of map projection. Figure 2.7 shows the world in Mercator projection and Gall-Peters projection. The Mercator map on the left shows the shape accurately but distorts size greatly toward high latitude. The Peters map on the right portrays area accurately but distorts shape. For example, you can determine the true shape of Alaska using a Mercator map and the true size of Alaska relative to other features using a Peters map. A caveat of using map projection is that at

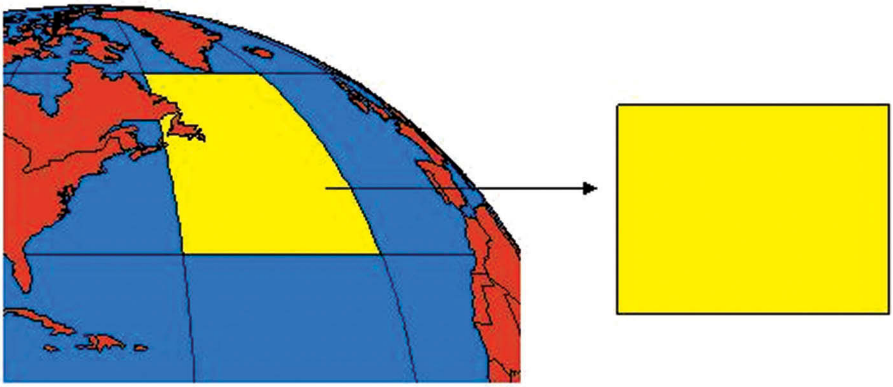


FIGURE 2.5
Map projection flattens the curved Earth's surface.

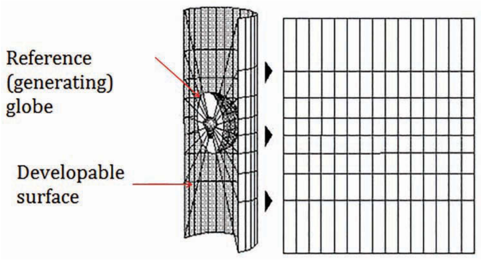


**FIGURE 2.6**
How map projection is constructed. (Adapted from ESRI, Understanding Map Projections, Retrieved from http://downloads2.esri.com/support/documentation/ao_/710Understanding_Map_Projections.pdf.)

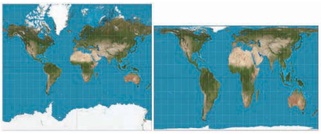


**FIGURE 2.7**
The world in Mercator projection vs. Gall-Peters projection.

least one geometric property is distorted. A Mercator map may suggest that Alaska and Brazil are similar in size, but Brazil is nearly five times larger than Alaska. The size of Alaska might be true to scale in the Peters map, but the shape of Alaska will be greatly distorted.

Thousands of map projections have been devised to serve different purposes and thus it is useful to classify them. There are three families of map projection: azimuthal, cylindrical, and conic. They use developable surface in different shapes. If the shape of a developable surface is a plane, cylinder, or cone, the map projection is called azimuthal, cylindrical, and conic, respectively, as shown in Figure 2.8. Cylindrical projection can show the entire world. Conic projection is good for showing areas with a longer east-west extent in mid-latitude. Azimuthal projection is commonly used to show hemispheres.

Map projection can also be classified in terms of a geometric property preserved in a map. The type of map projection that preserves shape true to scale is called conformal. The type of map projection that preserves area true to scale is called equal-area. Figure 2.9 shows how circles drawn on the reference globe change after map projection. The shape of the circles does not change in a Mercator map, which is conformal. The shape of circles changes from orthogonal circles to skewed circles, but size is preserved in Mollweide which is equal-area.

Distortion is unavoidable in projected maps that represent features on the curved Earth (Figure 2.10). Since the amount of distortion affects the

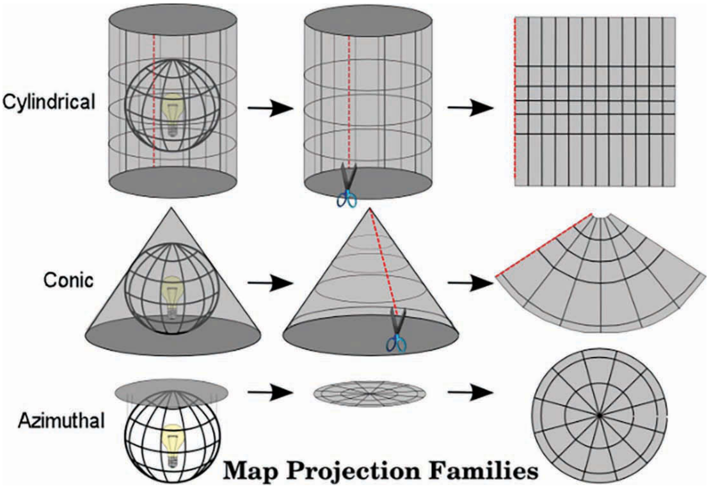


**FIGURE 2.8**
Family of map projection. (From QGIS 2.8 Documentation, retrieved from https://docs.qgis.org/2.8/en/docs/gentle_gis_introduction/coordinate_reference_systems.html)

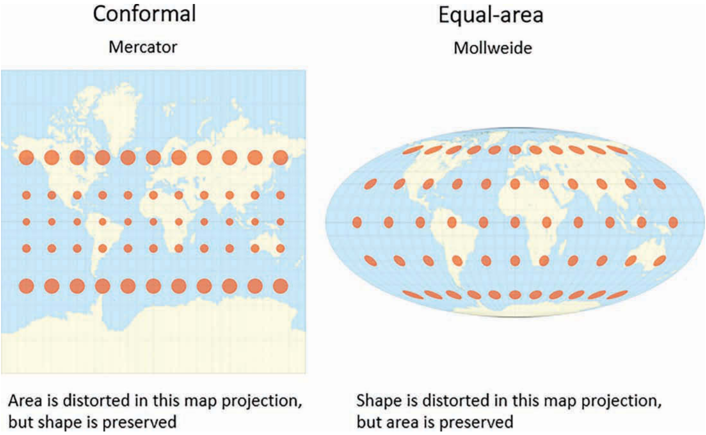

Conformal
Mercator
Equal-area
Mollweide
Area is distorted in this map projection, but shape is preserved
Shape is distorted in this map projection, but area is preserved

**FIGURE 2.9**
Properties of map projection. (From Eric Gaba, Wikimedia Commons User: Sting, retrieved from https://commons.wikimedia.org/wiki/File:Tissot_indicatrix_world_map_Mercator_proj.svg and https://commons.wikimedia.org/wiki/File:Tissot_indicatrix_world_map_Mollweide_proj.svg.)

positional accuracy of maps, it is important to understand the patterns of distortion in map projection. Figure 2.11 shows that the line that intersects a developable surface and reference globe—called the standard parallel—has the least distortion. The farther from standard parallel the more distortion present (i.e., less accurate). Letting the developable surface cut through the

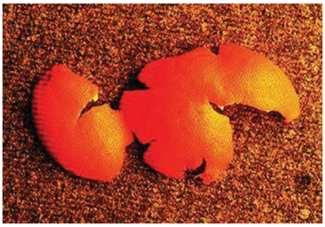


**FIGURE 2.10**
Distortion is unavoidable in map projection.

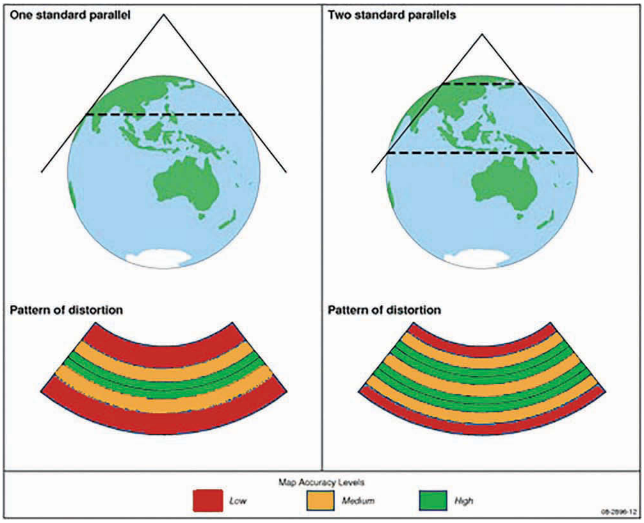
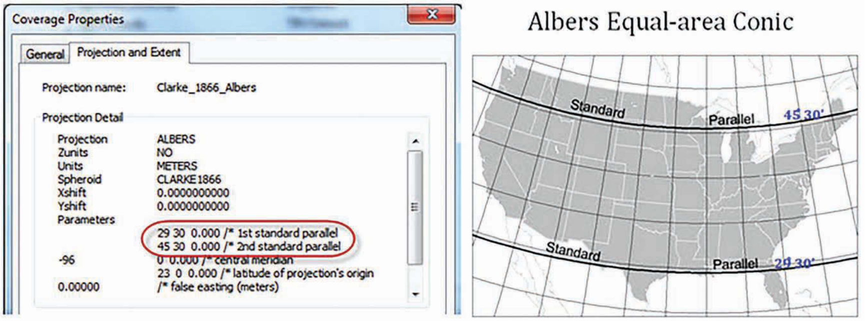


**FIGURE 2.11**
Pattern of distortion in map projection. (From Anzlic Committee on Surveying & Mapping, retrieved from https://www.icsm.gov.au/education/fundamentals-mapping/projections.)

reference globe results in two standard parallels (known as the secant case). Standard parallels are chosen so that the overall distortion is minimized in the areas to be projected. The same principle applies to other families of map projection.

The contiguous United States is commonly depicted in a type of map projection known as Albers Equal-Area Conic (Figure 2.12), which uses a cone as a developable surface (conic family) that cuts through the reference globe around two standard parallels (secant case) and preserves area true to scale (equal area). Figure 2.12 also shows that the first standard parallel is in 29°30′ and the second standard parallel is in 45°30′. These two lines of parallels fall within the areas of interest (contiguous United States). The aim is to minimize the overall distortion, which is lowest around standard parallels. Having two standard parallels within an area of interest reduces distortion more than having one standard parallel in the middle of an area of interest.

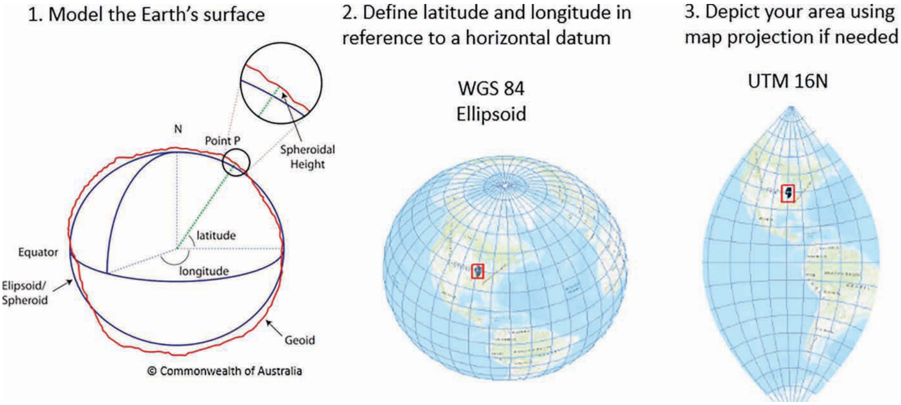


```
Coverage Properties
General | Projection and Extent

Projection name:        Clarke_1866_Albers

Projection Detail
  Projection            ALBERS
  Zunits                NO
  Units                 METERS
  Spheroid              CLARKE1866
  Xshift                0.0000000000
  Yshift                0.0000000000
  Parameters
                        29 30  0.000 /* 1st standard parallel
                        45 30  0.000 /* 2nd standard parallel
  -96                   0  0.000 /* central meridian
                        23  0  0.000 /* latitude of projection's origin
  0.00000               /* false easting (meters)
```

**FIGURE 2.12**
Parameters of Albers Equal-area map projection showing the contiguous United States; from Copyright © 2018 Esri, ArcGIS, ArcMap, and the GIS User Community. All rights reserved. With Permission.

## 2.4 Coordinate Systems for GIS

Coordinate systems used in GIS can be understood in terms of a progressive flattening of the Earth's irregular shape (Figure 2.13). First, the rugged Earth's surface is modeled as a geoid to measure altitude and as an ellipsoid to measure latitude and longitude. Second, latitude and longitude are defined with respect to a horizontal datum that fits the area of interest. Third, if needed, appropriate parameters of map projection (such as family, property, and standard parallels) are selected to depict an area of interest in a flat surface as realistically as possible while meeting the requirements of the map being created. GIS include tools that can display maps in any coordinate system on the fly or change coordinate systems permanently as a preliminary to spatial analysis based on Euclidean geometry, which assumes coordinates on a flat surface.



1. Model the Earth's surface
N
Point P
Spheroidal Height
latitude
longitude
Equator
Elipsoid/ Spheroid
Geoid
© Commonwealth of Australia
2. Define latitude and longitude in reference to a horizontal datum
WGS 84 Ellipsoid
3. Depict your area using map projection if needed
UTM 16N

**FIGURE 2.13**
Coordinate systems in terms of progressive flattening of the Earth's irregular surface.

There are two types of coordinate systems used in GIS—the geographic coordinate system (GCS) and the projected coordinate system (PCS). The GCS refers to coordinates defined on the curved Earth, and the PCS refers to coordinates defined on the flat (projected) surface as illustrated in Figure 2.14. A GCS is simply an alternative name for latitude and longitude, which are defined on the basis of a particular datum. (Note that the same latitude and longitude may differ depending on datum.) Longitude and latitude on the curved Earth may be rendered onto a flat surface using map projection, resulting in (*x*, *y*) in a Cartesian (projected) coordinate system if needed. A PCS is defined on the basis of a GCS. Many publicly available geospatial data are stored in GCS to allow the data to be displayed and projected as needed.

Since there is no way to display geospatial data stored in a GCS (a three-dimensional coordinate system) on a flat computer screen, it is displayed by default using Plate Carrée in a GIS. Plate Carrée represents one degree of latitude and longitude in one linear unit on a flat surface. While a line of longitude has constant spacing, one degree of spacing in a line of latitude gets smaller toward high latitude and becomes zero at the pole. The zero dimension at the pole is forced to expand to the equator, resulting in distortion in shape (extended in east-west direction) and size (larger) in high latitude. Figure 2.15 shows the contiguous United States in Plate Carrée on the left and in Albers Equal-Area Conic with standard parallels falling within an area of interest after the layer is projected on the right.

Latitude and longitude (or GCS) are commonly expressed in degrees, minutes, and seconds (DMS) format. In a GIS, decimal degrees are widely used to express latitude and longitude. Decimal degree expresses latitude and longitude in fractional degree by converting minutes and seconds to

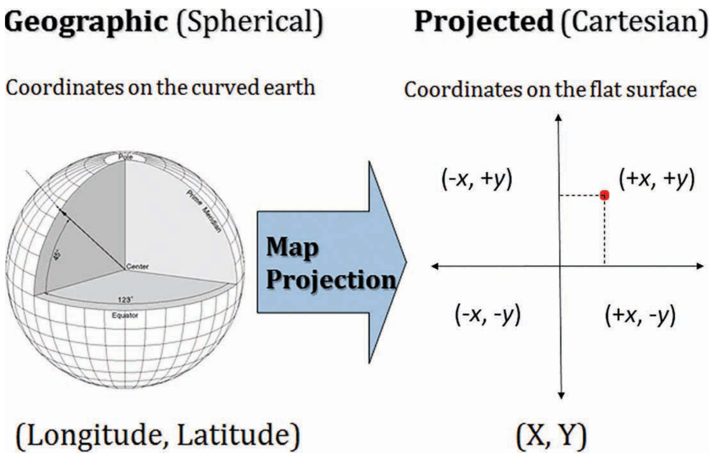


**FIGURE 2.14**
Geographic coordinate system vs. projected coordinate system.

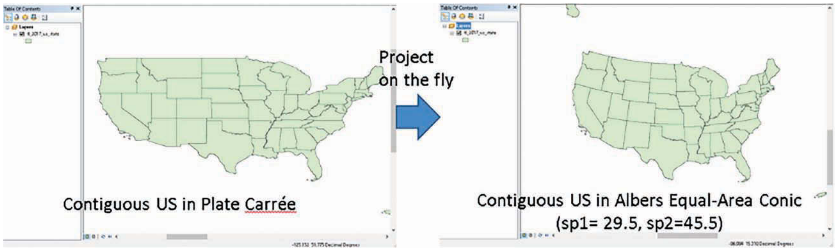

Project
on the fly
Contiguous US in Plate Carrée
Contiguous US in Albers Equal-Area Conic
(sp1= 29.5, sp2=45.5)

**FIGURE 2.15**
Data stored in GCS displayed in Plate Carrée vs. Albers Equal-Area Conic.

degrees. To convert minutes to degrees, you divide the number of minutes by 60. To convert seconds to degrees, you need to divide the number of seconds by 3600. For example, latitude 42°55′26.30″ N is equal to 42.92397 (= 42 + 55/60 + 26.30/3600) in decimal degrees. Decimal degree takes a negative value if the direction of latitude is south or the direction of longitude is west. Figure 2.16 shows longitude, latitude in quadrants of the geographic grid depicted in Plate Carrée. Longitude and latitude in the northern and eastern hemisphere will have positive values (+, +) in decimal degrees. Longitude and latitude in the southern and western hemisphere have negative values (–, –) in decimal degrees. For instance, longitude 87°39′19.67″ W is equal to –87.65546 (= –1 × (87 + 39/60 + 19.67/3600)) in decimal degrees since the direction of longitude is west.

One of widely used PCS in GIS is the Universal Transverse Mercator (UTM). The UTM was initially developed by the US Army Corps of Engineers and has been adopted by various mapping agencies as a standard PCS for topographic (reference) mapping all around the world. Many publicly available raster data (such as digital orthophoto quadrangles) are cast to the UTM. The UTM divides the world into 60 zones by six-degree longitudinal strips (Figure 2.17). The zone numbering increases eastward from the international dateline (180° longitude). While Mercator uses a cylindrical developable

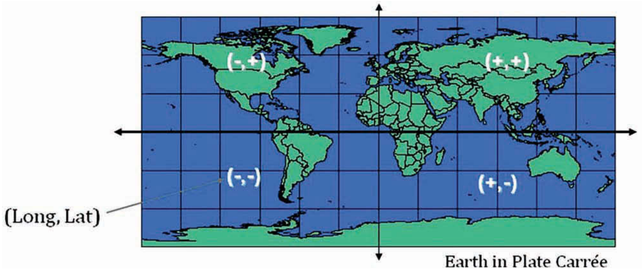


**FIGURE 2.16**
Signs of decimal degree in quadrants of geographic grid.

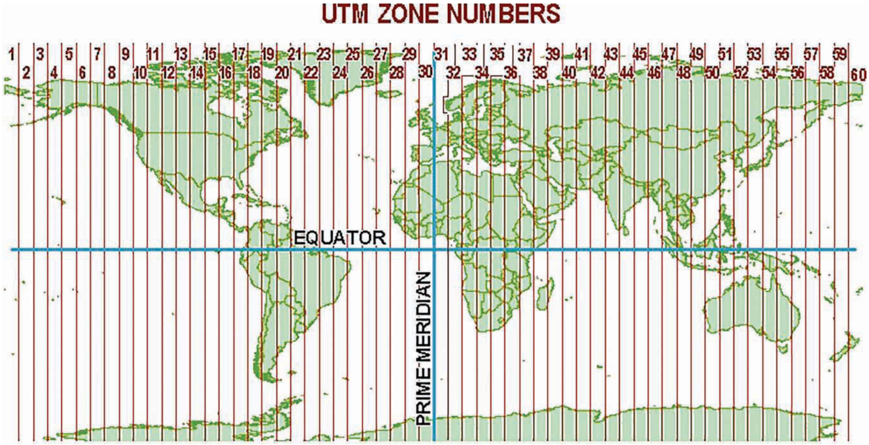


**FIGURE 2.17**
UTM zones of the world. (Reprinted with permission from National Geospatial-Intelligence Agency, retrieved from http://earth-info.nga.mil/GandG/coordsys/grids/grid1.html.)

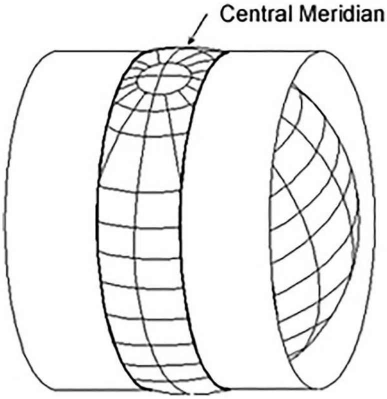


**FIGURE 2.18**
How UTM zones are defined. (From United Nations Statistics Division, UNGEGN-ICA webcourse on Toponymy, retrieved from https://unstats.un.org/unsd/geoinfo/UNGEGN/docs/_data_ICAcourses/_HtmlModules/_Selfstudy/S06/S06_03.html.)

surface placed around the equator, the UTM uses six-degree longitudinal strips as a developable surface placed around the meridian in the middle of each UTM zone (known as the central meridian; Figure 2.18).

Each UTM zone has its own Cartesian coordinate system, where the equator serves as an *x* axis and central meridian serves as a *y* axis. A position can be located using UTM zone and (*x*, *y*) offset from an origin of the Cartesian coordinate system. A meter is a measurement unit for the UTM. To avoid negative coordinate values, an origin shifts westward (500,000 m) and shifts southward (1,000,000 m) in the southern hemisphere. The amount of shift in the west and south directions is referred to as false northing and false easting, respectively. For instance, UTM coordinates 16N (445572, 4641636) means the coordinates are located 445,572 m east and 4,641,636 m north of the origin in UTM zone 16N.

Municipalities and jurisdictions govern local and regional assets such as land records and utilities and thus have a need for precisely measuring the locations of these assets using a specific coordinate system. State Plane Coordinate (SPC) systems were developed to meet the needs of administrative units to precisely survey and inventory properties in the United States. SPC systems divide the United States into zones that follow state boundaries and then subdivide each zone into smaller zones that are designated mostly by direction (West, East, North, South; Figure 2.19). For instance, Illinois has two SPC zones—IL East and IL West.

An SPC system is a projected coordinate system. Two map projections—Lambert conformal conic and transverse Mercator—are used for the SPC system (Figure 2.20). A developable surface cuts through two standard parallels in Lambert conformal conic, making it suitable to map areas that have a greater east-west extent. Transverse Mercator is good for showing areas with a great south-north extent as distortion is minimized around the central meridian. For instance, Kansas (with a greater east-west extent) uses Lambert conformal conic, and Illinois (with a greater south-north extent) uses transverse Mercator. The SPC system also use false northing and false easting to avoid negative coordinate values.

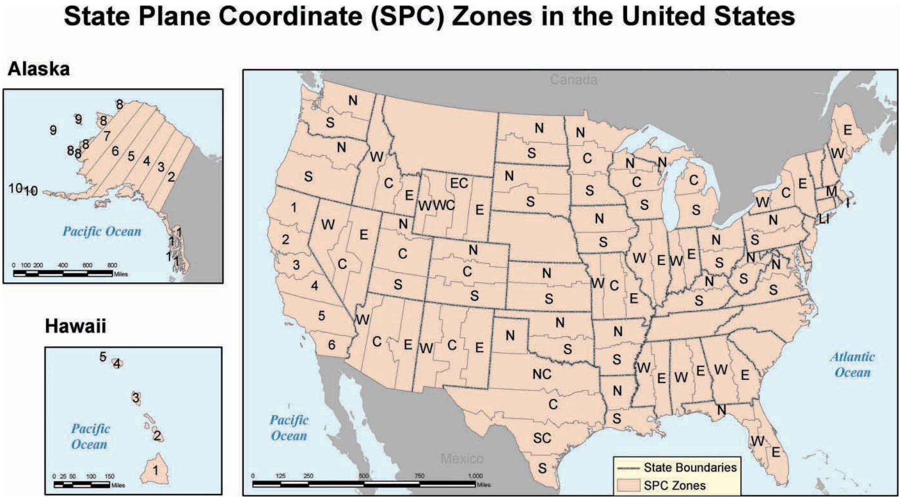


**FIGURE 2.19**
SPC system zones.

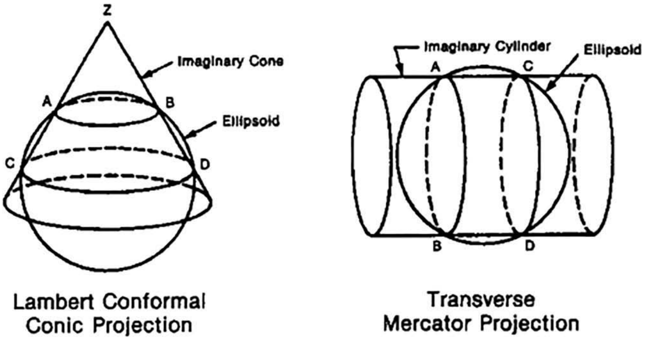

Z
Imaginary Cone
A
B
Ellipsoid
C
D
Lambert Conformal Conic Projection
Imaginary Cylinder
Ellipsoid
A
C
B
D
Transverse Mercator Projection

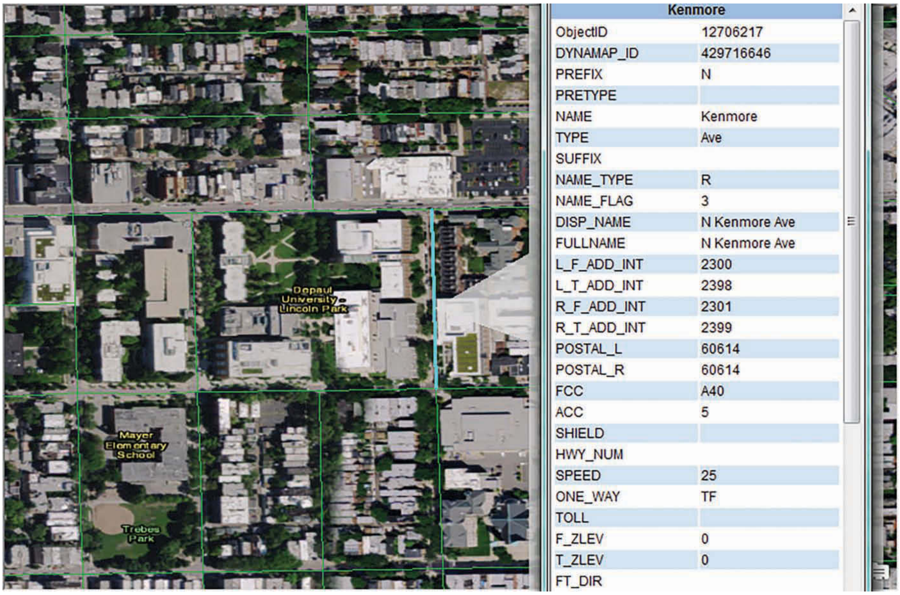
**FIGURE 2.20**
Map projections used for SPC systems. (From NOAA, retrieved from https://www.ngs.noaa.gov/PUBS_LIB/ManualNOSNGS5.pdf.)

## 2.5 Address Geocoding

As locations are commonly identified by street addresses, they often need to be converted to metric coordinates such as latitude and longitude. A standard GIS software comes with address geocoding. Geocoding refers to the process of turning the textual descriptions of locations (e.g., street addresses, place names) to metric descriptions of locations (e.g., decimal degrees). Provided that addresses are defined in reference to a street network, address geocoding requires three components: (a) input table containing street addresses; (b) reference database (transportation network) against which addresses can be matched; and (c) an algorithm that determines how addresses will be matched against the reference data. Today geocoding is commonly performed using an online geocoding service as it is convenient to store a large volume of reference data with frequent updates on a server.

For address geocoding to work, street addresses should be formatted in the order of housing number, prefix direction (if any), street name, street type, suffix direction (if any), city, state, and zip code (e.g., 1234 N. Main St, Chicago IL 60012) in the case of US streets. A reference database should be comprised of a road segment that contains the street name, address range, adjacent zip code, etc., in a predefined format. Figure 2.21 shows that a selected road segment falls within the address range between 2300 and 2399 in N Kenmore Ave. The fields named L_F_ADD_INT and L_T_ADD_INT contain the beginning and ending house number on the left-hand side (at the digitizing order) of the street centerline, respectively.

A geocoding algorithm goes through several steps—parsing and standardization, matching, and location assignment. It first splits street addresses into standardized values of components. For example, the street address "2350



Kenmore

| | |
|---|---|
| ObjectID | 12706217 |
| DYNAMAP_ID | 429716646 |
| PREFIX | N |
| PRETYPE | |
| NAME | Kenmore |
| TYPE | Ave |
| SUFFIX | |
| NAME_TYPE | R |
| NAME_FLAG | 3 |
| DISP_NAME | N Kenmore Ave |
| FULLNAME | N Kenmore Ave |
| L_F_ADD_INT | 2300 |
| L_T_ADD_INT | 2398 |
| R_F_ADD_INT | 2301 |
| R_T_ADD_INT | 2399 |
| POSTAL_L | 60614 |
| POSTAL_R | 60614 |
| FCC | A40 |
| ACC | 5 |
| SHIELD | |
| HWY_NUM | |
| SPEED | 25 |
| ONE_WAY | TF |
| TOLL | |
| F_ZLEV | 0 |
| T_ZLEV | 0 |
| FT_DIR | |

**FIGURE 2.21**
Reference data for address geocoding; from Copyright © 2018 Esri, ArcGIS, ArcMap, and the GIS User Community. All rights reserved. With Permission.

North Kenmore Avenue, Chicago IL 60614" is parsed into standardized components "2350 N Kenmore Ave, Chicago IL 60614" so that Avenue and Ave are treated equally. Second, it finds potential matches between an input street address and a road segment in a reference database by calculating a score that measures how well input data matches candidate road segments in reference data. The road segment with the highest score such as "2300–2399 N Kenmore Ave, Chicago, 60614" would be found as a match. Third, if a match whose score exceeds a minimum score is found, the location is estimated with the assumption that parcel size is equal along the matched road segment. That is, 2350 N Kenmore will be located in the middle of the matched road segment within the address range of 2300 to 2399.

Thus, the quality of address geocoding results depends on the quality of the input table, reference database, and geocoding algorithm. Geocoding can be customized as there are multiple parameters that can be adjusted including the minimum required score, misspelling sensitivity (how much to penalize typos in input table), and offset value (how far the output point can be located from the street centerline). GIS software provides interactive and manual geocoding tools to help deal with cases where either multiple matches or no match is found. Although network-based address geocoding is the most common, it may not be adequate for use cases that require highly precise geocoding results for emerging applications like self-driving.