BCB 503 GEOSPATIAL ANALYSIS WORKSHOP

INTRODUCTION AND CONCEPTS

SPRING 2020

IMCI
Institute for Modeling Collaboration and Innovation

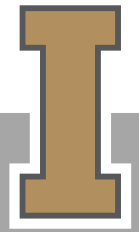
BCB503 Geospatial Workshop

WELCOME

erichs@uidaho.edu
208.310.2140

Overview

- Welcome to the quarantine zone!
- Instructor Introductions
- Zoom technical issues and setup
- Logistics and schedule for next two days
- Get to know each other exercise



IMCI
Institute for Modeling Collaboration and Innovation

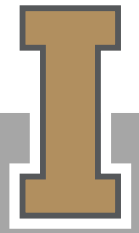
BCB503 Geospatial Workshop

WELCOME

erichs@uidaho.edu
208.310.2140

Overview

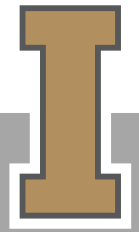
- Relax and let's have fun! We aren't testing you. This is about gaining knowledge you can use as a scientist.
- Be respectful, professional and kind. No inappropriate language or visuals.



IMCI
Institute for Modeling Collaboration and Innovation

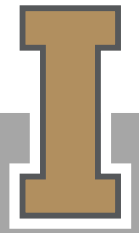
ADDITIONAL PREPARATIONS

- R and RStudio should be installed, with appropriate libraries
- Able to follow along with lessons from class web site (using a web browser)
  - https://erichseamon.github.io/2020-03-26-uidaho-geospatial/
- We will walk thru downloading data, but if you already have done that, great!
- **Be prepared for technical snafus. It's going to happen. Let's try to roll with it.**



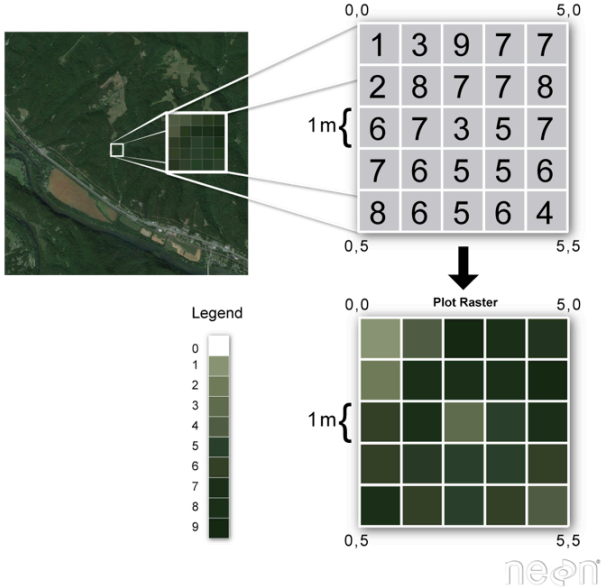
INTRODUCTION TO RASTER DATA

- Raster vs. Vector data
- Why do we store data in a raster format? Pros and cons
- Continuous vs. Categorical rater data. What is classification



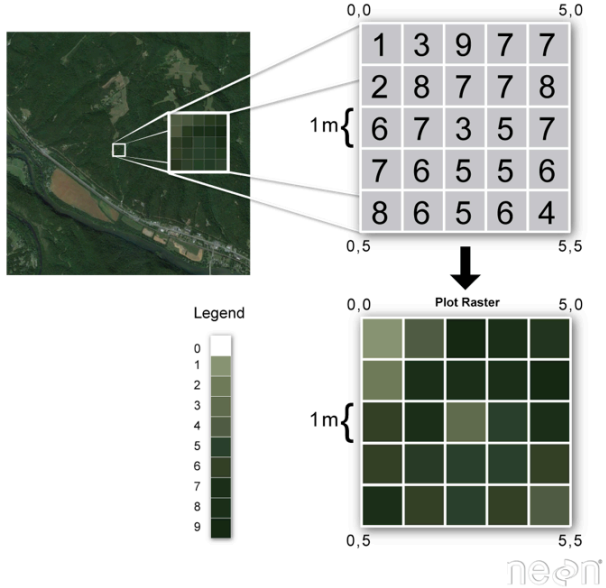
INTRODUCTION TO RASTER DATA

- The two primary types of geospatial data are **raster** and **vector** data.
- Raster data is stored as a grid of values which are rendered on a map as pixels. Each pixel value represents an area on the Earth's surface.
- Vector data structures represent specific features on the Earth's surface and assign attributes to those features.



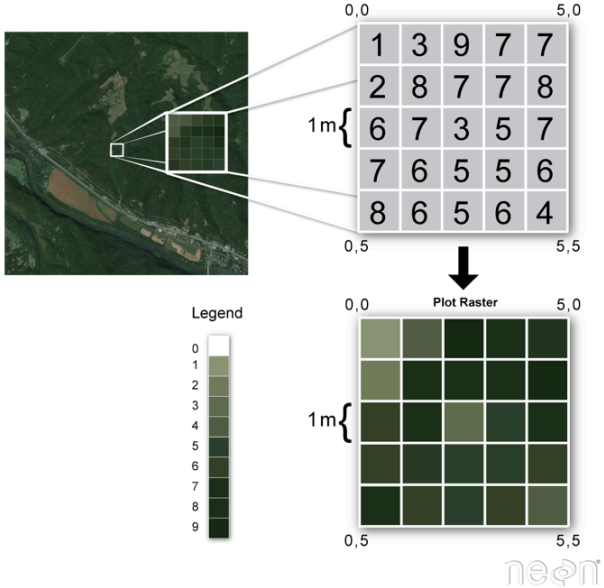
INTRODUCTION TO RASTER DATA

- Raster data is any pixelated (or gridded) data where each pixel is associated with a specific geographical location.
- The value of a pixel can be continuous (e.g. elevation) or categorical (e.g. land use).

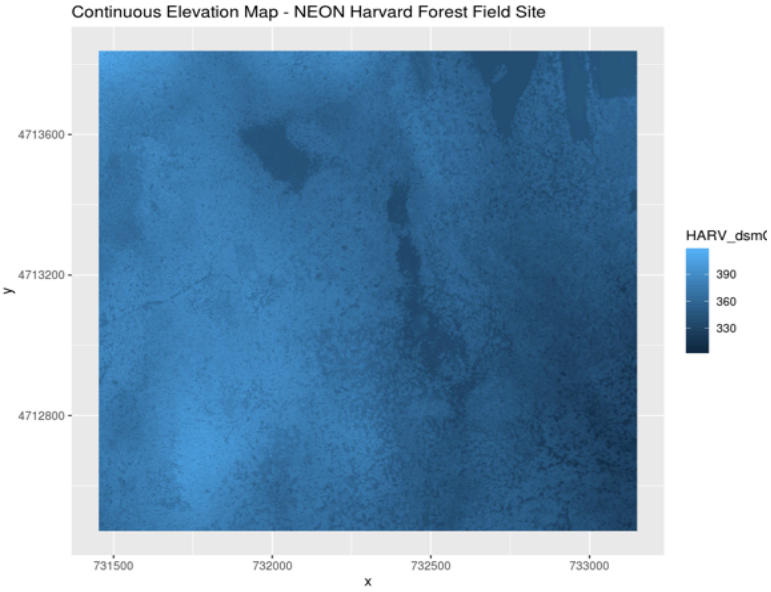


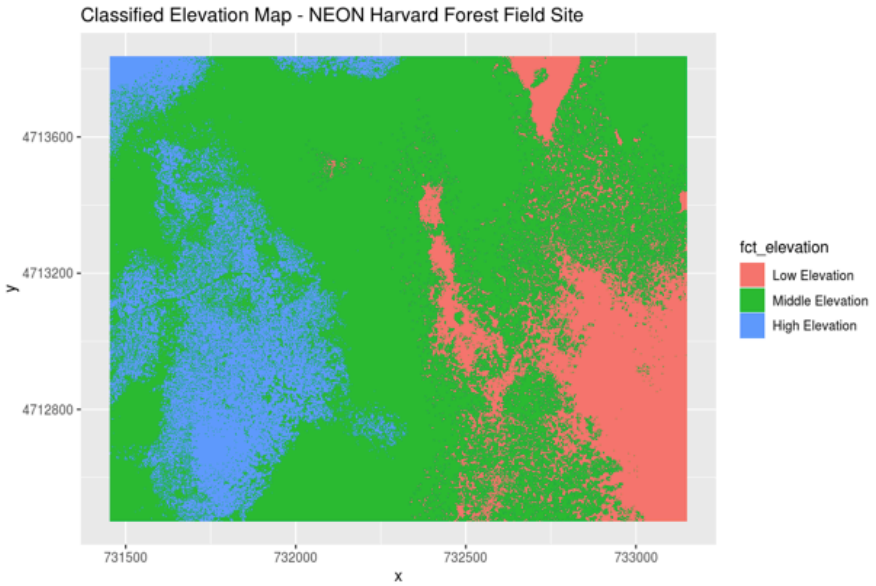
INTRODUCTION TO RASTER DATA

- A geospatial raster is only different from a digital photo in that it is accompanied by spatial information that connects the data to a particular location.
- This includes the raster's **extent** and **cell size**, the number of rows and columns, its **projection** and **coordinate reference system (or CRS).**



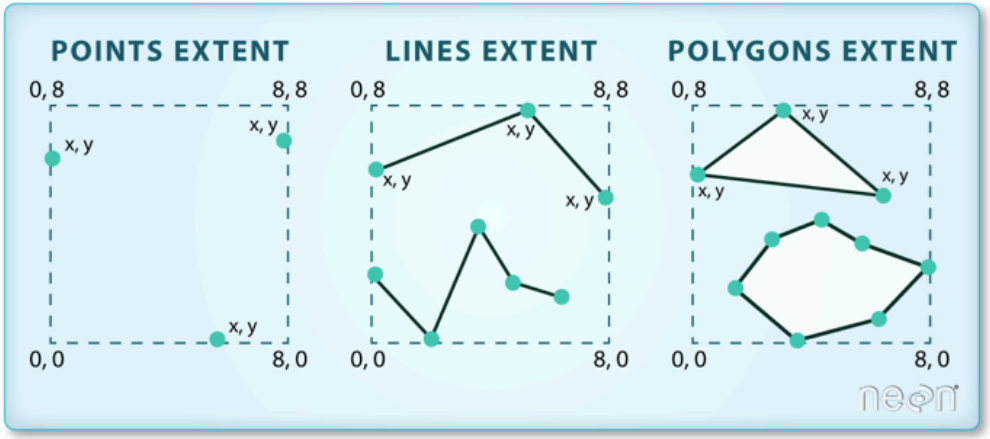
CONTINUOUS VS CATEGORICAL





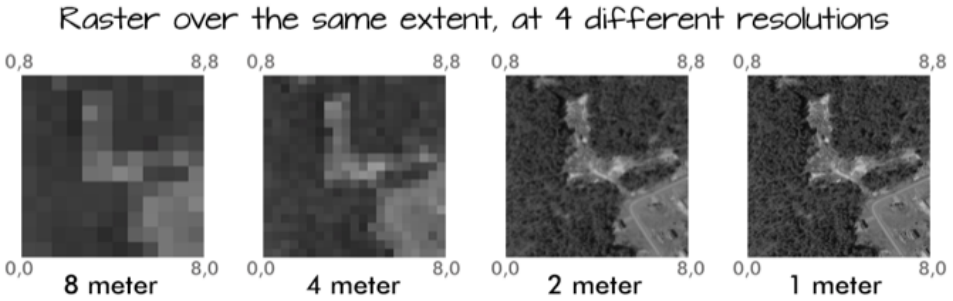
IMPORTANT ASPECTS TO RASTERS

- **Spatial extent**
- Resolution
  - We will focus on **spatial** resolution, but there are other forms of resolution, including **spectral**, **temporal**, and **radiometric**



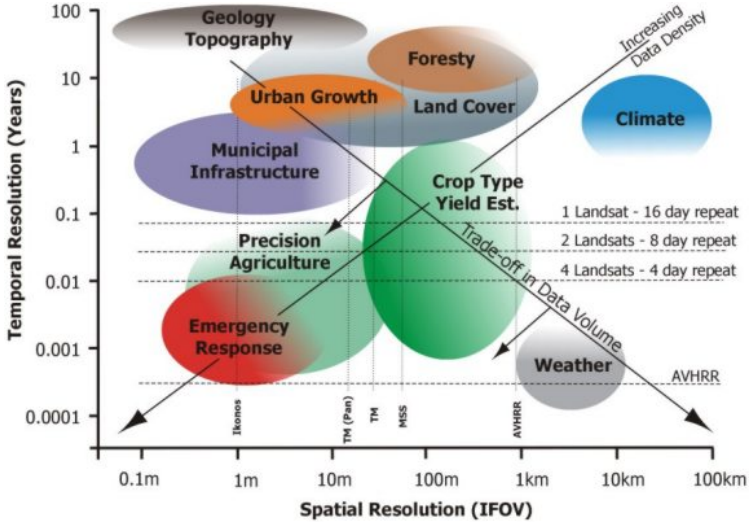
IMPORTANT ASPECTS TO RASTERS

- Spatial extent
- **Resolution**
  - We will focus on spatial resolution, but there are other forms of resolution, including **spectral**, **temporal**, and **radiometric**



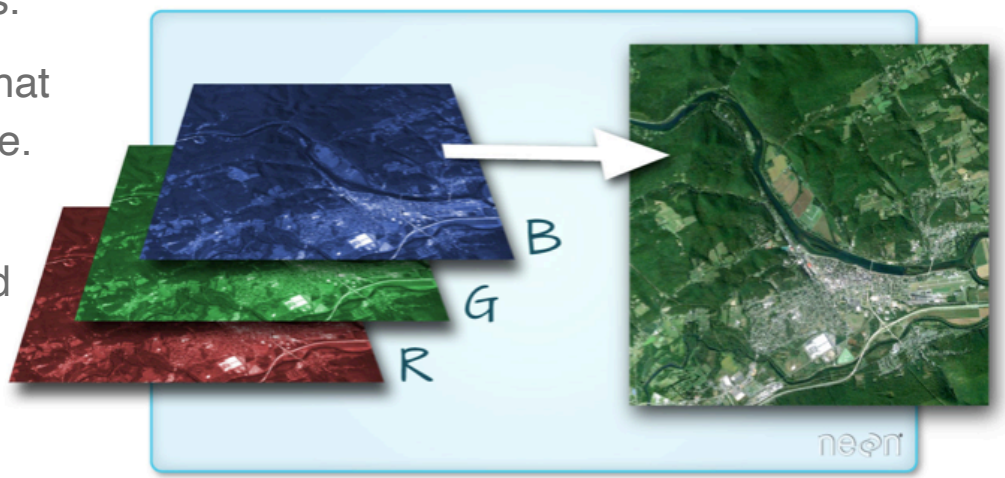
IMPORTANT ASPECTS TO RASTERS

- Spatial extent
- **Resolution**
  - We will focus on spatial resolution, but there are other forms of resolution, including **spectral**, **temporal**, and **radiometric**



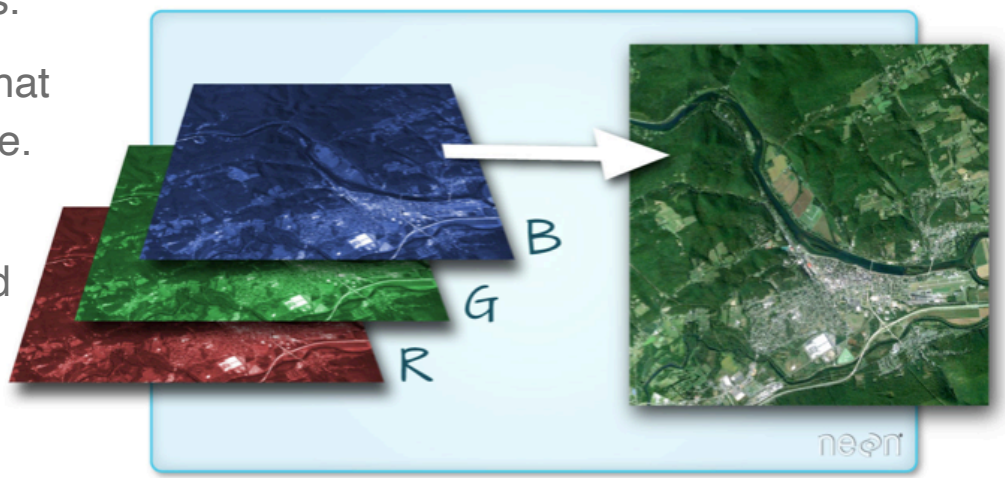
MULTI-BAND/MULTI-SPECTRAL DATA

- A raster can contain one or more bands.
- One type of multi-band raster dataset that is familiar to many of us is a color image.
- A basic color image consists of three bands: red, green, and blue. Each band represents light reflected from the red, green or blue portions of the electromagnetic spectrum.



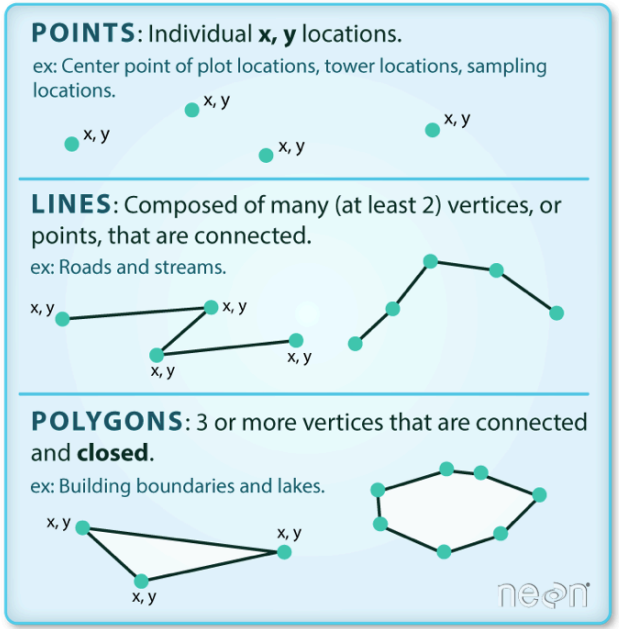
MULTI-BAND/MULTI-SPECTRAL DATA

- A raster can contain one or more bands.
- One type of multi-band raster dataset that is familiar to many of us is a color image.
- A basic color image consists of three bands: red, green, and blue. Each band represents light reflected from the red, green or blue portions of the electromagnetic spectrum.



INTRO TO VECTOR DATA

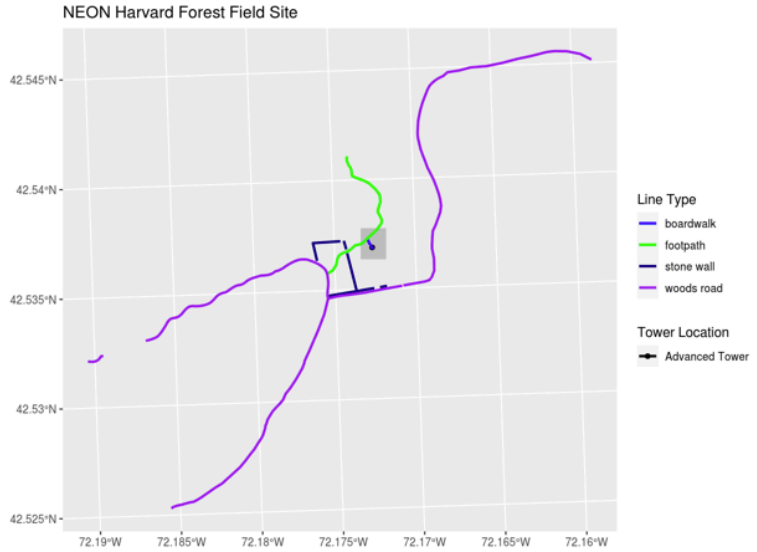
- Vector data structures represent specific features on the Earth's surface, and assign attributes to those features.
- Vectors are composed of discrete geometric locations (x, y values) known as vertices that define the shape of the spatial object.
- The organization of the vertices determines the type of vector that we are working with: point, line or polygon.



INTRO TO VECTOR DATA

Vector data has some important advantages:

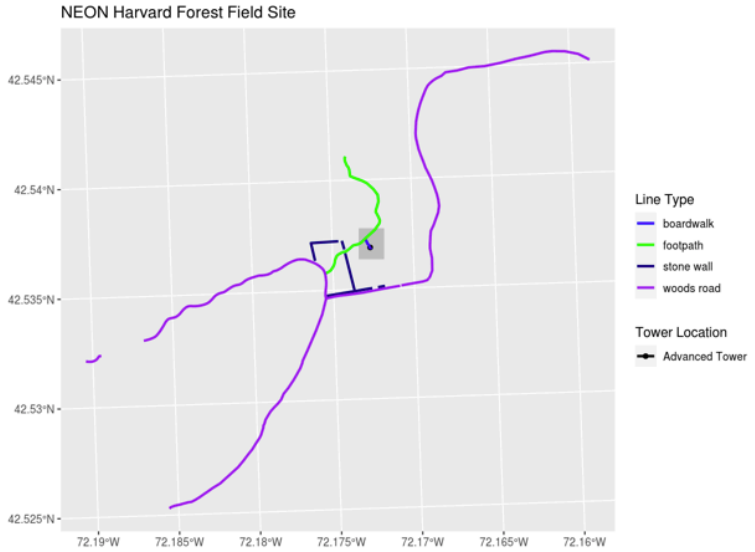
- The geometry itself contains information about what the dataset creator thought was important
- Each geometry feature can carry multiple attributes instead of just one, e.g. a database of cities can have attributes for name, country, population, etc.
- Data storage can be very efficient compared to rasters



INTRO TO VECTOR DATA

The downsides of vector data include:

- potential loss of detail compared to raster
- potential bias in datasets - what didn't get recorded?
- Calculations involving multiple vector layers need to do math on the geometry as well as the attributes, so can be slow compared to raster math.



INTRO TO VECTOR DATA

- Like raster data, vector data can also come in many different formats. For this workshop, we will use the **Shapefile** format which has the extension .shp. A .shp file stores the geographic coordinates of each vertice in the vector, as well as metadata including:
- **Extent, Object type,** and the **Coordinate reference system (CRS),** and **Other attributes**: for example, a line shapefile that contains the locations of streams, might contain the name of each stream.
- Shapefiles are an atomic collection (multiple files)

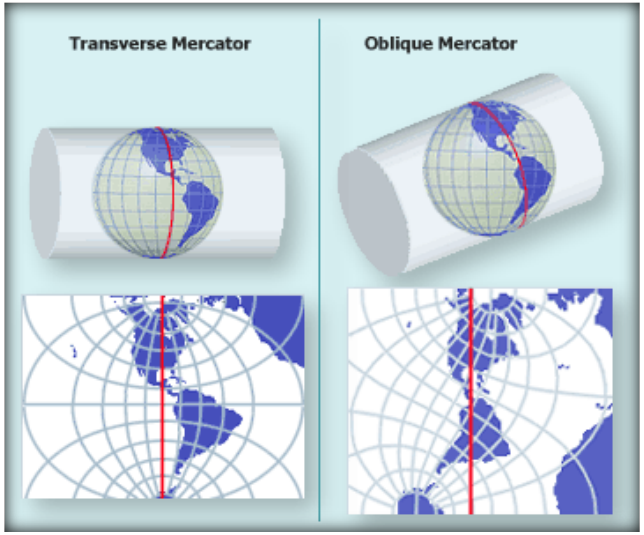
COORDINATE REFERENCE SYSTEMS

- The CRS associated with a dataset tells your mapping software (for example R) where the raster/vector is located in geographic space.
- It also tells the mapping software what method (projection) should be used to flatten or project the raster in geographic space.
- Key components of a CRS are:
  - Datum – a model of the shape of the earth (ex. WGS84, NAD83)
  - Projection – mathematical transform of angular measurements from spheroidal to flat. May include a zonal information if UTM
  - Ellipsoid - mathematical surface obtained by revolving an ellipse about the earth's polar axis. Selected to give a good fit to the geoid

COORDINATE REFERENCE SYSTEMS

Map Projections: to convert geodetic positions of a portion of the earth's surface to plane rectangular coordinates, points are projected mathematically from the ellipsoid to some imaginary developable surface - plane that can be rolled out flat

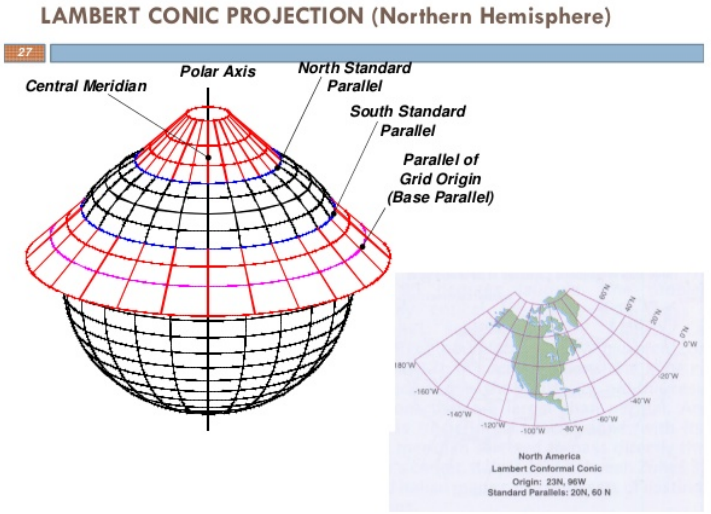
Coordinate Reference Systems: quantitative coordinate systems - based on mathematical projection models, often a cartesian coordinate system (i.e. x, y axes) representing relative positions within a particular map projection



COORDINATE REFERENCE SYSTEMS

Map Projections: to convert geodetic positions of a portion of the earth's surface to plane rectangular coordinates, points are projected mathematically from the ellipsoid to some imaginary developable surface - plane that can be rolled out flat

Coordinate Reference Systems: quantitative coordinate systems - based on mathematical projection models, often a cartesian coordinate system (i.e. x, y axes) representing relative positions within a particular map projection



COORDINATE REFERENCE SYSTEMS

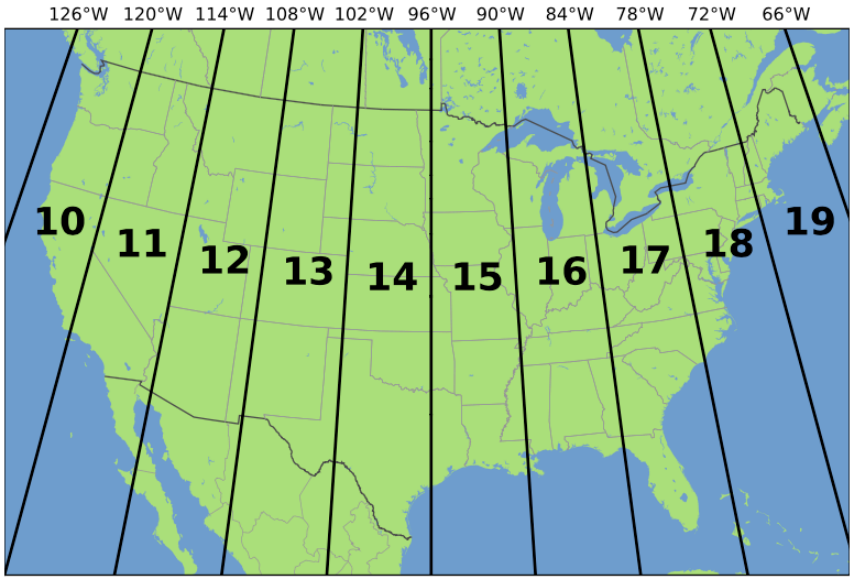
PROJ is an open-source library for storing, representing and transforming CRS information. PROJ.5 has been recently released, but PROJ.4 was in use for 25 years so you will still mostly see PROJ referred to as PROJ.4. PROJ represents CRS information as a text string of key-value pairs, which makes it easy to customise (and with a little practice, easy to read and interpret).

A PROJ4 string includes the following information:

- **proj=:** the projection of the data
- **zone=:** the zone of the data (this is specific to the UTM projection)
- **datum=:** the datum use
- **units=:** the units for the coordinates of the data
- **ellps=:** the ellipsoid (how the earth's roundness is calculated) for the data

Note that the zone is unique to the UTM projection. Not all CRSs will have a zone.

COORDINATE REFERENCE SYSTEMS

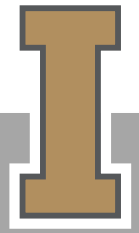


UTM Zones

```
+proj=utm +zone=18 +datum=WGS84 +units=m +no_defs +ellps=WGS84 +towgs84=0,0,0
```

COORDINATE REFERENCE SYSTEMS

```
+proj=utm +zone=18 +datum=WGS84 +units=m +no_defs +ellps=WGS84 +towgs84=0,0,0
```



COORDINATE REFERENCE SYSTEMS

- GDAL is a set of software tools that translate between almost any geospatial format in common use today (and some not so common ones).
- GDAL also contains tools for editing and manipulating both raster and vector files, including reprojecting data to different CRSs.

http://gdal.org

GEOSPATIAL LANDSCAPE

Commercial

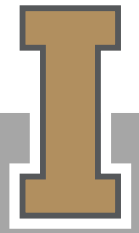
- ESRI
- MAPINFO
- Manifold
- Smallworld

Open source

- QGIS
- GRASS
- GDAL
- PostGIS/Postgres

Cloud

- Google Earth Engine
- ArcGIS Online



GEOSPATIAL LANDSCAPE

Geospatial Libraries: R: sf, sp, gdal, spplot, leaflet, spacetime, ncdf4

Geospatial Libraries: Python: shapely, geopandas, rasterio, gdal, rasterstats

Modeling: R: caret, mlr

Modeling: Python: tensorflow, keras, sci-kit learn, numpy/scipy, Pytorch

Visualization: R: ggplot2, ggpubr, seaborn, plotly

Visualization: Python: matplotlib

https://cran.r-project.org/web/views/Spatial.html