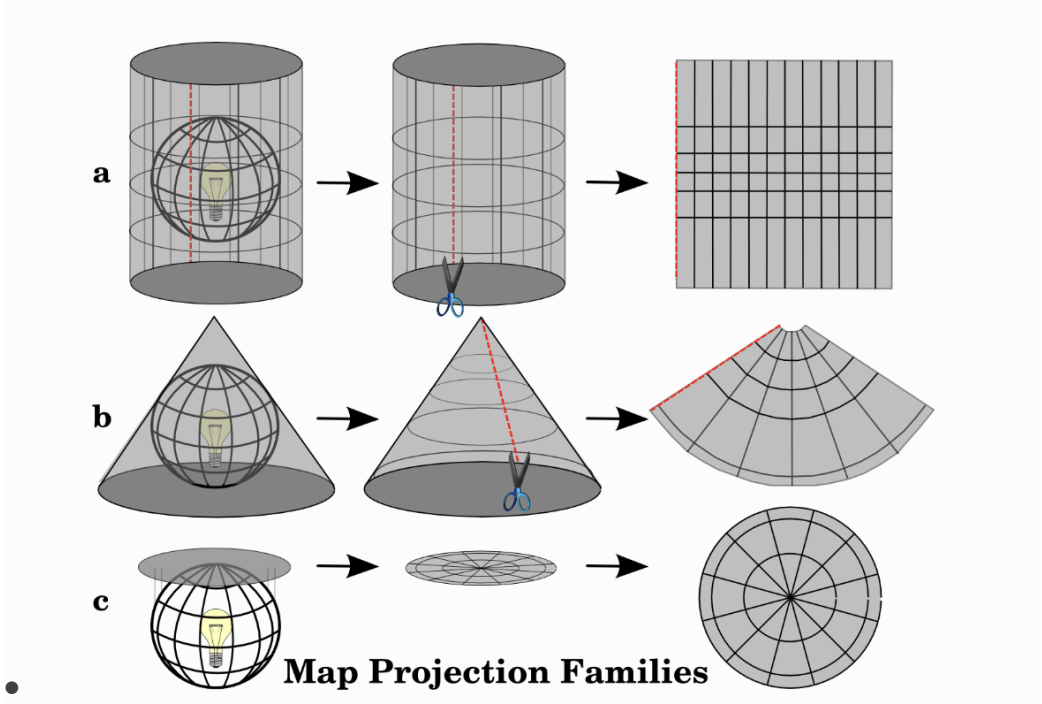
Geospatial Data Concepts

- Geo-Spatial concepts
    - Coordinate Reference System (CRS)
        - WGS 84, EPSG:26918 - NAD83
        - Precisely measure the locations on the surface of the earth as coordinates
        - A **coordinate reference system** (CRS) then defines how the two-dimensional, projected map in your GIS relates to real places on the earth.
        - Planar Projection/cylindrical projection/Conical Projections
        - 

        - Preserve accuracy of projections:
            - Angular conformity
            - Equal Distance
            - Equal area
        - Geographic Coordinate Reference vs Projected Coordinate Reference
            - GCRS: Latitude and Longitude
            - PCRS: XYZ plane
        - UTM CRS
            - To avoid too much distortion in angular conformity, the world is divided into **60 equal zones** that are all **6 degrees** wide in longitude from East to West.
            - define a two-dimensional coordinate within the **Area of Interest (AOI):**

        - The position of a coordinate in UTM south of the equator must be indicated with the **zone number** (35) and with its **northing (Y) value** and **easting (X) value** in meters. The **northing value** is the distance of the position from the **equator** in meters. The **easting value** is the distance from the **central meridian** (longitude) of the used UTM zone.
    - Shape File (.shp)
        - a common format for storing **vector** GIS data. Shapefiles store non-topological vector data along with related attribute data.
        - a shapefile is actually a collection of at least three basic files: .shp, .shx and .dbf.
        - Shp:(Shape) a direct access, variable-record-length file in which each record describes a shape with a list of its vertices.
        - *shx* : (hold the other two files together)Index file (mandatory). In the index file, each record contains the offset of the corresponding main file record from the beginning of the main file. The index file (.shx) contains a 100-byte header followed by 8-byte, fixed-length records.
        - *dbf* : (attributes) dBASE Table file (mandatory); a constrained form of DBF that contains feature attributes with one record per feature. The one-to-one relationship between geometry and attributes is based on record number. Attribute records in the dBASE file must be in the same order as records in the main file.
    - Raster File (.tif)
        - a rectangular array of regularly sampled values, known as pixels. Each pixel (picture element) has one or more numbers associated with it, specifying a color which the pixel should be displayed in.
        - Simplest representation of a pixel: three 8 bit (24 bits total) color values (ranging from 0-255) defining the amount of red, green, and blue respectively in each pixel. (can take a lot of disk space)
        - GIF: uses 8 bits for one pixels, reduce colors from 766 colors to 256 colors

FILE NAME	# OF BYTES	# OF BITS PER PIXEL	COMPRESSION	RASTER OR VECTOR
sample.ppm	6336015	24	NONE	RASTER
sample.ps	2926048	SPECIAL/24	SPECIAL	VECTOR
sample.rgb	1440512	24	LOSSLESS	RASTER
sample.png	140557	24	LOSSLESS	RASTER
sample.gif	61346	8	LOSSLESS	RASTER
sample.jpg	59575	24	LOSSY	RASTER

        -

JPEG - The JPEG format was developed as a highly (and lossy) compressed format, optimized for photographs. The format supports 24 bit images and although **data is discarded**, it is done in such a way that it will not be visible to most people. This format is extremely popular on the Web and in any situation where the file size is of significant importance (such as storing large numbers of pictures on a digital camera's memory card). Should be avoided however for any situation where you want the very best quality image, such as in printed publication graphics. Also, if you are working with an image to manipulate it in some way (such as with **Photoshop**), you should work with it in a non-lossy format (such as TIFF or, with Photoshop, PSD) and then, only when you are done, save the file as a **JPEG**.

PNG - This format is the third most commonly viewable format on the Web and is supported by most modern web browsers. It supports full lossless compression 24-bit color and is the format of choice for web images that must be of the highest quality. The compression quality of PNG files is also excellent, as it uses a generally more efficient algorithm than other formats such as GIF or TIFF.

GIF - The GIF format is also very popular for graphics on the web but only supports 8-bit color and so should be avoided for most applications. There are generally only two situations where we would recommend using the GIF format. The first is if your image has 256 or less colors and you want the maximum web compatibility. The second is for animated sequences of images on the Web, where the Animated GIF format is the standard for users not willing or able to use such applications as Adobe Flash.

TIFF, RGB, PPM - All of these are 24-bit color high quality formats and commonly seen. Most web browsers will not be able to view these images except through an external applciation but for non-web based applications, all of these formats are quite useful.

PSD - This format is the internal format used by the popular image editing program Adobe Photoshop. It stores more than just the image, such as layers information, text, etc... Use this format while working in Photoshop but when done, if you wish to share the image with people who you do not know have Photoshop, save it in another format, such as TIFF, PNG, or JPEG.

RAW - This is actually not a single format, but rather a set of internal (and each different) formats used by some digital cameras. Although some applications will read some RAW formats, storing your images in this format once you get them off of your camera is not recommended. Convert them to a more general format such as PNG or JPEG for saving for the long term.

- Bands

- *Coastal band was designed to highlight shallow water, measuring changes in ocean color, and detecting fine aerosol particles in the atmosphere.*
- *The Red, Green, Blue bands cover the range of 400 – 700 nanometers which corresponds to the visible spectrum. It is used in combination with other spectral bands to visualize what we cannot normally see.*
- *Red + Green + Blue, together, make up a Pan band (panchromatic, which stands for "all colors"). An image in Pan usually appears black & white and has a higher spatial resolution compared to most other bands. Thus, combining a panchromatic with any other spectral band makes the final composite image "sharper," highlighting more details.*
- *The Near-infrared (NIR) wavelengths are just beyond the visible red range, and are reflected off the leaves and canopy of healthy plants much stronger than in the blue, green, and red bands. The NIR band is ideal for monitoring plants and contrasting water bodies against the surrounding features.*
- *Images taken in the Cirrus band reveal high-altitude clouds that are invisible in most other bands.*
- *Using the LWIR band (stands for long-wave infrared region, between 8 and 14 microns), we can detect radiant heat emitted by land and water surfaces. This means we can measure temperatures of these surfaces. Landsat-8's two spectral bands are in the long-wave infrared region enabling this satellite to measure temperatures.*
  These two bands are: Thermal Infrared (TIRS):
  Band 10 – Thermal Infrared (TIRS) 1 (10.6 – 11.19 microns)
  Band 11 – Thermal Infrared (TIRS) 2 (11.50 – 12.51 microns)

- NDVI

  NDVI = (NIR - VIS)/(NIR + VIS)

  - Measure difference between nearly-infrared (which vegetation strongly reflects) and red lights (which vegetation strongly absorbs)
  - If NDVI approaches negative values, it's likely it's water; when close to 0, generally correspond to barren areas of rock, sand, snow;when close to +1, it's likely it's a dense forest
  - NDVI=0, meaning urbanized area

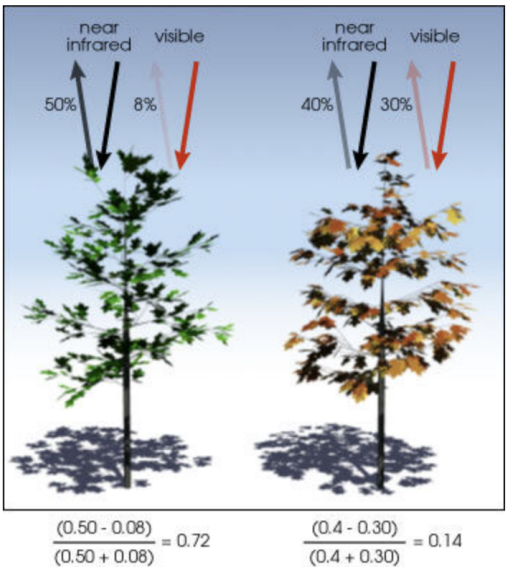


Image courtesy of NASA.

  -

- QGIS: import tif file and build statistics
  - Tool_Processing: Raster Analysis - Raster_Calculator: NDVI
  - Band distribution for different surface of area: tree/water/urbanized area
    - Areas with abundant trees generally have higher band4 value (>=120)
    - Areas with water/pool has low values for all bands (~50)
    - Areas with house/urbanized structures has high values for all bands (>100)

        - Using raster calculator (band4 > 110)

    - Raster layer statistics for four bands:
        - **Band1**: Minimum value: 19, Maximum value: 234, Range: 215, Sum: 1360570649, Mean value: 92.82813495479733, Standard deviation: 48.73311035854704, Sum of the squares: 34808852360.09144
        - **Band2**: Minimum value: 22, Maximum value: 227, Range: 205, Sum: 1461734205, Mean value: 99.73025667539841, Standard deviation: 45.34521836351554, Sum of the squares: 30137306747.14453
        - **Band3**: Minimum value: 31, Maximum value: 228, Range: 197, Sum: 1321109029, Mean value: 90.13577304798471, Standard deviation: 44.44977630320106, Sum of the squares: 28958802743.40999
        - **Band4**: Minimum value: 60, Maximum value: 235, Range: 175, Sum: 1834537042, Mean value: 125.1656077099093, Standard deviation: 37.05374408740325, Sum of the squares: 20123598263.7307

- QGIS: import csv file and generate statistics
    - Select features using expressions by opening the attribute table

Select Species from the csv files (example: “pin oak”, yellow point in the image)

Color the Top Three Species By Colors around Central Park Area: