DCN Planetears

Geospatial Data Curation:
an introduction
Module: GIS Introduction

Curriculum Development Team

Leighton L Christiansen
Centers for Disease Control and Prevention

https://orcid.org/0000-0002-0543-4268 - LChristiansen@cdc.gov

Kelly Grove
Florida State University

0000-0001-7926-169X
kegrove@fsu.edu

Melinda Kernik
University of Minnesota

kerni016@umn.edu

Timothy Norris
University of Miami

0000-0002-0898-3027
tnorris@miami.edu

Module List

- 0_Environment_Setup
- 1_GIS_Introduction
- 2_Ethics_and_GISData
- 3_Common_GIS_Data_Types
- 4_GIS_Metadata
- 5_Transformations

Suggested Primers

- Geodatabase Primer
- GeoJSON Primer
- GeoTIFF Primer
- netCDF Primer

DCN Planetears

Geospatial Data Curation:
an introduction
Module: GIS Introduction

Module Objectives: GIS Introduction

This module has six objectives. At the end of the module, learners should be able to:

1. Describe a geographic information system (GIS)
2. Identify common tool sets for working with GIS data
3. Describe characteristics of vector GIS data
4. Describe characteristics of raster GIS data
5. Understand the three principal components of a Coordinate Reference System
6. Conceptualize structure of a GIS project

The module has Lecture, Activity, and Quiz components to help reinforce new information.

Defining GIS

A collection of tools, data, hardware, and people used to locate and describe features on the Earth's surface.

The system of organization can be digital (computers and software) or analog (hand drawn on paper).

The actual process of organization is done with a purpose by human beings.

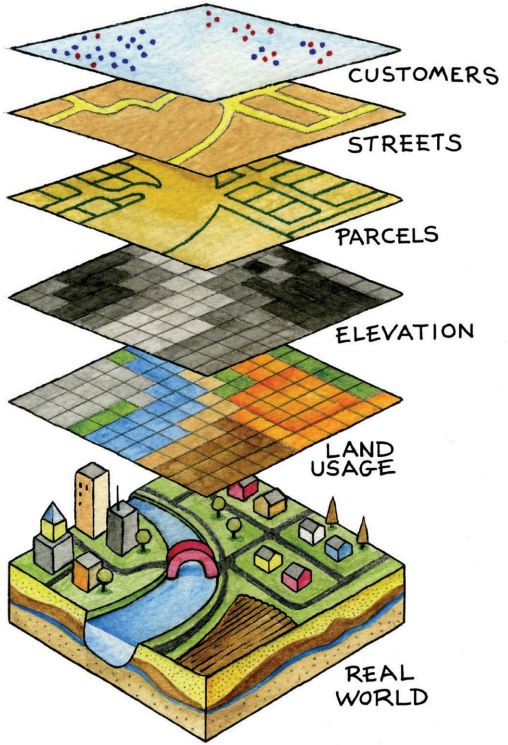


Image adapted from: https://saylordotorg.github.io/text_essentials-of-geographic-information-systems/index.html

Defining GIS

- Geographic Information System
  - collections of tools, data, hardware, and people
- Geographic Information Science
  - systematic inquiry into research questions about the relationship between GIS and socio-natural systems
- Geographic Information conStruction
  - tool building for storage, collection or analysis of geospatial data

Wright, D. J., Goodchild, M. F., & Proctor, J. D. (1997). Demystifying the Persistent Ambiguity of GIS as 'Tool' versus 'Science'. *Annals of the Association of American Geographers*, 87(2), 346-362. doi:10.1111/0004-5608.872057

To help you understand all the terms and jargon, these modules are accompanied by a Glossary_of_GIS_Terms

Useful Abstractions

Image credit: Timothy Norris

Physical Model
(files on disk)

Logical Model
(data structures - vector and raster)

Conceptual Model
(discrete or continuous data)

Reality
(the world out there)

Key elements of GIS Assemblages

- Tool set
  - collection of software programs
  - in some cases referred to as a stack
- Physical [and logical] data models
  - Relational databases (tables) [vector]
  - Extensible data models (trees) [vector]
  - Raster data (grids and images) [raster]
- Scripting/programming languages

Common Desktop GIS Assemblages

Environmental Systems Research Institute (ESRI)

- Tool set
  - ArcGIS Pro *NEW*
  - ArcMap Desktop *legacy*
  - ArcGIS Online
  - ArcCatalog
- Physical [and logical] data models
  - shapefiles (.shp) [vector]
  - geodatabases (.gdb) [vector and raster]
  - geotiffs (.tif) [raster]
  - database servers (SQL) [vector and raster]
- Scripting languages
  - python

Download/Login Instructions for ArcGIS Pro

1. Check to see if your institution has a license for Esri ArcGIS Pro
    a. If you do not have an account try using QGIS
2. Request an account from the department holding the license
3. Follow the download instructions for the program from Esri

Common Desktop GIS Assemblages

Quantum GIS (QGIS)
[Free Open Source Software]

- Tool set
  - QGIS - http://www.qgis.org/en/site/
  - base: Geospatial Data Abstraction Library (GDAL)
  - base: Open GIS Reference (OGR)
- Physical and logical data models
  - shapefiles (.shp) [vector]
  - geopackages (.gpkg) [vector and raster]
  - json (.geosjon) [vector as trees]
  - geotiffs (.tif) [raster]
  - relational database servers (SQL) [vector and raster]
- Scripting languages
  - python

Download Instructions for QGIS

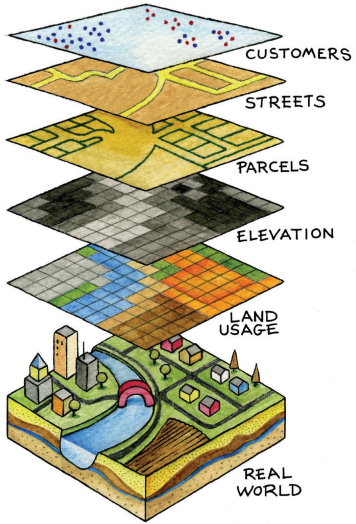
1. Go to the Download QGIS website
2. Locate your operating system in the list of download options
3. Please download the latest long term release. The long term release version of the program and the most stable at this time.

Common Code-based GIS Assemblages

- Tool set
  - Python - https://www.python.org/
  - R - https://www.r-project.org/
  - Packages: GDAL/OGR, geopandas; sp, ggplot, maps
  - leaflet for visualization
- Physical and logical data models
  - raw text, csv (tables) [vector]
  - raw text, json, geojson (trees) [vector]
  - raw text, grids (raster) [raster]
  - Relational databases (sql) [vector and raster]
- Scripting languages
  - Python / R

Geospatial Data

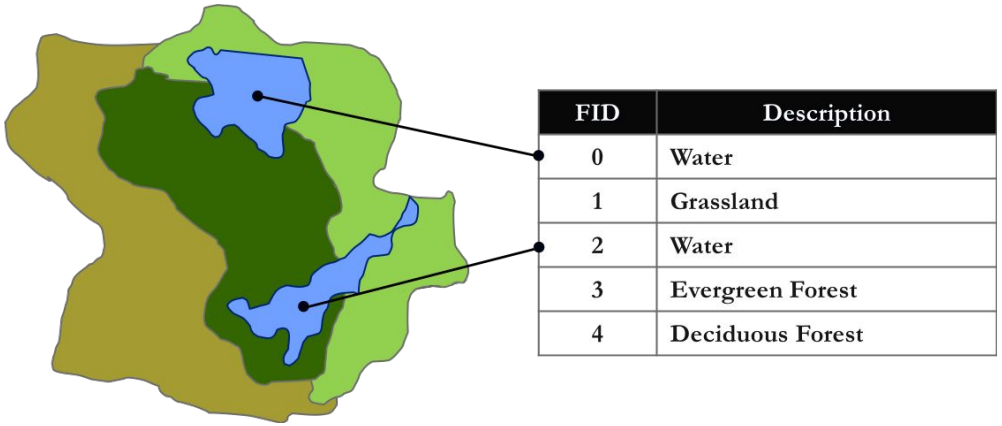
Physical Model (files on disk)	**Raster Structure** geotiff, grid	**Vector Structure** shapefile, geodatabase, json
Logical Model (vector and raster)	Raster Model	Vector Model
Conceptual Model (discrete or continuous)	Object View Continuous	Object View: Discrete
Reality (the world out there)	Entity: Temperature, Topography	Entity: Trees, Houses, Streets



adapted from: https://saylordotorg.github.io/text_essentials-of-geographic-information-systems/index.html

Vector Data Model
Points, Lines, and Polygons

- all based on x,y,(z) coordinate pairs of geographic data
- lines and polygons are built from groups of points
- attribute data is linked to points, lines, or polygons (features)
- each feature is associated with a unique record in an attribute table



FID	Description
0	Water
1	Grassland
2	Water
3	Evergreen Forest
4	Deciduous Forest

Image adapted from original by; Barry Nickel

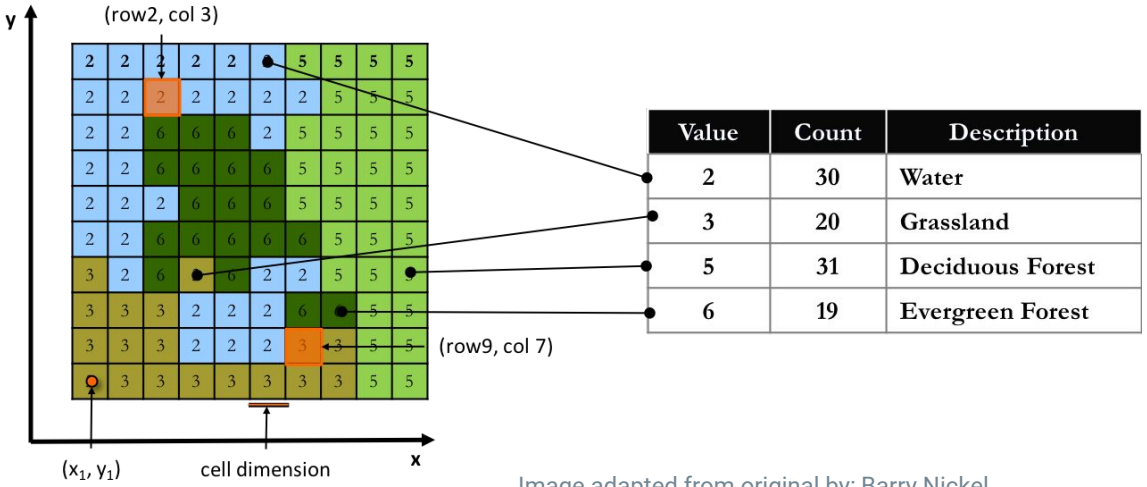
Vector Data Model
File Formats

Common File Formats

Name	Extension	Source
shapefile	.shp *	ESRI
geojson	.json	open
geodatabase	.gdb	ESRI
google earth	.kml .kmz	open
autoCAD	.dxf .dwg	AutoDesk

* .shp is the main file extension, each shapefile is actually a collection of files that include: dbf, sbn, shx, sbx, prj,, xml (and more - be careful!!)

Raster Data Model
Grids and Images



Value	Count	Description
2	30	Water
3	20	Grassland
5	31	Deciduous Forest
6	19	Evergreen Forest

- each cell represents an x,y coordinate
  - each cell has a specific size on the surface of the earth (scale)
  - cell scale is based on the resolution of the image
- each cell has only one value (color, numeric, or categorical)

Image adapted from original by; Barry Nickel

Raster Data Model
Resolution

- Original resolution of the collected data limits spatial accuracy
- Can't improve by replicating cells to create smaller size cells
- Location implied rounded to cell coordinate (center of cell)

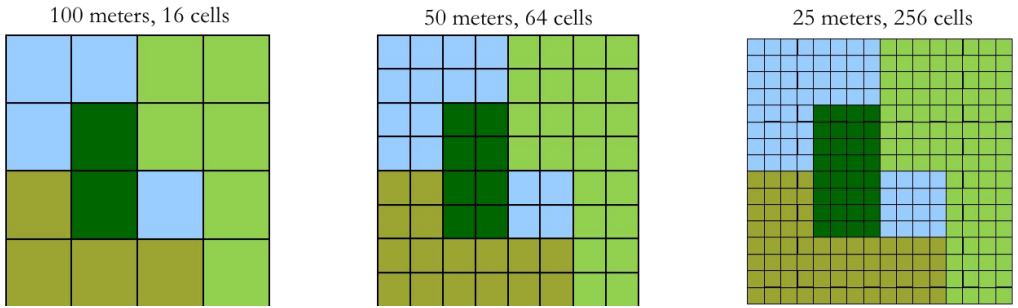


Image adapted from original by; Barry Nickel

Raster Data Model

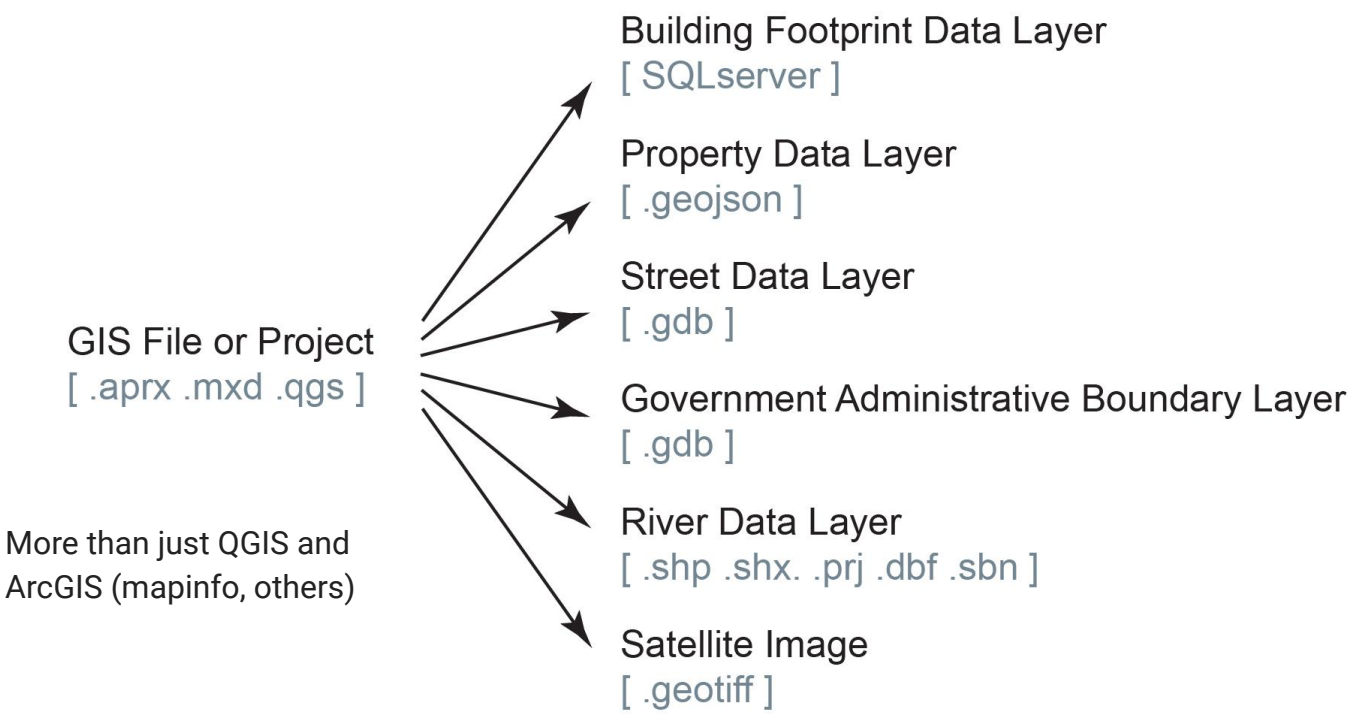
File Formats

Common File Formats

Name	Extension	Source
geotiff	.tiff .tif	open
jpg	.jpg	open
Arcinfo GRID	*	ESRI
ERDAS imagine	.img	ERDAS

* grids are stored in directories (folders) with many files all with the extension .adf; all files must be present.

GIS Projects



C U R A T E D

Map Projections and Coordinate Reference Systems

Image: D.M. Swart
Artistic cartography: creative ways to peel the globe

Coordinate Reference Systems (CRS) Basics

1. **Datum**
2. **Geographic coordinate reference system**
3. Projected coordinate reference system

➔ Geospatial data must have: **datum + geographic coordinate system**

➔ Projected coordinate systems are optional (needed for measurement)

CURATED

Datum

- Center of the Earth?
- That which is given?

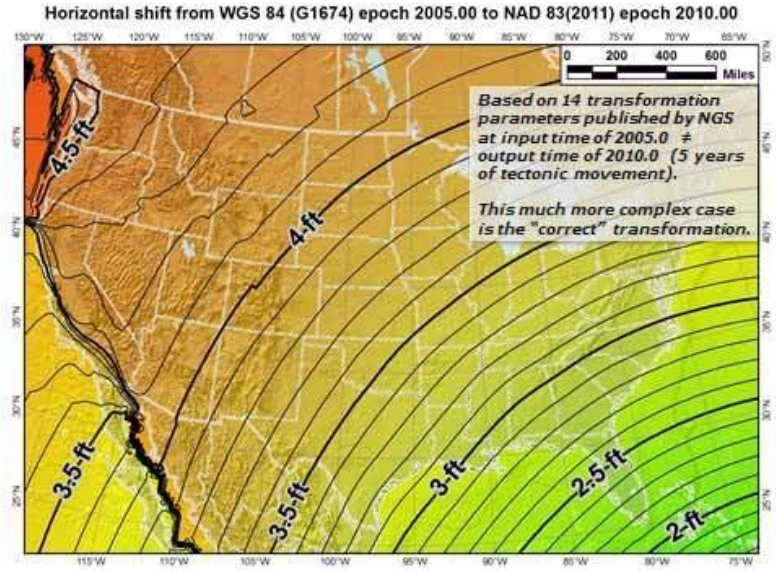
Ellipsoid - mathematically defined surface approximating the shape of the Earth

Geoid - surface of the Earth's gravity field approximating sea level

Image credit: Timothy Norris

Geographic Coordinate Systems

- Latitude and Longitude - *spherical coordinates*
- Very common, but cannot be used for measurement
- Things to remember:
  - **EPSG** - European Petroleum Survey Group
  - **WGS 84** - most common globally - ESPG:**4326**
    i. World Geodetic System 84
  - **NAD 83** - most common in the USA - ESPG:**4269**
    i. North American Datum



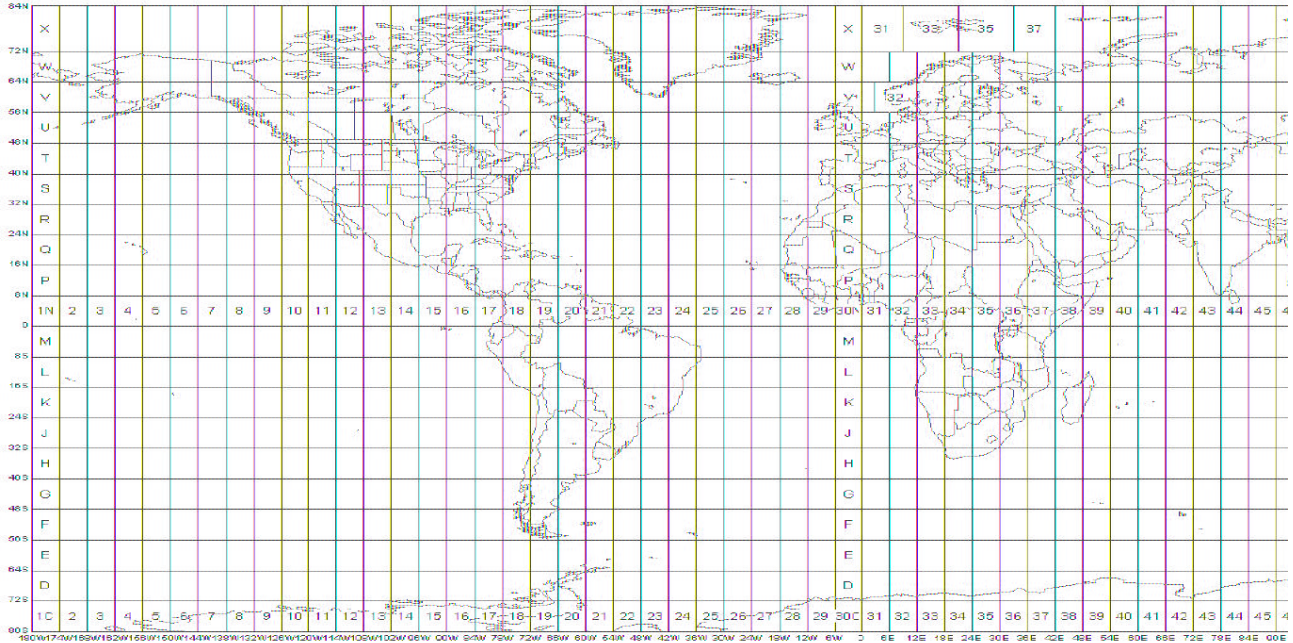
Expected difference between WGS84 and NAD 83 in 2022 Michael Dennis - National Geodetic Survey

Projected Coordinate Systems

- meters or feet - *Cartesian coordinates*
- Used for measurement and mapping
- Things to remember:
  - Projected Coordinate Systems are *specific* to the area being mapped
  - In the USA: the "State Plane System"
  - Around the Globe: the "UTM Grid"
  - For web based mapping systems
    WGS 84 Web Mercator - ESPG:3857

CURATED

Universal Transverse Mercator (UTM)



The 60 UTM coordinate system zones, each divided into north and south halves at the equator

UTM Grid Zones of the World compiled by Alan Morton. https://www.dmap.co.uk/utmworld.htm.

State Plane Coordinate System (SPCS)

The State Plane Coordinate System (SPCS) … in which each individual state has between one to six zones, depending on the state's size and shape. This coordinate system's high level of accuracy is achieved through the use of relatively small zones.

USGS definition

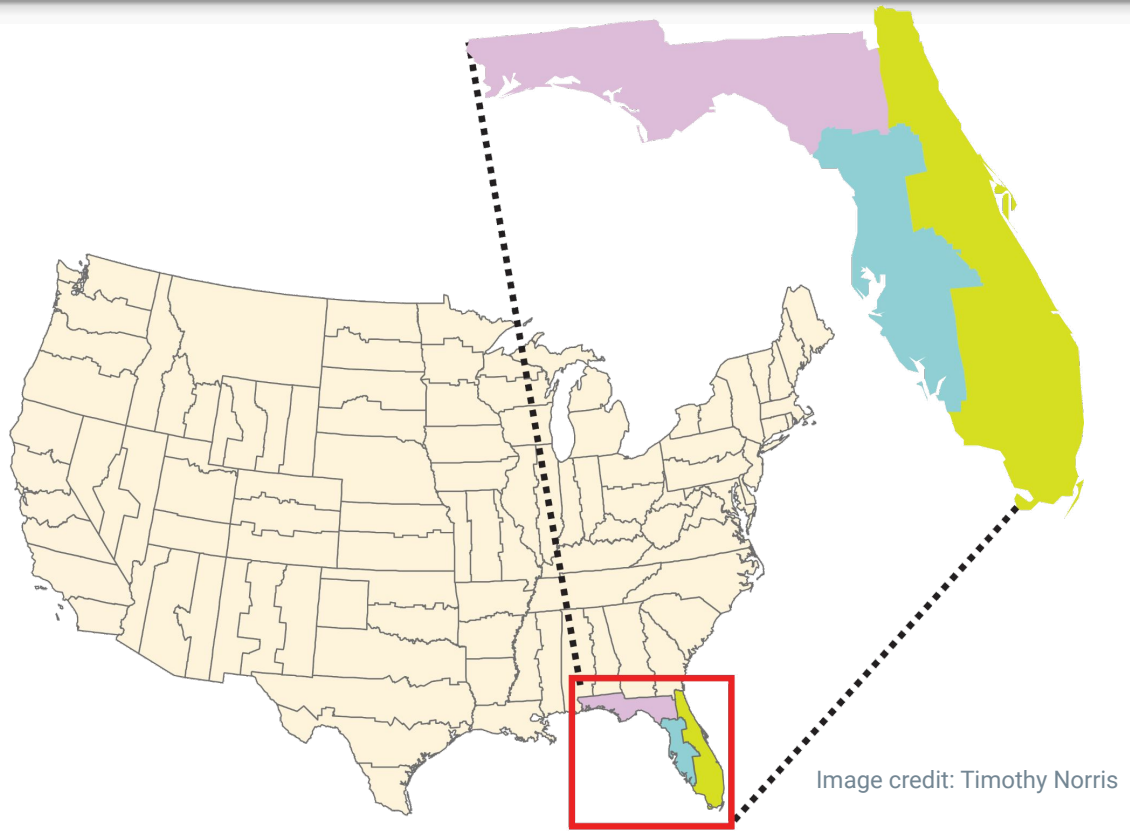


Image credit: Timothy Norris

Review and Skills Checking: Introduction to GIS

This module has six objectives. At this point, you should be able to:

1. Describe a geographic information system (GIS)
2. Identify common tool sets for working with GIS data
3. Describe characteristics of vector GIS data
4. Describe characteristics of raster GIS data
5. Understand the three principal components of a Coordinate Reference System
6. Conceptualize structure of a GIS project

Next, we will now have an Activity, and a Quiz.

Activities

Using the data in Dataset_No_ReadMe (this is a subfolder of the 1-4_Excersise_Dataset, folder):

1. Create a new GIS project in QGIS and load one raster and one vector layer

    - Review Layers and Browser panes
    - Look at Information Panel: find projection, pixel dimension, other metadata
    - Locate export layer panel: explore options and CRS selector

GIS Introduction: QUIZ 1

Quiz Questions 1 to 5 (of 9)

1. Describe, in your own words, what is a GIS?
    a. Answer:
2. Describe, in your own words, what are the principal components that make up a GIS?
    a. Answer:
3. What are two common GIS software assemblages?
    a. Answer:
4. What are two common GIS data models?
    a. Answer:
5. Yes or no. Are data layers stored in the GIS project file?
    a. Answer:

GIS Introduction: QUIZ 2

Quiz Questions 6 to 9 (of 9)

6. Yes or no. All geospatial data must have a datum defined?
    a. Answer:
7. What are the names and EPSG codes for two common datums used in North America?
    a. Answer:
8. Yes or no. All geospatial data must have a coordinate reference system (CRS) defined?
    a. Answer:
9. Name two common projected coordinate reference systems used in North America?
    a. Answer: