MICHIGAN STATE UNIVERSITY

Brief Intro to (Geo)spatial Data

AFRE 891

Spring 2024

Geographic Data

A single unit of geographic data includes:

1. Location (where): Anatone
2. Attributes (what): *data that describe the location*

Also great to have metadata:

When: 2003
Who: Anatone 4-H
How: local census?

Geospatial Data

Encodes geographical location geometrically with coordinates

Anatone, WA: 46.130479, -117.134167

Challenge: how do we know that those coordinates reference this specific location?

Coordinate Reference System (CRS)

Challenge: how do we know that those coordinates reference the desired location?

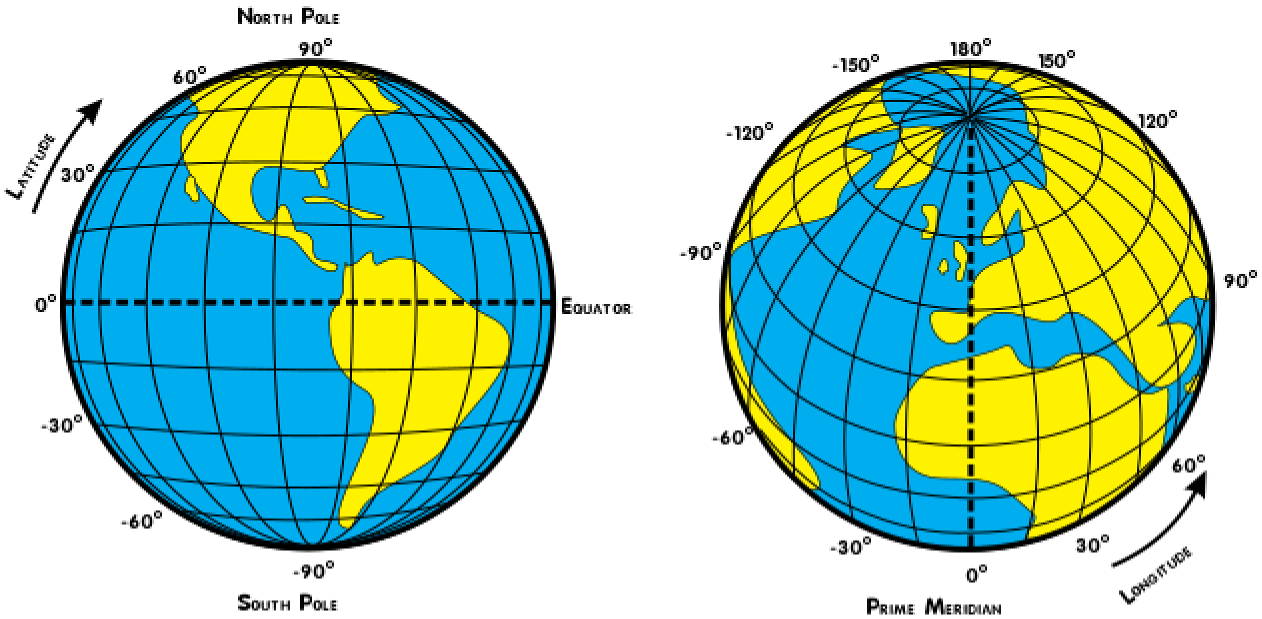
Solution: Coordinate Reference System (CRS)

- System for associating coordinates with a **specific, unambiguous** location on the earth's surface

New Orleans = 30N, 90W

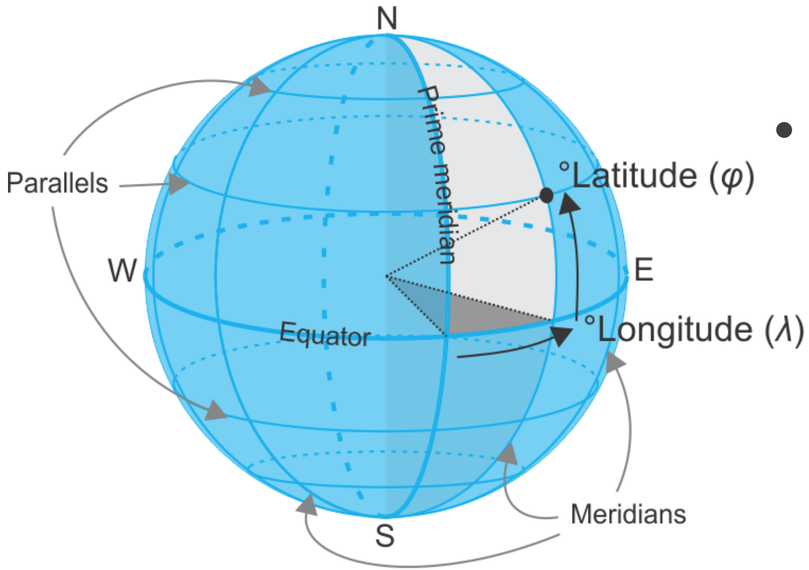
Coordinate Reference System (CRS)

Geographic coordinates: **Latitude** and **Longitude** pair



http://latitude-longitude.net

Coordinate Reference System (CRS)

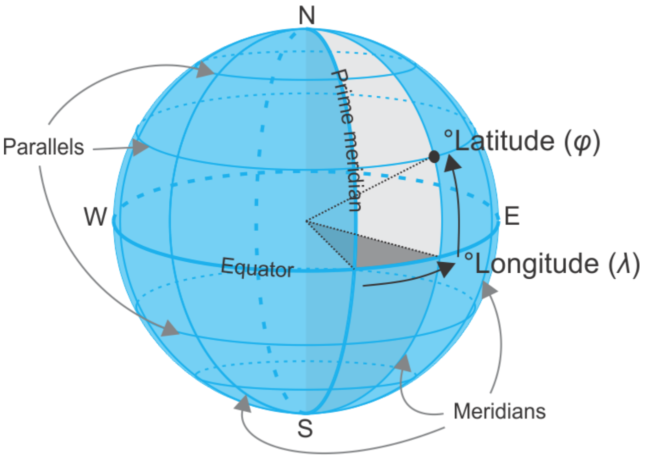


Complication: there are **many CRSs**

- Earliest known: Hipparchus 190-120 BCE
- Knowledge of earth and our ability to measure its shape improved over time…

Two Types of Coordinate Reference Systems

Geographic CRS



Angular units = Degrees (DMS or DD)

Good for storing data

Projected CRS

Cartesian units = Feet or Meters

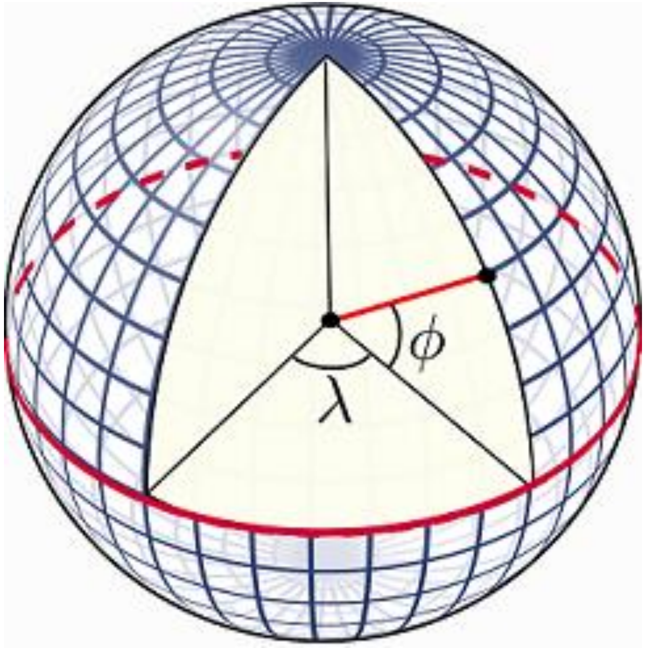
Good for local/regional mapping and analysis

Geographic Coordinate System (GCS)

Latitude and longitude coordinates are an **angular** system

Latitude: angle between line passing through equatorial plane and line between the point and the center of the earth

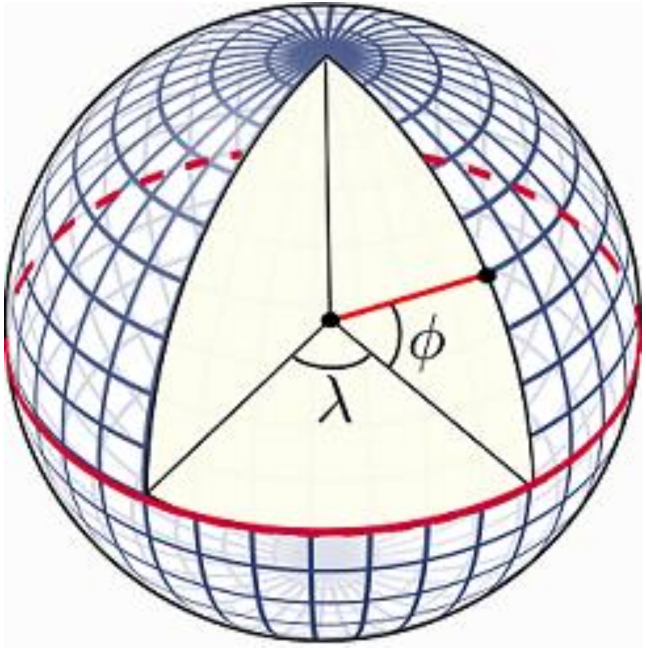
Longitude: angle from a reference meridian (line of constant longitude) to a meridian passing through the point



https://rspatial.org/raster/spatial/

Geographic Coordinate System (GCS)

Problem: can't measure these angles directly



https://rspatial.org/raster/spatial/

Geographic Coordinate System (GCS)

➔ Use a **datum**, a model of the earth's shape

2 Most Common:

WGS84 (EPSG: 4326)

Based on satellites, used by cell phones, GPS

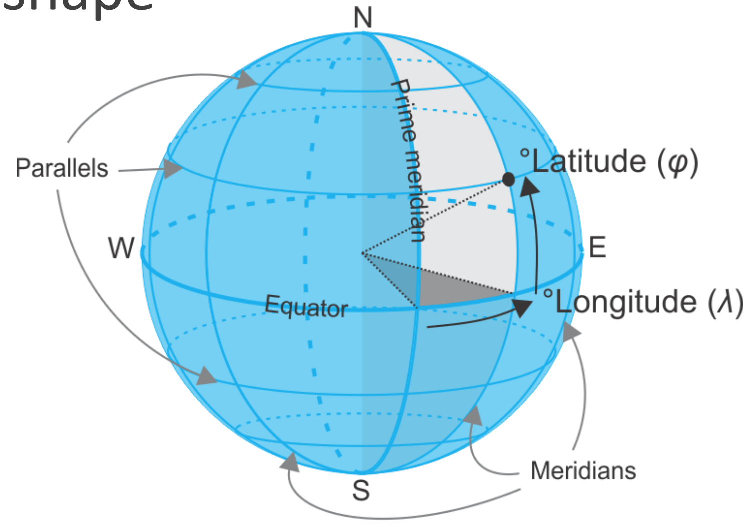
Best overall fit for most places on earth

NAD83 (EPSG: 4269)

Based on satellites and survey data

Best fit for USA

Used by many federal agencies (Census)



CRSs are referenced in software by numeric codes, often called **EPSG codes**

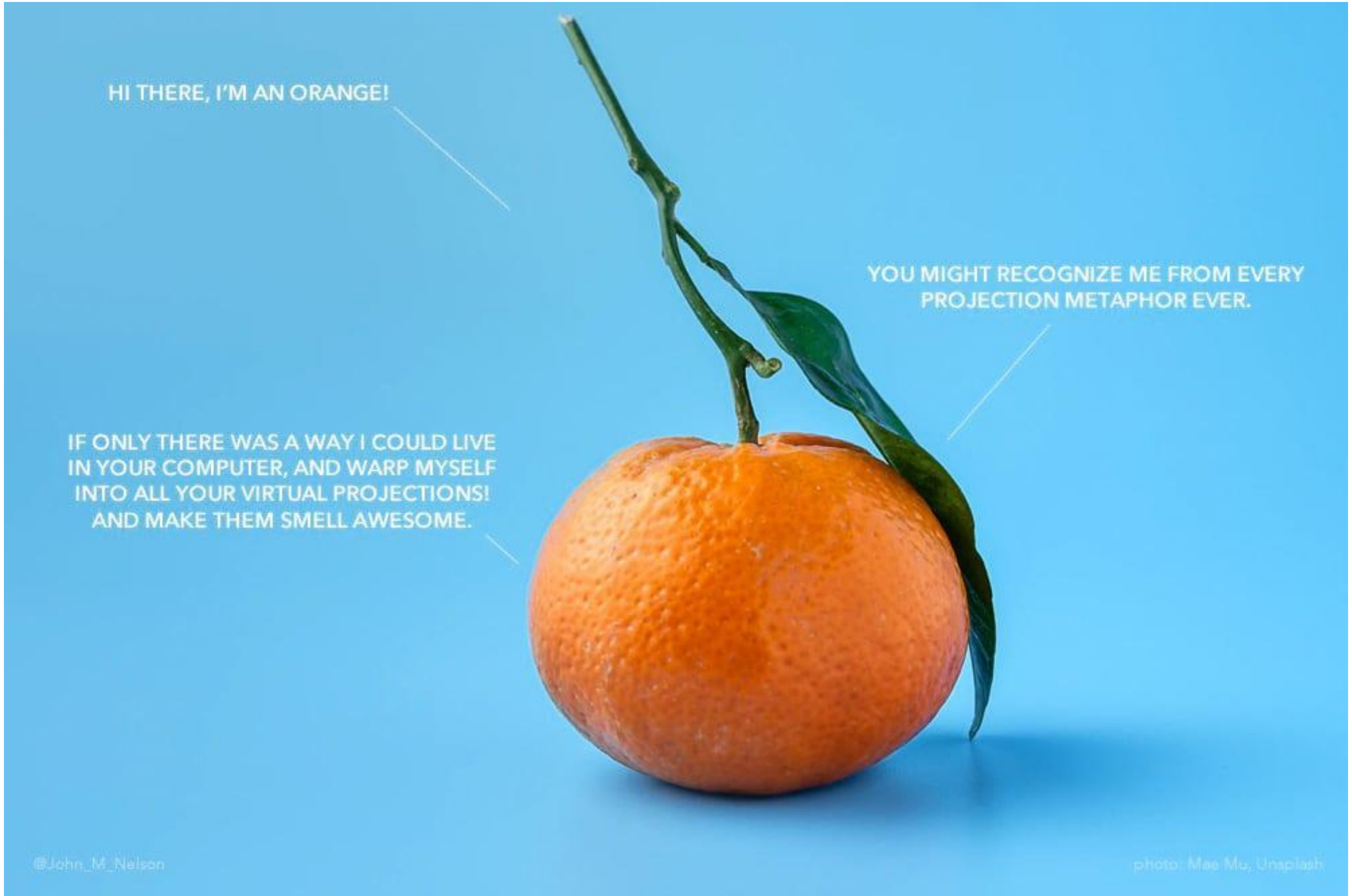
Map Projections: 3D to 2D

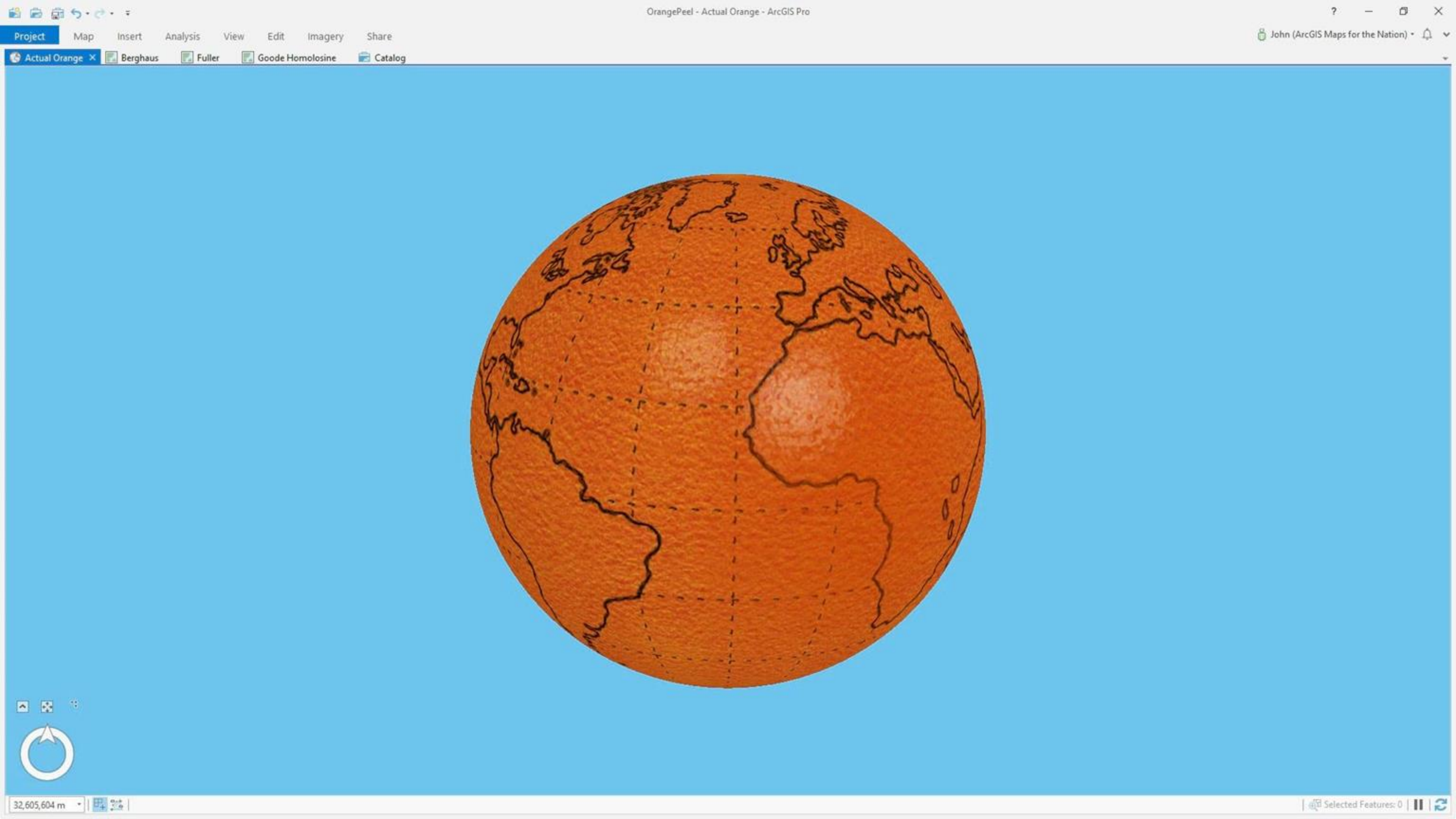
The earth isn't flat (spoiler), but maps and (most) computer screens are

We use **map projections** to transform a datum (3D spherical model of the earth) to a 2D plane.

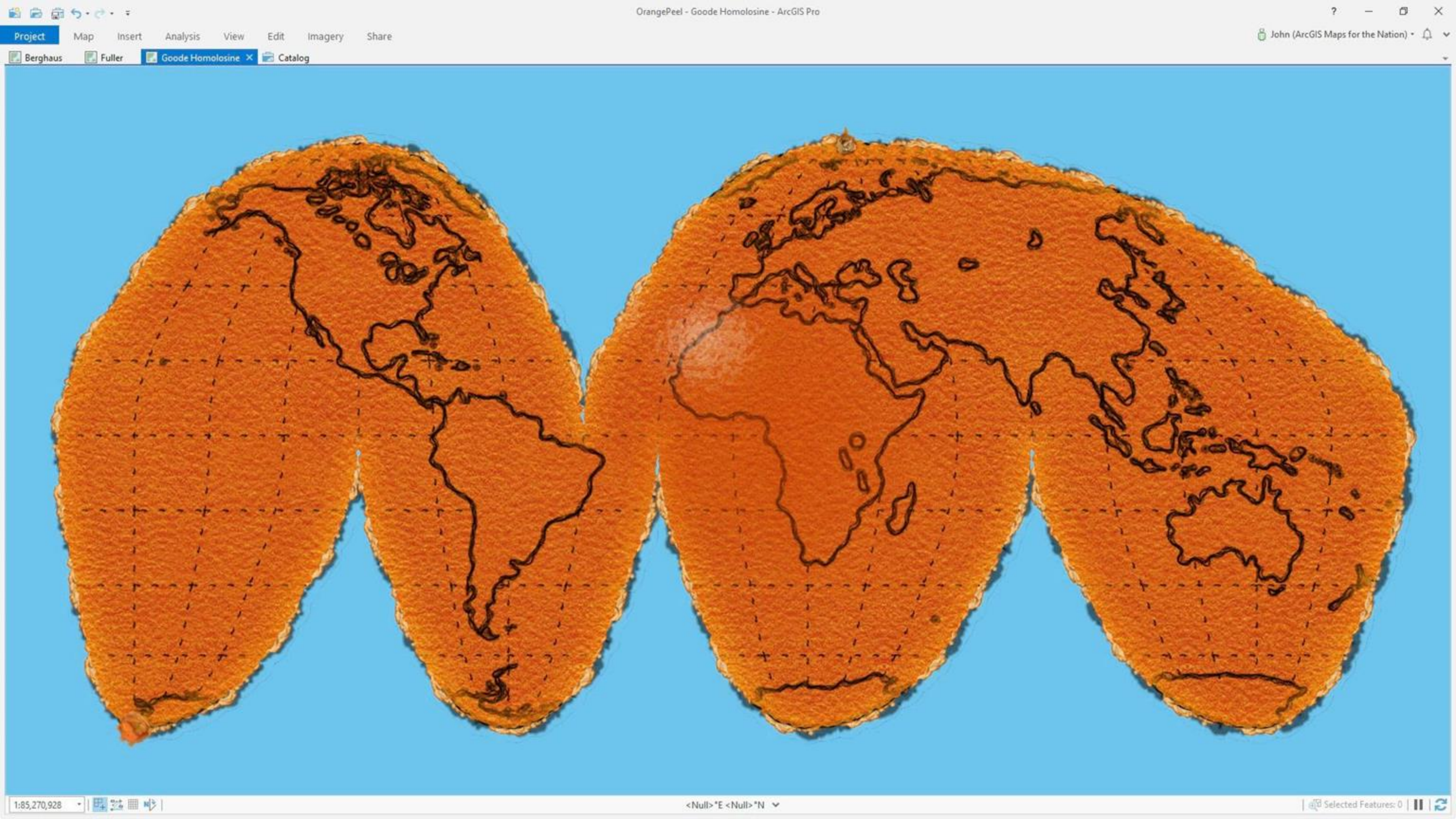
This necessarily introduces **distortion**, which might impact your maps and spatial analyses!

Map Projections: 3D to 2D

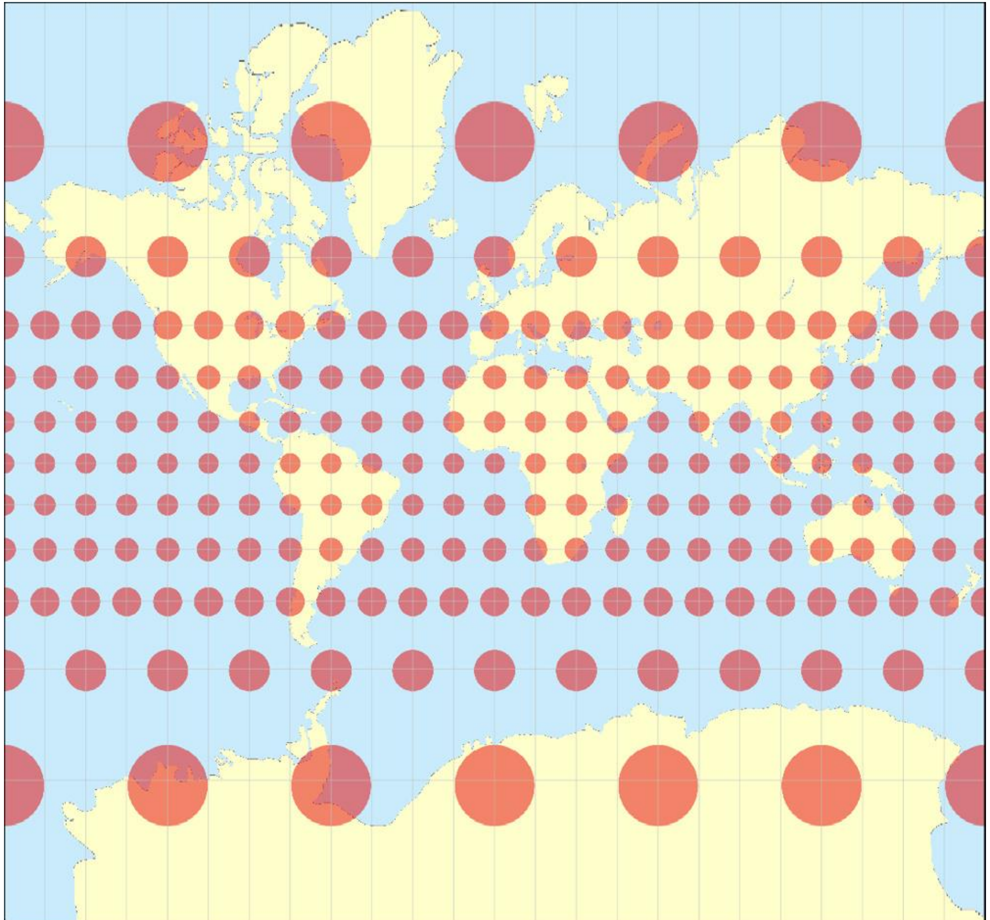




OrangePeel - Actual Orange - ArcGIS Pro
Project
Map
Insert
Analysis
View
Edit
Imagery
Share
John (ArcGIS Maps for the Nation)
Actual Orange
Berghaus
Fuller
Goode Homolosine
Catalog
32,605,604 m
Selected Features: 0

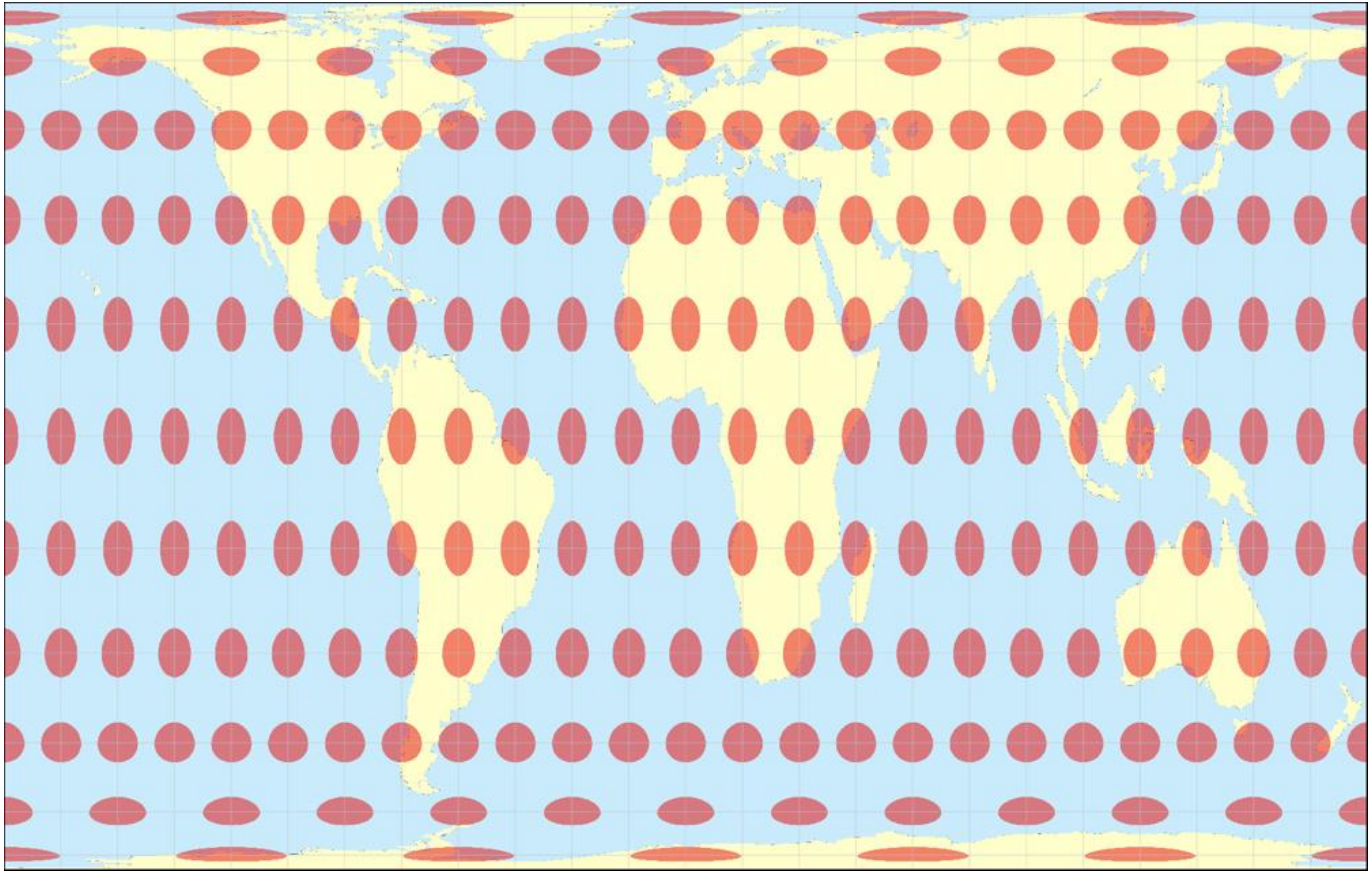

OrangePeel - Goode Homolosine - ArcGIS Pro
Project
Map
Insert
Analysis
View
Edit
Imagery
Share
John (ArcGIS Maps for the Nation)
Berghaus
Fuller
Goode Homolosine
Catalog
1:85,270,928
<Null>°E <Null>°N
Selected Features: 0

Mercator Projection: Preserves Angles

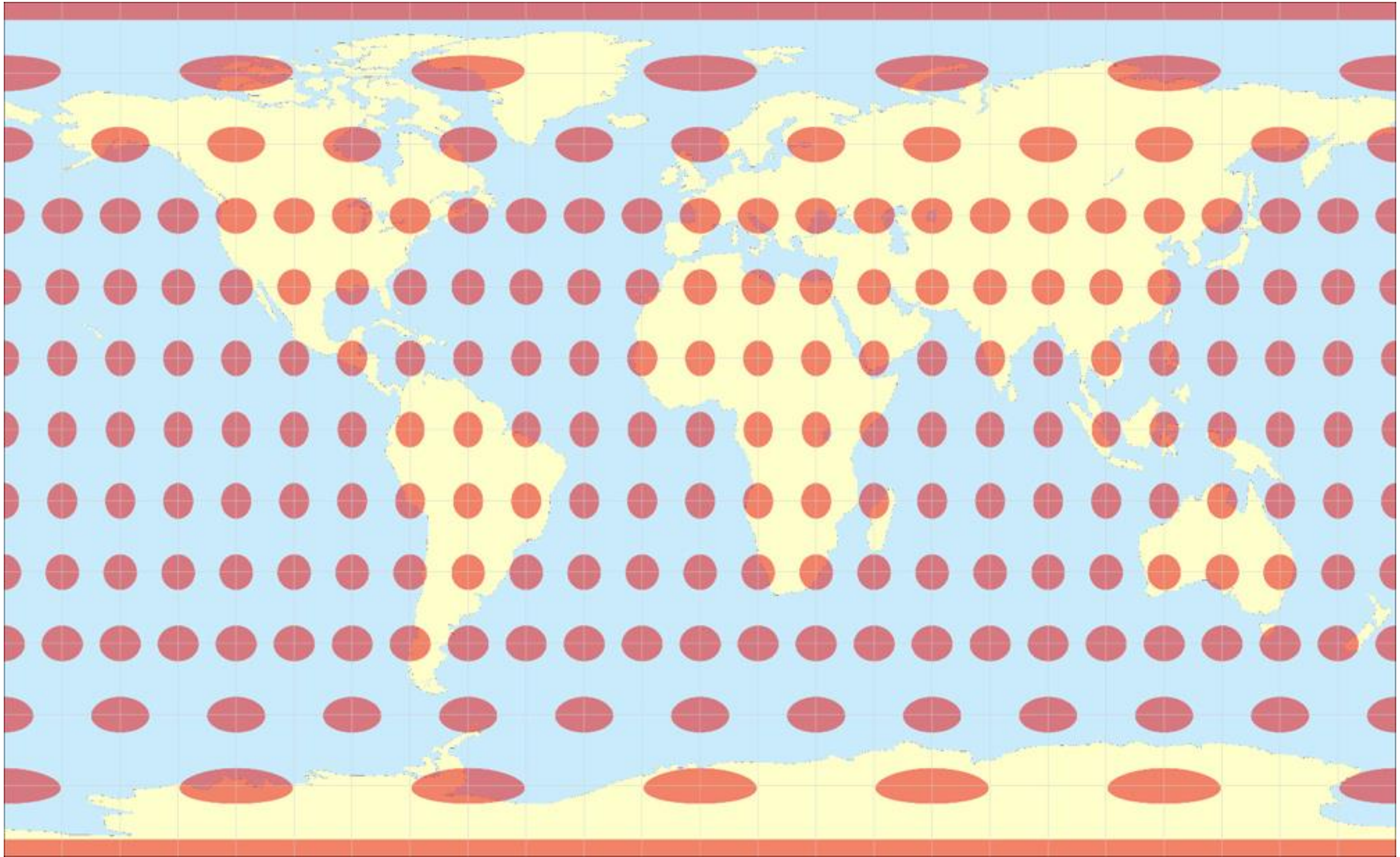


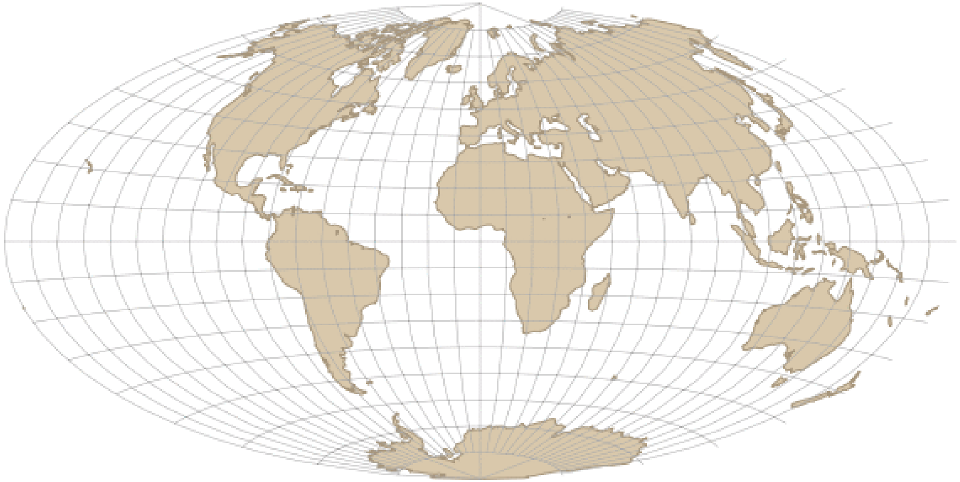
Gall-Peters Projection: Preserves Area

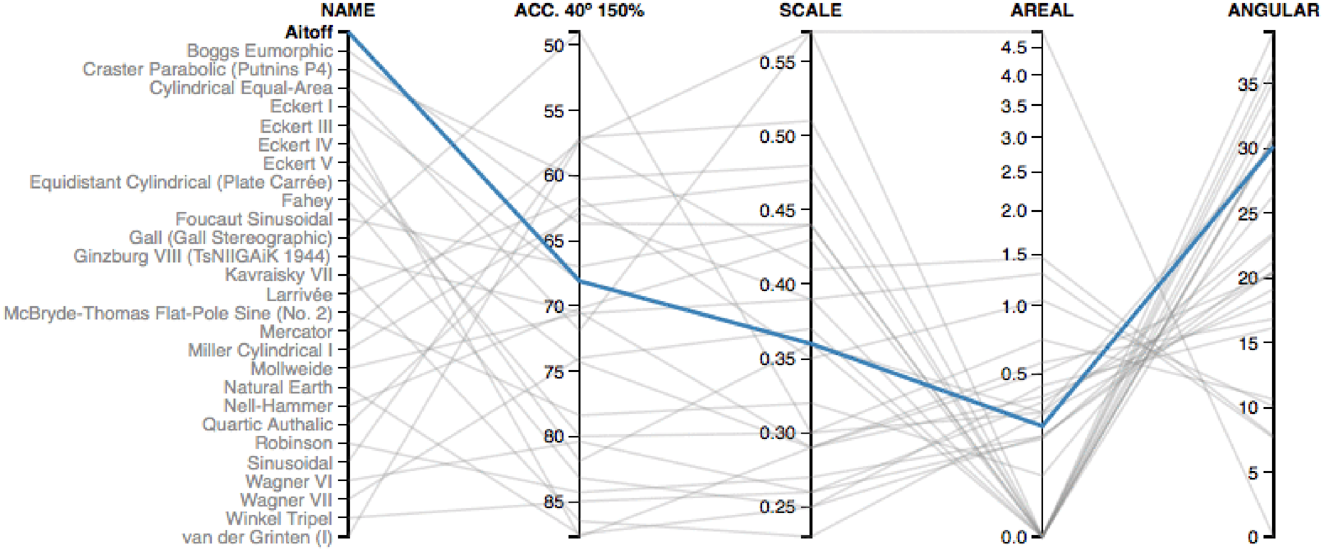
Relevant XKCD comic (977)



Equirectangular Projection: Preserves Distance





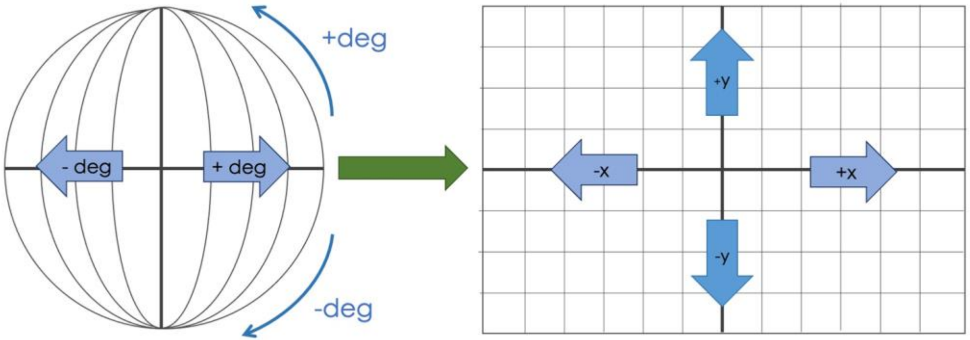


NAME
ACC. 40° 150%
SCALE
AREAL
ANGULAR
Aitoff
Boggs Eumorphic
Craster Parabolic (Putnins P4)
Cylindrical Equal-Area
Eckert I
Eckert III
Eckert IV
Eckert V
Equidistant Cylindrical (Plate Carrée)
Fahey
Foucaut Sinusoidal
Gall (Gall Stereographic)
Ginzburg VIII (TsNIIGAiK 1944)
Kavraisky VII
Larrivée
McBryde-Thomas Flat-Pole Sine (No. 2)
Mercator
Miller Cylindrical I
Mollweide
Natural Earth
Nell-Hammer
Quartic Authalic
Robinson
Sinusoidal
Wagner VI
Wagner VII
Winkel Tripel
van der Grinten (I)
50
55
60
65
70
75
80
85
0.55
0.50
0.45
0.40
0.35
0.30
0.25
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
35
30
25
20
15
10
5
0

Projected Coordinate System (PCS)

In short:

- Cartography is bonkers
- Map projections are needed to go from a 3D datum to 2D coordinates (i.e. making maps)
- They introduce distortion in at least area, shape, distance, or direction

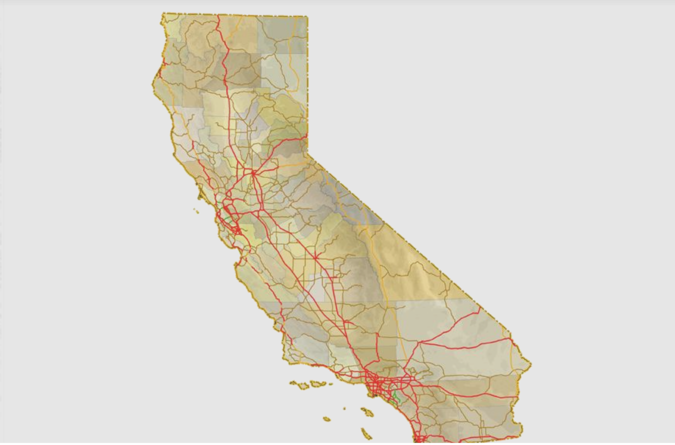


Projected Coordinate System (PCS)

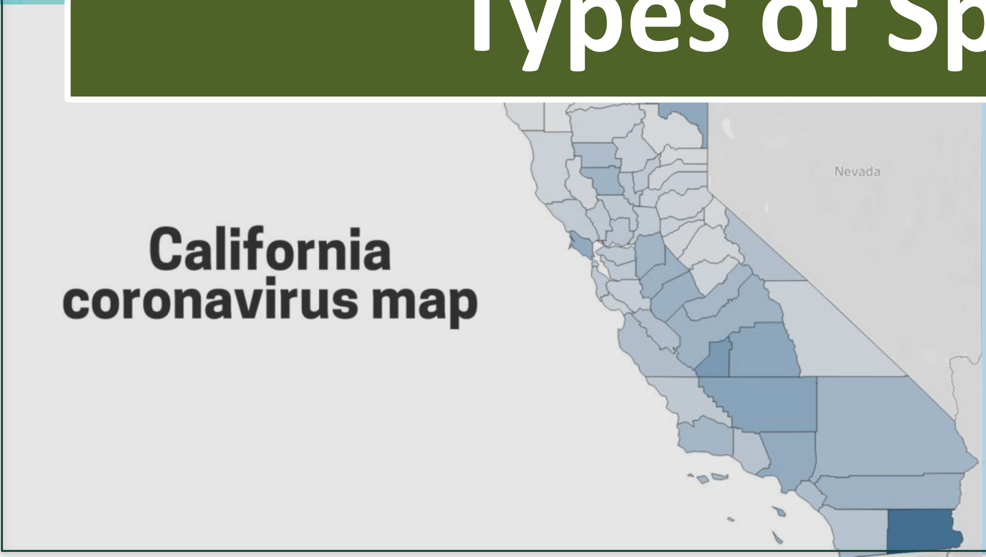
- Fortunately for us, the `sf` package in R performs most geometric operations using spherical geometry, so ignores these distortion issues!

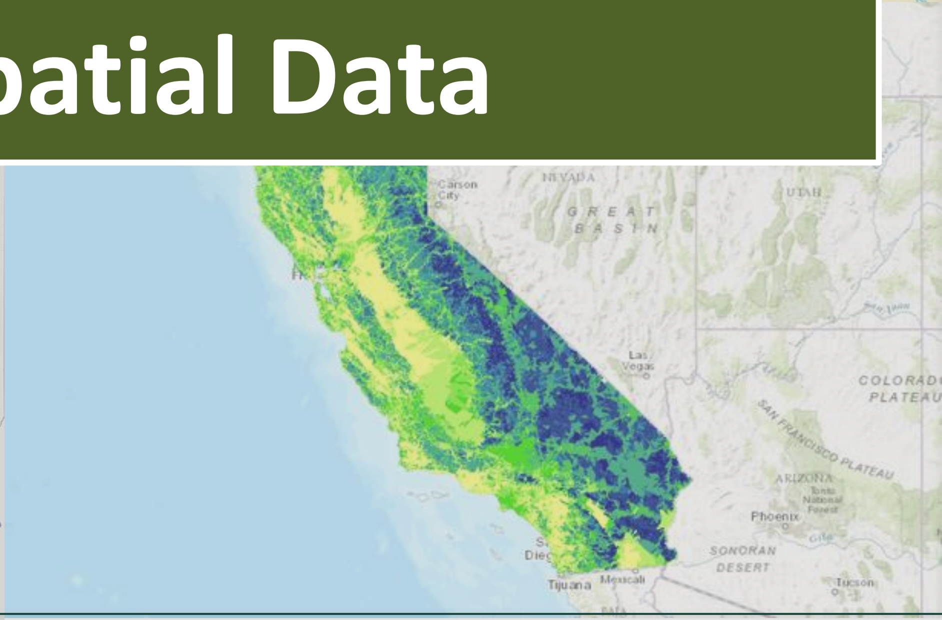
We'll still need to

- Know the CRS of your input data
- Select the CRS that is most suitable for your data and application (we'll learn how to transform to other CRSs)
- Use a projection for making maps



Types of Spatial Data





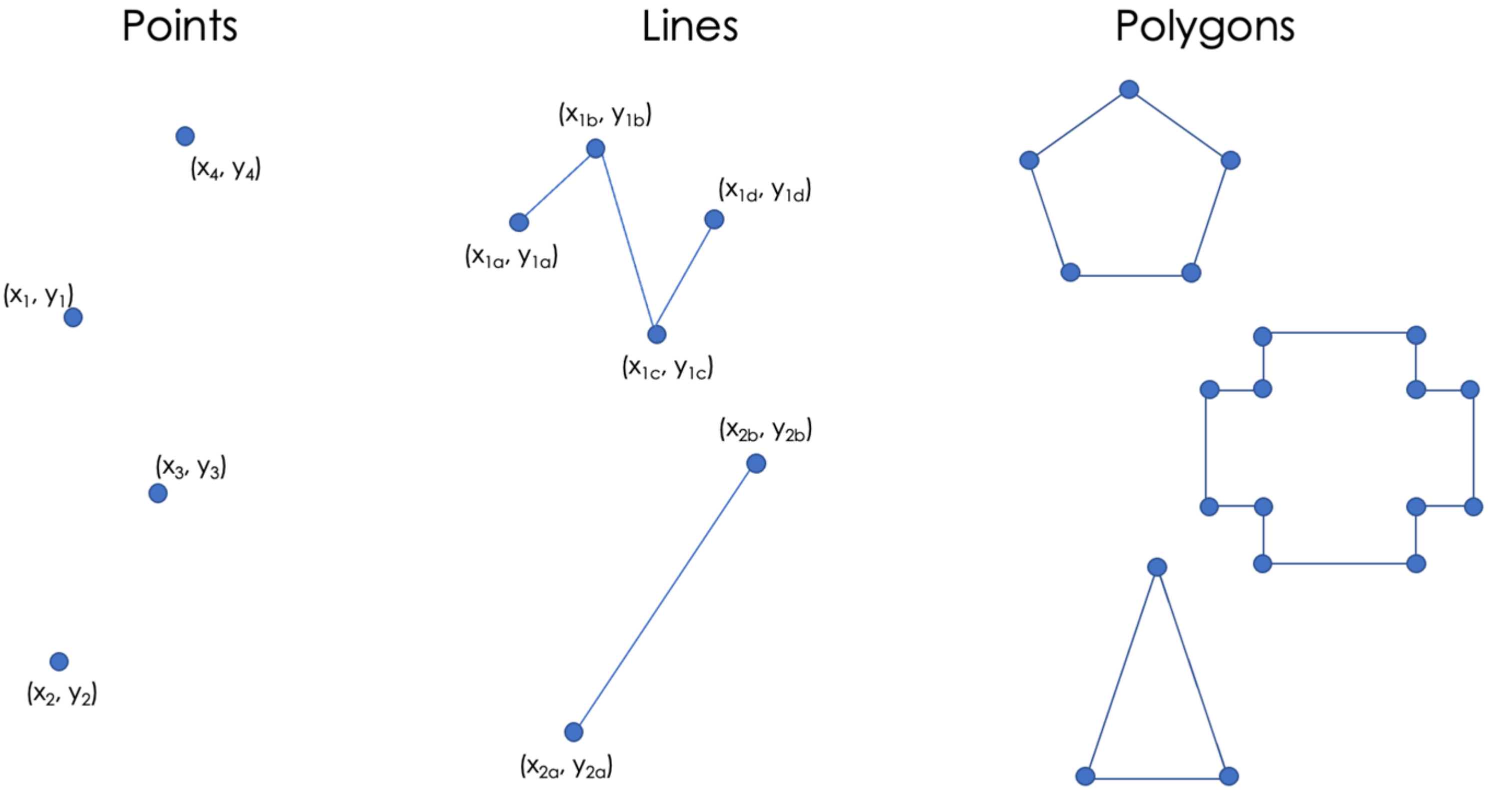
Types of Spatial Data

There are **two** main spatial data structures:

Vector

Raster

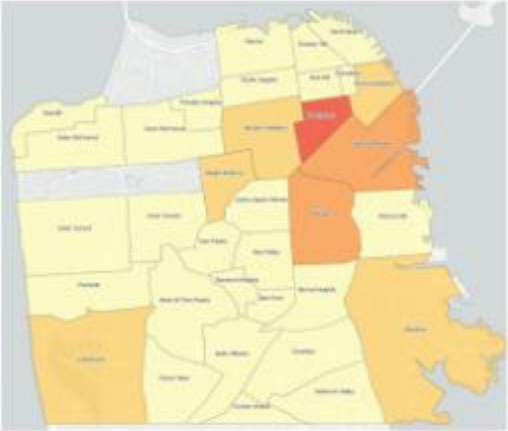
Vector Data: “Connect the Dots”



Vector Data: Points, Lines, and Polygons

Crime locations

City freeways



Neighborhoods

Vector Data with Attributes

Each row represents one geospatial **feature**

ID
1

Vector Data with Attributes

Each row represents one geospatial **feature**

Attributes describe the features (fields or columns)

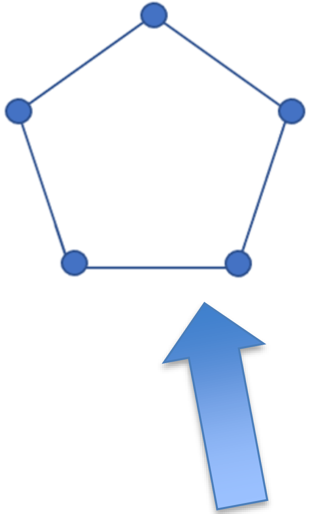
ID	City	Population
1	Detroit	645,658

Vector Data with Attributes

Each row represents one geospatial **feature**

Attributes describe the features (fields or columns)

Each feature has an associated **geometry** or **geometry collection**



ID	City	Population
1	Detroit	645,658

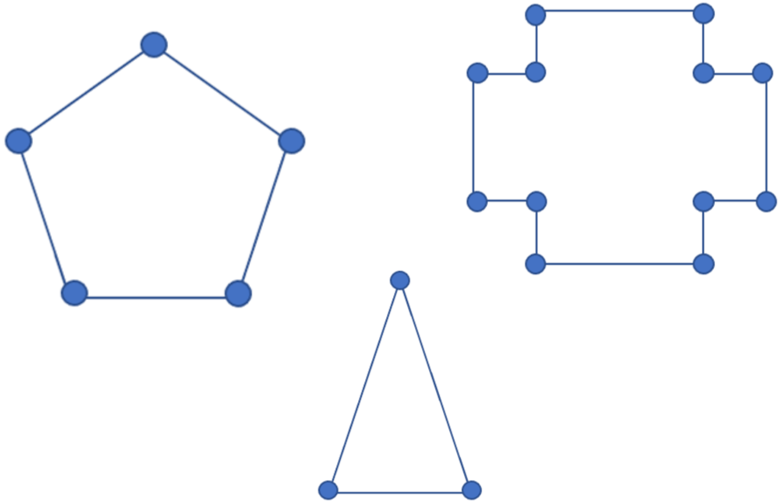
Vector Data with Attributes

Each row represents one geospatial **feature**

Attributes describe the features (fields or columns)

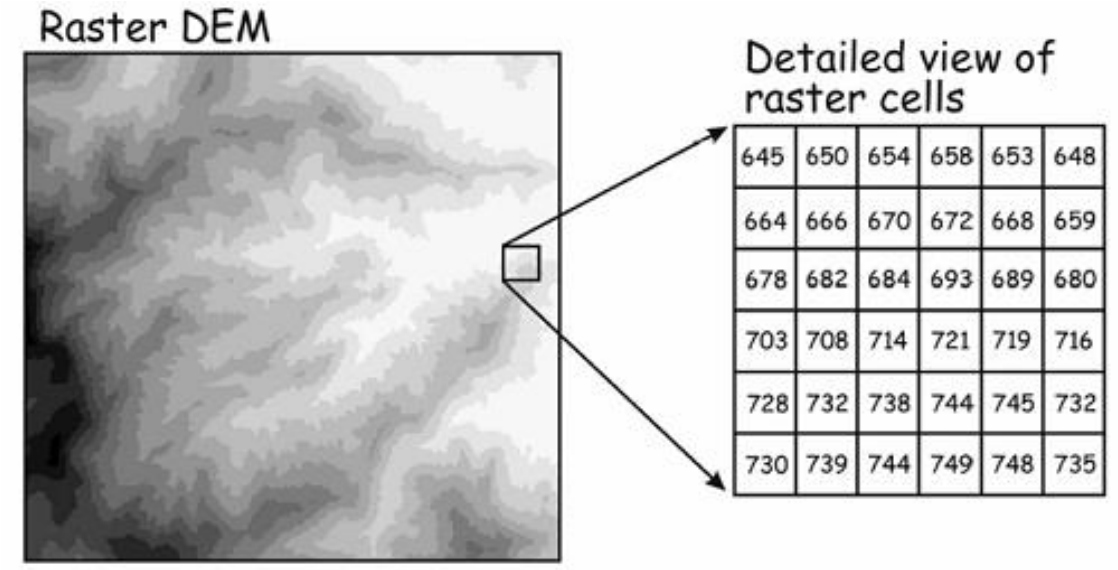
Each feature has an associated **geometry** or **geometry collection**

A group of features is called a **layer**



ID	City	Population
1	Detroit	645,658
2	Lansing	113,592
3	East Lansing	47,427

Raster Data: regular grids describing continuous phenomena



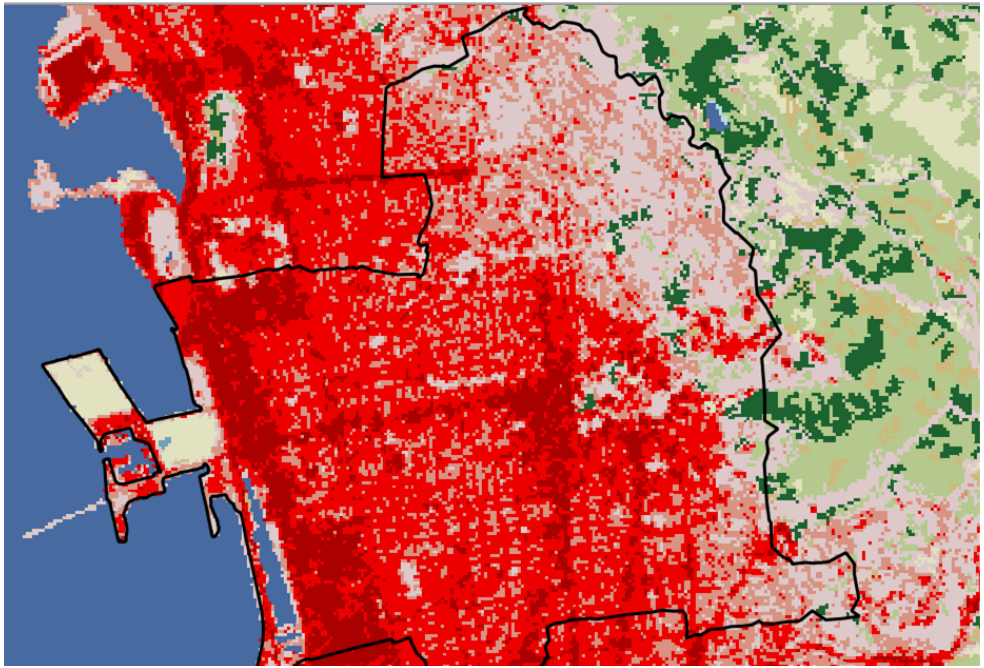
A location is divided into rectangular **cells**

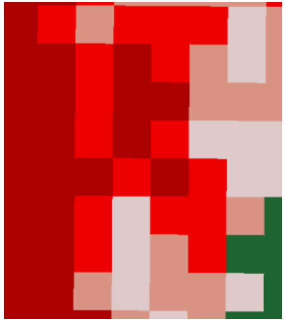
Cells have regular size (e.g. 20m x 20m)

Grid has a fixed number of rows and columns

Each cell has a value that represents the attribute of interest (e.g. elevation)

Categorical Raster Data





30m pixels

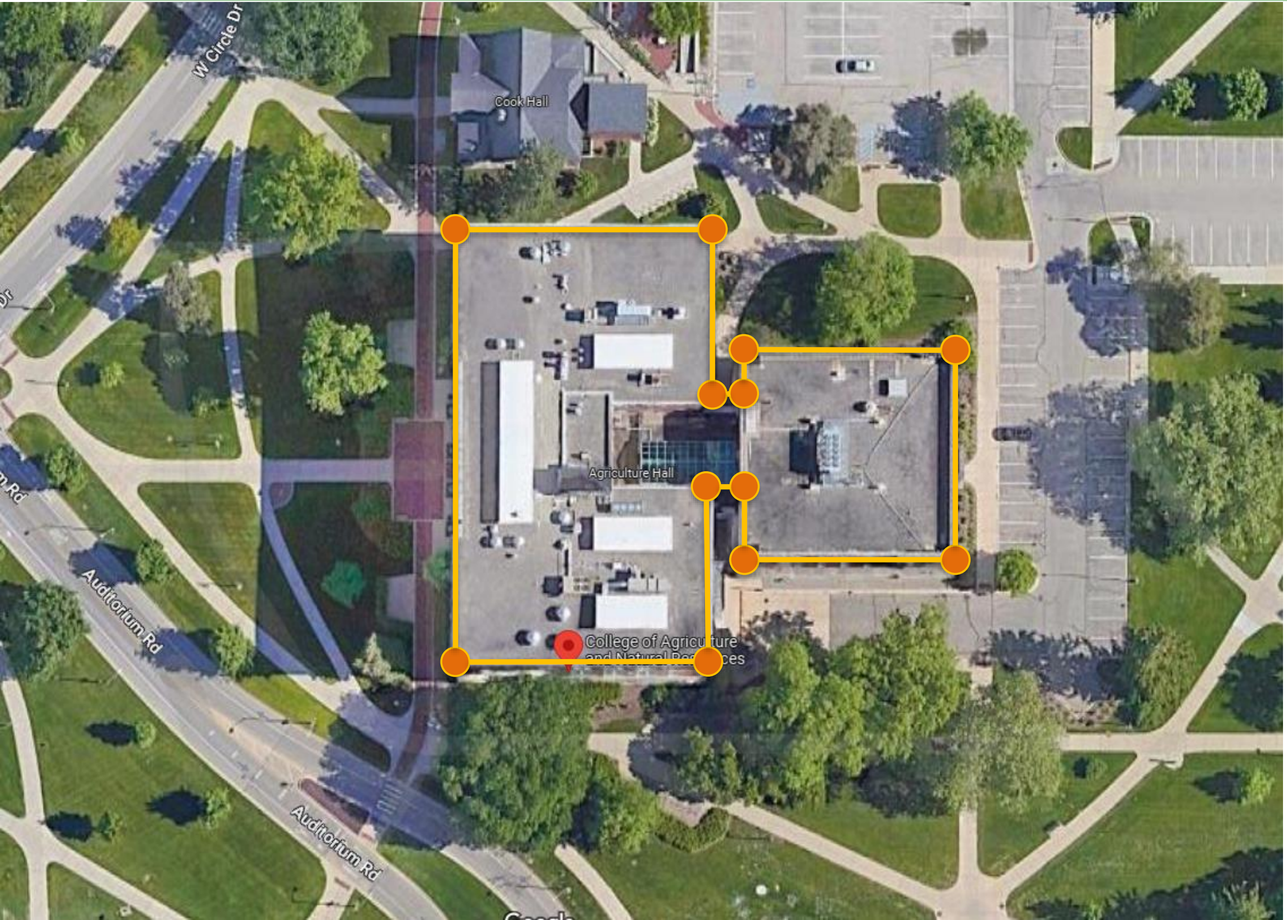
NLCD Land Cover Classification Legend

- 11 Open Water
- 12 Perennial Ice/ Snow
- 21 Developed, Open Space
- 22 Developed, Low Intensity
- 23 Developed, Medium Intensity
- 24 Developed, High Intensity
- 31 Barren Land (Rock/Sand/Clay)
- 41 Deciduous Forest
- 42 Evergreen Forest
- 43 Mixed Forest
- 51 Dwarf Scrub*
- 52 Shrub/Scrub
- 71 Grassland/Herbaceous
- 72 Sedge/Herbaceous*
- 73 Lichens*
- 74 Moss*
- 81 Pasture/Hay
- 82 Cultivated Crops
- 90 Woody Wetlands
- 95 Emergent Herbaceous Wetlands

* Alaska only

Satellite Imagery Data are Raster Data

Satellite Imagery Data are Raster Data



Note: aerial or satellite imagery and other remotely-sensed data are commonly used as **sources** of vector data

I.e. the Ag Hall building footprint can be digitized and converted to vector data.

Vector vs. Raster

Vector data are **better for discretely bounded** data

e.g. political boundaries, waterways, parcels

Raster data are **better for continuous data**

e.g. temperature, rainfall, elevation, night lights

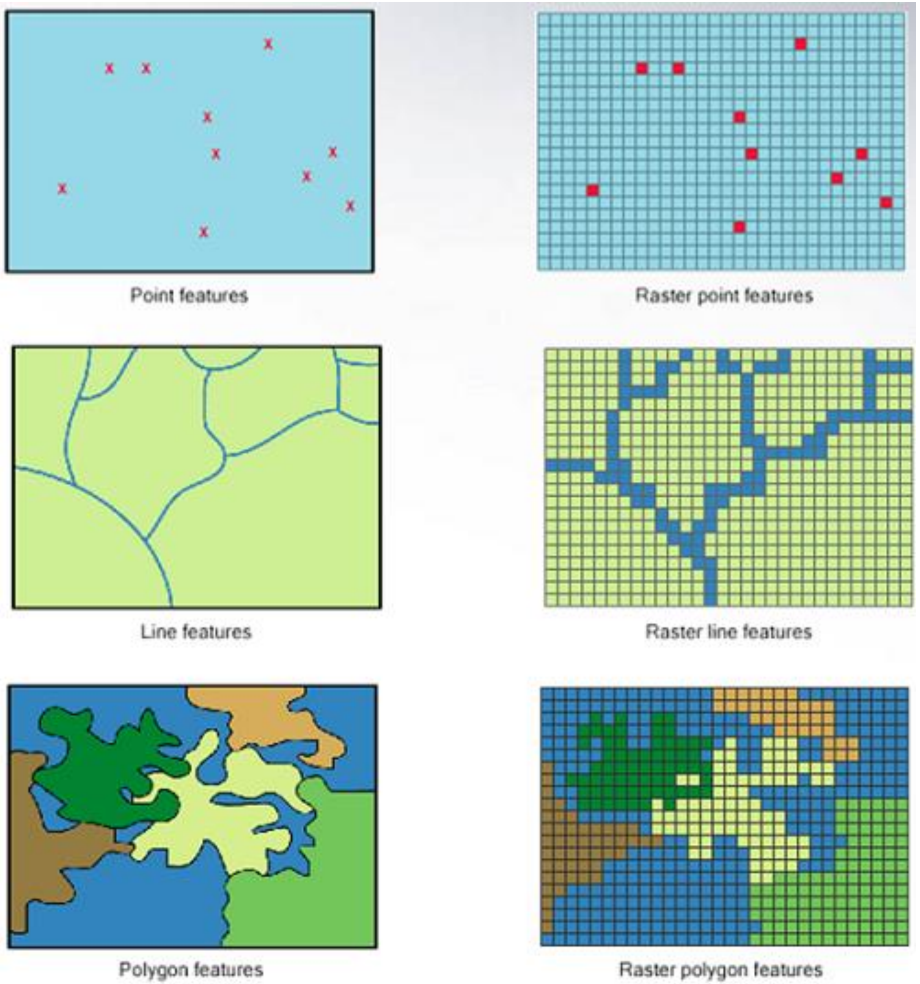


Image: http://gsp.humboldt.edu/

Common Spatial File Formats

Vector

- Shapefile (.shp)
- GeoJSON, JSON
- KML
- GeoPackage

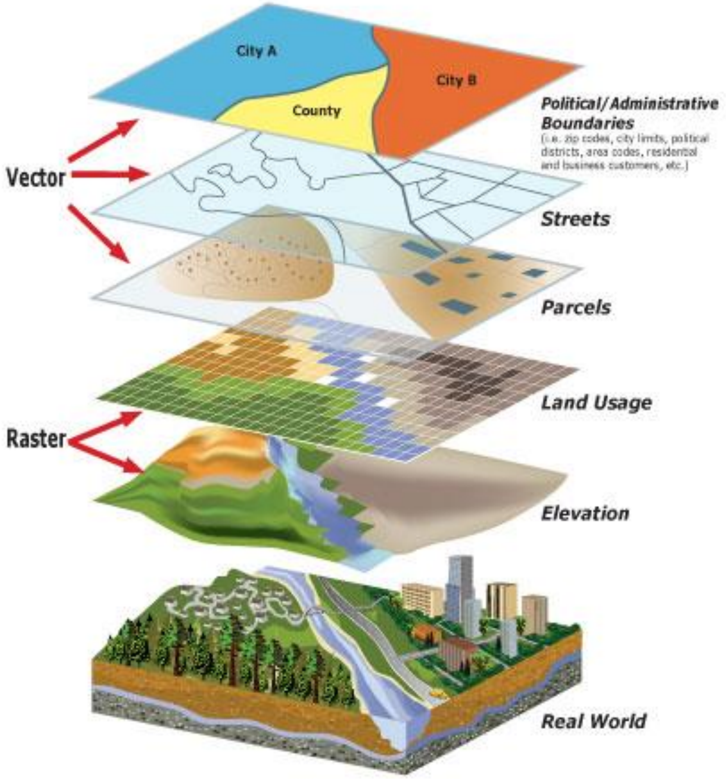
Raster

- GeoTIFF (.tif)
- netCDF
- DEM

Georeferencing

Data layers in the **same coordinate reference system** can be linked dynamically to explore associations and build complex models of the real world

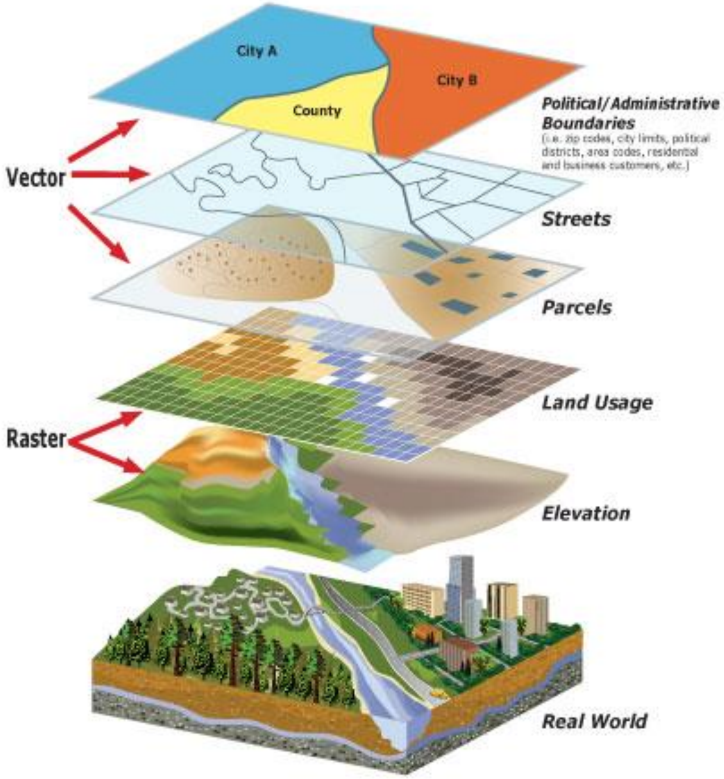
- Can link rasters and vectors



The Power of Spatial Data

What does this mean for us?

- Add vector data to create maps of features and attributes
- Combine, split, and perform operations on vectors
- Extract raster values for the vector features, across space and/or time



The Power of Spatial Data

A recent workflow of mine:

- Link vector parcel data for two Northern CA counties to water billing records
- Extract raster temp/precip data for each parcel/billing cycle
- Read in vector data of pricing zones, calculate distance of each parcel to boundary
- Use spatial regression discontinuity models
- Make a bunch of maps