Brief Intro to Geospatial Data

Team Geo

D-Lab

About me

Questions about you!

What is your department or program?

Are you a graduate or undergraduate student? Faculty? Staff? Affiliate?

Have you worked with

- Geospatial data?
- ArcGIS or QGIS?
- Other? If yes what?

How would you describe your level of R programming knowledge? New, beginner, advanced beginner, intermediate, competent....

Outline

- What’s the D-Lab?
- Workshop Introduction
- Geospatial Data
- Coordinate Reference Systems
- Types of Spatial Data
- GIS Software
- Geospatial Workflows

What is the D-Lab?

https://dlab.berkeley.edu/

What is the D-Lab?

- The DLab provides support for Berkeley students, faculty, staff and affiliated scholars who use **computational tools and methods** in their research.
- We do this primarily by offering **workshops** and **consulting** services on programming tools, software packages, data analysis, data management, and related techniques and technologies.
- See: https://dlab.berkeley.edu
- Subscribe to our **newsletter**!

At every stage of the research process

- Research design
- Survey methods / sample design
- Data acquisition, cleaning & management
- Statistical methods and evaluation of results
- Qualitative data analysis / mixed methods
- Data visualization - communicating results
- Professional Development

Our Moto: It's OK Not 2 Know!

REALLY!

We aim to be

- inclusive,
- supportive,
- non-judgy!

In that same spirit, we're always trying to improve!

Make sure to fill out the feedback form at the end of this workshop using this form!!!

You can also find that link in the calendar invite for this workshop.

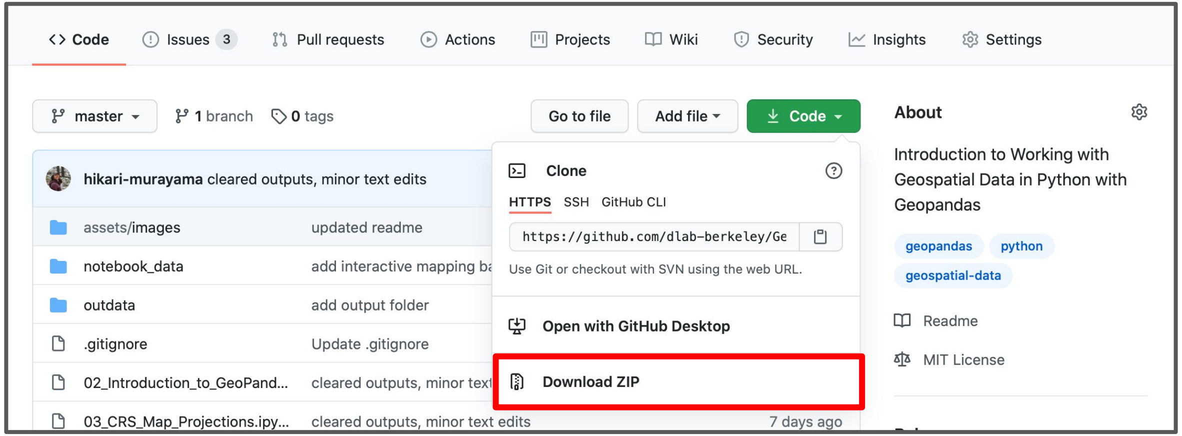
Thanks!

Our Workshop Today


START

How to follow along

- Find our workshop Github page on: https://github.com/dlab-berkeley
- Download workshop materials (I'll post the link in the chat)
- Click on the "**Code**" button to download the files (example below)



Let’s take a couple minutes to make sure we are all set

Information is on the workshop Github page:
https://github.com/dlab-berkeley/R-Geospatial-Fundamentals

- In the readme and the participant-instructions files

Today, you will need to have the following R packages installed:

- `sf, ggplot2, here, units, tmap, mapview, tidyverse`
- install.packages(`‘package_name’`)

Feel free unmute or drop a line in the chat to ask any one of us if you need help

Berkeley
UNIVERSITY OF CALIFORNIA

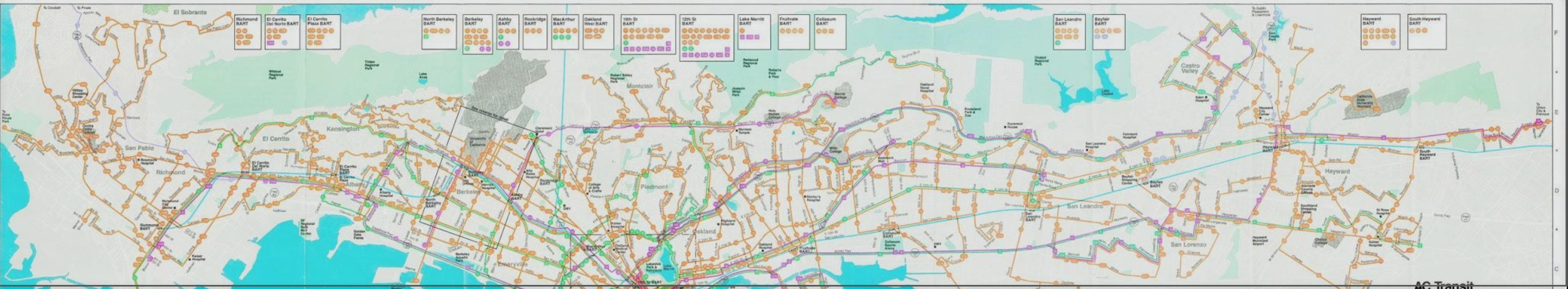
Workshop Goals

Introduction to working with geospatial data

- Geospatial data files and formats
- Reading and writing geospatial data in R
- Coordinate reference systems and Map Projections
- Working with different types of vector data
- Combining spatial and aspatial data
- Mapping geospatial data
- Asking questions of spatial data - or spatial queries
- Familiarity with common geospatial workflows

Geospatial Data

Lab



Geographic data

A single unit of geographic data includes:

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

Also great to have metadata:

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

Geospatial data

Encodes location geometrically with coordinates:

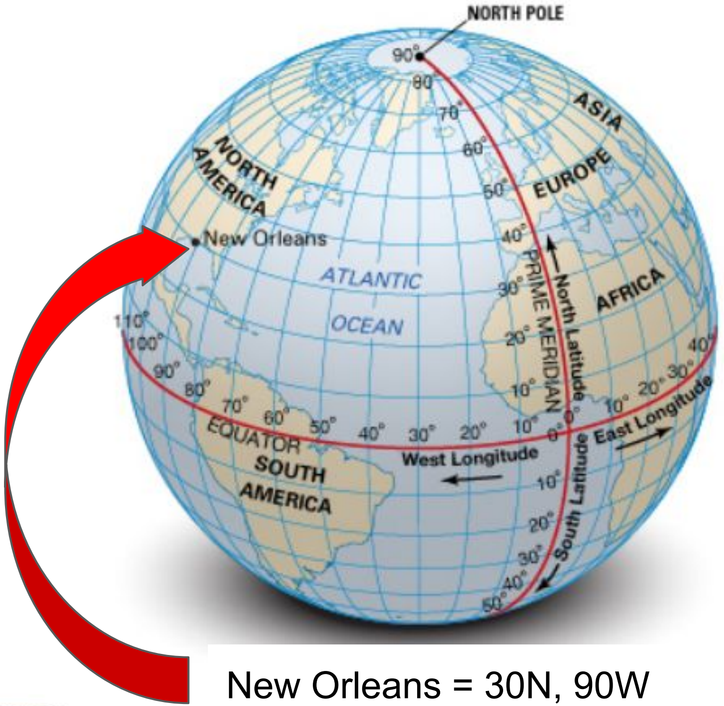
`Anatone, WA: 46.130479, -117.134167`

BUT, how do we know that those coordinates reference that specific location?

Berkeley
UNIVERSITY OF CALIFORNIA

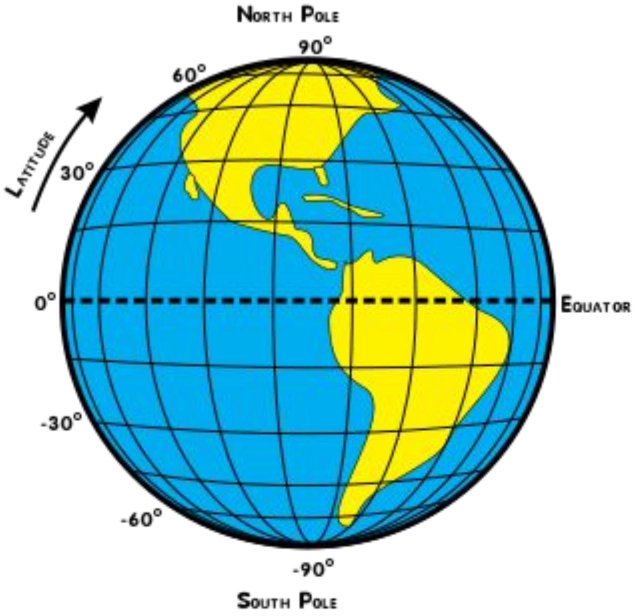
Coordinate Reference Systems (CRS)

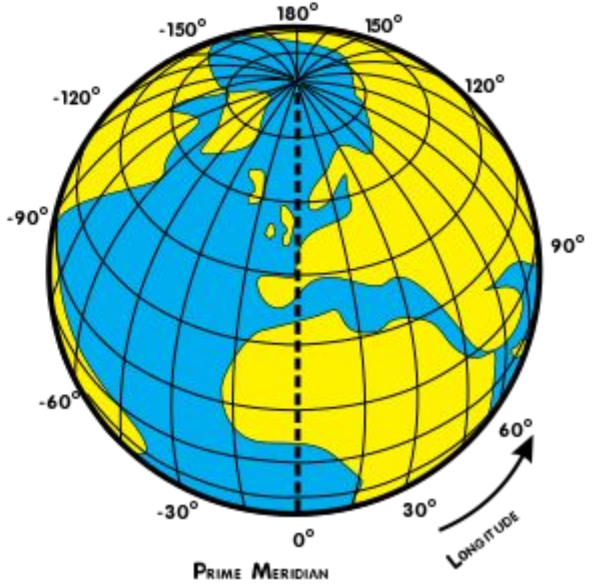
A Coordinate Reference System, or CRS, is a system for associating coordinates with a **specific**, **unambiguous** location on the surface of the Earth.



New Orleans = 30N, 90W

Coordinate Reference Systems (CRS)

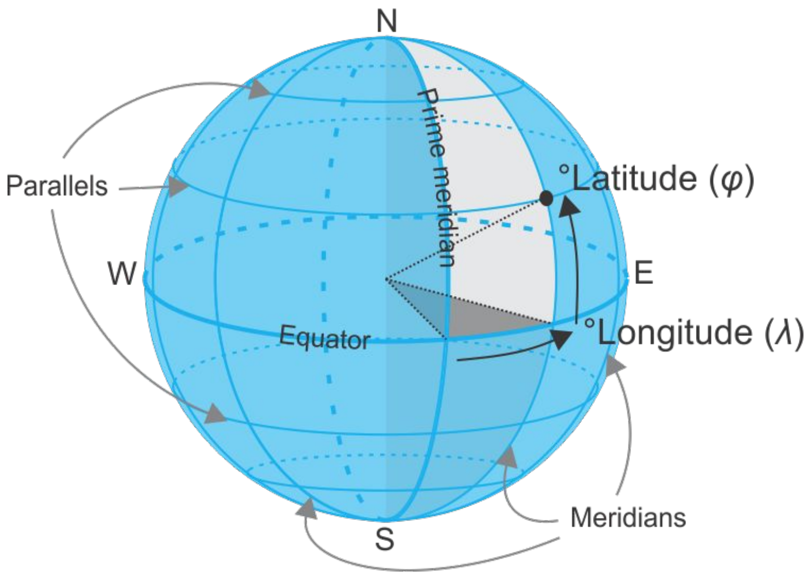




Geographic Coordinates: *Latitude* and *Longitude*

http://latitude-longitude.net

Coordinate Reference Systems (CRS)

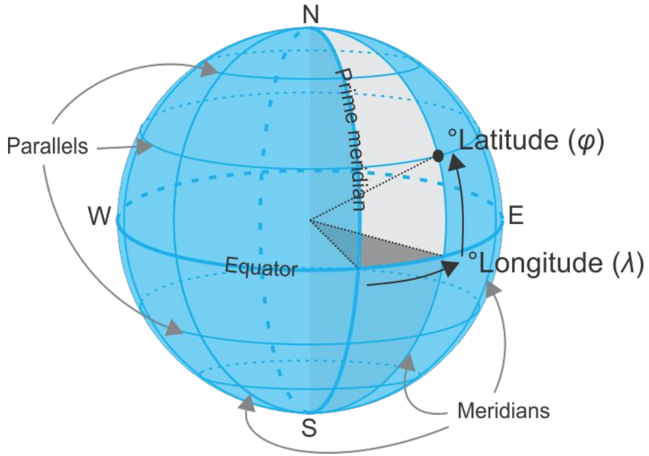


There are **many** CRSs, not just one!

Why? Because our understanding of and ability to measure the shape of the earth has changed over time.

Two Types of Coordinate Reference Systems

Geographic CRS



Angular Units = Degrees (DMS or DD)

Projected CRS

Cartesian Units = Feet or Meters

Good for local & regional mapping & analysis

Geographic Coordinate Systems (GCS)

Widely used! Expressed as latitude & longitude

WGS84 (EPSG: 4326)

Based on satellites, used by cell phones, GPS

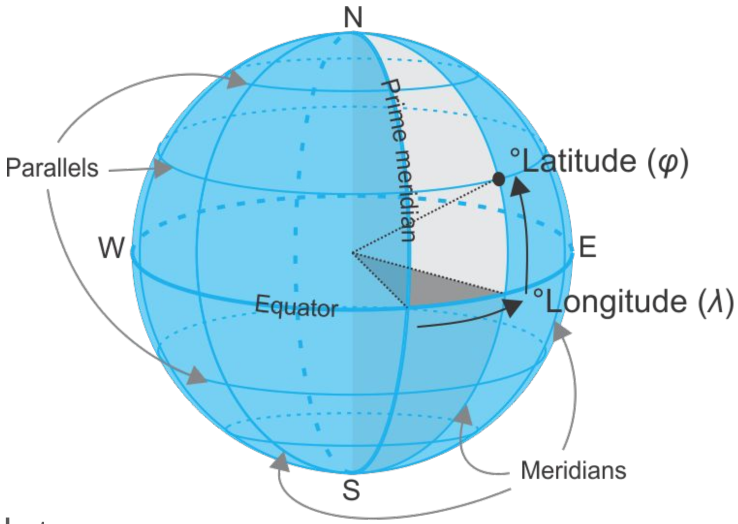
Best overall fit for most places on earth

NAD83 (EPSG: 4269)

Based on satellites and survey data

Best fit for USA

Used by many federal data products, like Census data



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

Berkeley UNIVERSITY OF CALIFORNIA Lab

Projected Coordinate Systems (PCS)

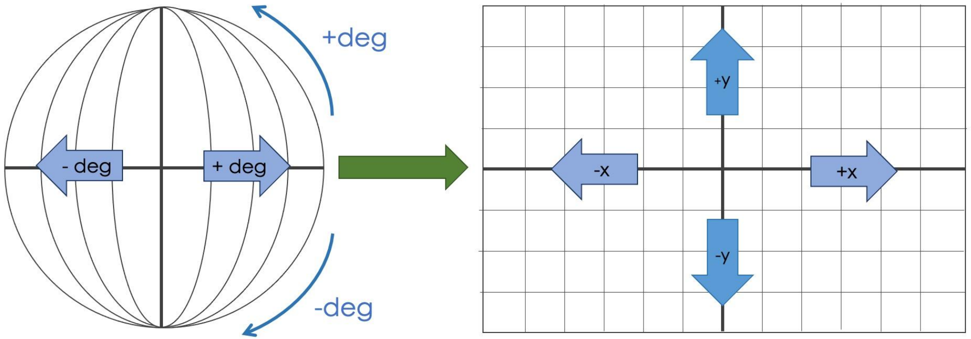
Map Projections transform geographic coordinates (lat/lon) to 2D coordinates (X/Y)

All map projections introduce **distortion** in area, shape, distance or direction.

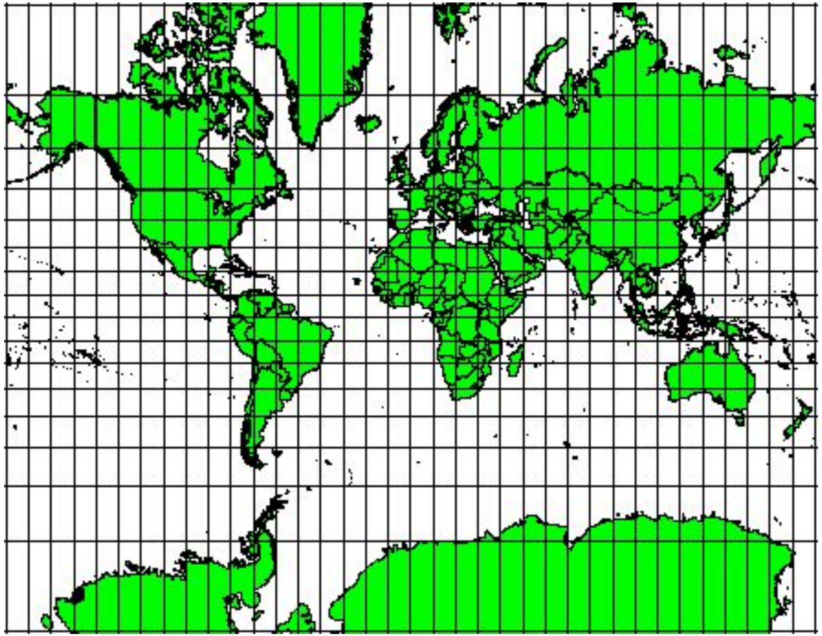
Specific map projections minimize distortion in one or more properties

You need to know the coordinate reference system of your input data

You need to select the CRS that is most suitable for your data and application.



The **Mercator projection**, for example, is used where angular relationships are important (shape, direction), but areas are distorted, especially as you move away from the equator.



Source: [QGIS Gentle introduction to GIS]

In short...

The Earth is not Flat!

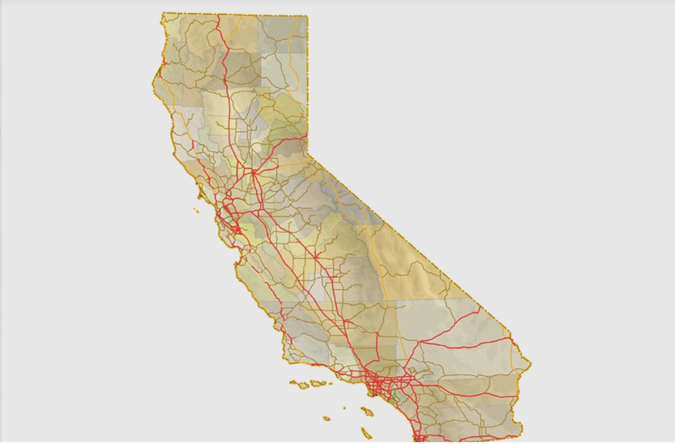
But computer screens, paper maps and most of the math underlying spatial analysis methods assume it is!

We use **map projections** to transform geographic coordinates from a 3D model of the earth to a 2D plane.

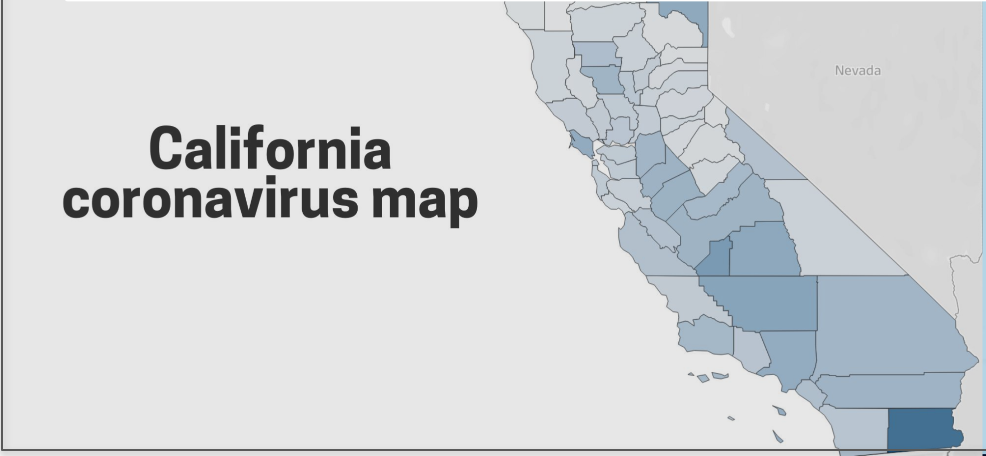
This makes location data much more usable *BUT*

It introduces distortion.

You need to understand how this distortion impacts your maps and spatial analysis!



Types of Spatial Data



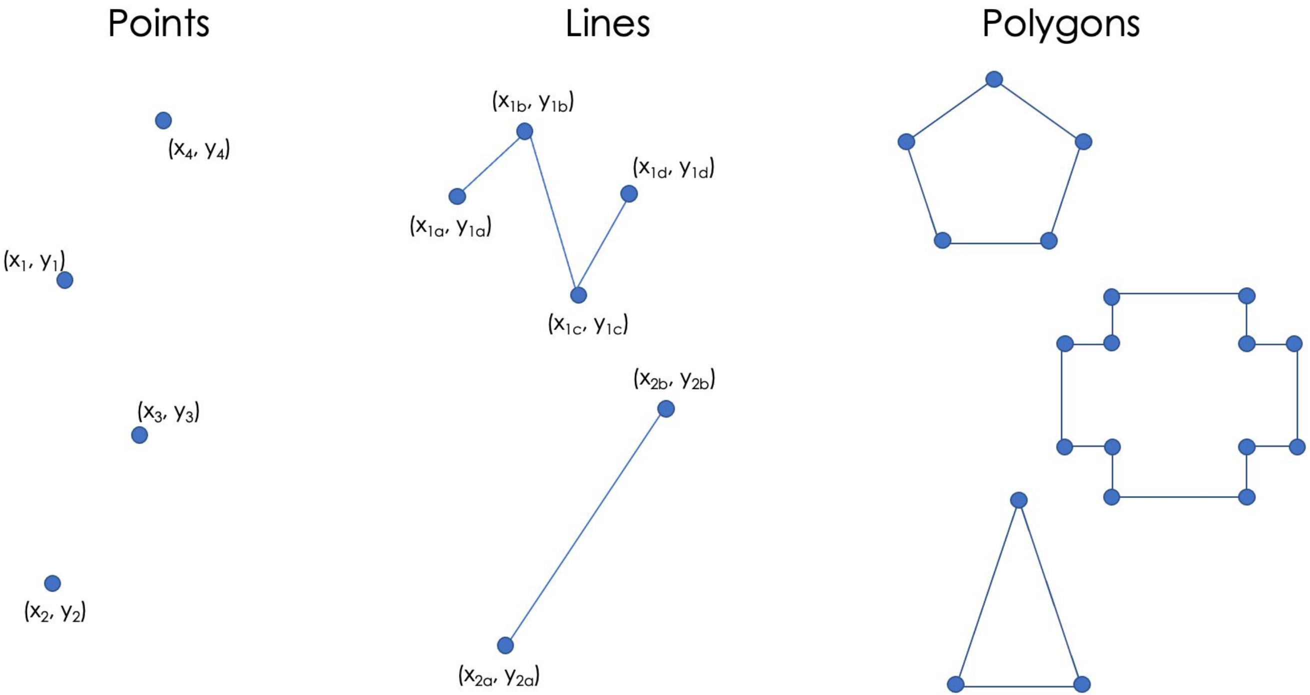
Types of Spatial Data

There are two fundamental spatial data models:

- Vector
- Raster

Vector Data

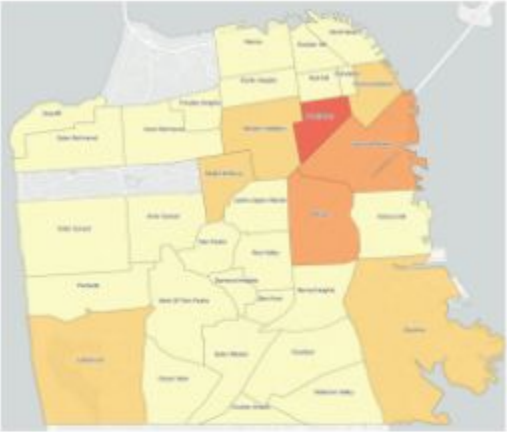
"Connect the dots"



Points, Lines, Polygons

Crime locations

City freeways



Neighborhoods

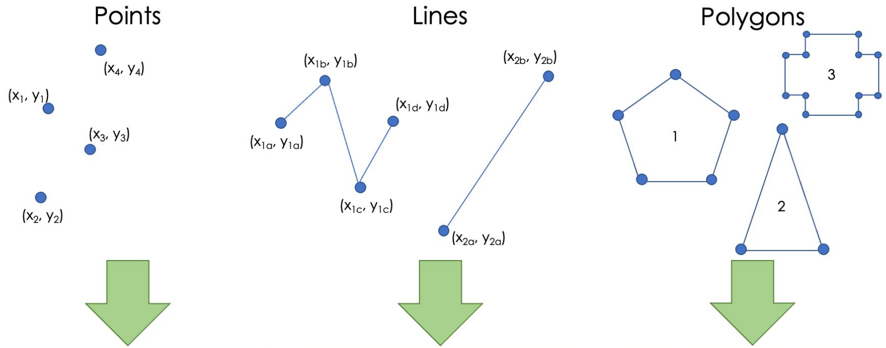
Vector Data with Attributes

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**

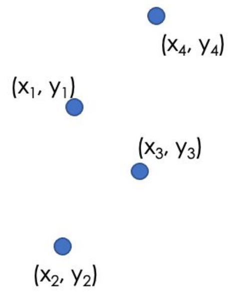


ID	School	# of Students
1	Berkeley High School	1392
2	MLK Middle School	421
3	Walkerville Elementary	130

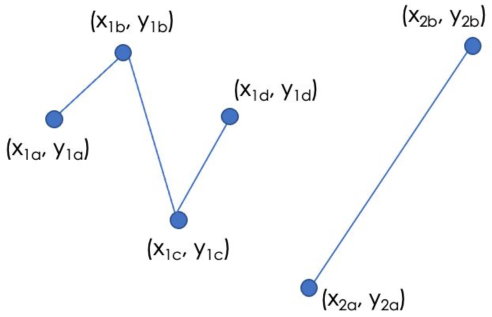
ID	Bridge	Toll Amount
1	Golden Gate Bridge	$7
2	Bay Bridge	$6
3	Rainbow Bridge	Free

ID	City	Mayor
1	Berkeley	Jesse Arreguín
2	Oakland	Libby Schaaf
3	Narnia	The White Witch

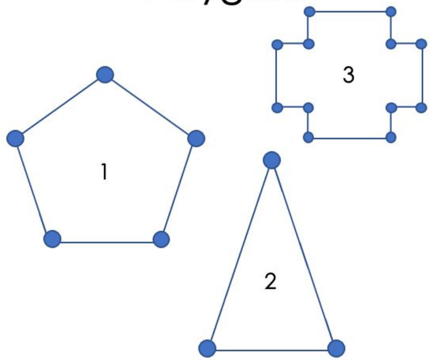
Points



Lines



Polygons

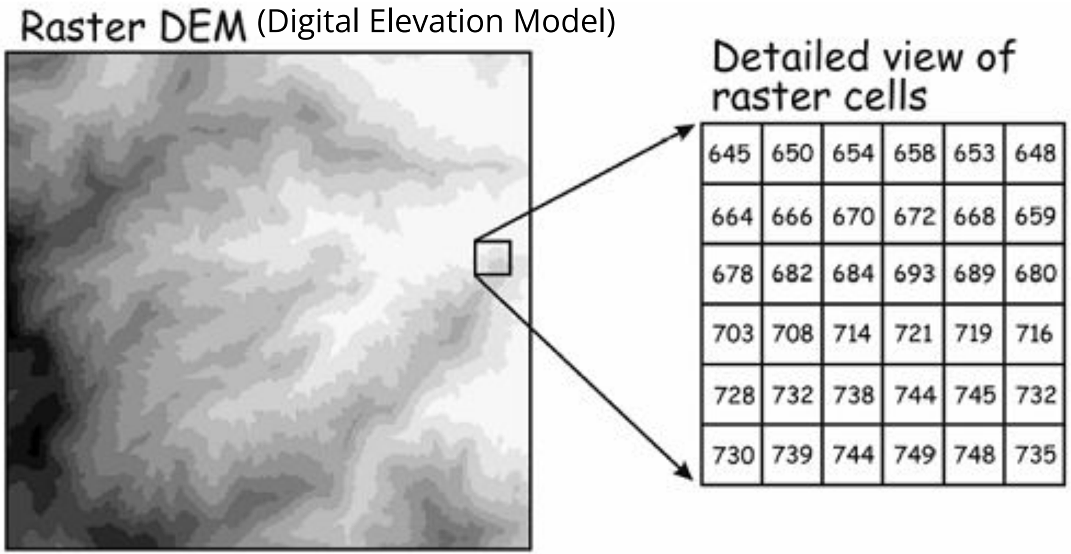


ID	School	# of Students
1	Berkeley High School	1392
2	MLK Middle School	421
3	Walkerville Elementary	130

ID	Bridge	Toll Amount
1	Golden Gate Bridge	$7
2	Bay Bridge	$6
3	Rainbow Bridge	Free

ID	City	Mayor
1	Berkeley	Jesse Arreguín
2	Oakland	Libby Schaaf
3	Narnia	The White Witch

Raster Data - regular grids



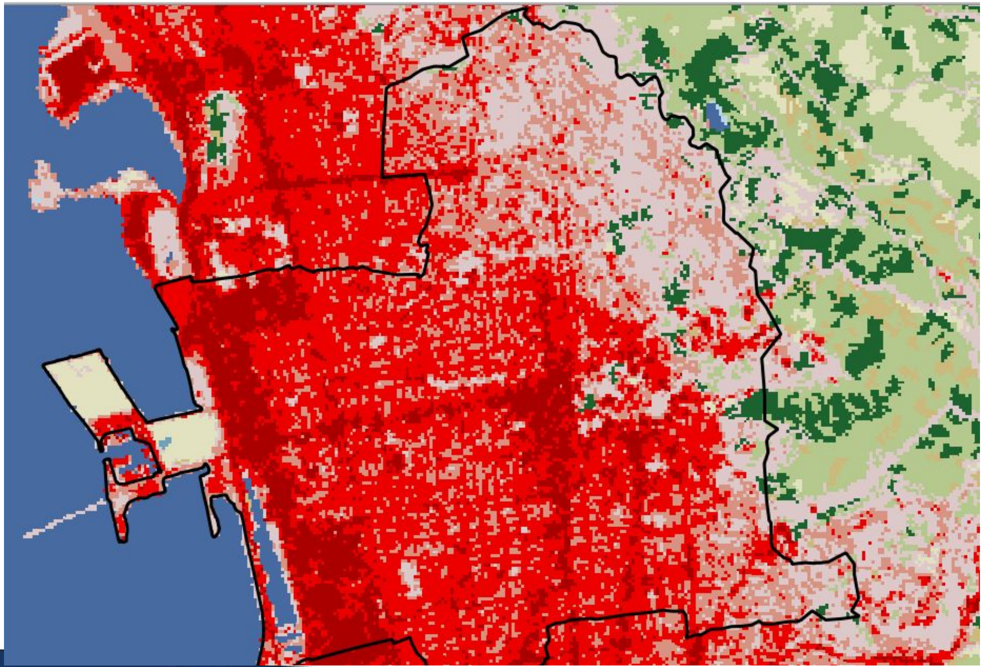
A location is represented by a grid cell

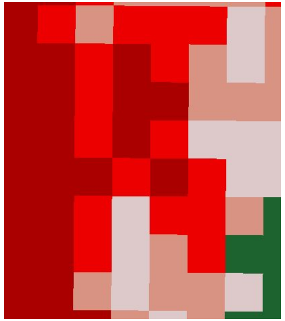
Cells have regular size, eg 30x30m

Grid has dimension - 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

Imagery Data are Raster Data

Note:
Aerial imagery, satellite data and other remotely sensed geographic data are commonly used as sources of vector data

In other words, the building footprints or streets can be digitized off of the imagery and saved as vector data.

Berkeley
UNIVERSITY OF CALIFORNIA

Source: BING Aerial imagery

Vector vs Raster

<u>Vector data are better for **discretely bounded** data</u>
e.g. political boundaries, fire hydrants, rivers, roads, etc.

<u>Raster data are better for **continuous** data</u>
e.g. temperature, elevation, rainfall, etc.

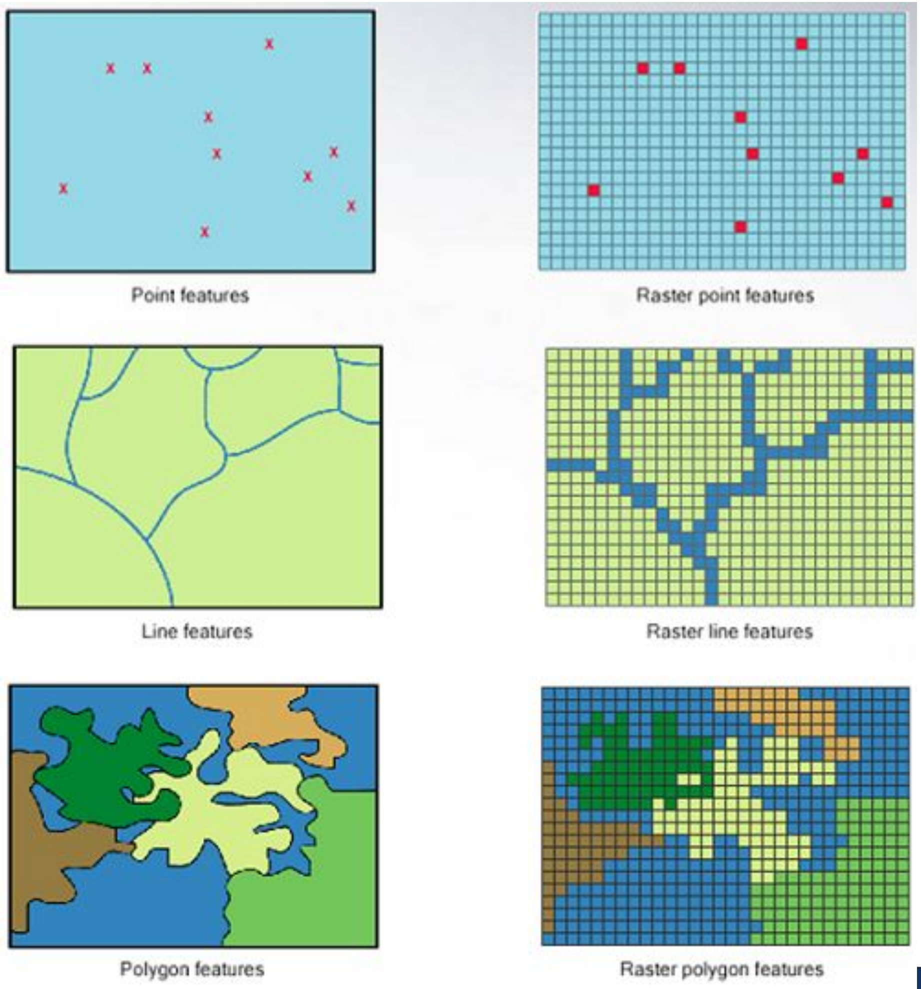


Image: http://gsp.humboldt.edu/

Some Common File Formats

Vector Data

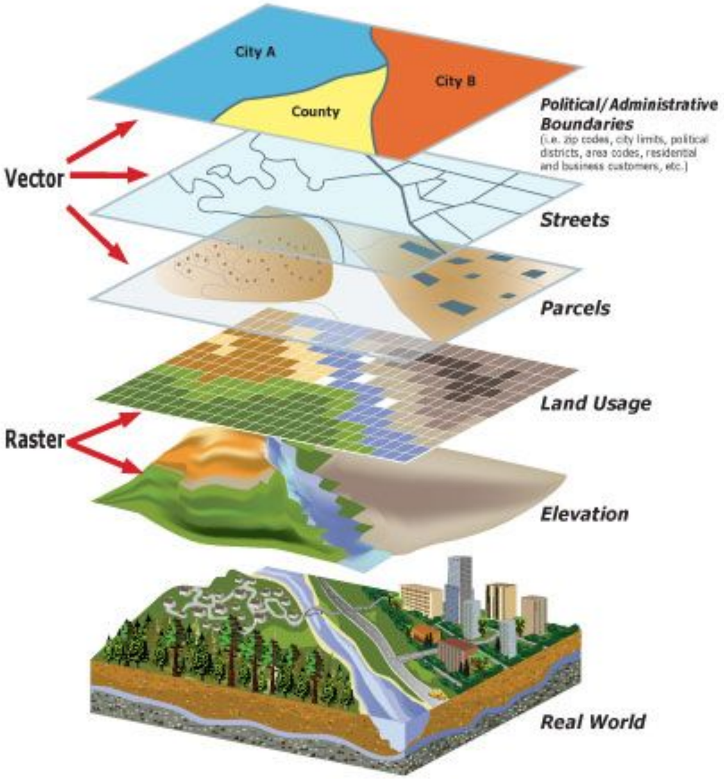
- Shapefile (.shp...)
- GeoJSON, JSON
- KML
- GeoPackage

Raster Data

- GeoTIFF
- netCDF
- DEM

Georeferenced

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



GIS Software

Geographic Information Systems (GIS)

GIS is software that integrates many types of data that includes spatial location and includes geometric operations for working with spatial data.

Spatial data vs. Geospatial data

Berkeley
UNIVERSITY OF CALIFORNIA

GIS stands for **Geographic Information System**

GIS is software, but also a set of technologies and methods

-- a platform for visualizing **geospatial data.**

-- and making different kinds of maps

-- transforming / editing geospatial data -- **geoprocessing***

-- and performing spatial analysis

**75% of what folks do with GIS*

Desktop GIS - ESRI’s ArcGIS

Powerful

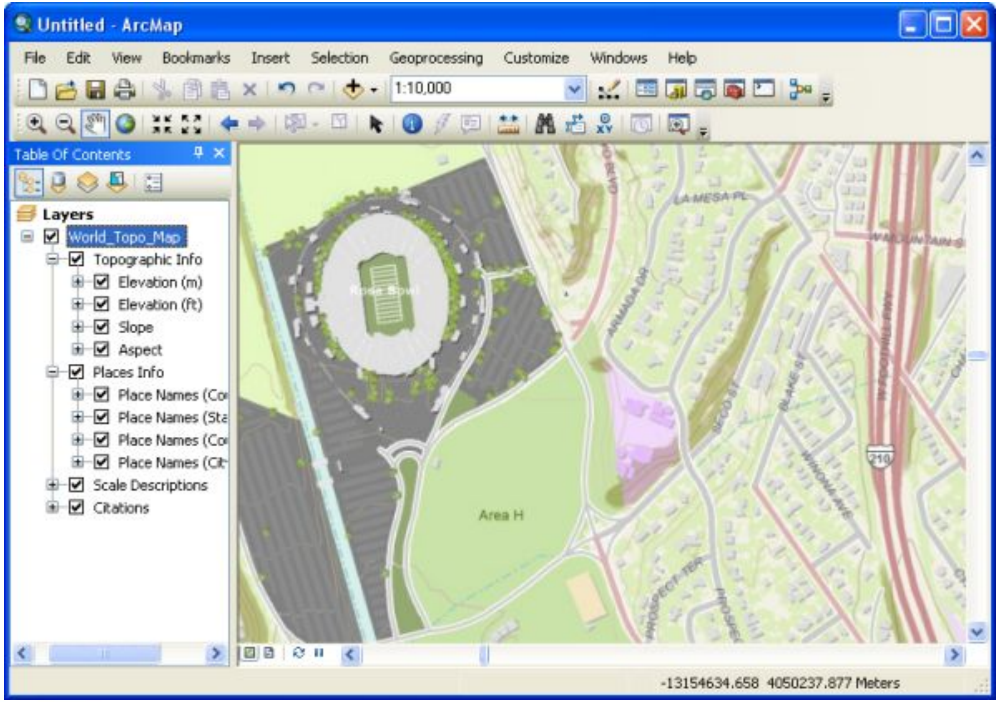
Full-featured

Robust

UCB site license

Slow learning curve

Only runs on MS Windows OS



Desktop GIS - QGIS

Free and Open Source

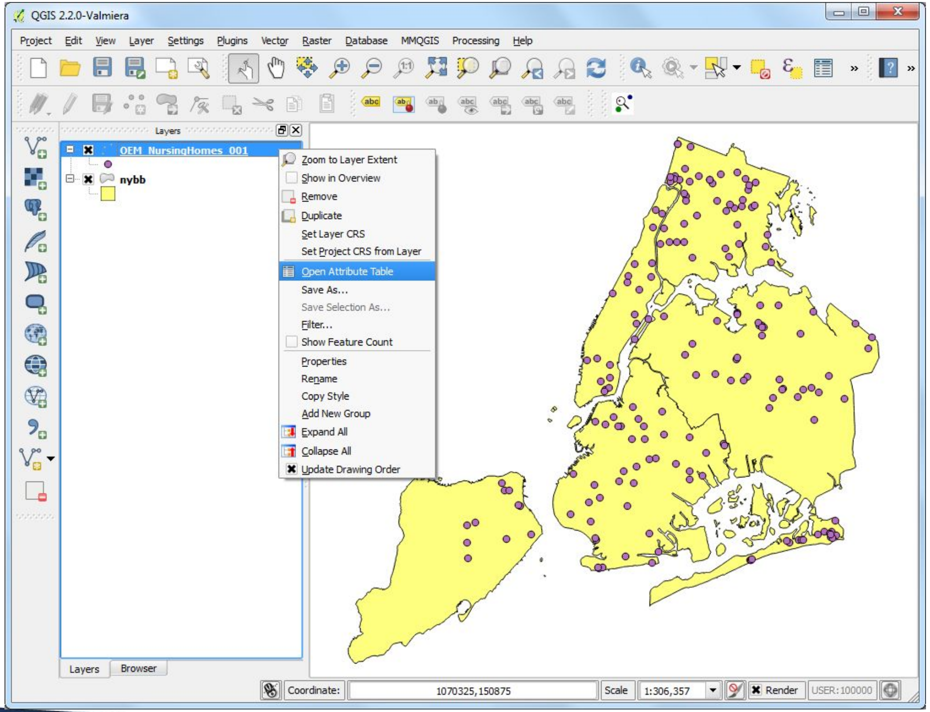
Runs on Macs & PCs

Pretty Stable

Lots of functionality

Slow learning curve

Not as good as ArcGIS



Programming languages with geospatial data support

Others

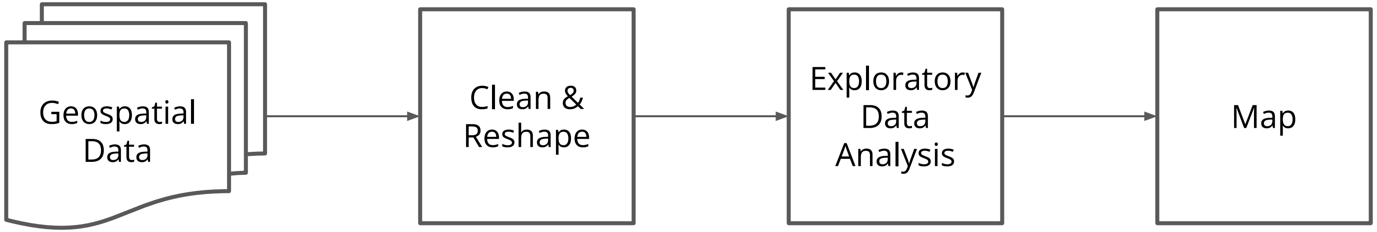
Spatial Databases - `PostgreSQL/PostGIS`

Web-based GIS - `ArcGIS Online`, `CARTO`

Software geospatial data support - `Tableau`

Geospatial Workflows

Generalized Geospatial Workflow



Motivating Questions

After introducing the basic building blocks of geospatial data, mapping & analysis, we'll wrap everything up by showing you how to answer questions like:

What is the total grocery-store sales volume of each census tract?

Any questions?