Overview of Geospatial Data

Patty Frontiera, Drew Hart, Hikari Murayama
Team Geo
D-Lab

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?

What is the D-Lab?

- 2013 UC-Berkeley initiative to provide services and support to conduct cutting-edge, data intensive research in social sciences.
- Offers consulting services on research design, data analysis, data management, and related techniques and technologies.
- Host of networks to connect with users of social science data in and off-campus.

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 :)

Our Workshop Today


START

About me

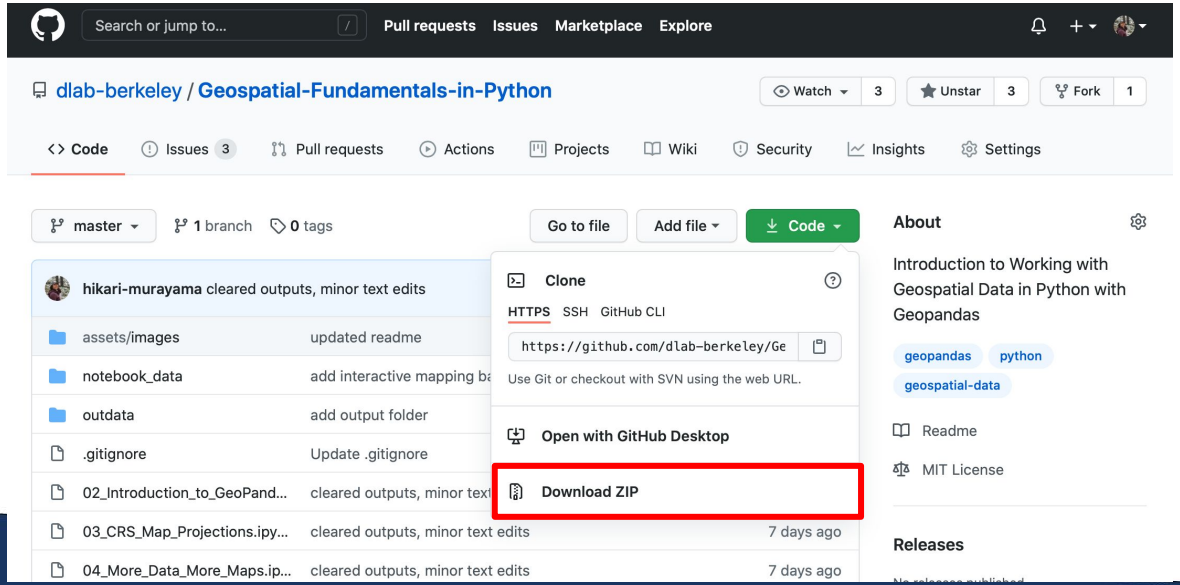
About you

Who are you?

Why are you here?

How to follow along

- Download workshop materials (I'll post the link in the chat)
- Click on the green "Code" button to download the files



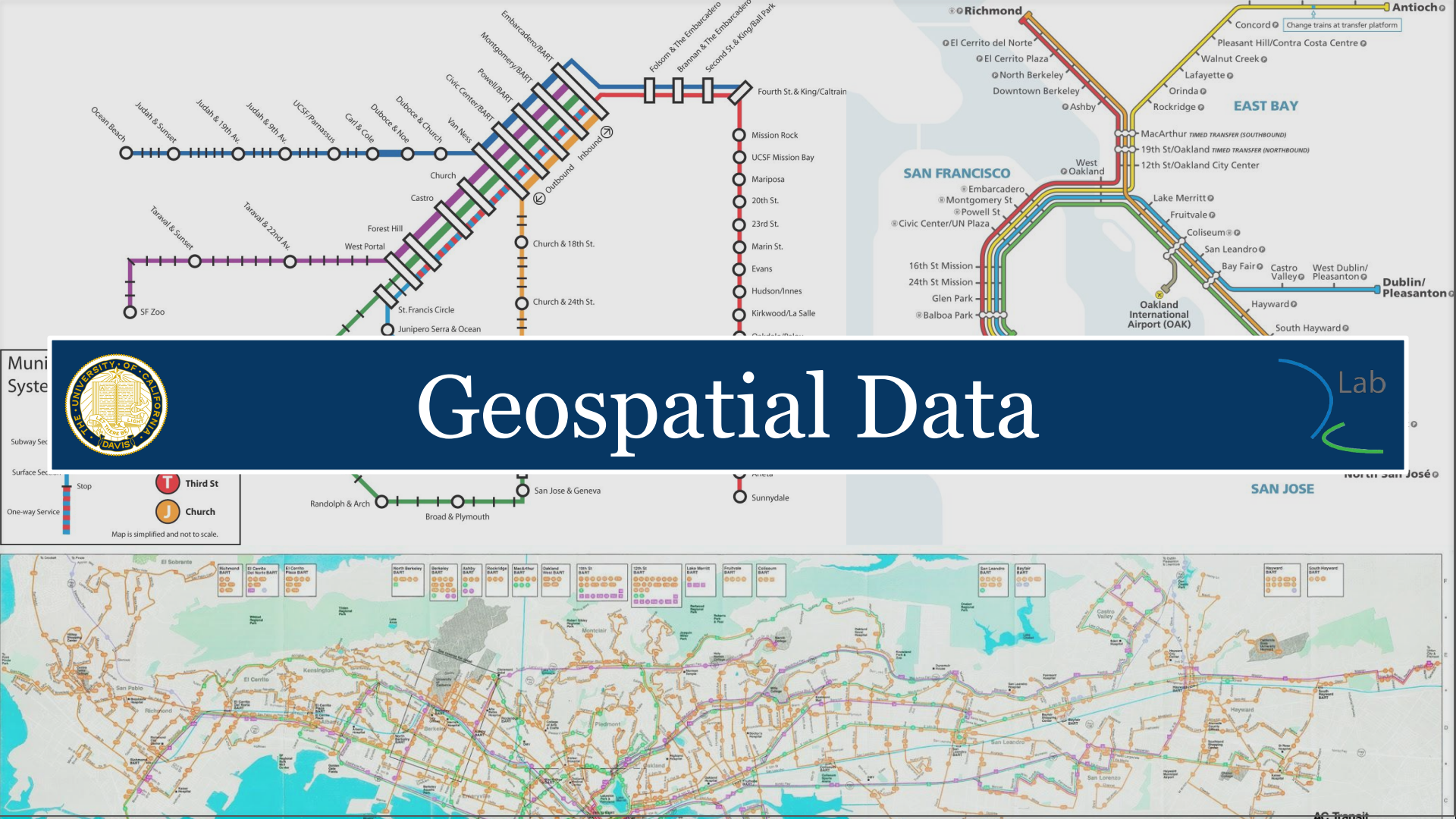
Let’s take a couple minutes to install everything we need

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

Workshop Goals

Introduction to working with geospatial data

- Geospatial data files and formats
- Loading geospatial data
- Coordinate reference systems
- Working with different types of vector data
- Combining spatial and aspatial data
- Mapping geospatial data
- Familiarity with geospatial workflows



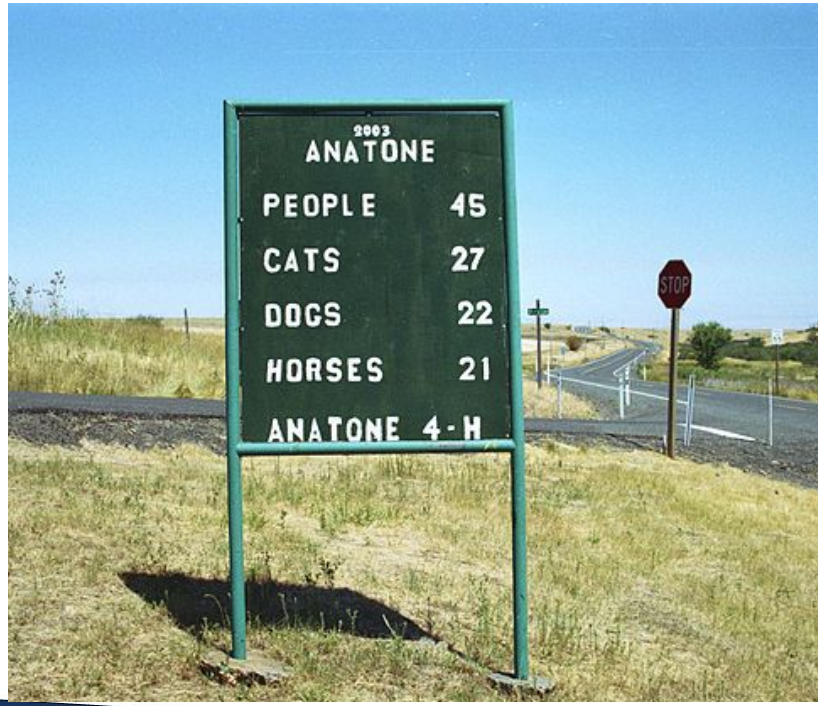
Geospatial Data

Lab

Geographic Information Systems (GIS)

GIS integrates many types of data, which includes a spatial location.

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

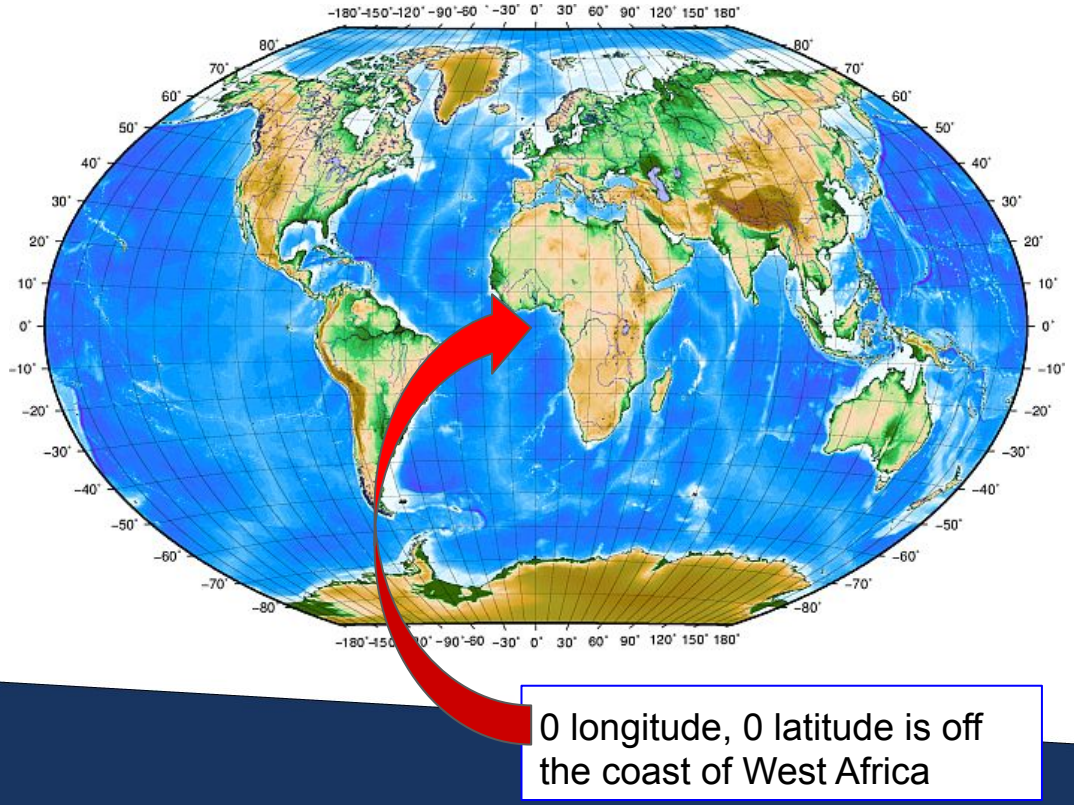
Encodes location geometrically with coordinates:

`Anatone, WA: 46.130479, -117.134167`

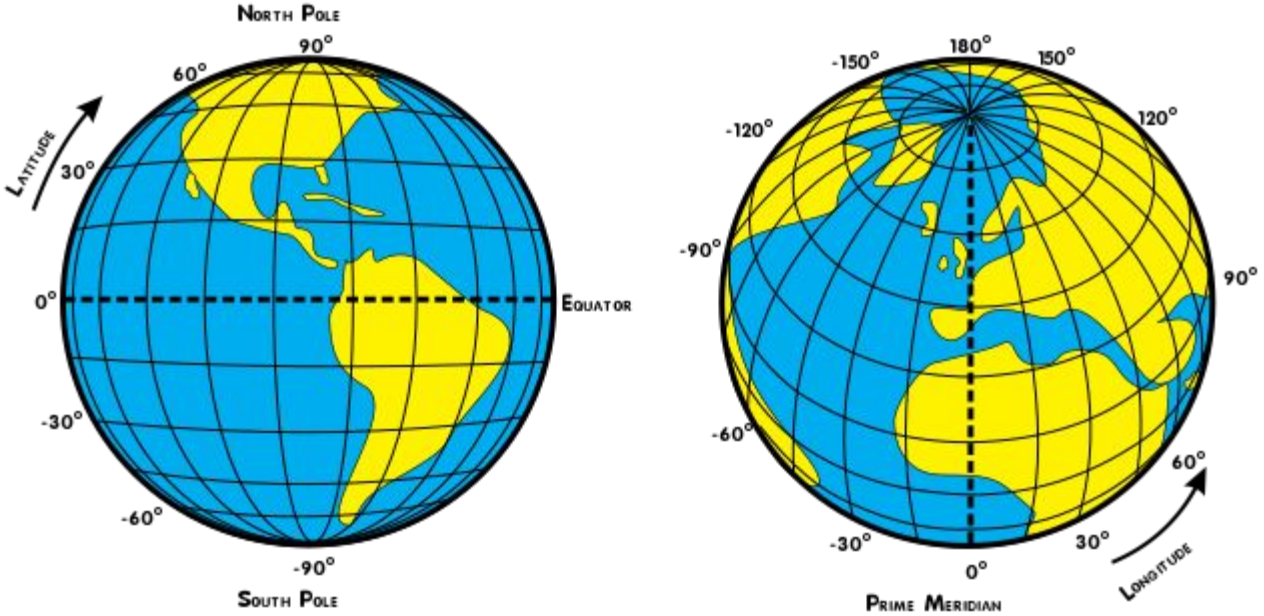
Coordinate Reference Systems

Coordinate Reference Systems (CRS)

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



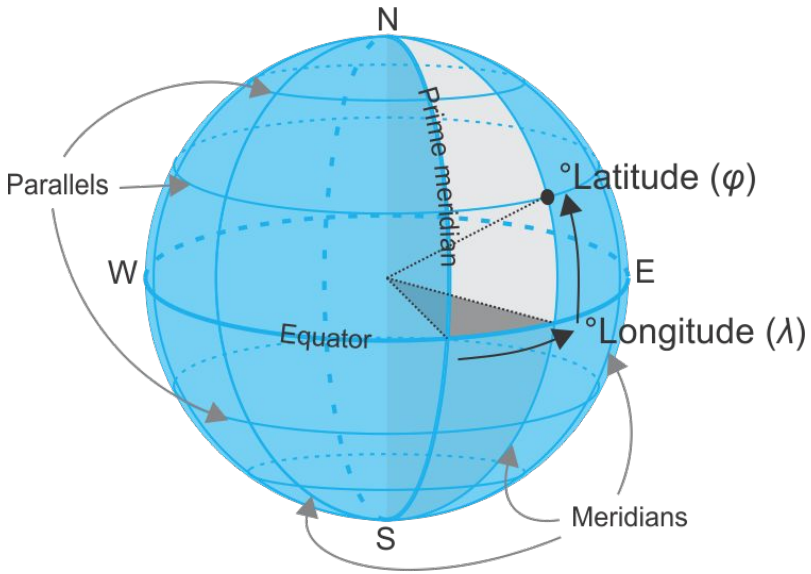
Coordinate Reference Systems (CRS)



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.

Let’s think about it as an orange peel

(trust me, it’ll make sense in a second)

Imagine having a globe as a lamp...

Depending on:

- How you tilt your lamp
- How your walls are oriented
- Other factors...

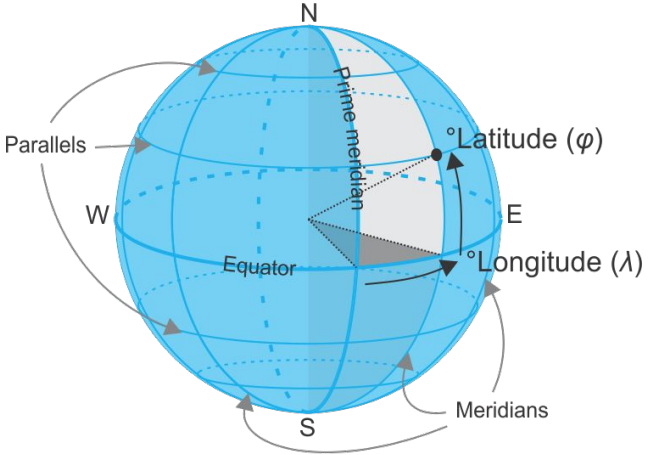
The map image you get on your wall will change

In short…

Since the earth isn’t flat, our maps are distorted in appearance to appear on our maps!

Two Types of Coordinate Reference Systems

Geographic CRS



Unit = Degree (DMS or DD)

Good for storing data

Projected CRS

Unit = Feet or Meters

Good for mapping & analysis

Geographic Coordinate Systems (GCS)

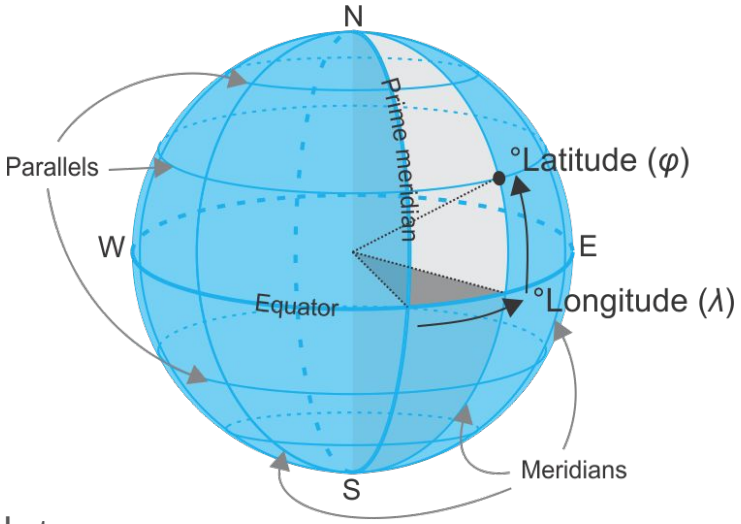
Widely used! Expressed as latitude & longitude

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**

Projected Coordinate Systems (PCS)

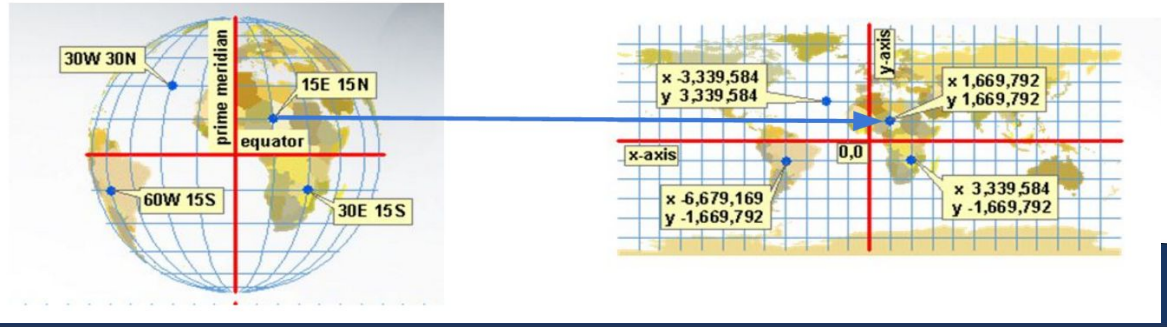
Map Projections transform geographic coordinates (lat/lon) to 2D coordinates (X/Y)

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

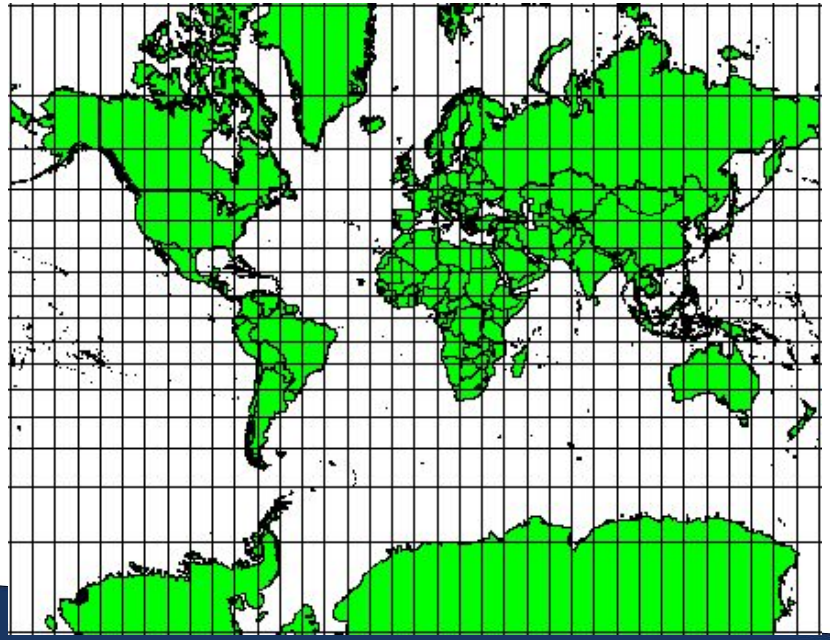
Specific map projections minimize distortion in one or more properties

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

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

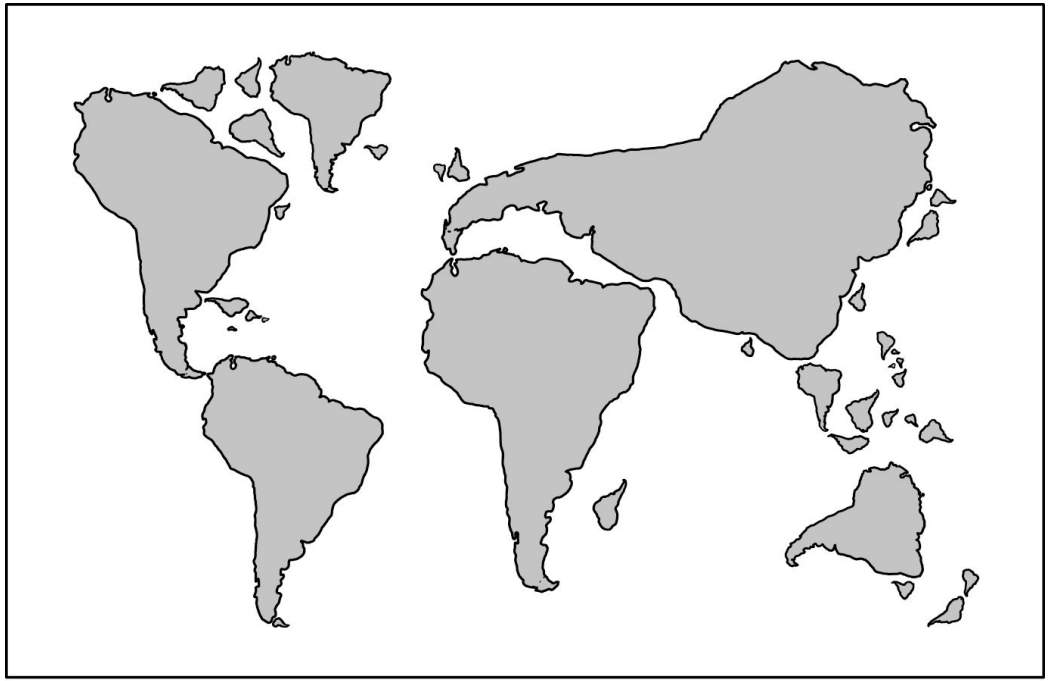


The **Mercator projection**, for example, is used where angular relationships are important (shape, direction), but areas are distorted, especially as you move away from the equator.



Source: QGIS Gentle introduction to GIS

Some projections are worse than others

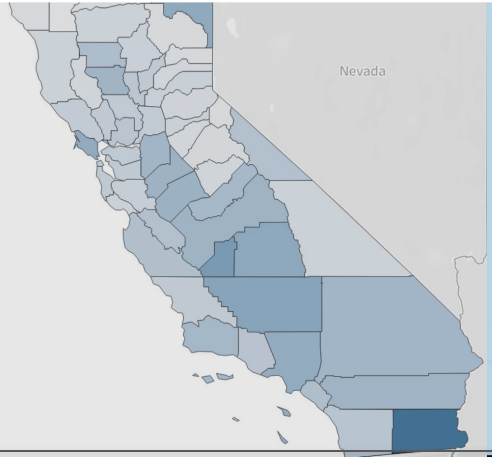


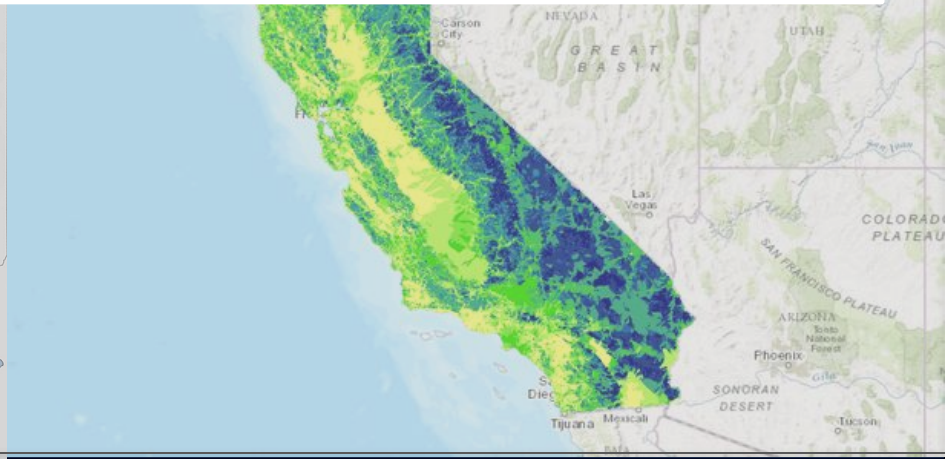
BAD MAP PROJECTION #358: OOPS, ALL SOUTH AMERICAS!

Types of Spatial Data

Lab

California coronavirus map



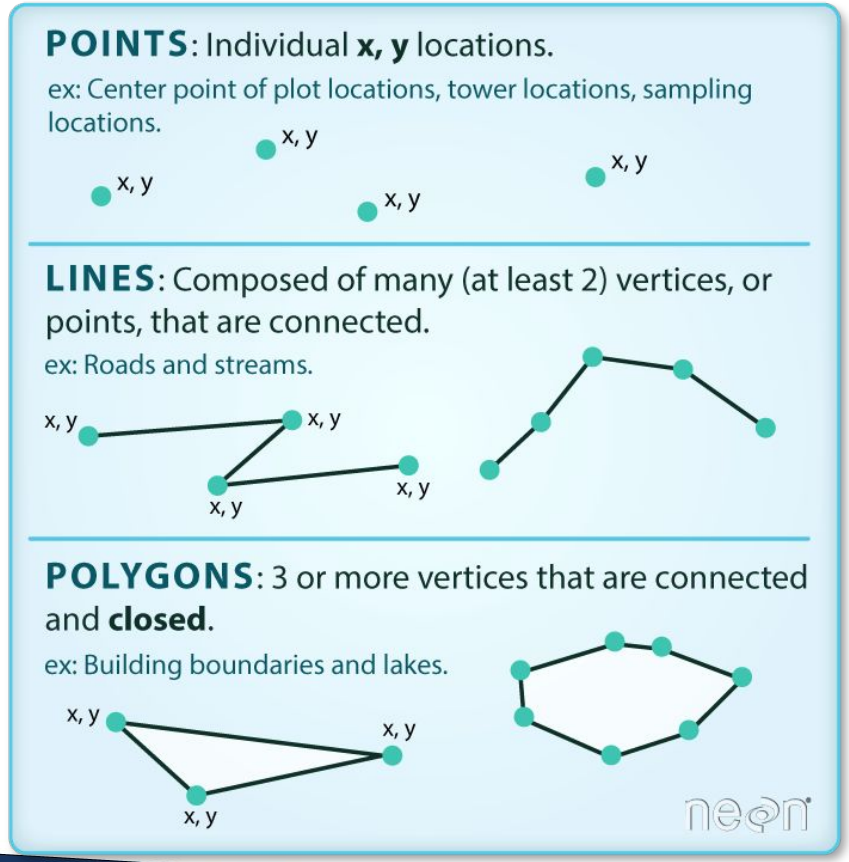


Types of Data

- Vector
- Raster

Vector Data

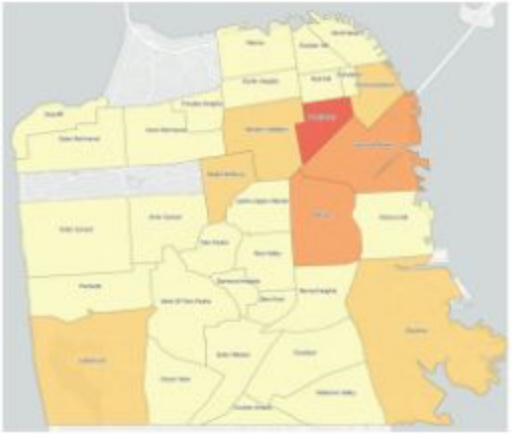
"Connect the dots"



Points, Lines, Polygons

Crime locations

City freeways



Neighborhoods

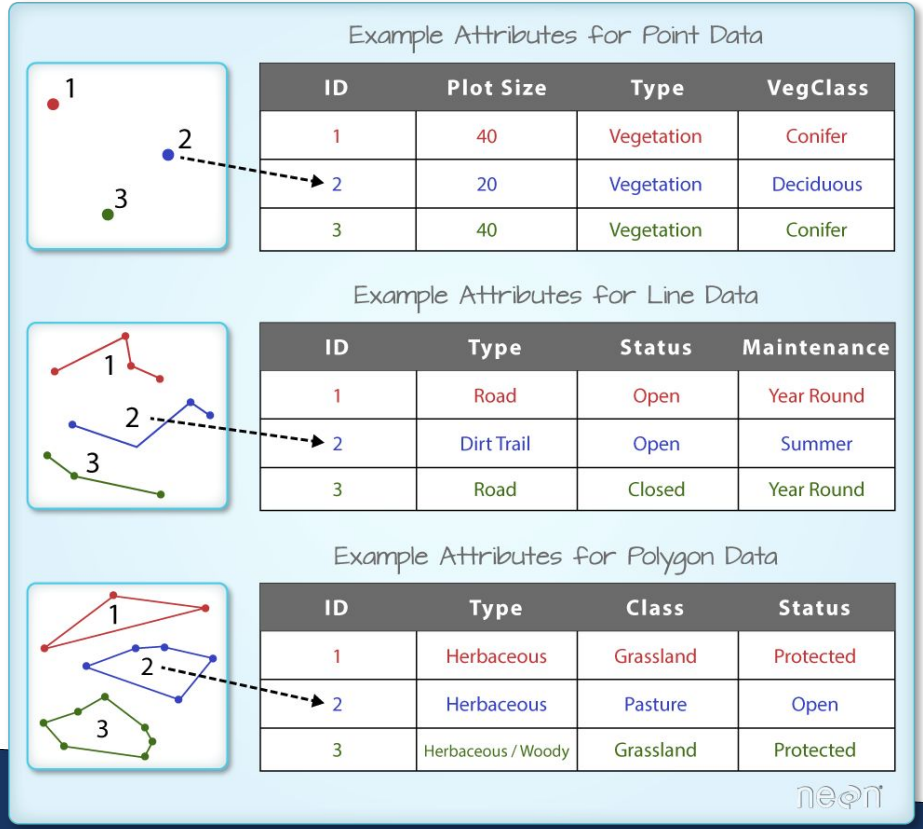
Vector Data with Attributes

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**

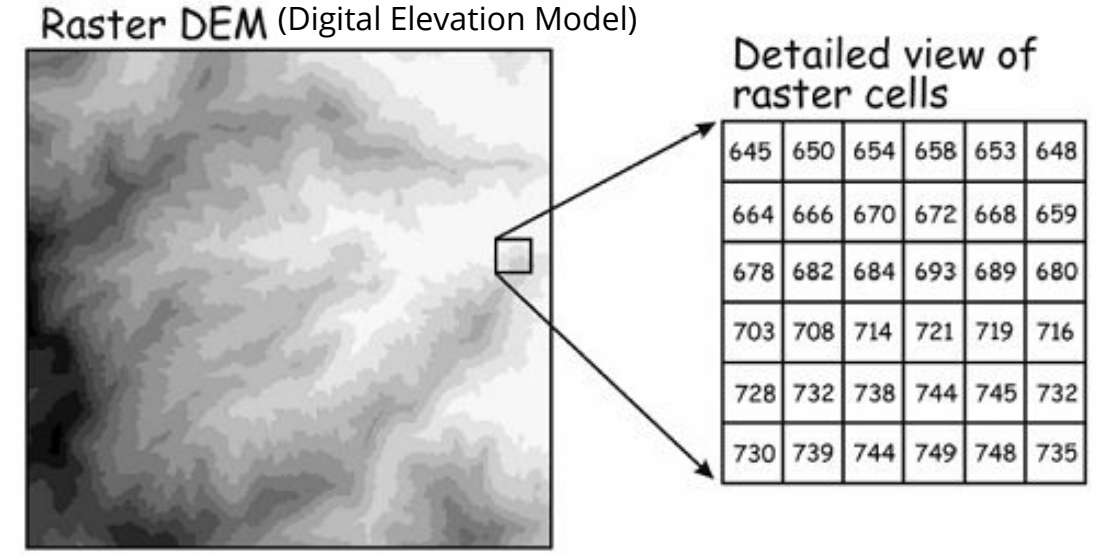


ID	Plot Size	Type	VegClass
1	40	Vegetation	Conifer
2	20	Vegetation	Deciduous
3	40	Vegetation	Conifer

ID	Type	Status	Maintenance
1	Road	Open	Year Round
2	Dirt Trail	Open	Summer
3	Road	Closed	Year Round

ID	Type	Class	Status
1	Herbaceous	Grassland	Protected
2	Herbaceous	Pasture	Open
3	Herbaceous / Woody	Grassland	Protected

Raster Data - regular grids



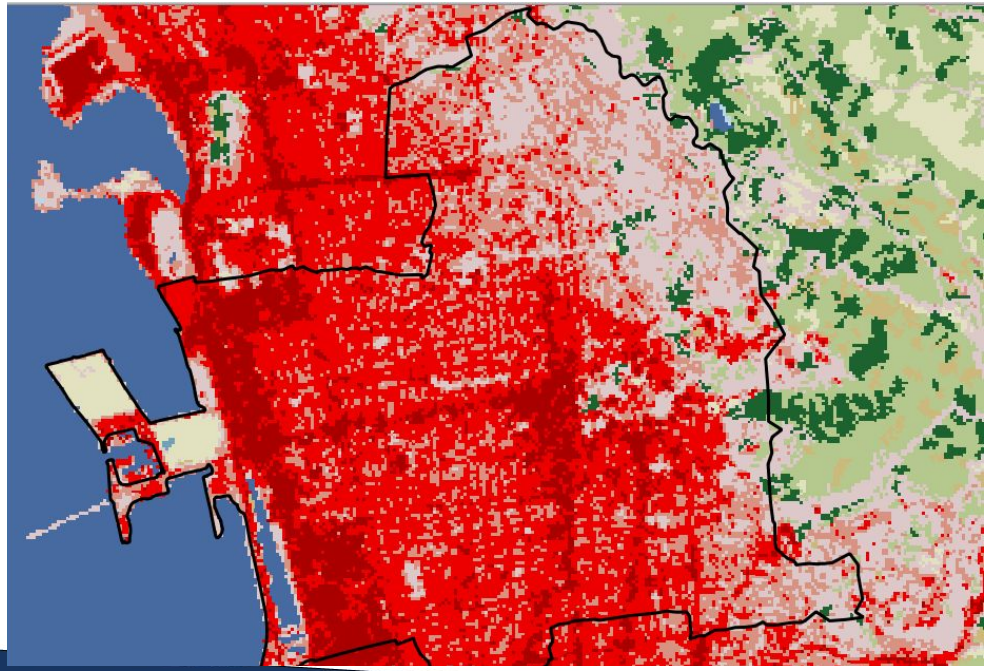
A location is represented by a grid cell

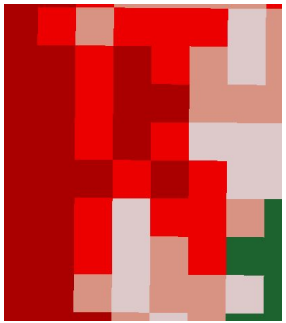
Cells have regular size, eg 30x30m

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.

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.

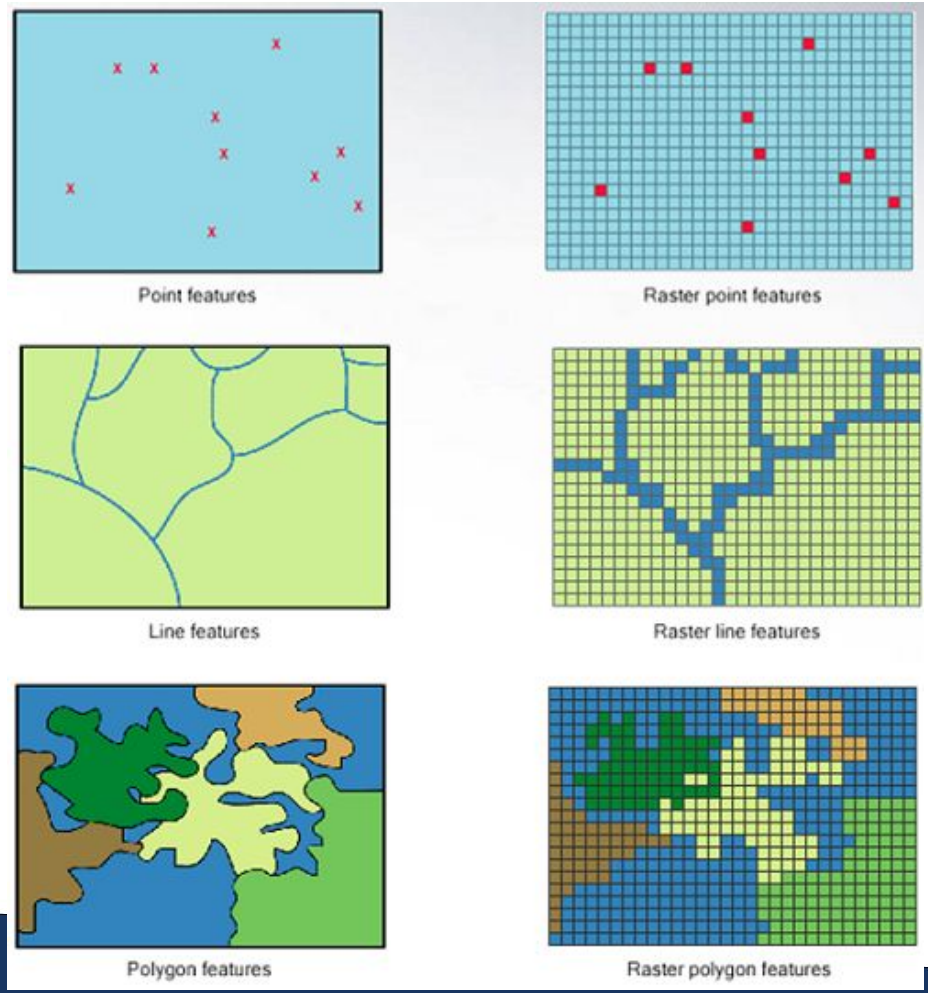


Image: http://gsp.humboldt.edu/

Some Common File Formats

Vector Data

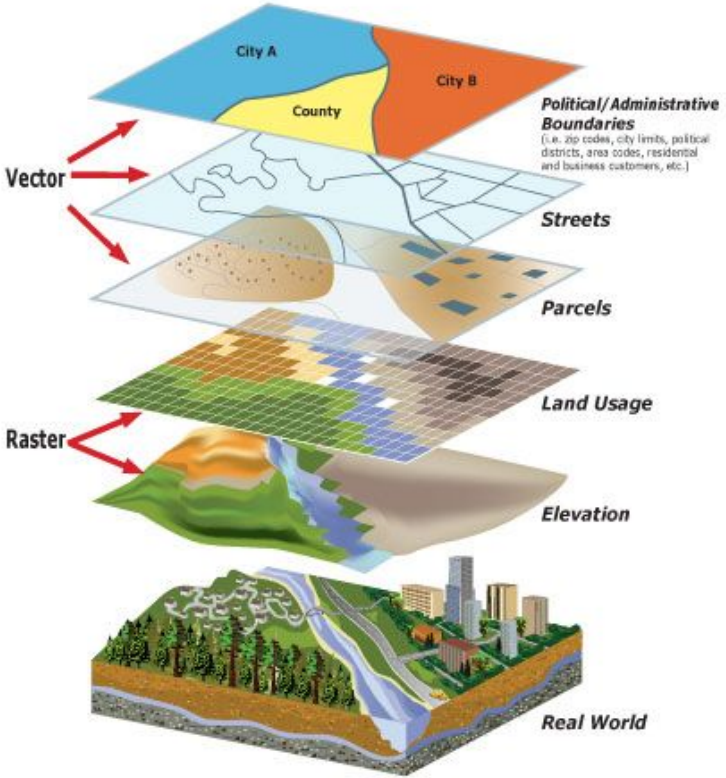
- Shapefile (.shp...)
- GeoJSON, JSON
- KML
- GeoPackage

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

Lab

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

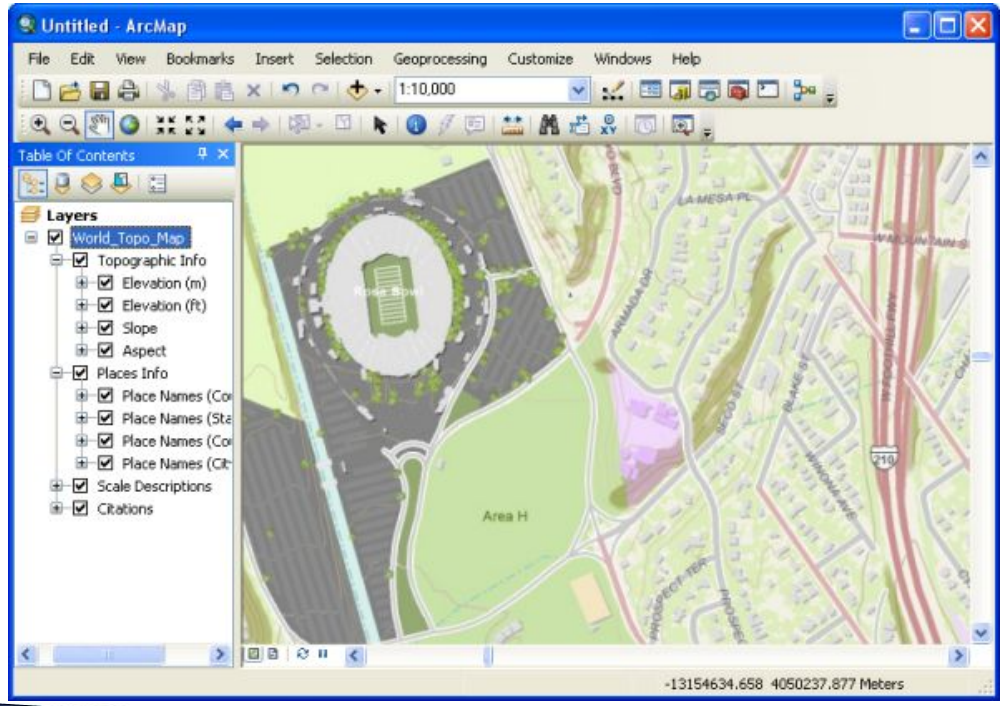
Full-featured

Robust

UCB site license

Slow learning curve

Only runs on MS Windows OS



Desktop GIS - QGIS

Free and Open Source

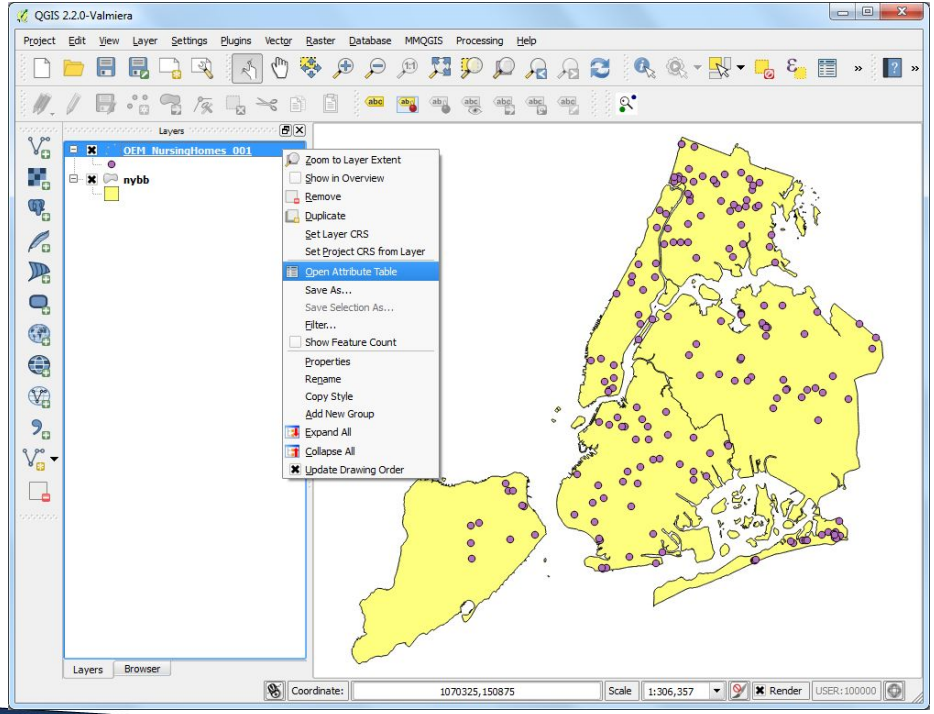
Runs on Macs & PCs

Pretty Stable

Lots of functionality

Slow learning curve

Not as good as ArcGIS



Programming languages with geospatial data support

Others

Spatial Databases - `PostgreSQL/PostGIS`

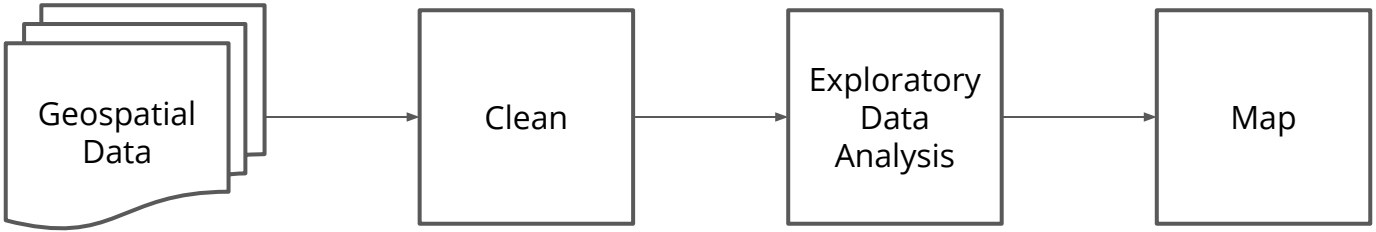
Web-based GIS - `ArcGIS Online`, `CARTO`

Software geospatial data support - `Tableau`

Geospatial Workflows

Lab

Generalized Geospatial Workflow



Motivating Questions

At the end we'll wrap everything up and answer questions like:

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

Any questions?