CP 255
Introduction to Geospatial data

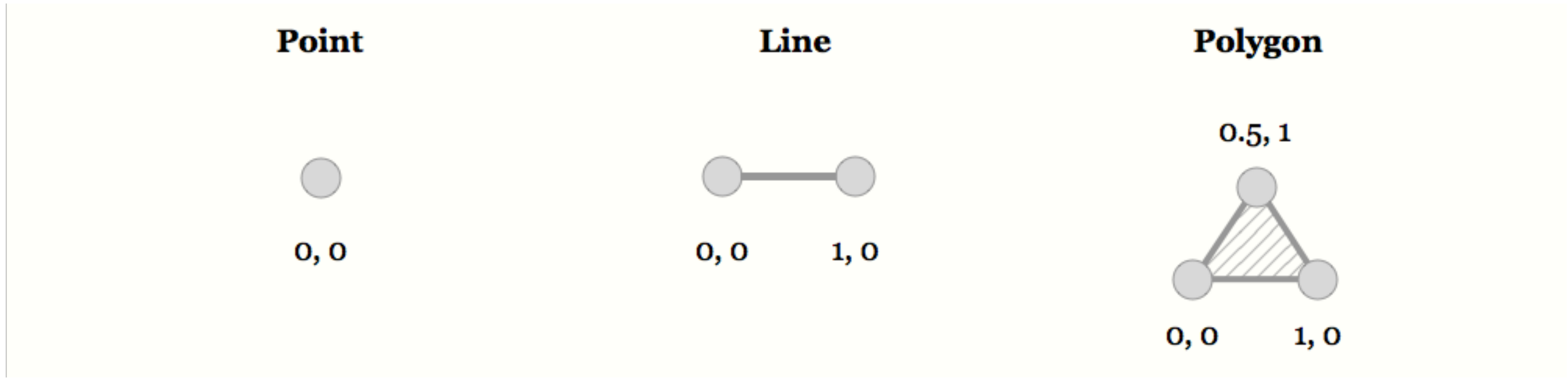
Acknowledgments to Samuel Maurer for the initial version of this presentation

Geospatial data analysis and visualization

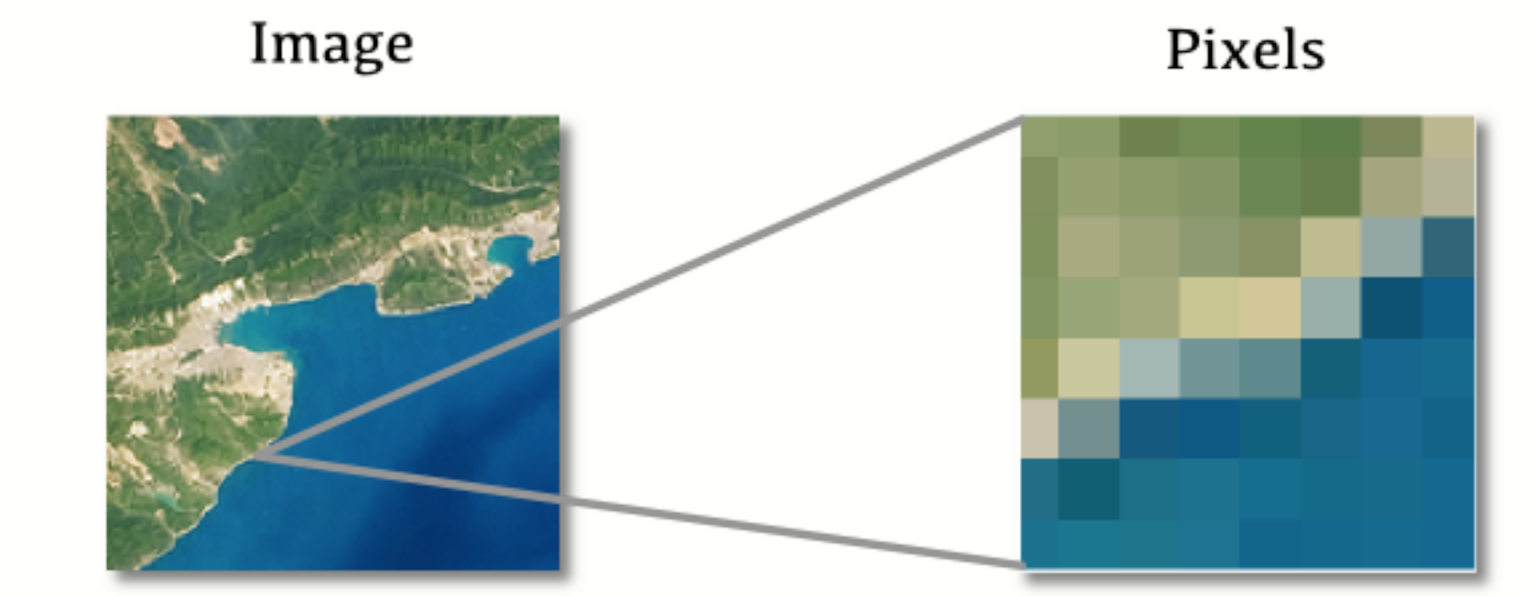
- Geospatial data is data about locations in physical space, specifying both the locations themselves and other associated attributes
- GIS = Geographic/Geospatial Information Systems/Science
    - Standard structures and methods for analyzing and visualizing geospatial data
    - Often done with desktop software packages: ESRI ArcGIS; open source QGIS
- Increasingly, there's a broader world of geospatial data tools that do similar things but don't necessarily frame themselves as being "GIS"
    - Online map platforms, Python libraries, web cartography tools, etc.

Types of location data

- **Vector**: typically from administrative or GPS-type sources



- **Raster**: typically from photographic-type sources



Problem: Earth's surface is not a coordinate plane

- Coordinates like latitude and longitude are different from purely mathematical coordinates like in a scatter plot
- Relational metrics like *distance* or *area* can only be calculated after you make particular assumptions about how the coordinates match up with the Earth's surface

—> One important piece is the *projection* of the data
—> Another is the *datum*

Map projections

- Map projections can optimize various properties:
  - “Equal area” projections (e.g. Albers)
  - “Conformal” projections maintain shapes
  - “Equidistant” projections maintain distance from a point or along meridians and parallels
  - Others are optimized for local navigation, data collection, visual appearance, etc.

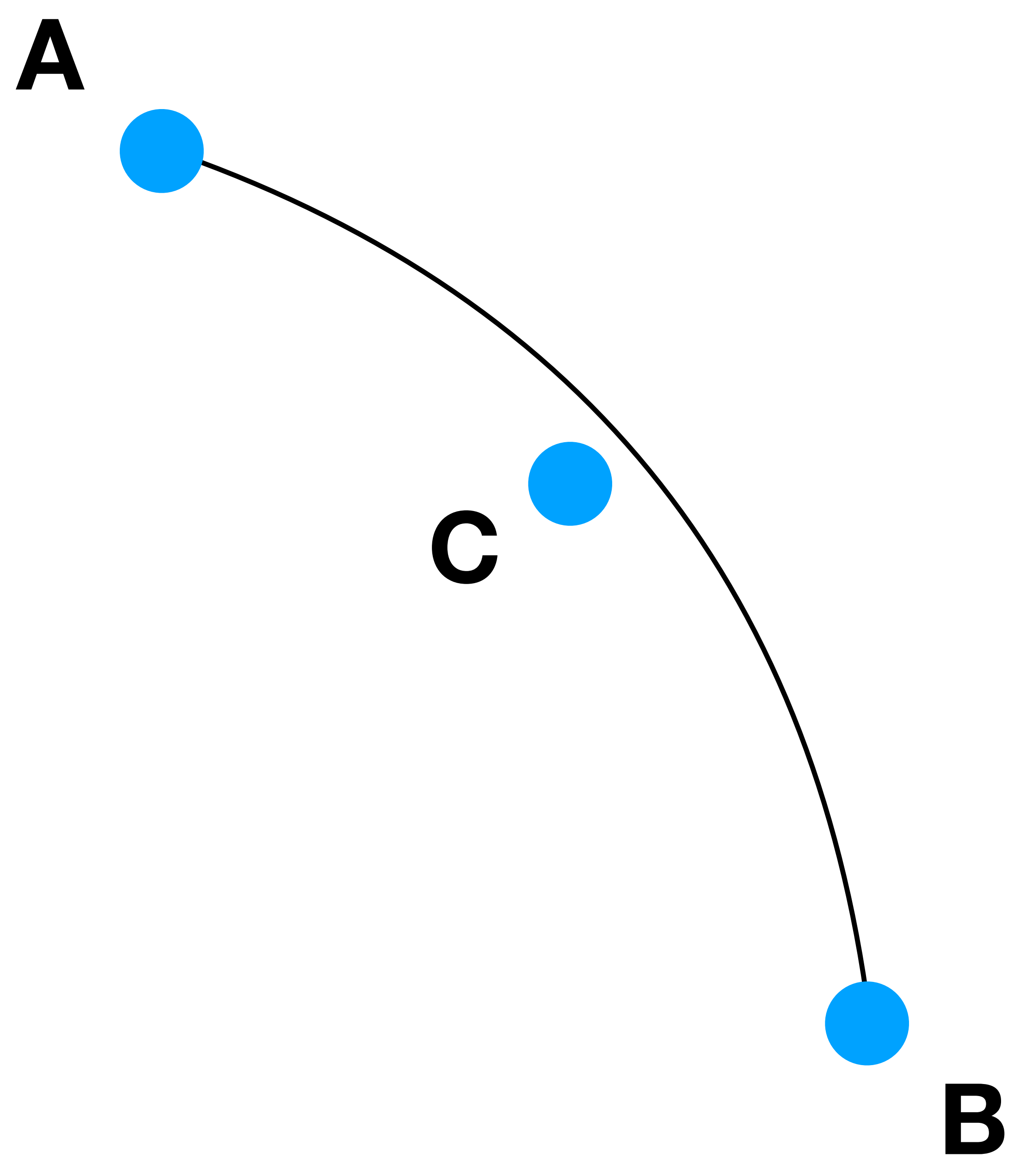
- Some projections have their own coordinate systems
  - e.g., Universal Transverse Mercator projections use local (x, y) distances in meters from an origin point

Albers

Transverse Mercator

Pseudo-Mercator

https://en.wikipedia.org/wiki/Map_projection



Even if you're not making a map, if the data is in a projected coordinate system it can affect your analysis.

Is point C to the left or right of a "straight line" between A and B?

It might depend on the projection, especially if the spatial extent of the data is large.

Datums

- But it gets more complicated! Earth is not completely spherical, nor even an “oblate spheroid”

- A *datum* refers to a detailed approximation of how the earth is shaped (ignoring terrain), resulting in a particular match-up of coordinates to physical spots on the ground

- All coordinates — including latitude and longitude — implicitly or explicitly refer to a particular datum standard

What you need to know

- Geospatial datasets will have an associated **Coordinate Reference System (CRS)**
    - This always includes a datum, and *may* include a projection if the CRS uses projected coordinates like meters instead of lat/lon
    - If a dataset is missing a CRS, you'll have to guess what it's supposed to be in order to work with the data!

Things you can do with geospatial data

- Spatial joins
  - Merge datasets based on things like whether the geometries intersect
- Distance and network calculations
  - How far are things from each other?
- Topological calculations
  - Which shapes touch each other?
- Geoprocessing
  - Buffers, unions and intersections of shapes, etc.