Spatial Data Analysis in R

Deck 1: Introductions

Eco 697DR – University of Massachusetts, Amherst – Spring 2022
Michael France Nelson

Welcome to Spatial Data Analysis in R!

In This Slide Deck

- Introductions
- Course Overview
- Review of Key Concepts

Introductions: About Me

A bit about me…

- Lecturer, Department of Environmental Conservation at UMass
- Experience teaching and using spatial data and analyses in R and ArcGIS
- Background in plant biology and invasion ecology
- I love computers and programming, but I miss working in the field!

Introductions: About You

- What are your research projects and interests?
- What are your learning goals for this course?
- Why are you interested in spatial analysis?

About The Course

Course Structure Overview	**We will closely follow the text's Part I: Chapters 1 - 6**

• Part 1: Concepts and Techniques
  • Readings
  • Lab exercises
• Part 2: Main Projects
  • Dataset
  • Proposal
  • Methods/results
  • Report
	• Ch. 1: Introduction
	• Ch. 2: Scale, distance concepts
	• Ch. 3: Land-cover concepts
	• Ch. 4: Spatial statistics: point patterns
	• Ch. 5-6: Spatial dependence
	• Selected material from later chapters

My Expectations

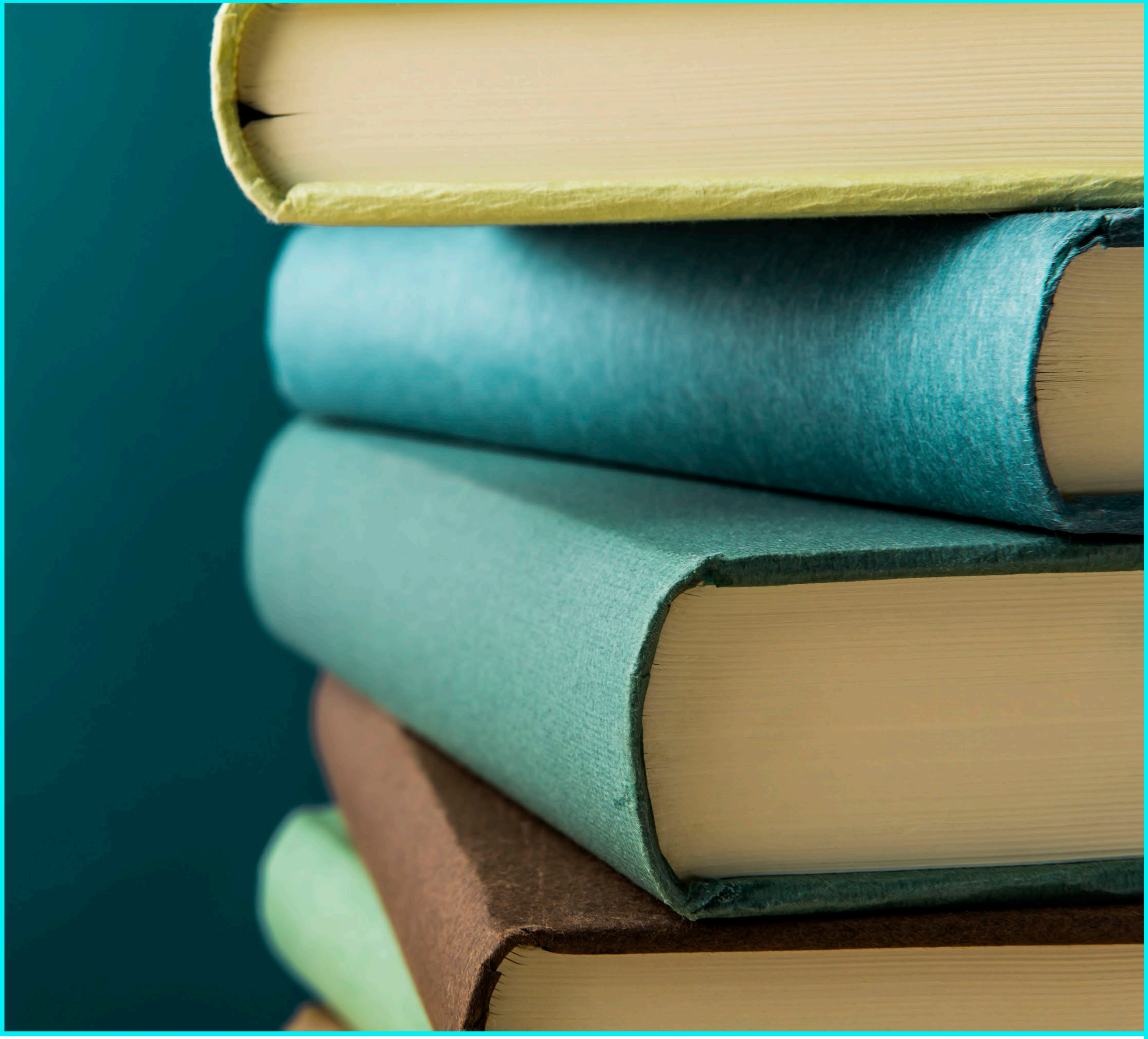
How to succeed in this course

- Active contributions to class!
- Engagement in peer learning and teaching
- Course feedback to me
- Critical thinking
- I don’t expect perfection. I’m more interested in your thoughtful learning.
- Contribute to the course Wiki
  - We’ll go over how to use the Wiki in a future lab session

About The Text

Spatial Ecology and Conservation Modeling, Robert Fletcher and Marie-Josée Fortin

- This is a great text, but it’s also a first edition
- There are some mismatches in the book code, online supplemental material, and online data
  - Updated code from Robert Fletcher is available in Moodle



About Spatial in R: The sp to sf Transition

A note about spatial data and packages in R

- sp and raster are the core packages for most spatial work in R… at least historically
- R evolves:
  - sf (simple features), terra, and stars (rasters) are the new way
  - New packages play nicely with ggplot and tidyverse
- We will use the sp/raster world in this course
  - The text uses these packages and it's good to know how they work
- Be aware of the transition to the sf world in the coming years.

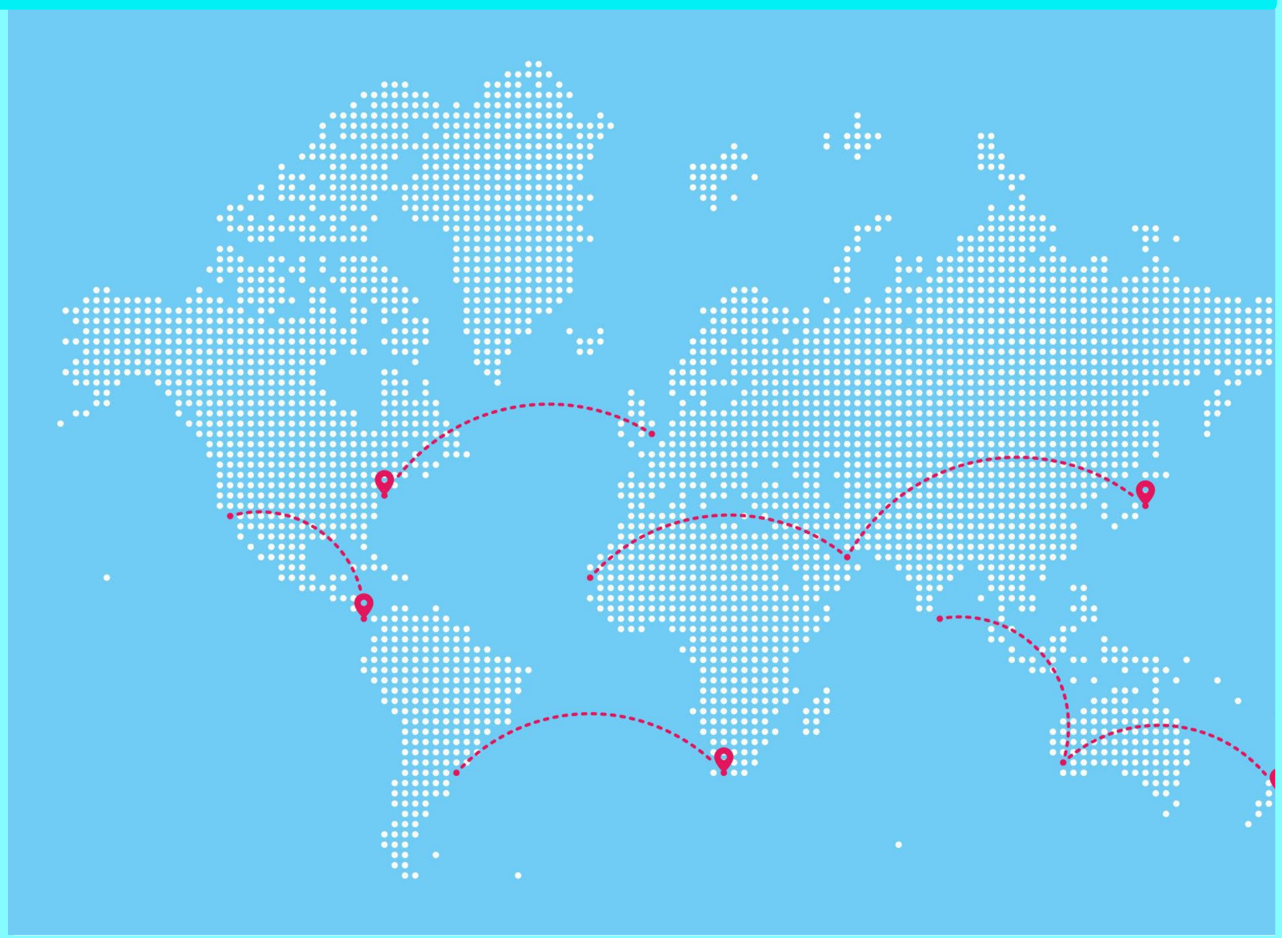
About The Labs

- Labs are meant to reinforce and supplement the concepts in the text and lectures.
- Labs are loosely based on the book code, with added material to reinforce concepts that I think are important or useful.
- You will need to find some of your own data!
  - Be sure to share any good finds in the Wiki.
- Consider due dates tentative: we'll see how things go and assess as needed.

Starting Points

We need to remember some foundational concepts from

- GIS and Geography
  - Spatial data types
  - Coordinates systems and projections
- Statistics
  - Inferential and descriptive statistics
  - Populations and samples
  - Parameters and statistics
  - Frequentist paradigm
  - Probability theory
  - Hypothesis testing
  - Uncertainty



Course Software and Computer Concepts Review

Course Software

R and Rstudio

GitHub account, GitHub desktop application

- You may set up git outside of github desktop, but this can be difficult, especially on Mac computers.
- GitHub desktop is an easy work-around.

Word processing and spreadsheet software

Key computer concepts

- Paths: absolute and relative
- How to browse files and folders on your OS
- File naming conventions
- R-coding best practices
  - Use code comments
  - Be concise
  - Aim for readability and reproducibility

Review of Statistical Concepts

Sample Statistics

Measures of Center	Measures of Spread

• Mean
  • Great for symmetric data
  • Sensitive to outliers
• Median
  • Better for skewed data
  • Robust to outliers
• Mode
  • Helpful for categorical data
  • Can be problematic if there are ties
	• Min and max
	• Range, size of the range
	• Standard deviation
	  • This is most useful for well-behaved data, sensitive to outliers
	• Interquartile range (IQR)
	  • More robust to outliers

Regression Models

Basic Concepts	Advanced Concepts
• Response variable and predictors (covariates). • Intercept and slope coefficients • Model residuals • Basic model diagnostics • Model comparison • Model residuals • This is where you find a lot of spatial autocorrelation!!!	• Fixed and random effects • Nested effects • GLMs, especially logistic regression • Mixed models

Probability Theory and Distributions

Important Concepts

- Parametric distributions
  - Normal and Poisson
  - Distribution functions
- Independent events
- Law of total probability
- Variance and covariance
- Correlation

Don’t Panic!

We’re in this together

- This course is not meant to be too theoretical.
- I don’t expect you to be an expert on all the concepts on the last several slides.
- We’ll review all relevant concepts as needed, and as requested.

Geography/GIS Concepts Review

Key Concepts

- Data types: raster and vector
- Coordinates in raster and vector data
- Vector data types: polygons, lines, points
- Scale: grain and extent
- Projections and coordinate references
- Distance measurements: e.g. Manhattan, Euclidean

Tobler's First Law of Geography

- "Everything is related to everything else, but near things are more related than distant things."
- A key theme in this course is **spatial autocorrelation**. Tobler's law encapsulates the idea of spatial autocorrelation. We'll spend a lot of time learning how to understand, quantify, and deal with spatial autocorrelation.

Spatial Autocorrelation

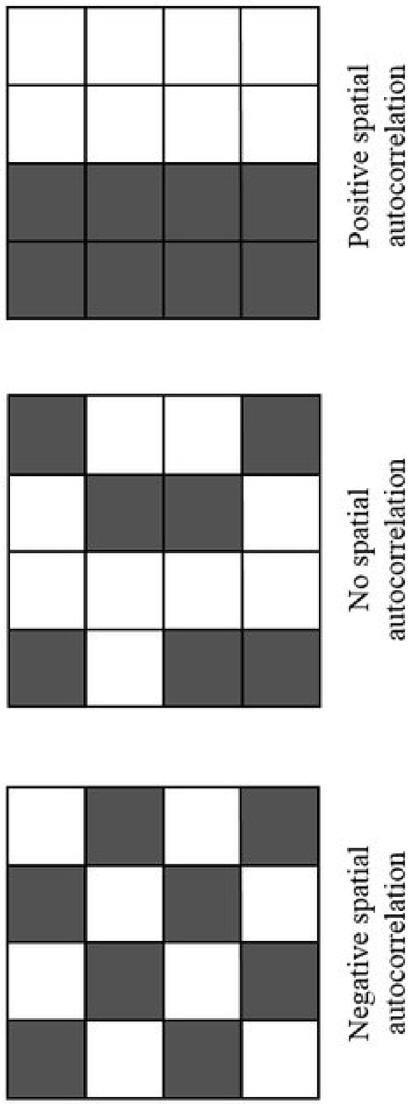


Figure 3 from (Amaya-Gómez et al. 2020)

- Informal definition of spatial autocorrelation: nearby things are more similar than expected by chance alone.
- Spatial autocorrelation of a variable can be the result of (for example)
  - Biotic interactions
  - Gradients
  - Similar abiotic conditions like substrate type, soil chemistry, etc.
  - Genetic similarity
- Spatial autocorrelation is the rule, rather than the exception and we'll spend lots of time on it later!

Types of Spatial Data

Spatial Data Paradigms

Raster	Vector
• Data are stored in a rectangular grid. • Cells • Coordinates are implicit	• Data are stored as points. • Each point has a set of coordinates • Points, lines, polygons • Coordinates are explicit

Vector Data

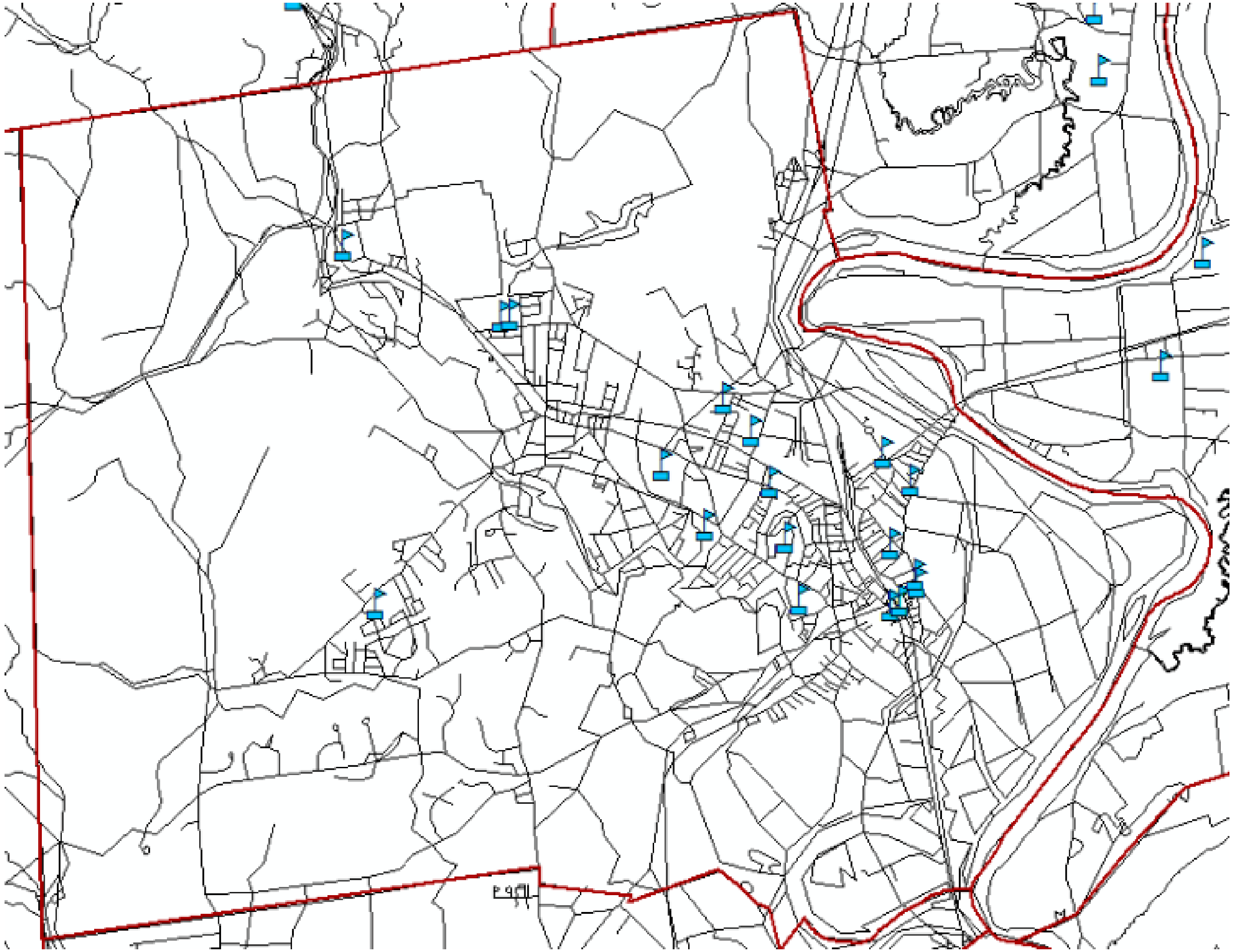
Vector data represents features as points, lines, and polygons and is best applied to discrete objects with defined shapes and boundaries.

Features have a precise shape and position, attributes and metadata, and useful behavior.

Source: Zeiler, M. 1999. *Modeling Our World: The ESRI® Guide to Geodatabase Design*. Redlands, CA: ESRI Press. 199 pp.

Vector (Feature) Data

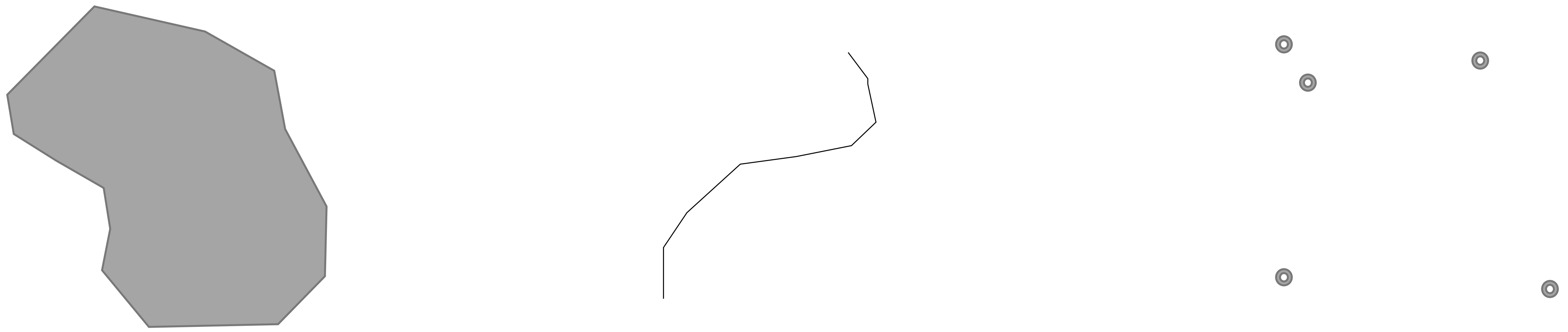
- Vectors can represent:
  - Points
  - Lines
  - Polygons
- All vector data are built from points
  - Each point has a coordinate



Vector Data

- The vector data paradigm associates features with attributes.
- Feature: stores the spatial information.
  - Each vertex in a feature has explicit x- and y-coordinates. This has important consequences!
- Attribute table: stores the associated data values
- Key points:
  - Features and attribute tables are different data structures.
  - They’re often stored in separate files.
  - The vector data paradigm associates a particular feature with a particular row in the attribute table.
- Raster data is a fundamentally different way of encoding spatial information.

How should you represent a feature?



Raster Data

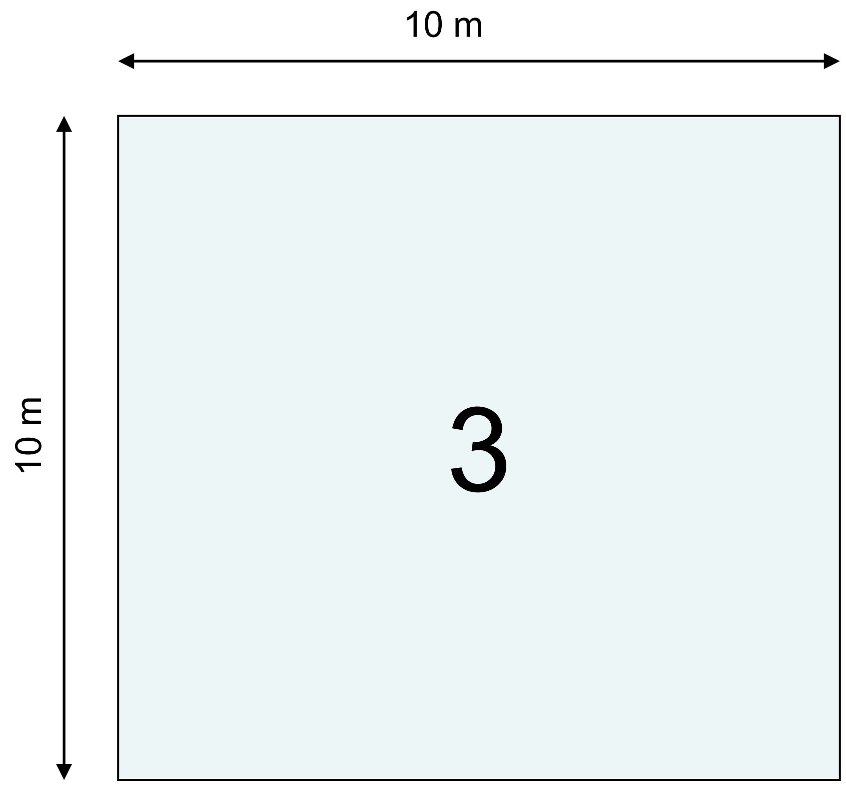
Surface of values composed of square **pixels (or cells)** each with a specific **value**

2	2	5	3	3
2	2	5	3	3
2	5	5	3	3
4	5	4	4	3
4	4	5	4	6
1	1	4	5	6
1	1	1	1	5

- cnty_grid_v2
- cnty_grid_v3
- county_grid
- dens2
- dist_roads

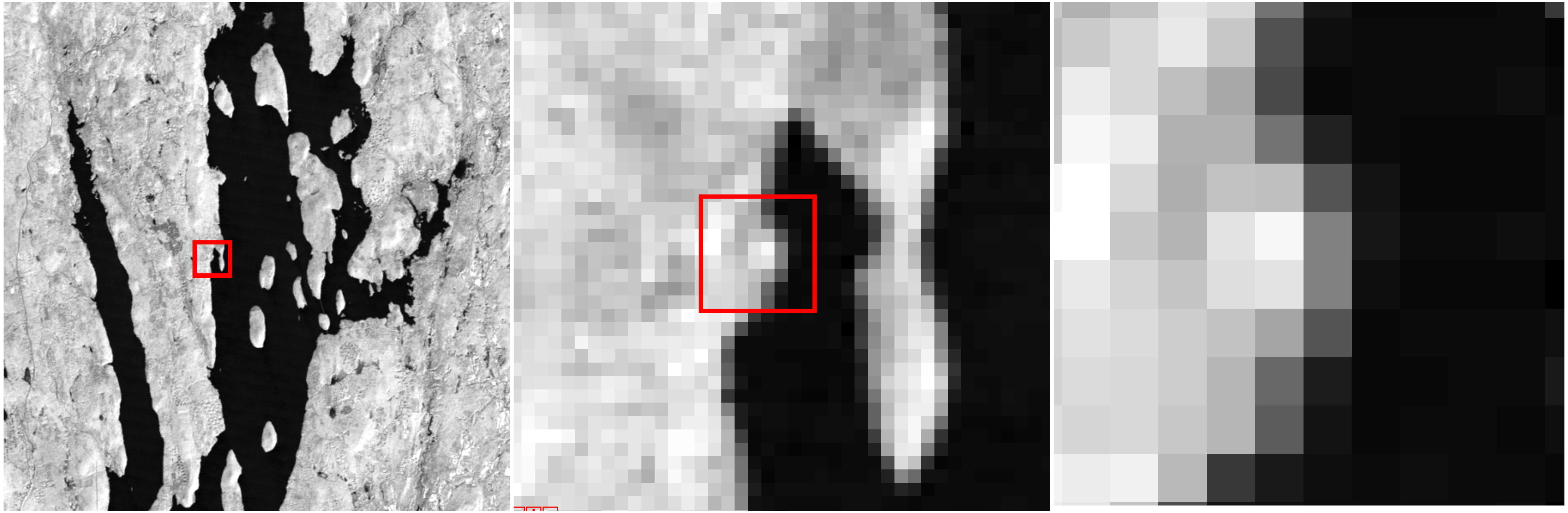
A raster layer is composed of <u>pixels</u>

- Surface of values composed of [usually] **square** pixels each with a specific value
- Pixels have a specific length/width size (ex. 10 m).
- Pixels may only hold **one *value.***
  - Compare to vector **attributes**



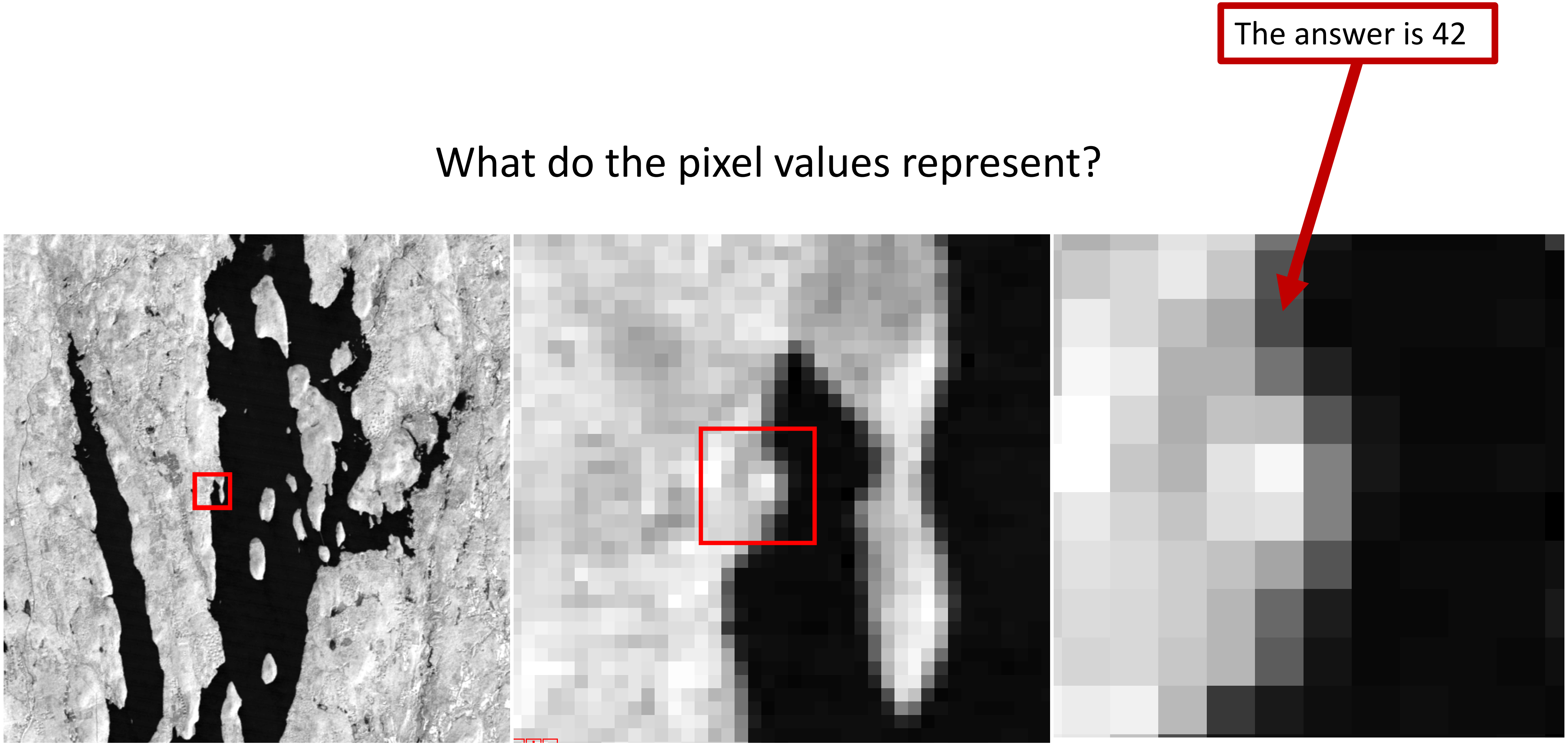
Raster Data

What do the pixel values represent??



Raster Data

What do the pixel values represent?

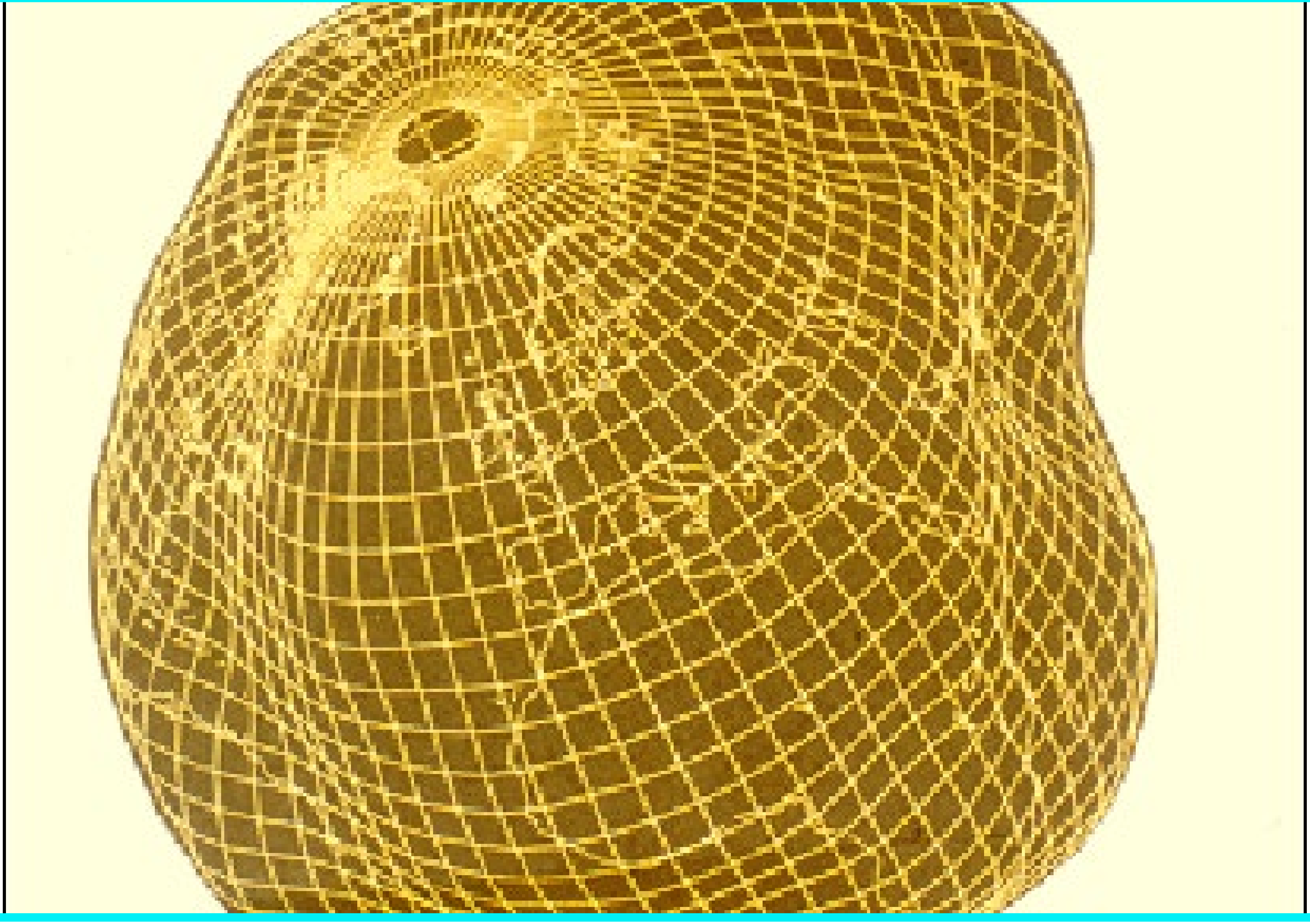


The Earth’s Shape

What is Earth's shape?

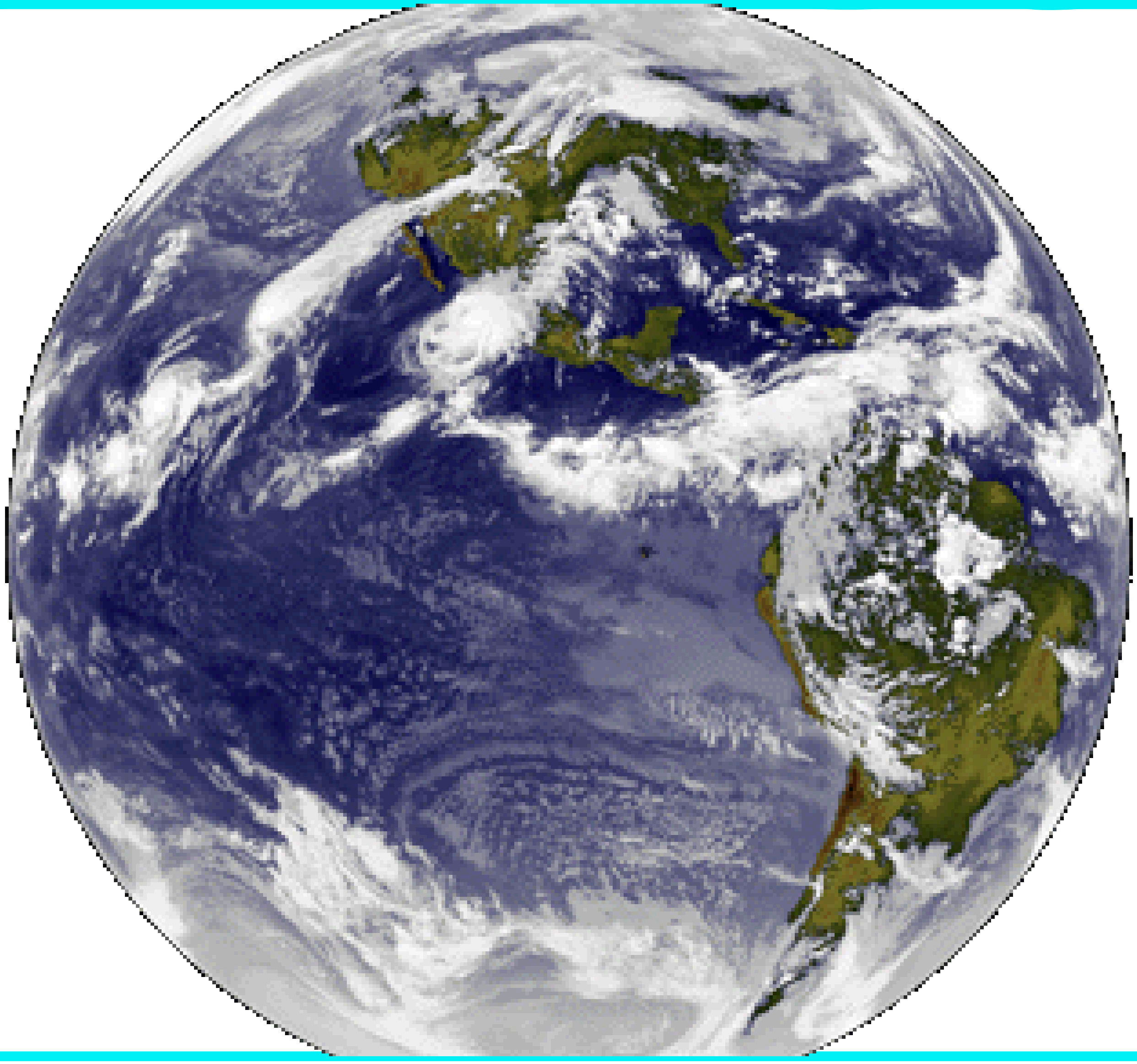
Some Possible Models

- Sphere
- Ellipsoid
- Lumpy space potato
- Geoid



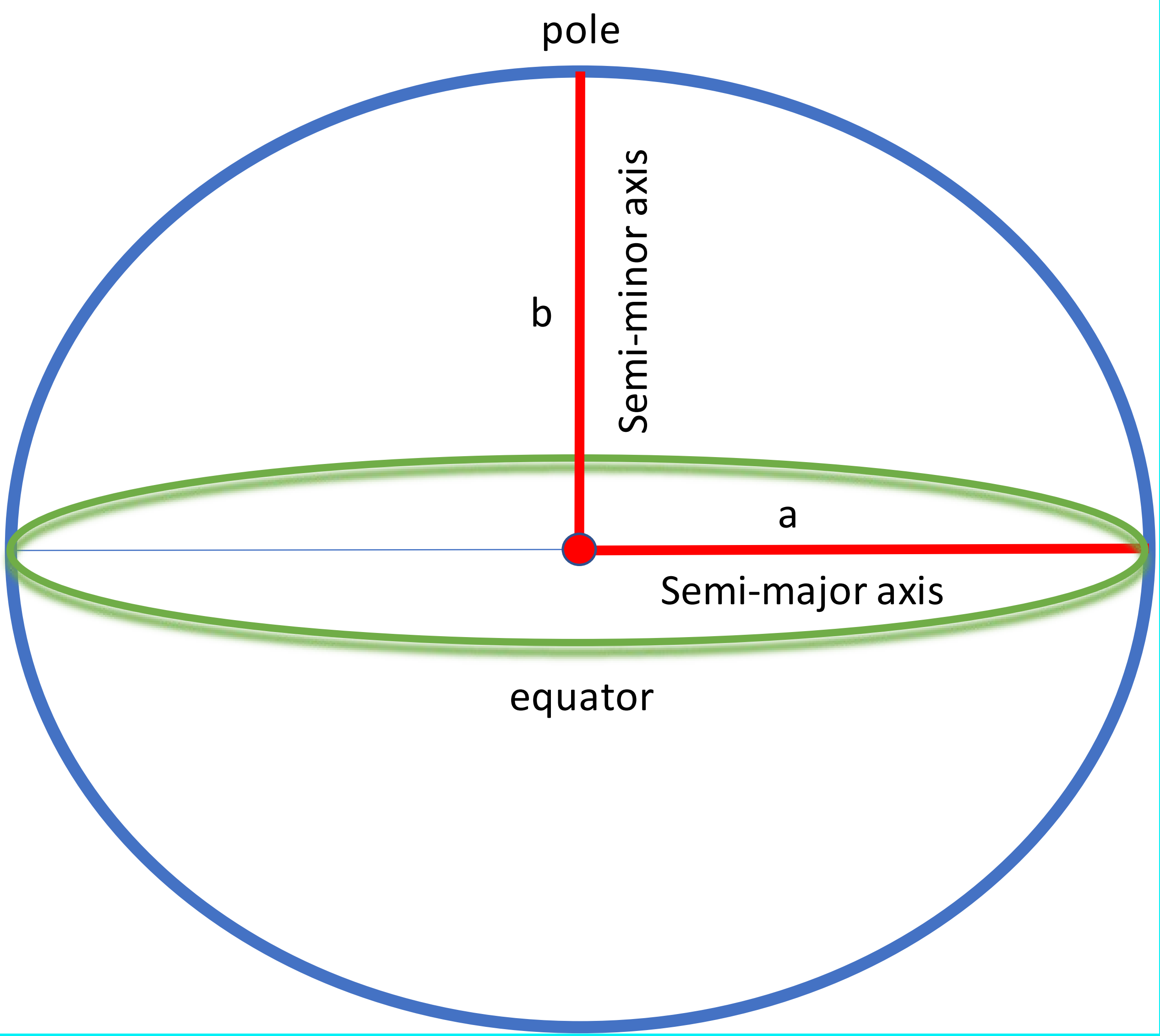
Spherical Model

- A useful conceptual simplification.
- But it’s a little too simple: The earth has a large equatorial bulge due to rotation around its axis.
- The bulge is enough to matter for coordinate systems:
  - Earth’s equatorial bulge is about 40km



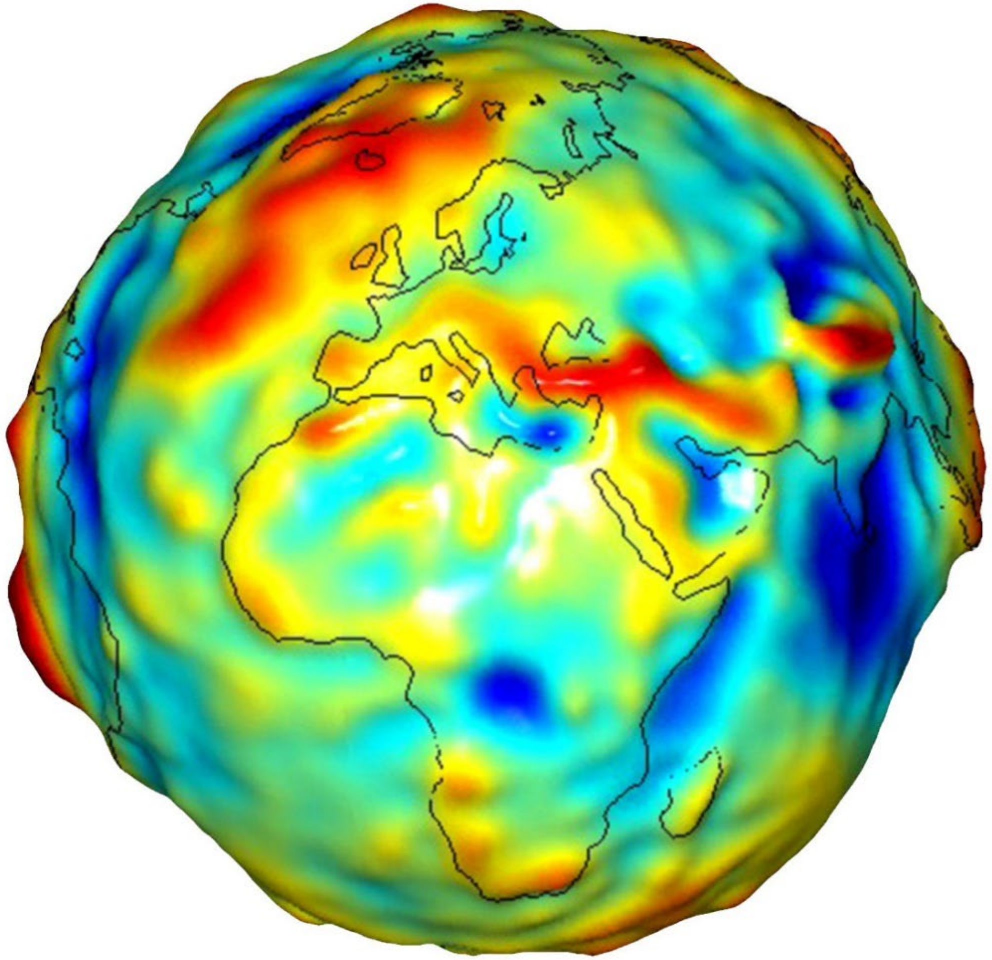
Ellipsoid Models

- $A\ measure\ of\ flattening$:
- $f = \frac{a-b}{a}$



Lumpy Space Potato

- The true shape of the earth is more like a lumpy potato with undulations from the ellipsoid as much as 100 m.
- There is also a large bulge in the earth of 10 to 15m in the Southern Hemisphere giving rise to the description of earth as **pear shaped.**

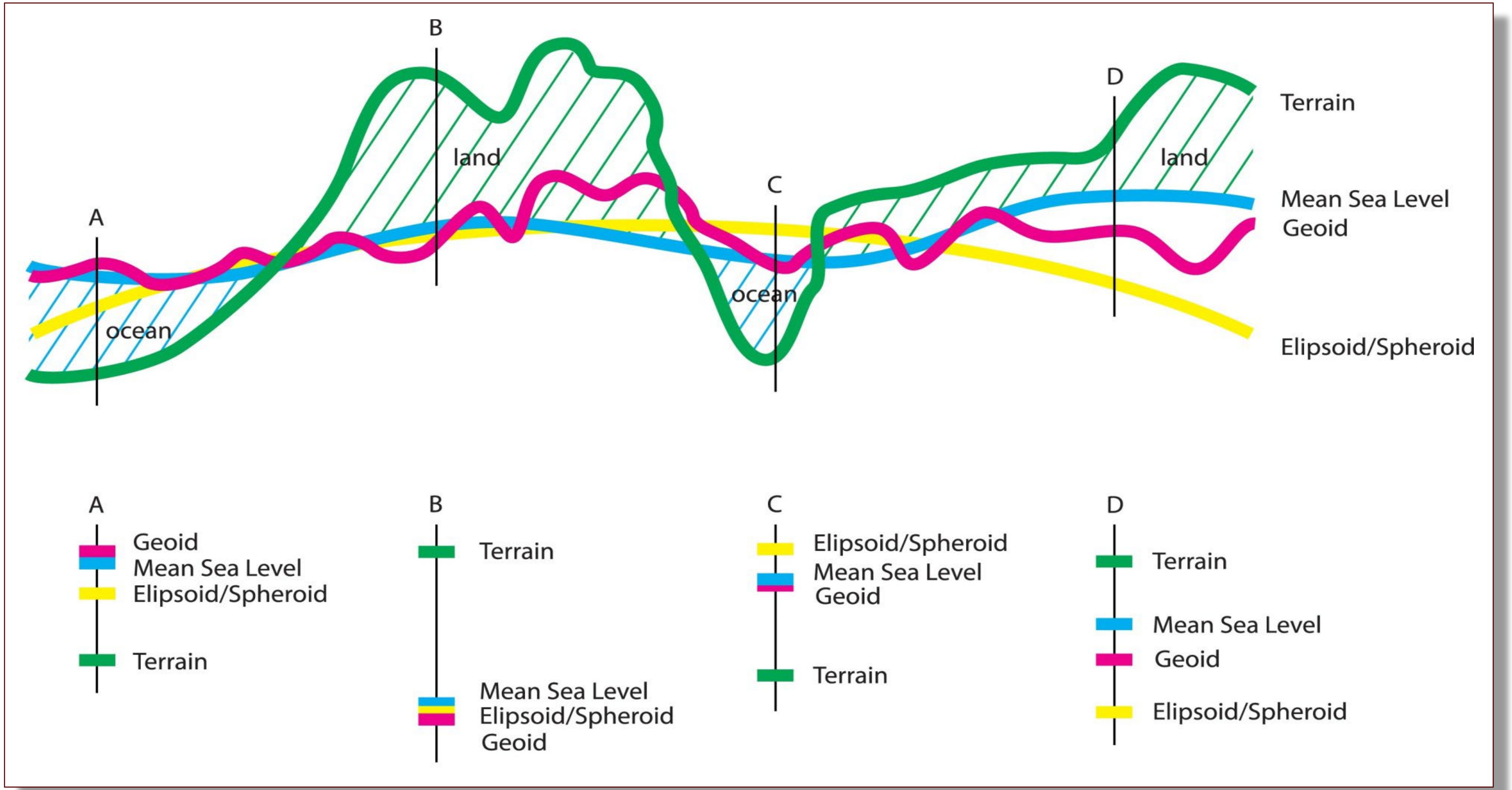


Source: Paul Bolstad. 2012. GIS Fundamentals – A first text on Geographic Information Systems. 4th ed.

What is Earth's true shape?

- The actual shape of the Earth is a Geoid, literally "Earth Shaped".
- The Geoid is determined by gravitational measurements.
- The Geoid is similar to the Earth's **mean-sea-level** surface.
    - For land, MSL is height to which water would rise in a well that is connected to the ocean.

Terrain, Ellipsoid, and Geoid



http://www.icsm.gov.au/mapping/web_images/cross_section.jpg

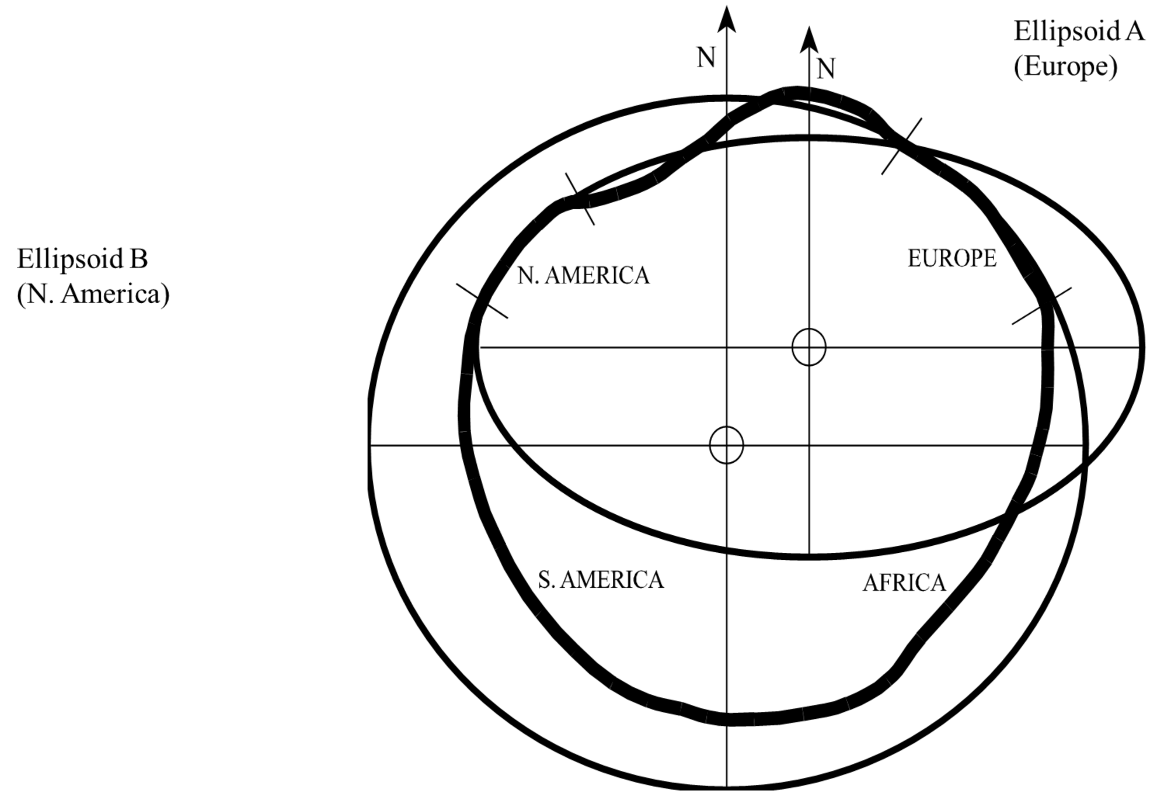
http://www.esri.com/news/arcuser/0703/geoid1of3.html

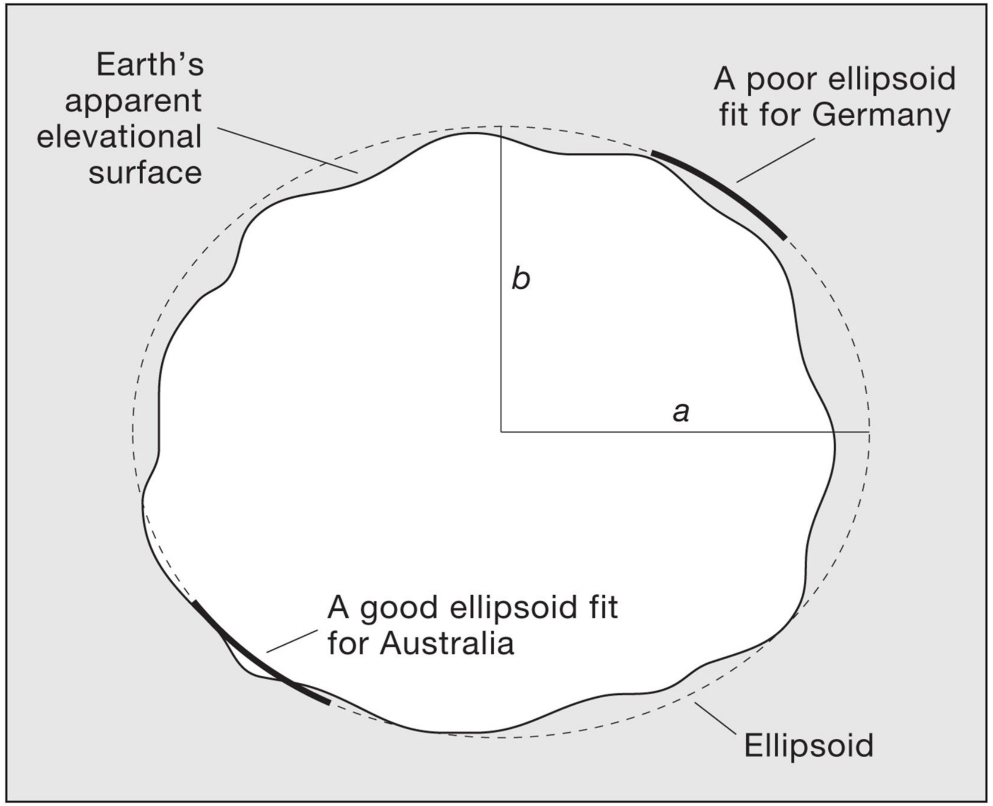
Limiting Complexity: Tradeoffs

- All models are wrong, some models are useful.
- The Geoid, while a much simpler shape than the earth's topographic surface is still very complex.
- For most uses, the simpler ellipsoid works well.
- But… How do we choose the "best" ellipsoid?

Local Ellipsoids

Different Ellipsoids are developed to fit the area of interest accurately over the area of interest



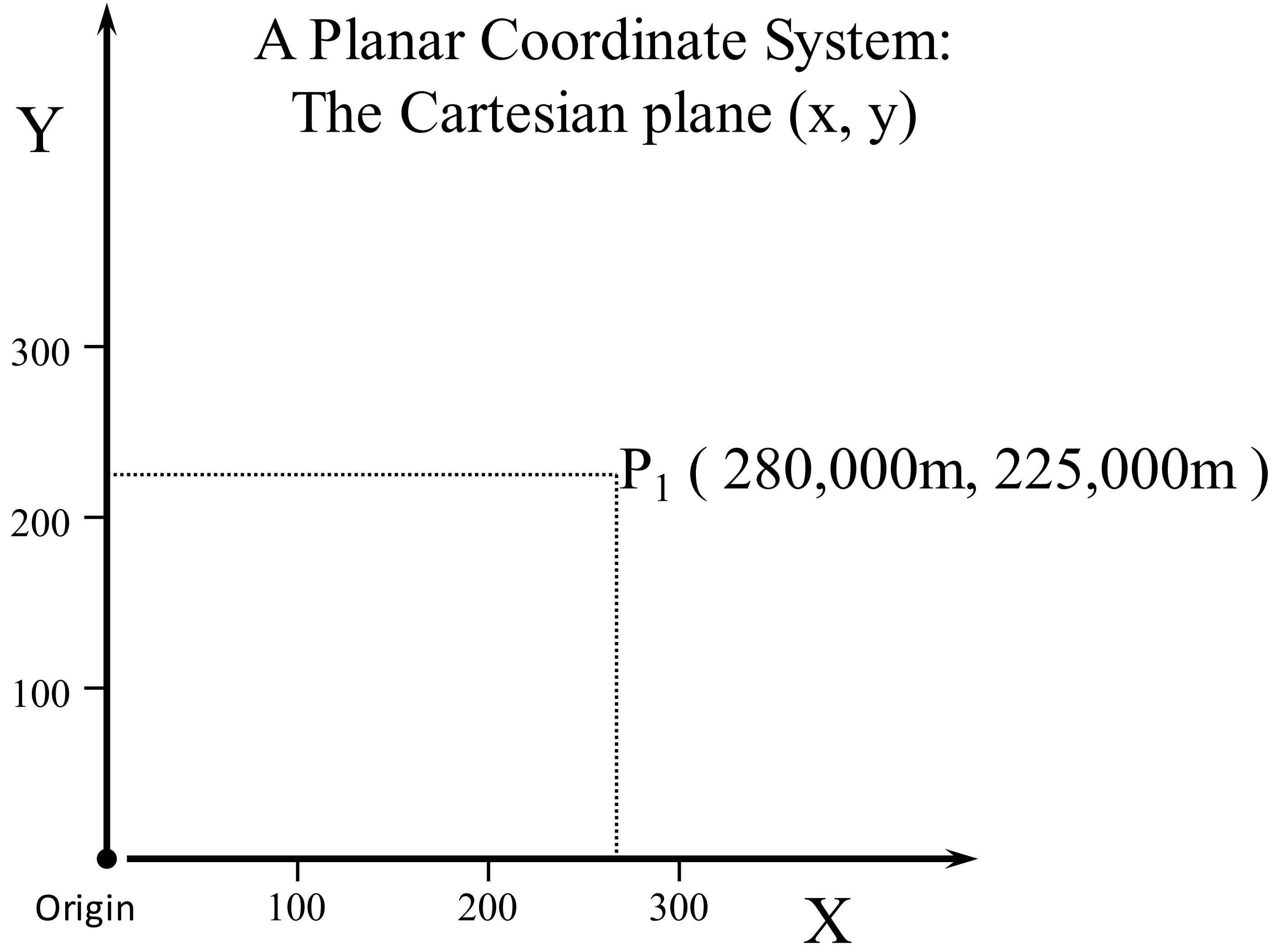


Coordinate Systems

Geographic Coordinates and Datums

Coordinate Systems

- To be meaningful, spatial data (whether raster or vector) must be associated with a location.
- Coordinate systems are used for the location or registering of those data



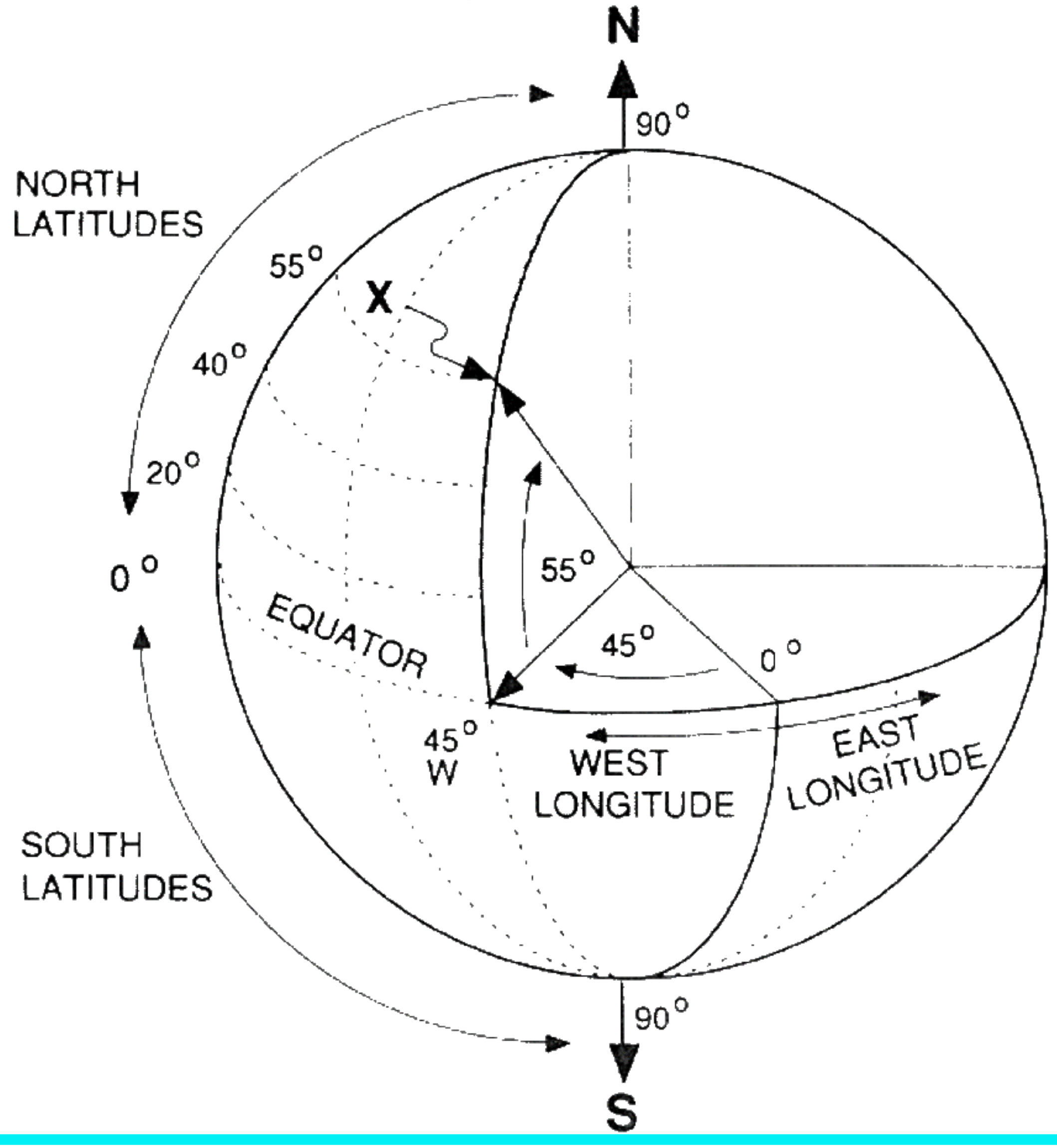
Spherical Coordinate System (2D)

We can use non-Euclidean, spherical geometry to define a 2D coordinate system on a 3D object (the ellipsoid).

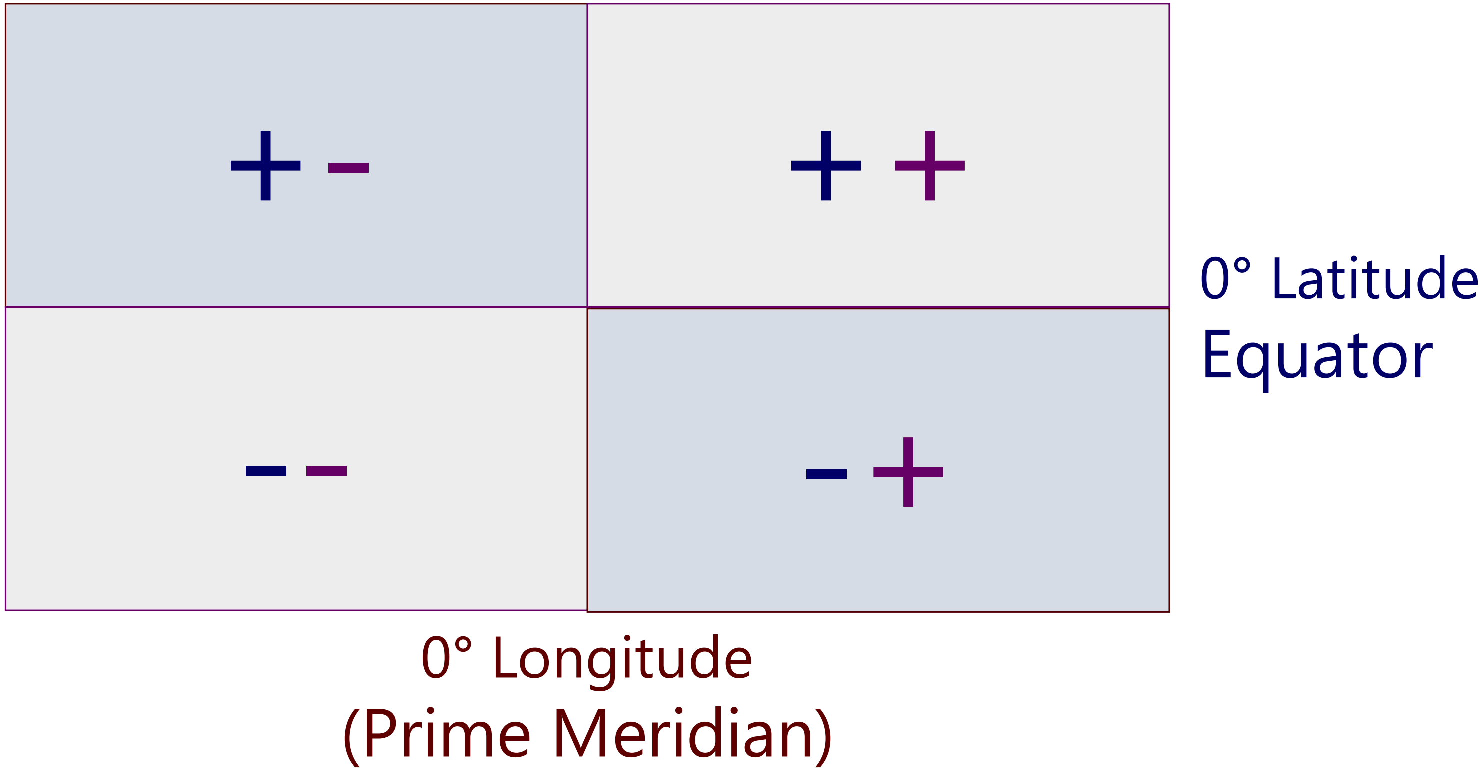
- Don’t worry, we’re not going to learn how to do spherical trigonometry!

Latitude: degrees (°) North or South of the Equator

Longitude: degrees (°) East or West of The Prime Meridian

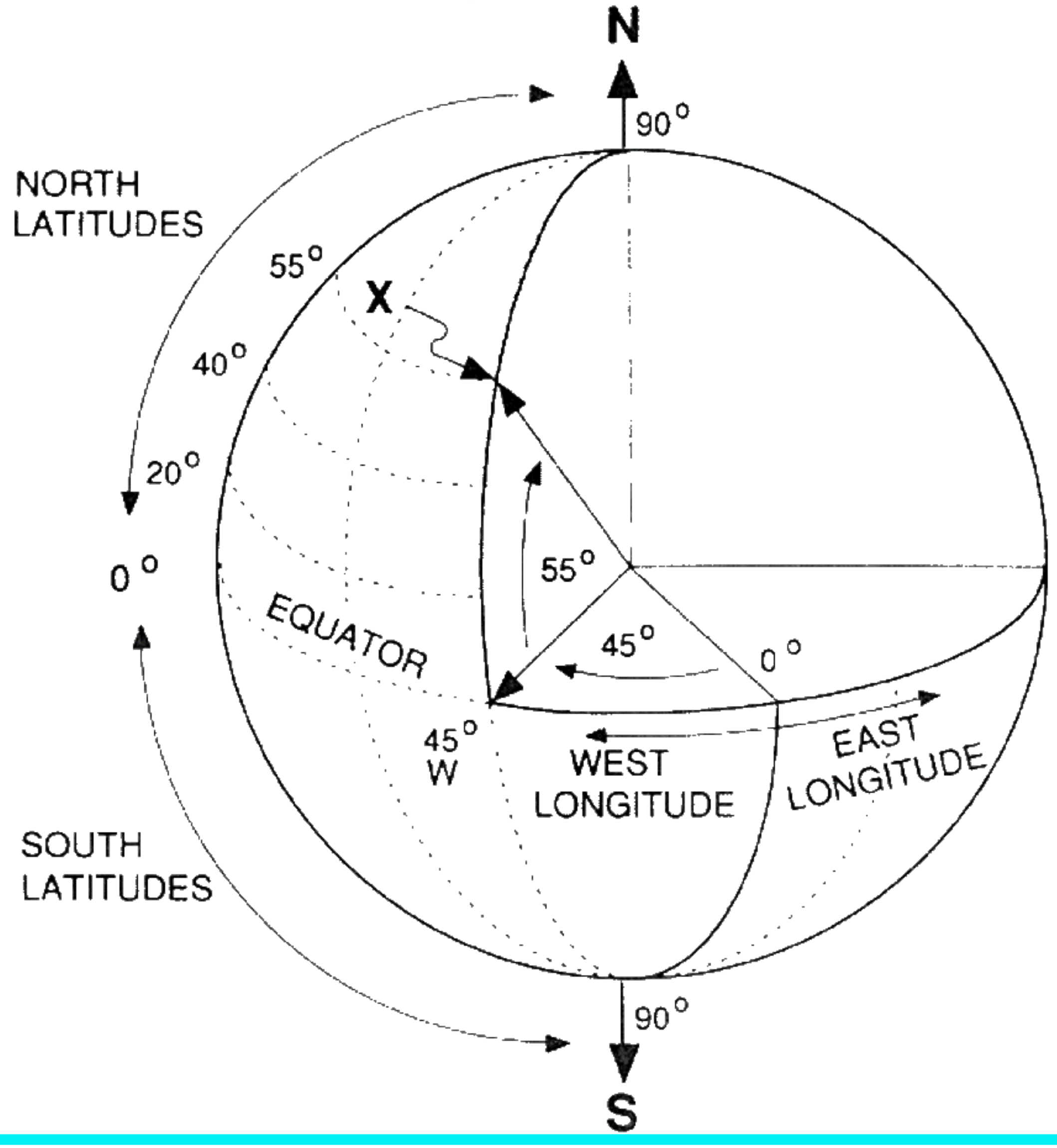


Sign Convention



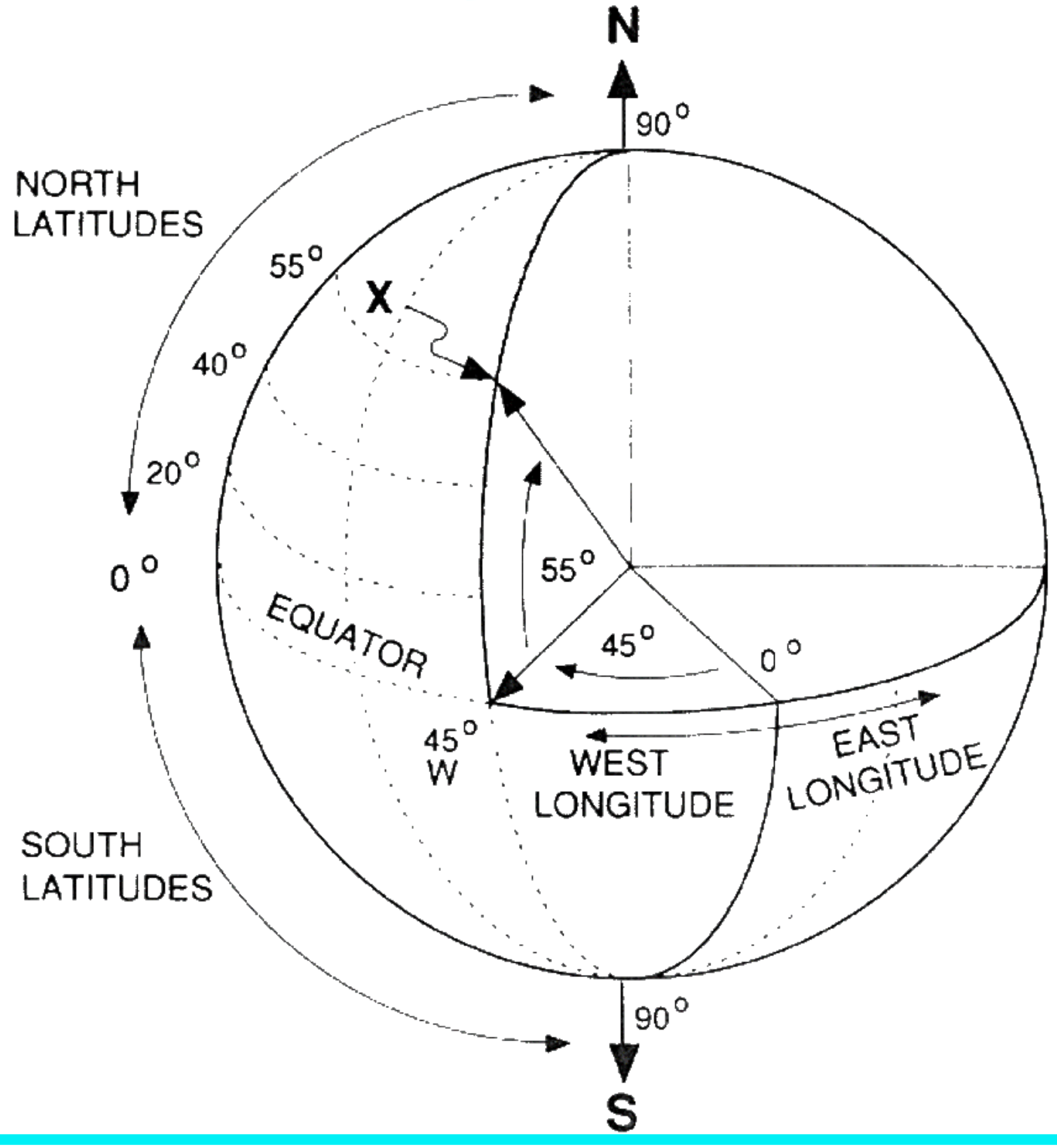
Spherical Coordinate System (2D)

- The x- and y-coordinates on an ellipsoid model comprise a Geographic Coordinate System (GCS).
- Any time you use degrees, longitude, or latitude you are talking about a GCS.
  - Why do I say a GCS, and not the GCS?
- A GCS lives on the mathematically ideal surface of an ellipsoid model.



Spherical Coordinate System (2D)

- A GCS is not enough to specify locations on Earth's surface, for that we need to anchor the ellipsoid model to specific locations.
- Remember local ellipsoids?
- The pairing of an ellipsoid model (and its GCS) with anchor points is a **datum**.



What is a Datum?

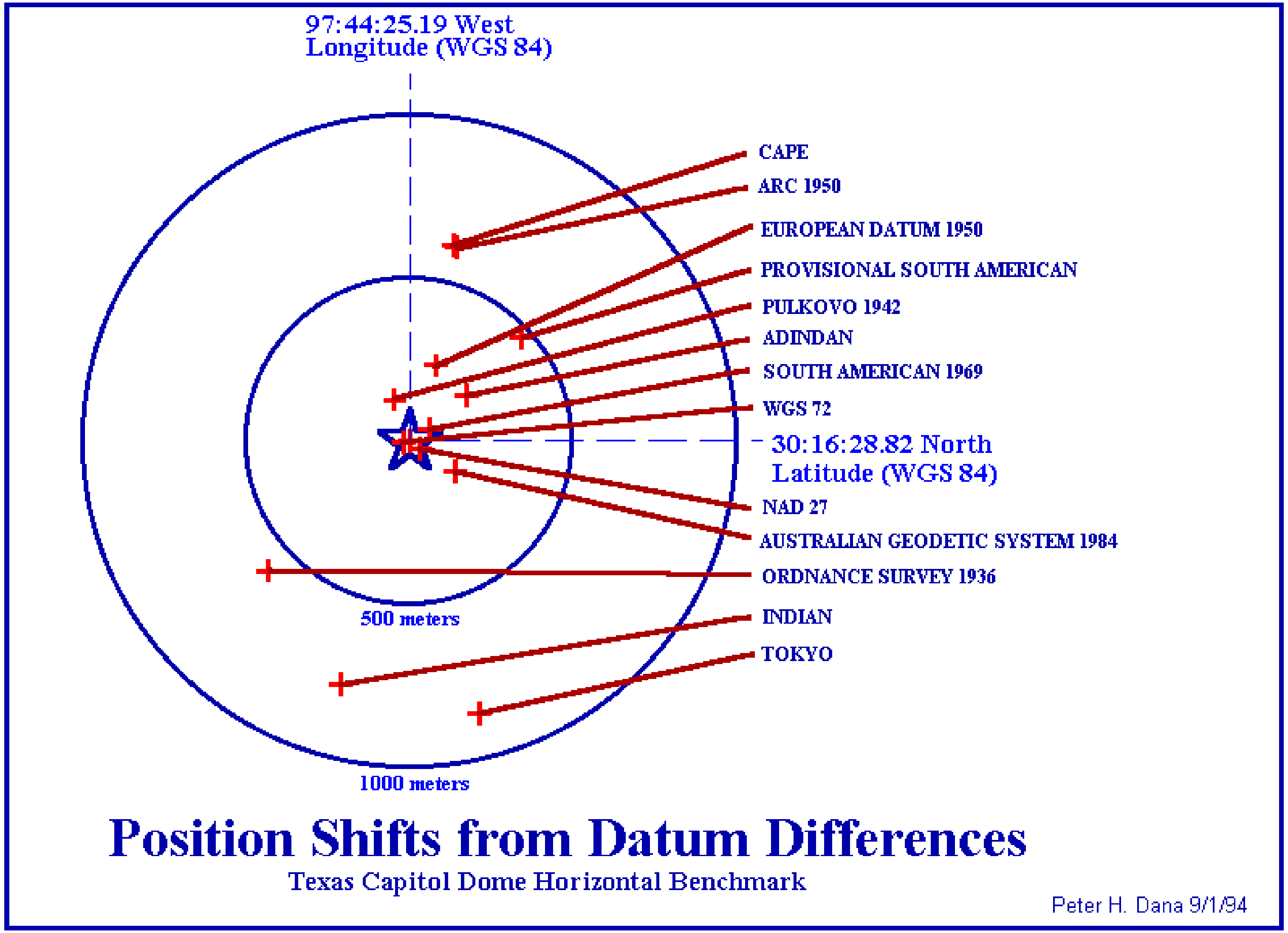
*In surveying and geodesy, a **datum** is a reference point or surface against which position measurements are made, and an associated model of the shape of the earth for computing positions*

http://en.wikipedia.org/wiki/Geodetic_system

- A geodetic datum is a mathematical model of the earth upon which geodetic computations are based.
- A datum is a reference system with two components:
  - A specified **elliposid with a spherical coordinate system** and an **origin**
  - A set of highly accurate surveyed **points** and lines to anchor the ellipsoid
- There are *Regional* and *Global* Datums.

Does it Make a Difference?

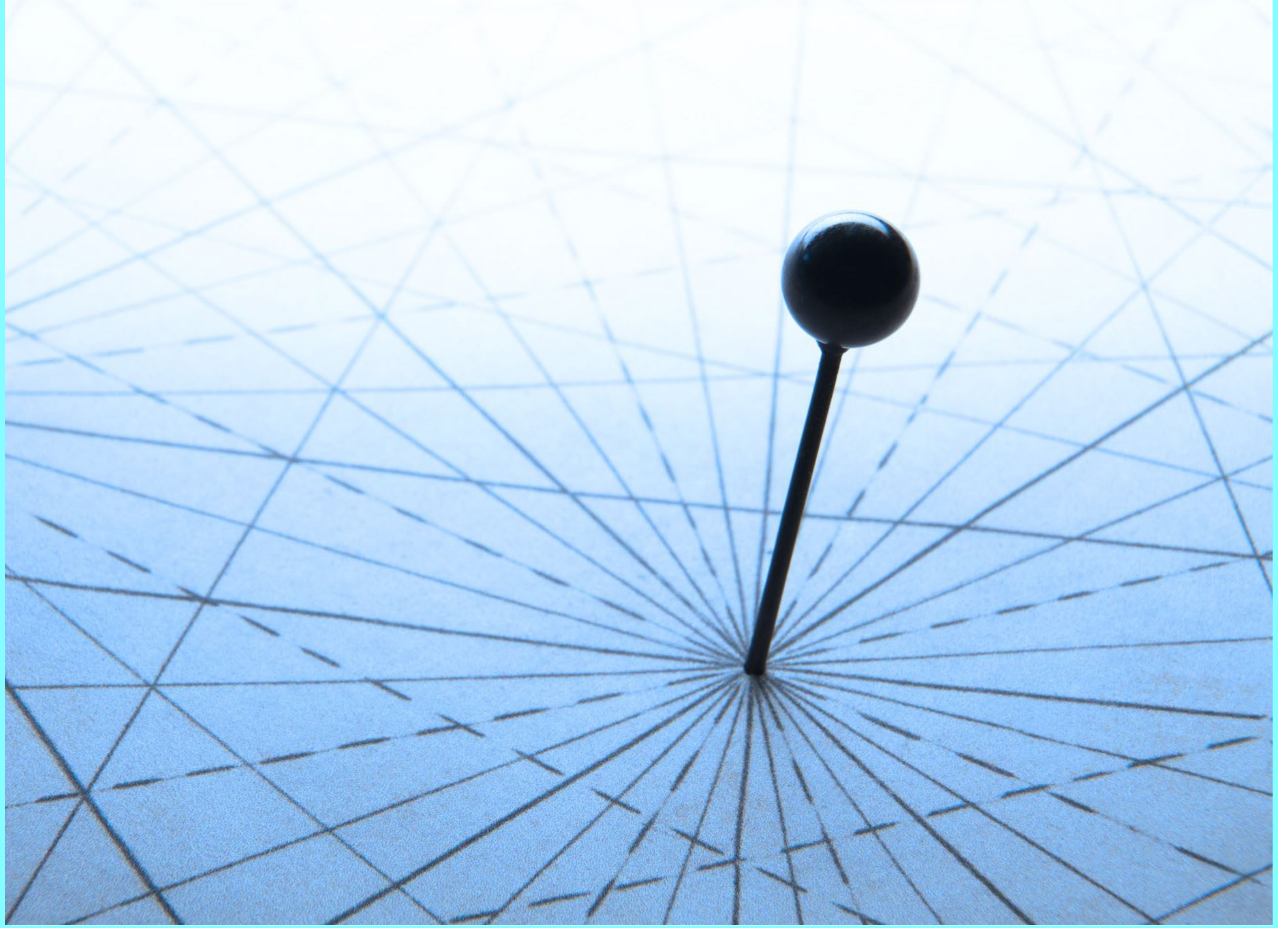
- Short answer: yes!
- Long answer: if you don't know which datum your data are tied to, you can be off by 1km or more.
  - This might not matter for some purposes.
  - This could be terribly important for others – it all depends on the scale.



Geographic Coordinate Systems

Recap: so…. What are GCSs and datums?

- A GCS is a set of 2D coordinates, based on spherical geometry/trigonometry, that can uniquely specify a location on an ellipsoid model.
- A datum is a combination of an ellipsoid (with its GCS) and a set of anchor points on the surface of the Earth.



Projected Coordinates

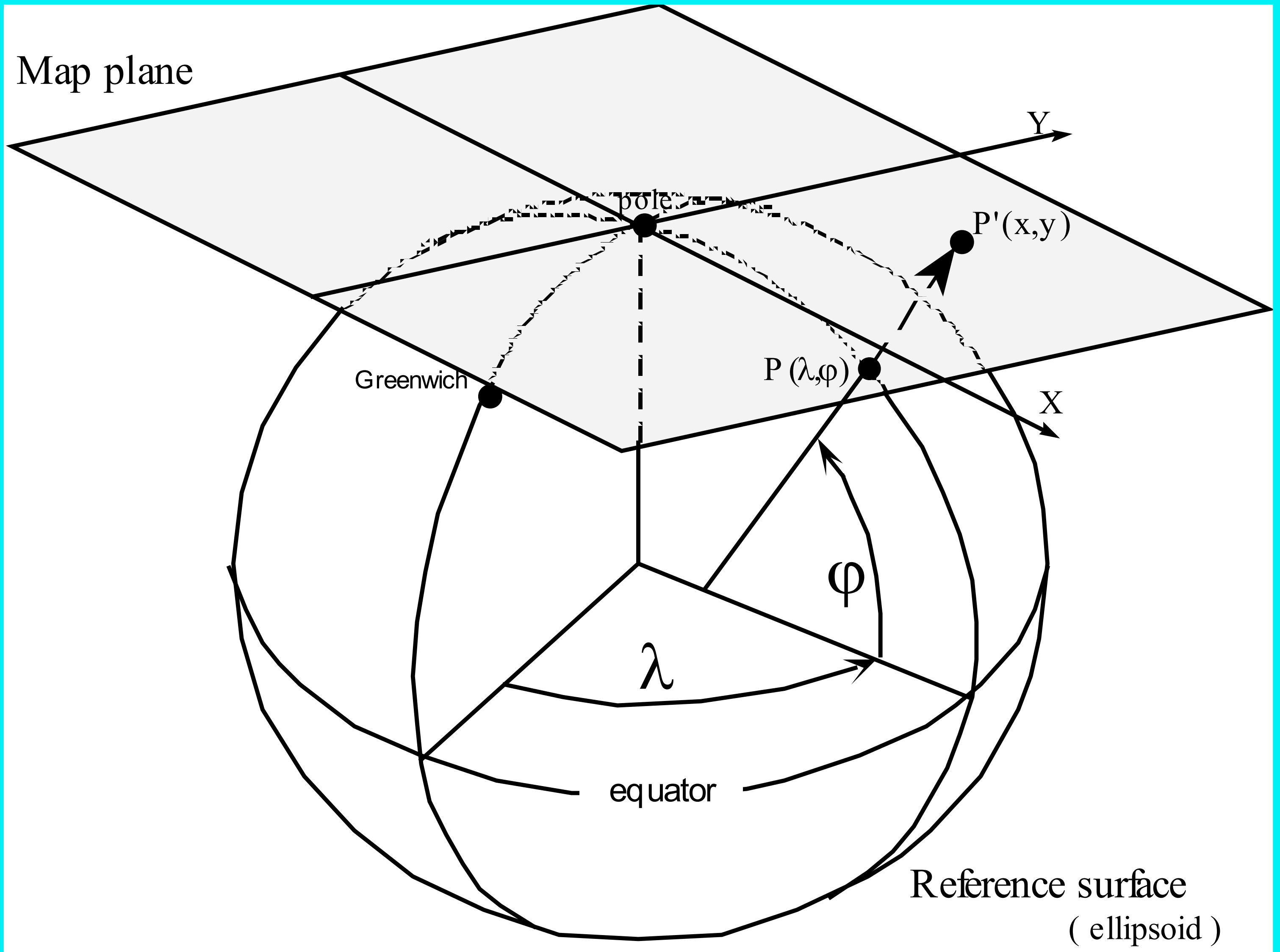
Projected Coordinates and Map Projections

What’s a Projected Coordinate System?

- Problem: we need to represent the curved surface of Earth on a 2D flat surface.
- Solution: an algorithm for translating 2D spherical coordinates to 2D flat coordinates.
- Another problem: there are many such algorithms, and all of them cause distortion
- Solutions: different projection types can minimize distortion of shapes, distances, or direction (at the expense of the other attributes).

The Map Projection Principle

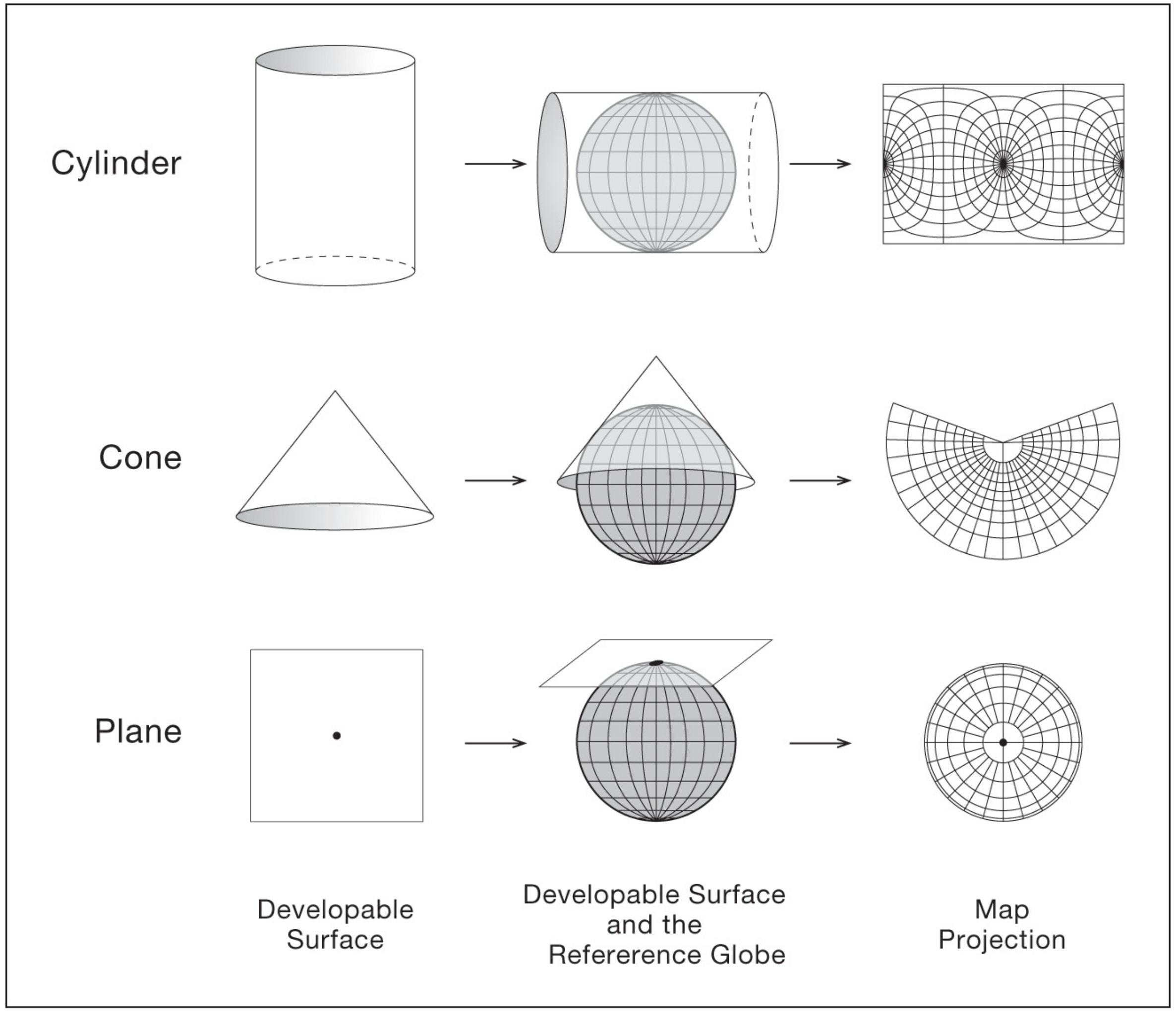
1. Reference globe
2. Developable surface
   - ➢Cylinders
   - ➢Cones
   - ➢Planes



Projection Techniques

There are 3 primary projection methods:

- Cylindric
- Conic
- Planar

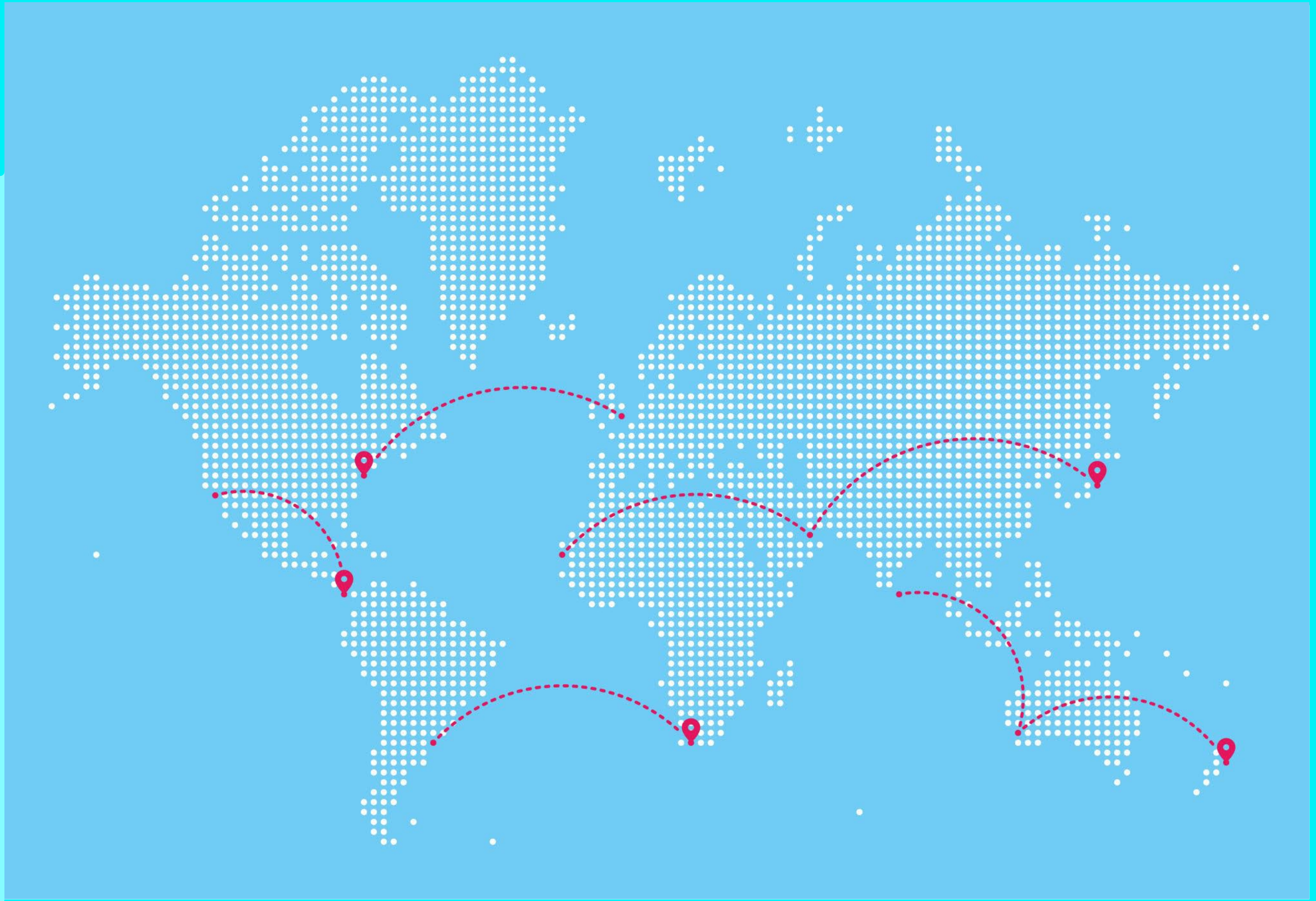


Map Projection Type

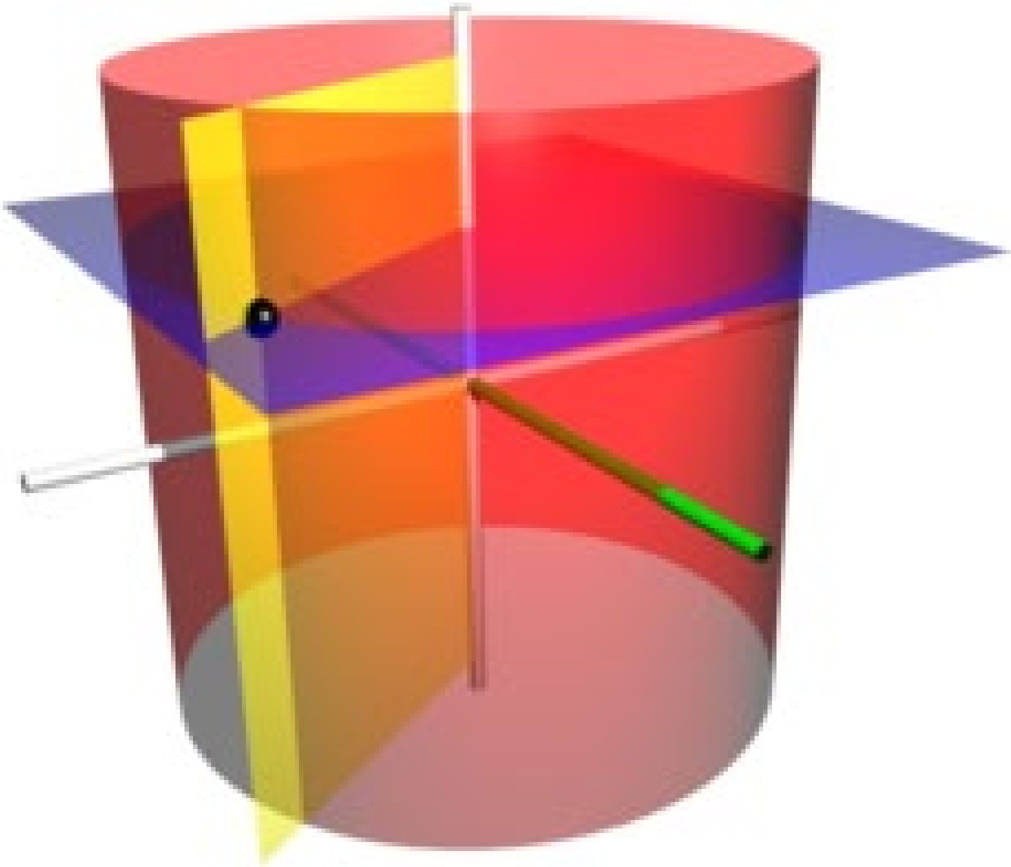
Map projections can preserve:

- Shapes – conformal map
- Areas – equal area maps
- Angles
- Distances
- Directions
- Compromise of

- But… none of the properties are perfectly preserved

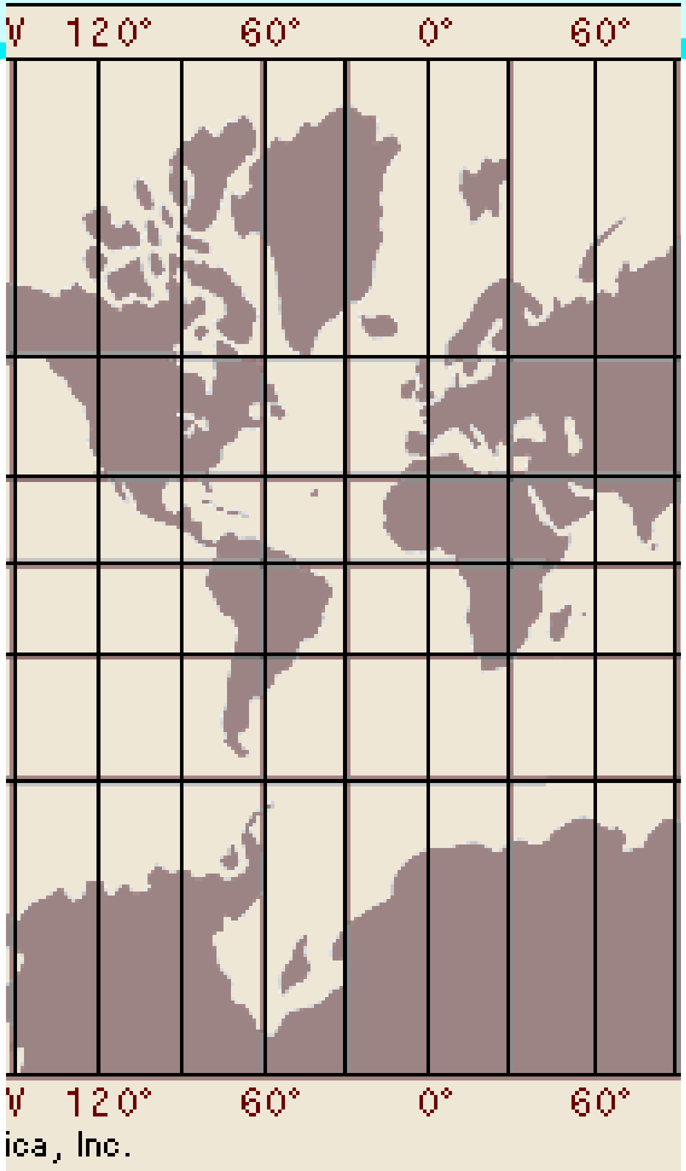


Cylindrical Projections

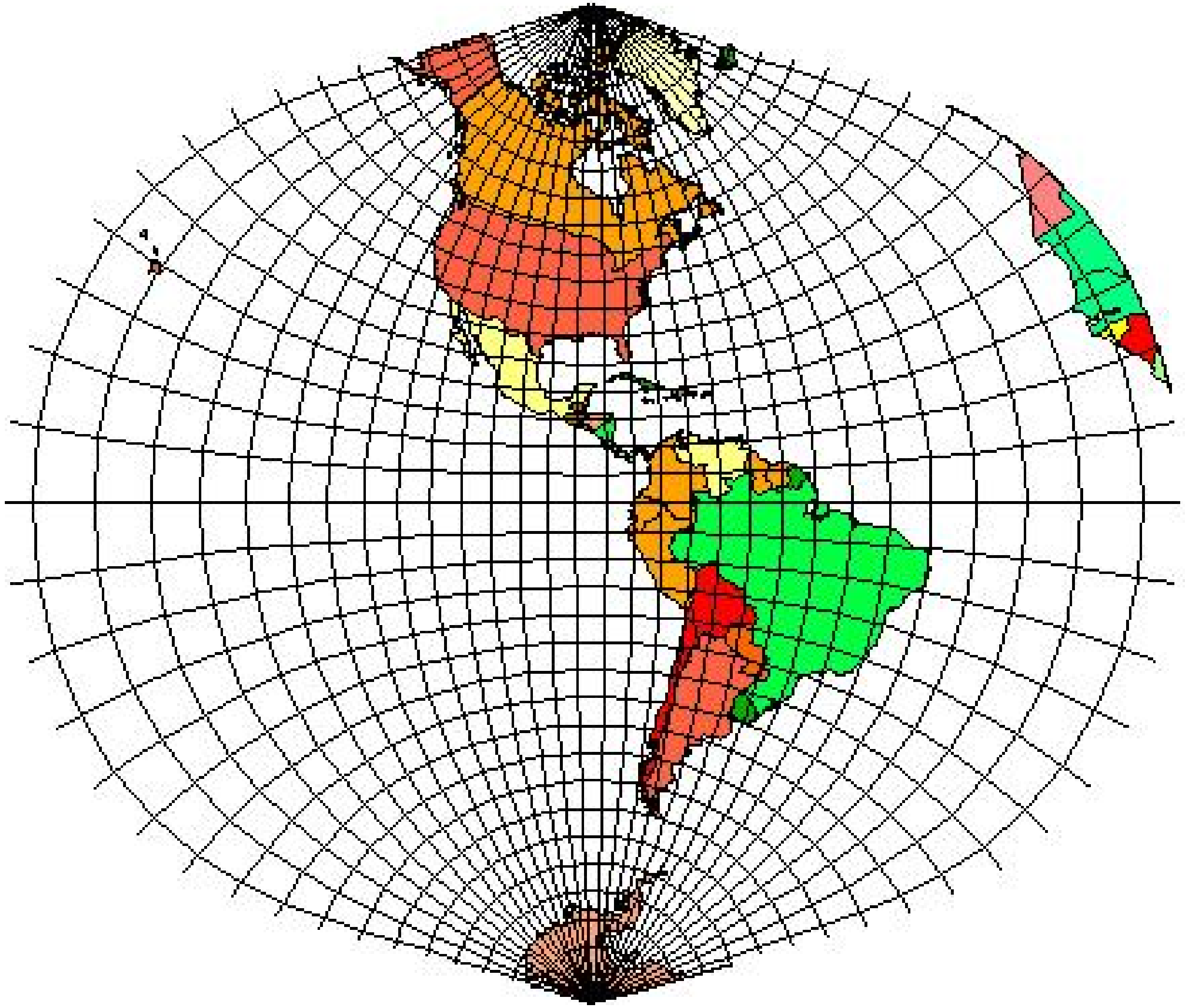


Mercator Projections

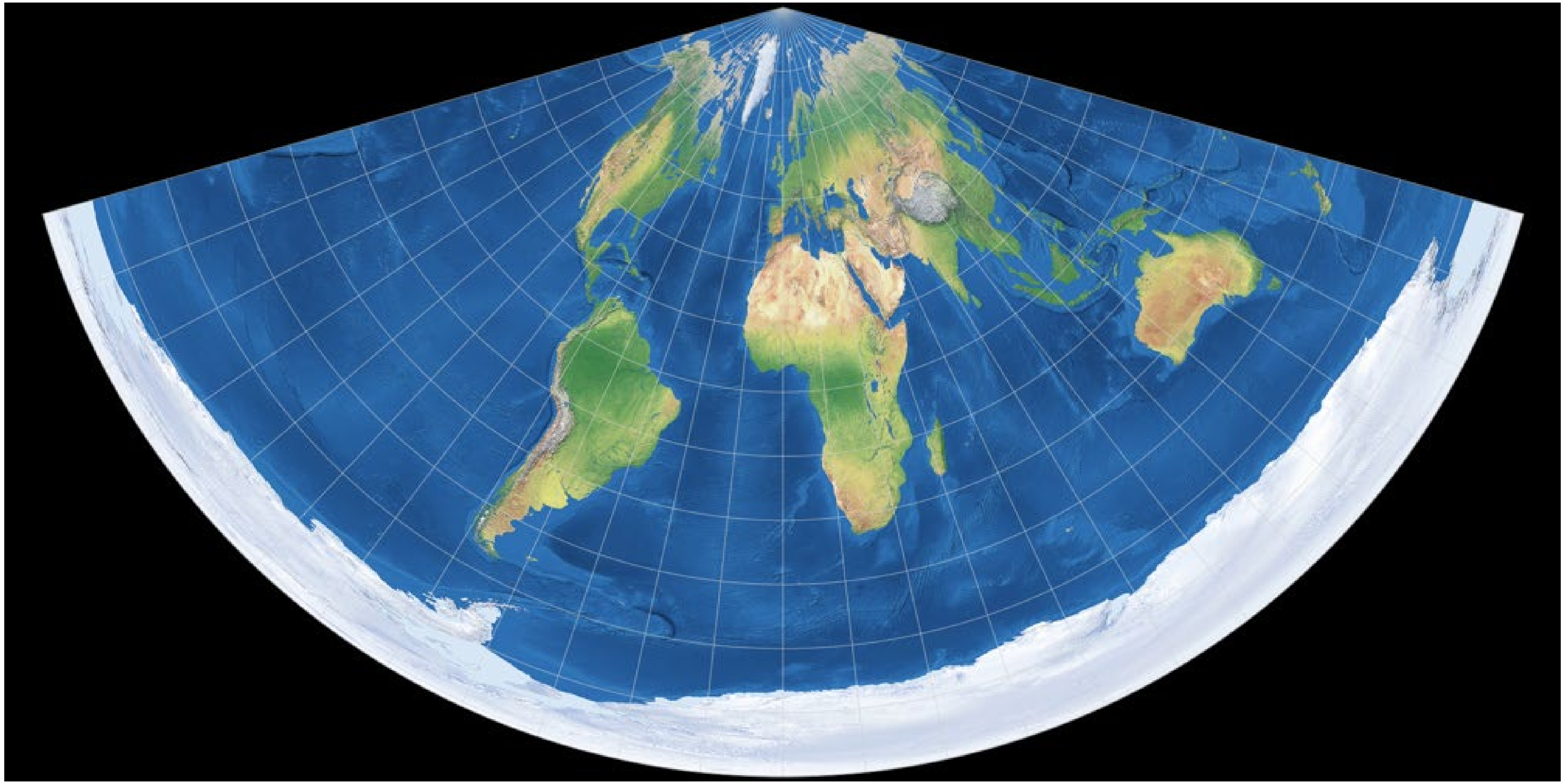
- Cylindrical projection developed by Dutch cartographer Gerardus Mercator in 1569
- Preserves shape & direction
- Used widely for navigation charts because direction is preserved.



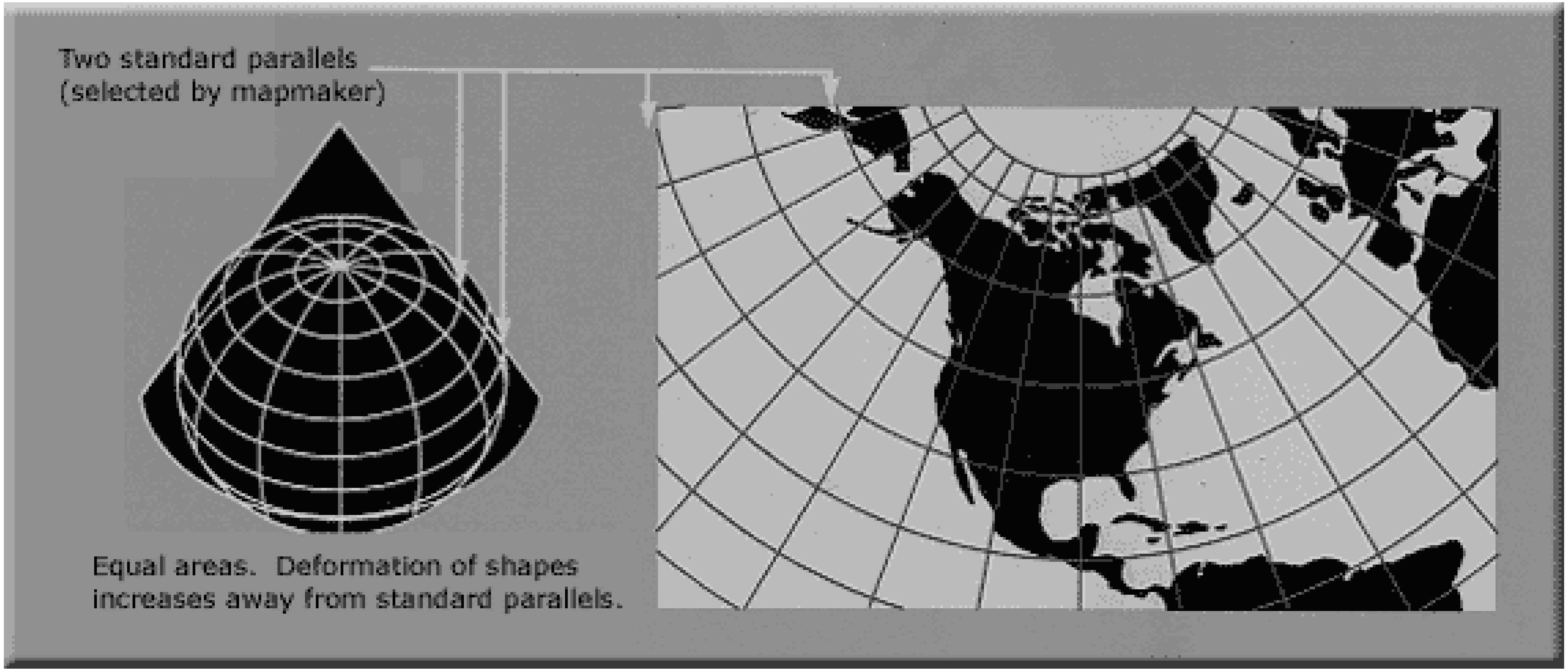
When would we want to use transverse Mercator?



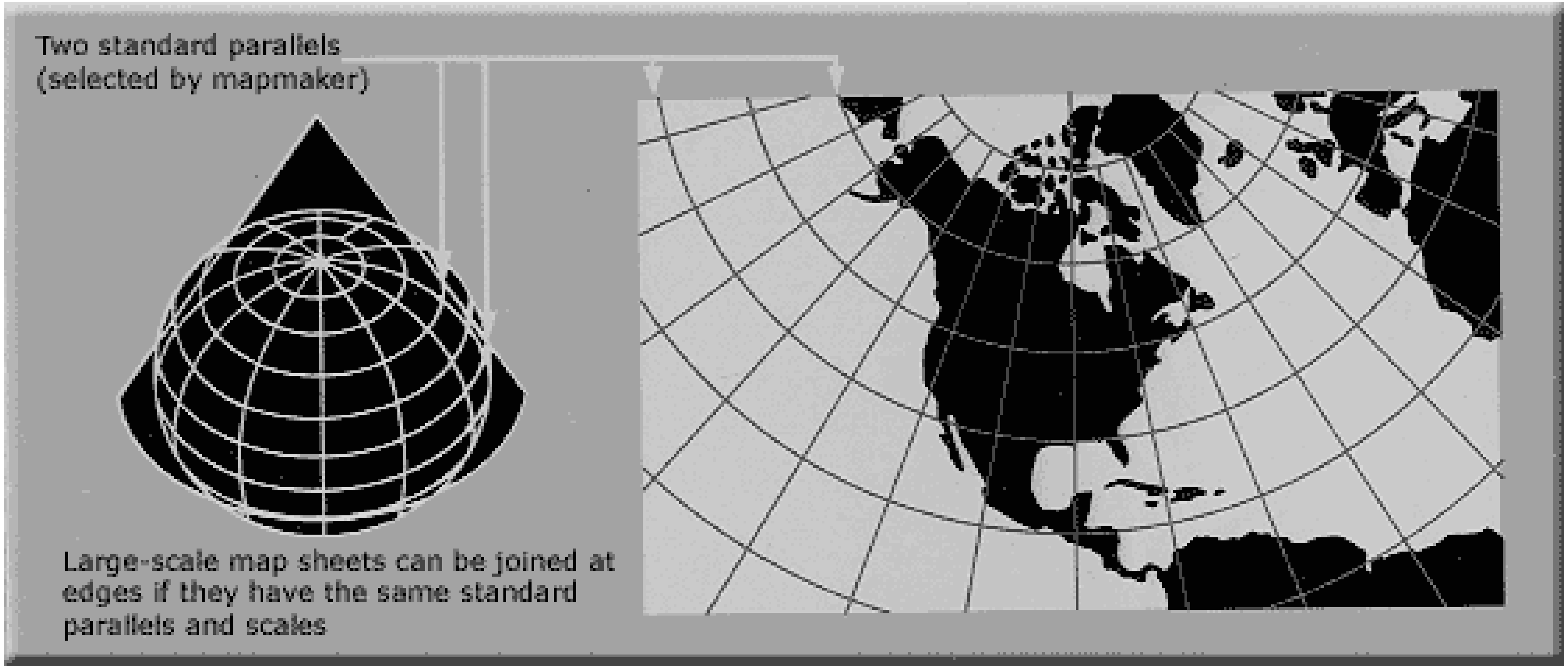
Conic Projections



Albers Equal Area Conic

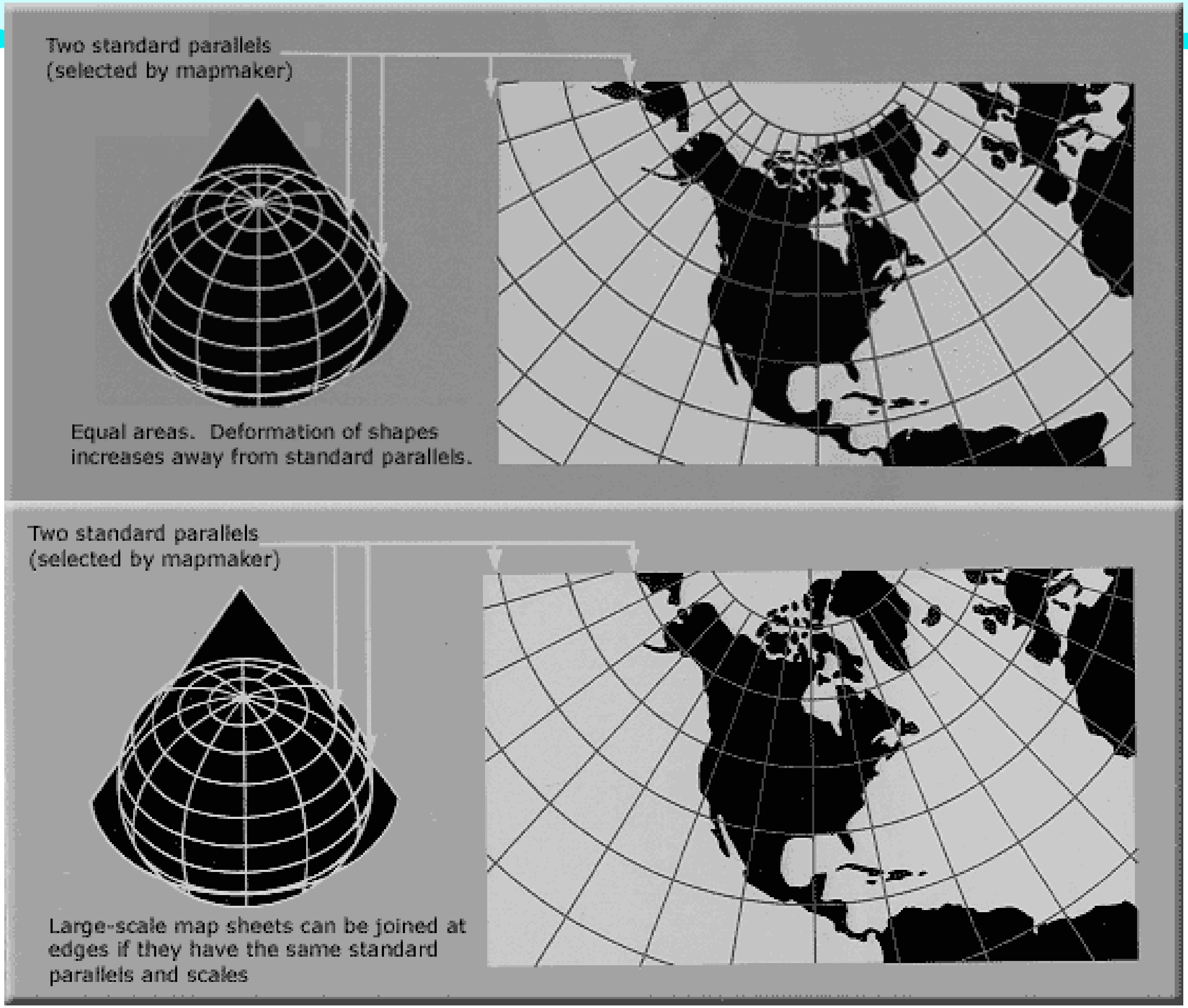


Lambert Conformal Conic

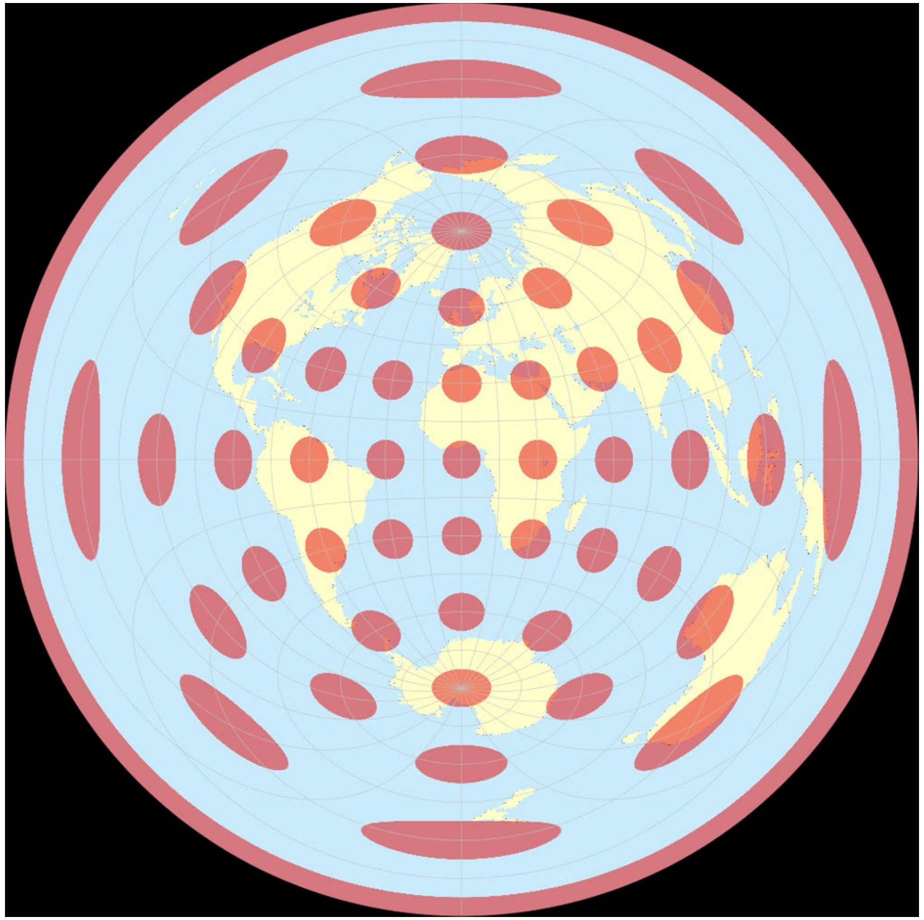


Conic: Conformal or Equal Area

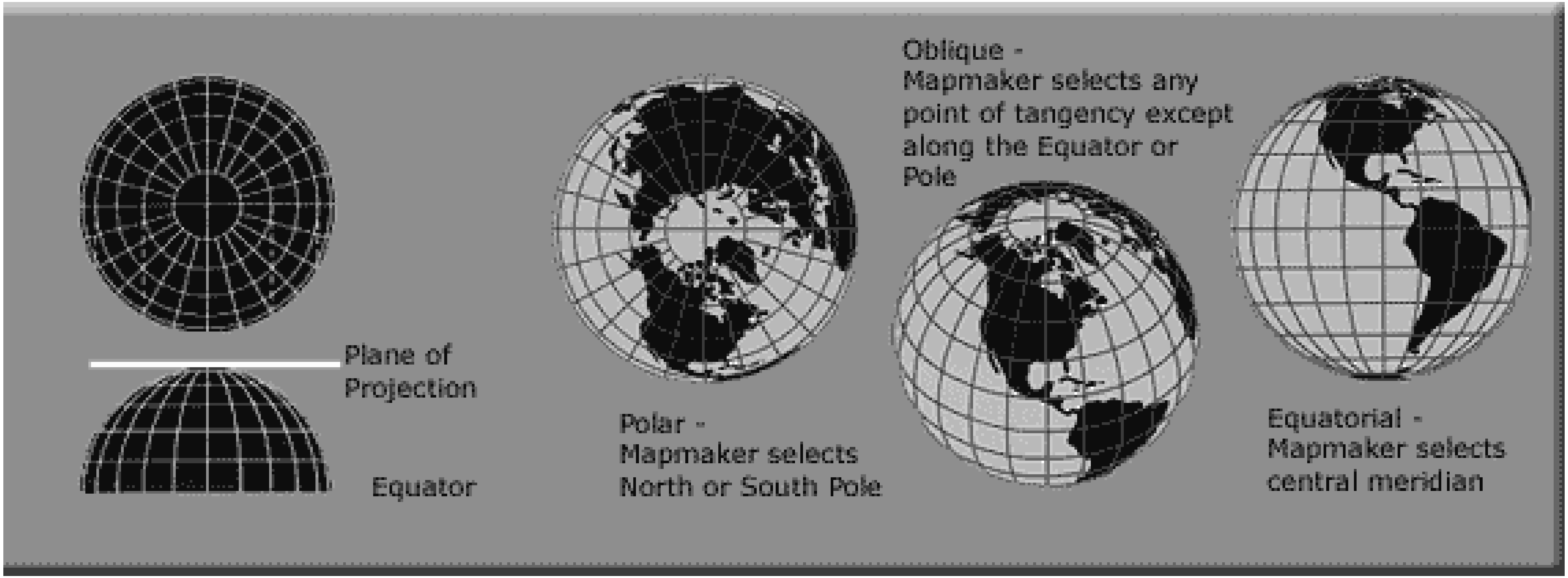
- Equal area: Areas of shapes are (mostly) preserved)
  - North and south parallels are squished
- Conformal: Shapes of objects are (mostly) preserved.
  - Central parallels more closely spaced



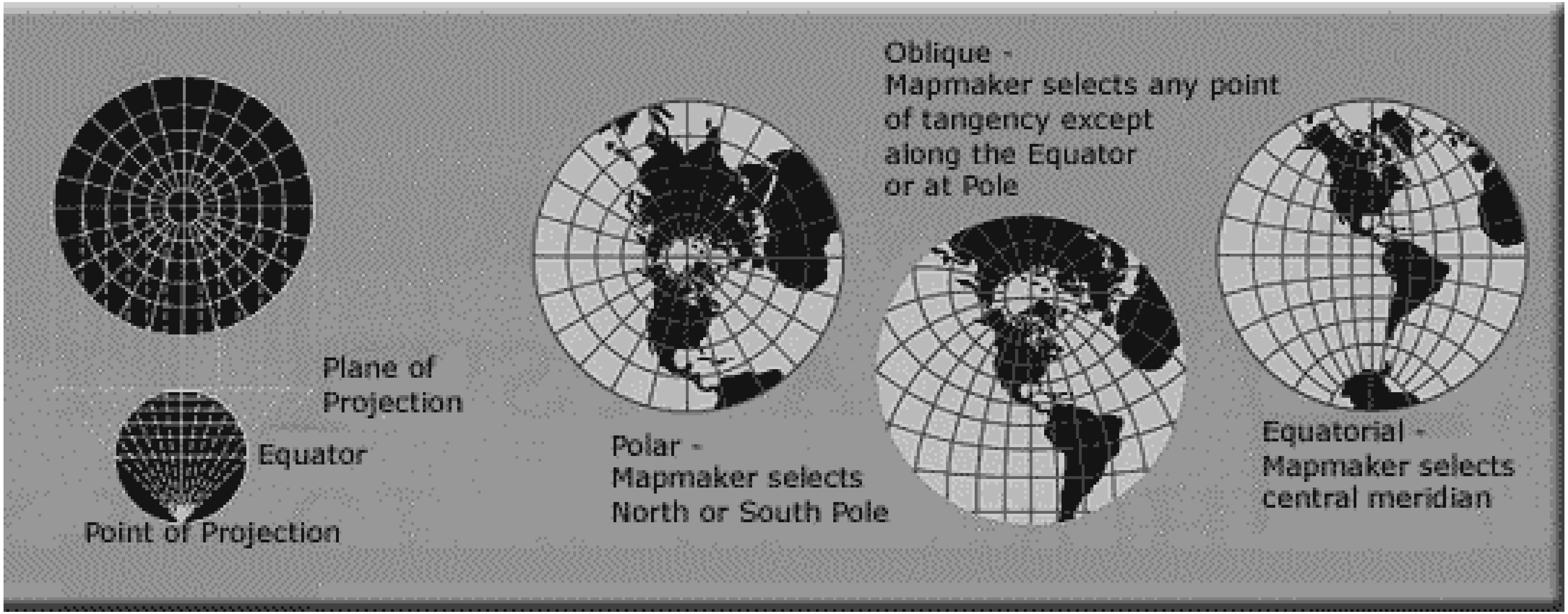
Azimuthal (planar) Projections



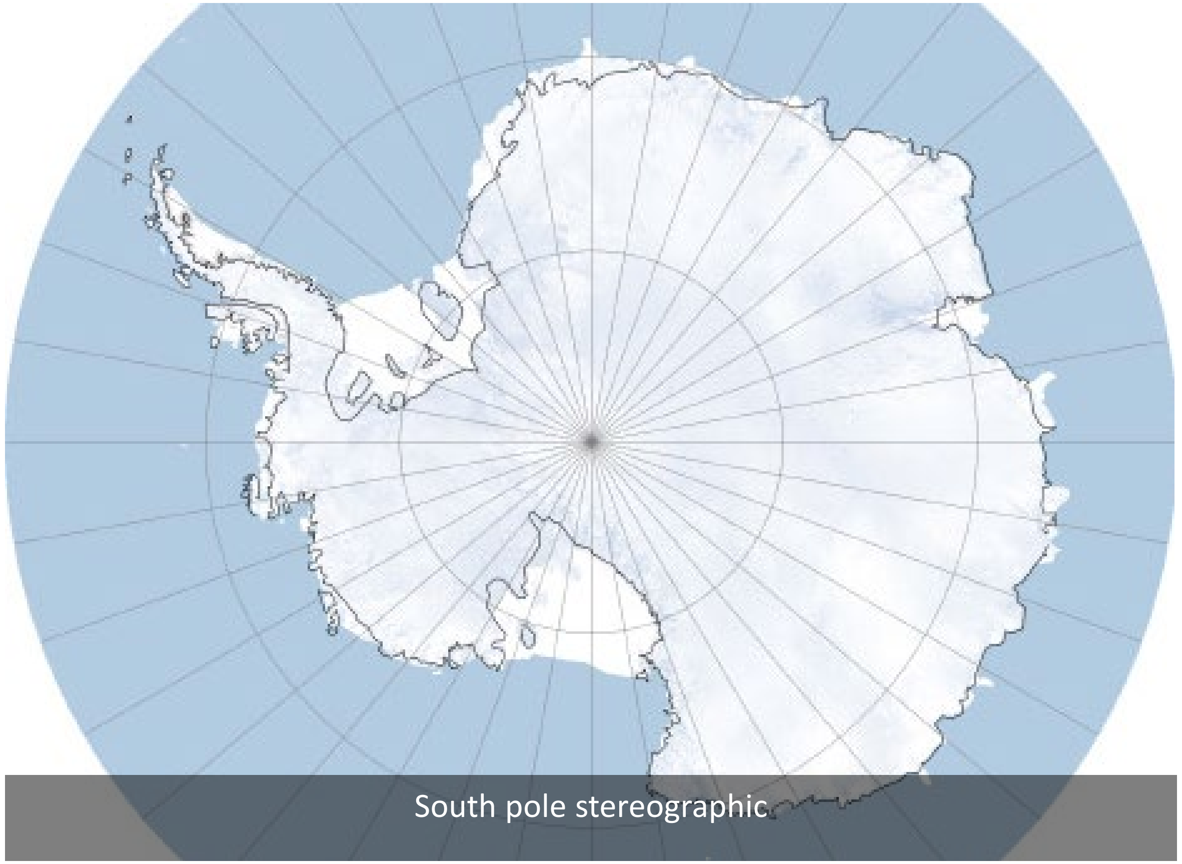
Orthographic (Azimuthal)



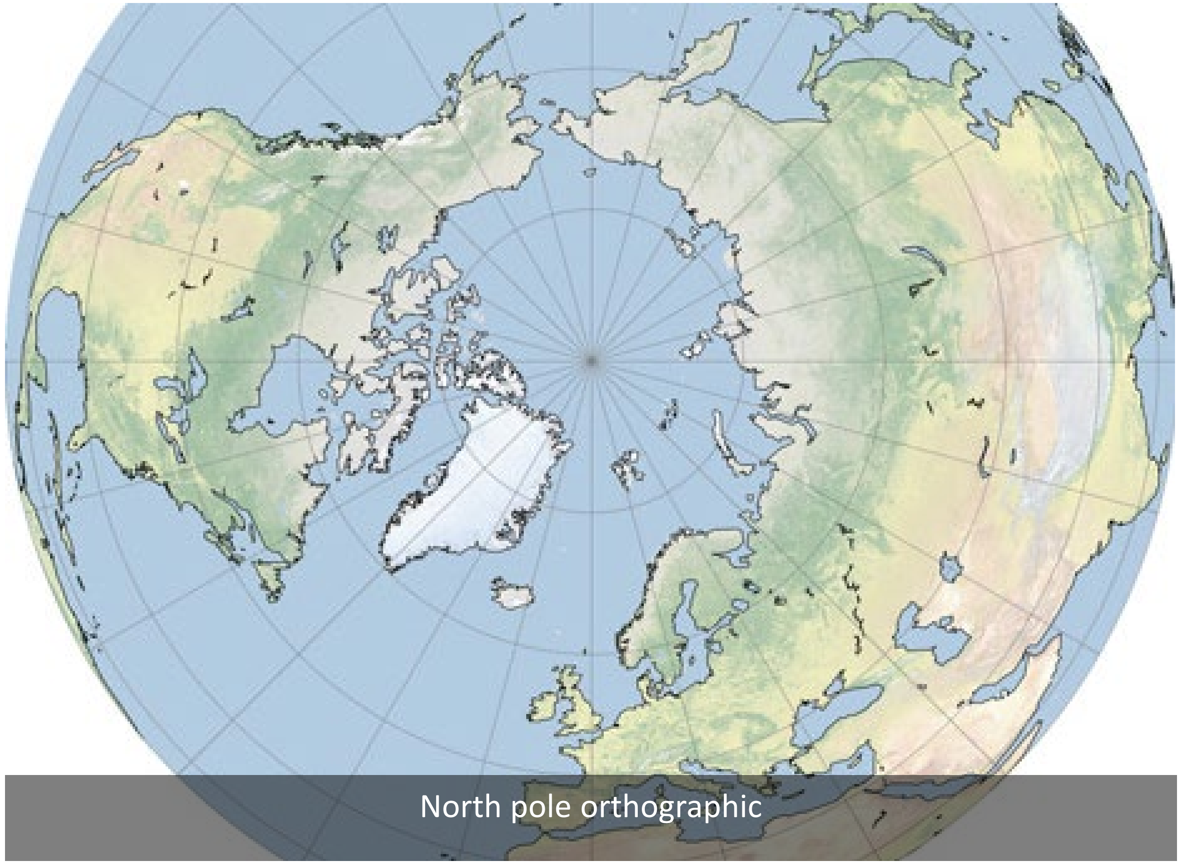
Stereographic (Azimuthal)



When would we want to use an Azimuthal projection?



South pole stereographic

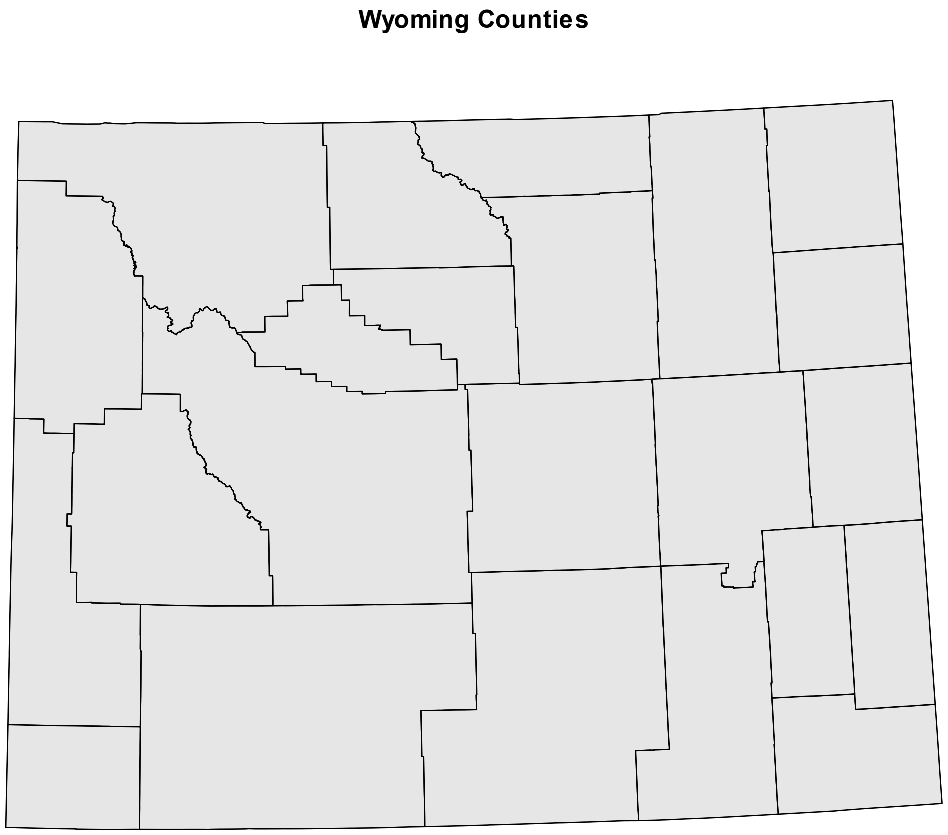


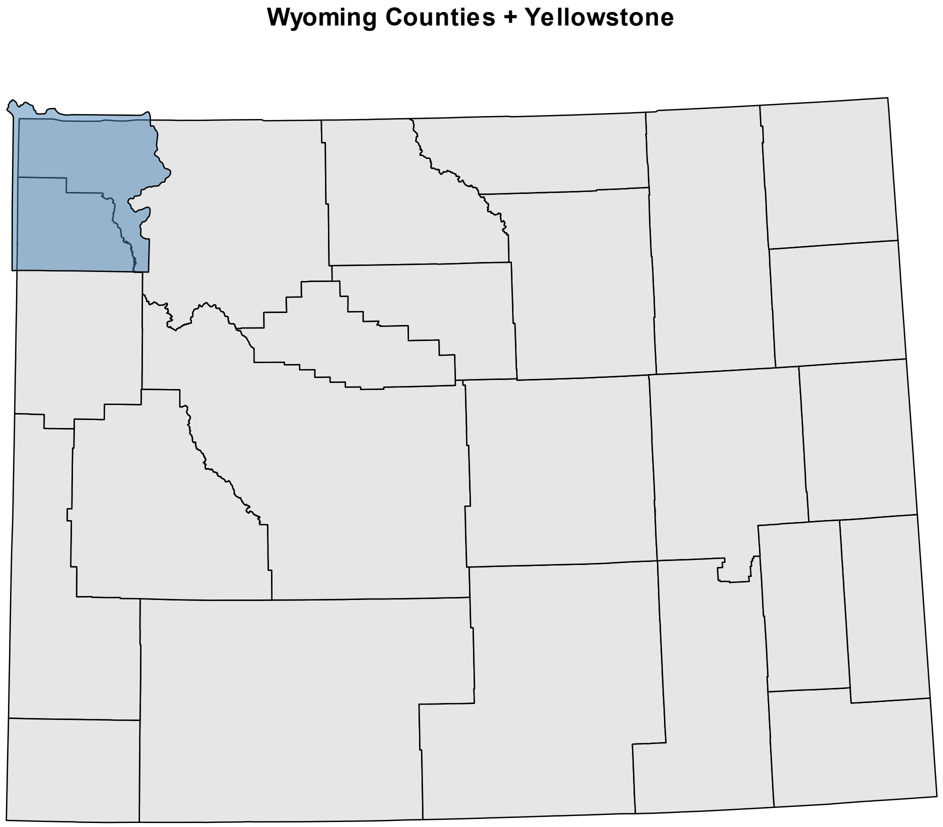
North pole orthographic

Spatial Operations

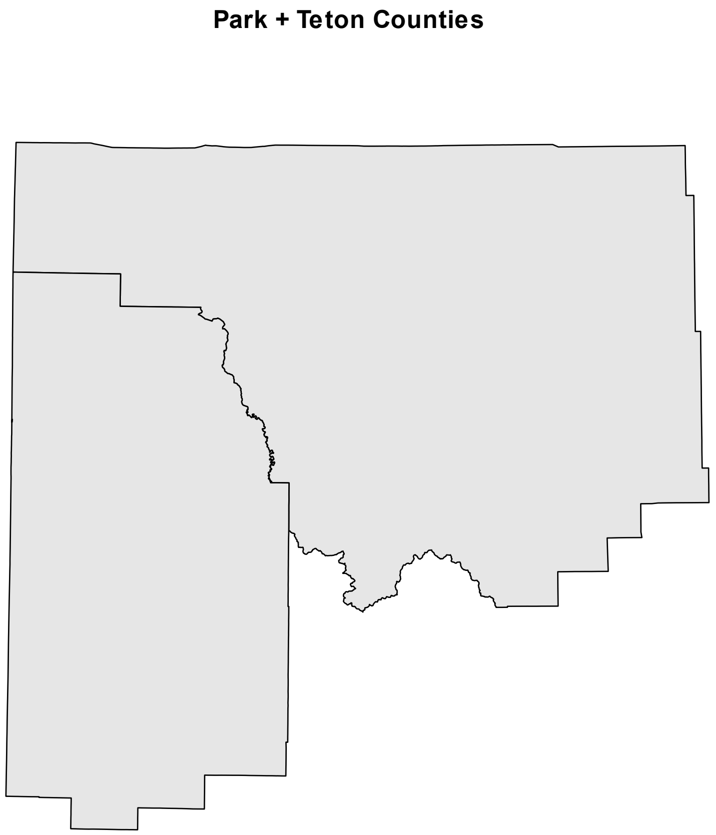
- Union
- Intersection
- Buffer
- Tessellation
- Masks
- Reprojecting

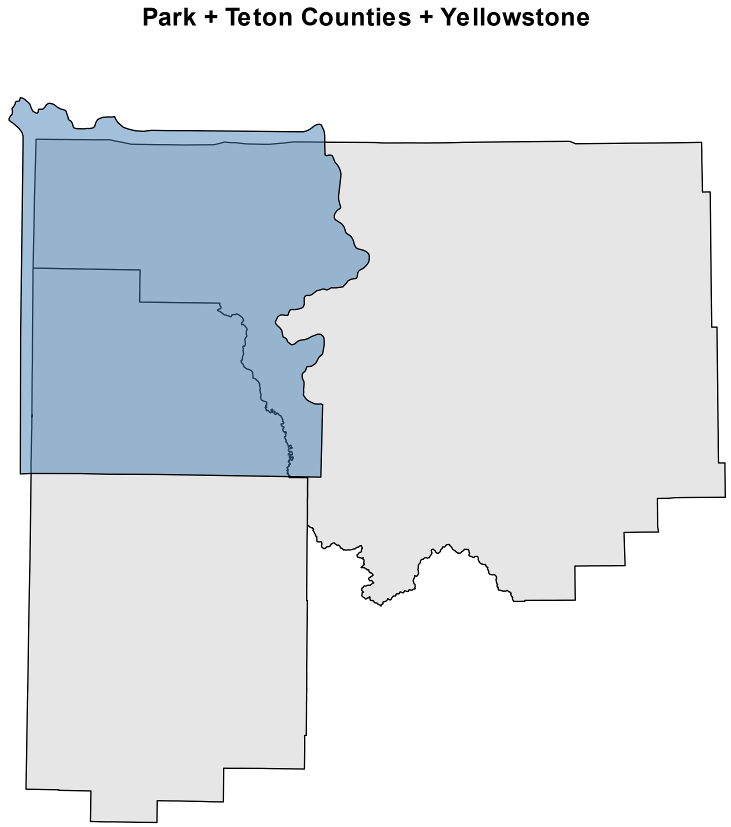
Example Area



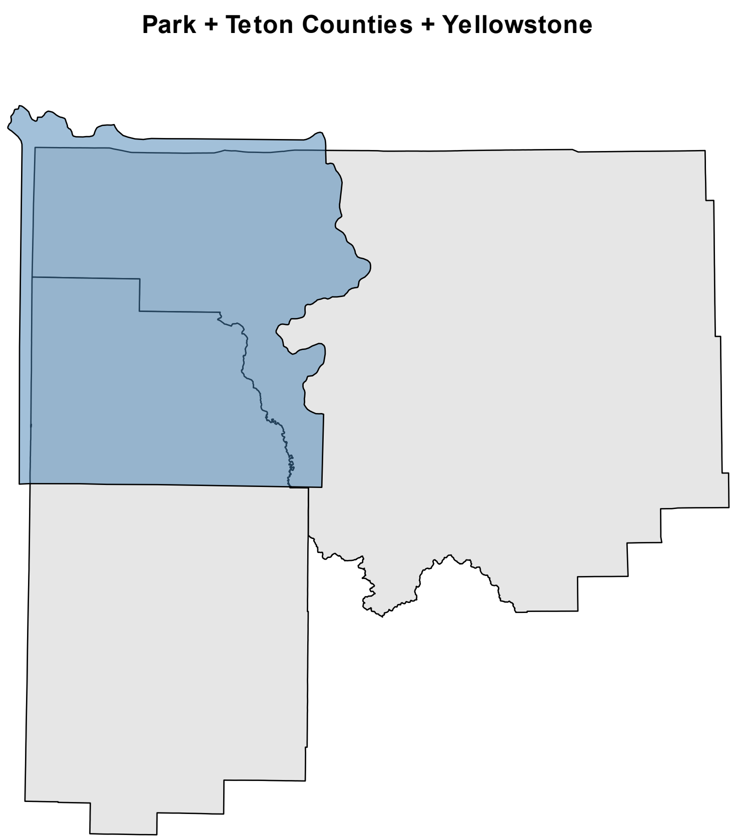


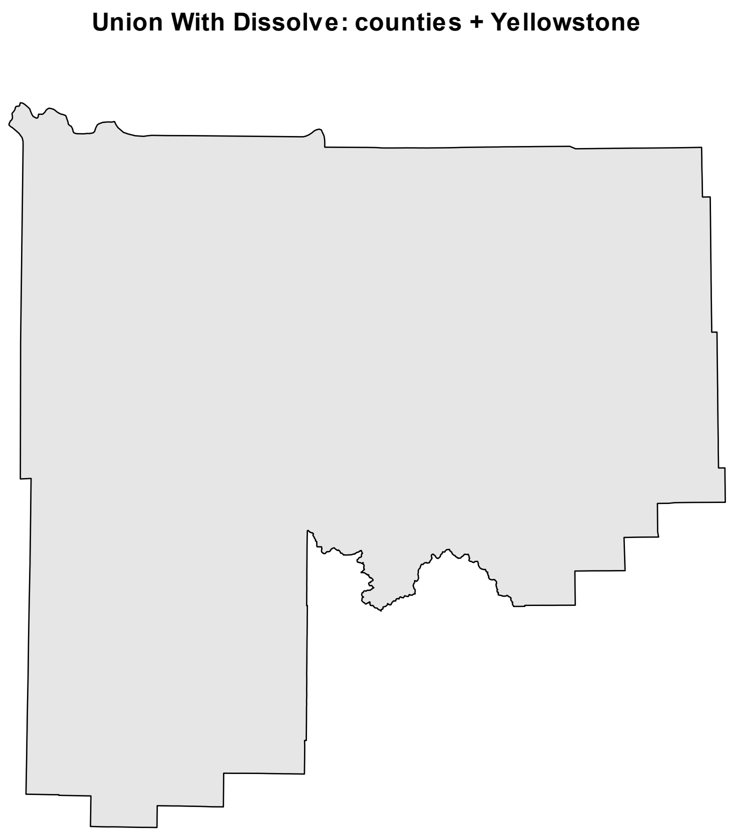
Zoom In: Yellowstone + Counties





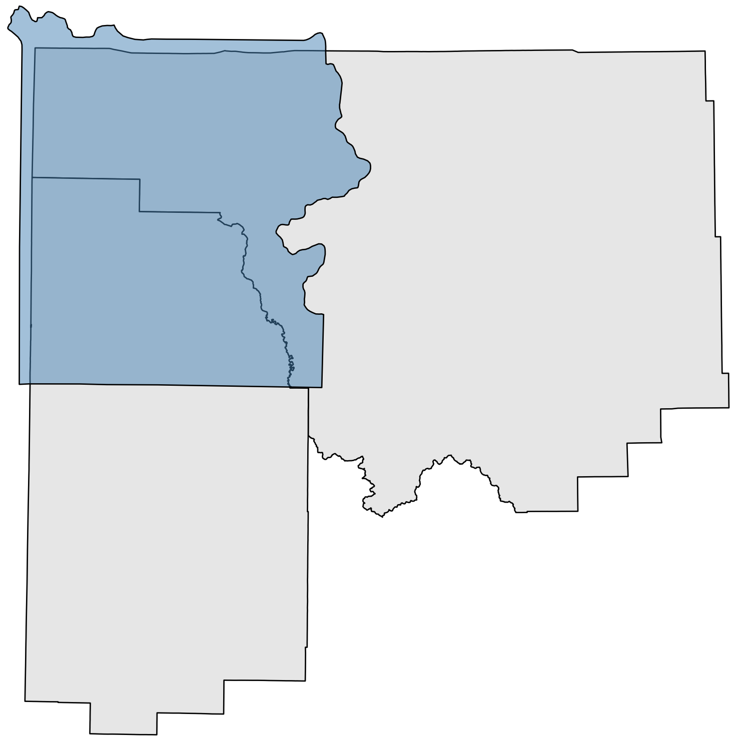
Union + Dissolve



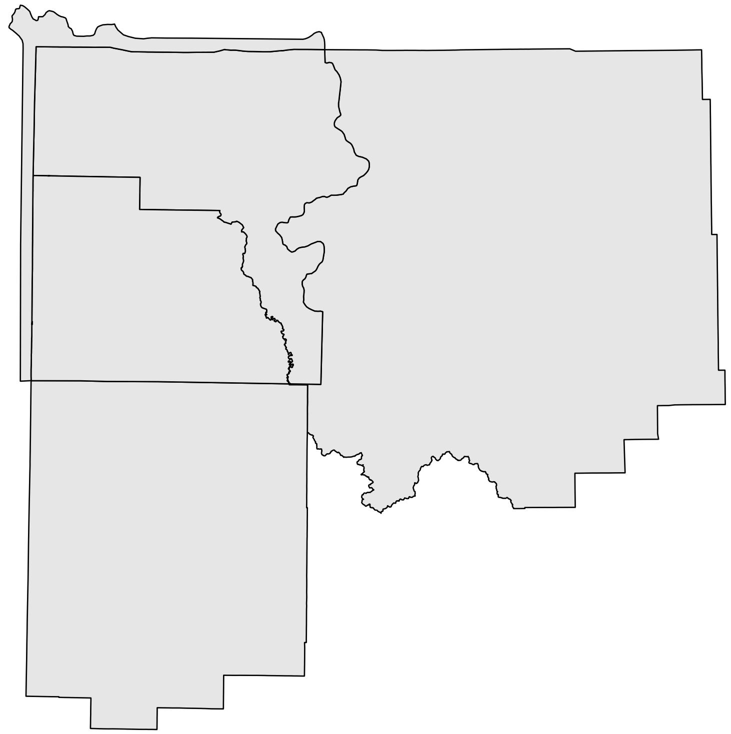


Union

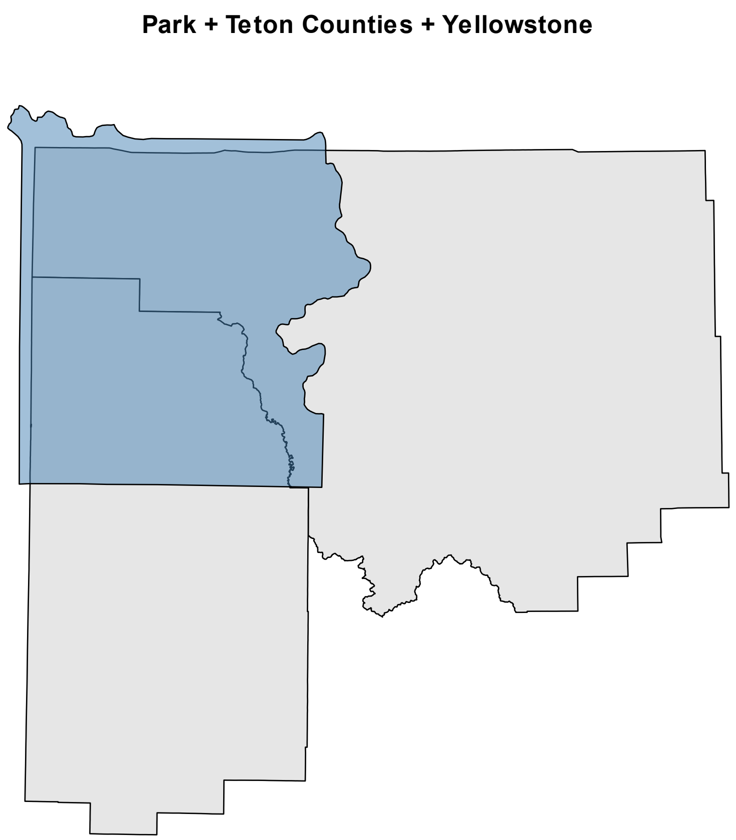
Park + Teton Counties + Yellowstone

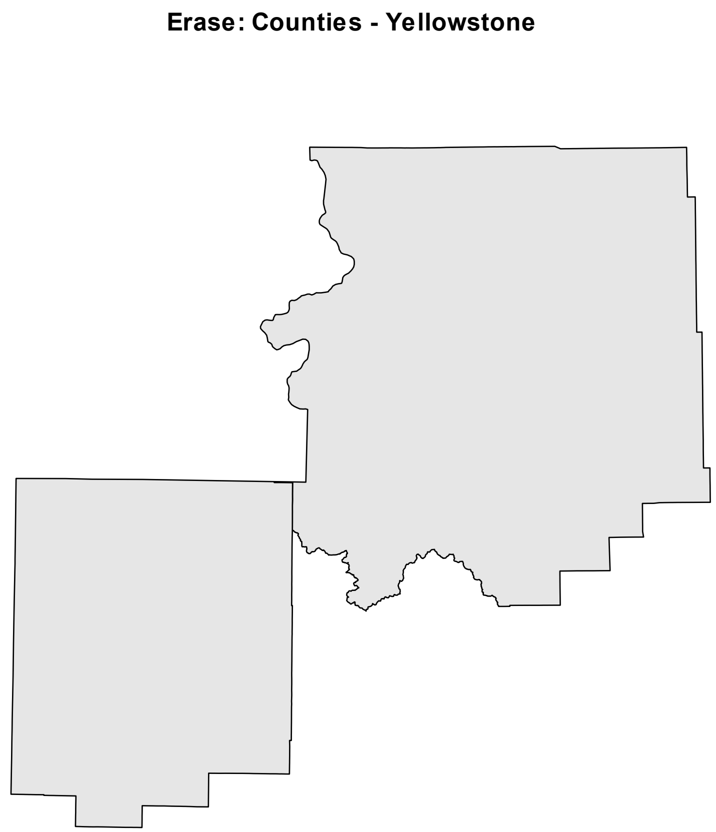


Union Without Dissolve: counties + Yellowstone

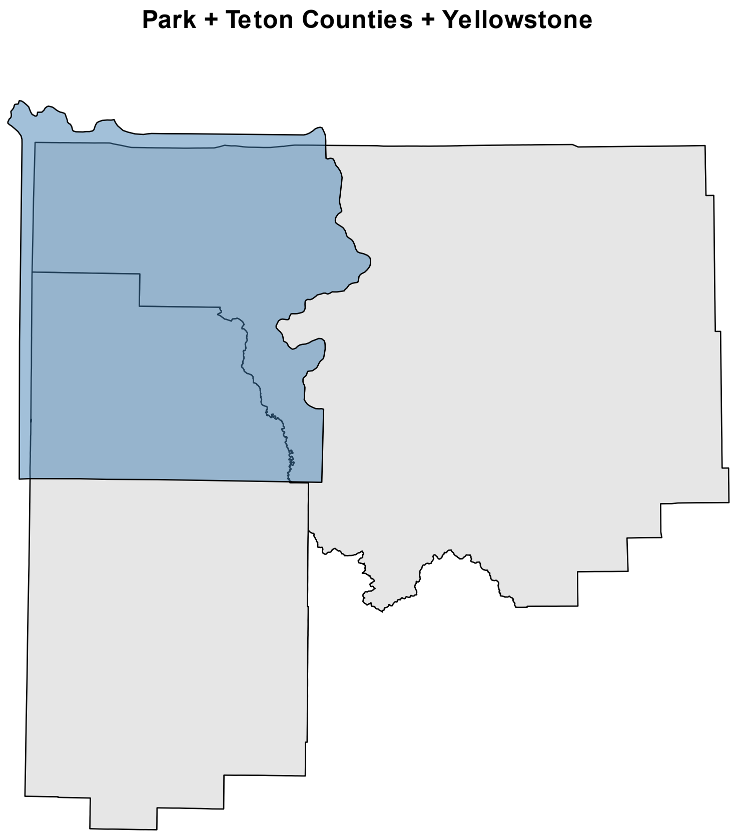


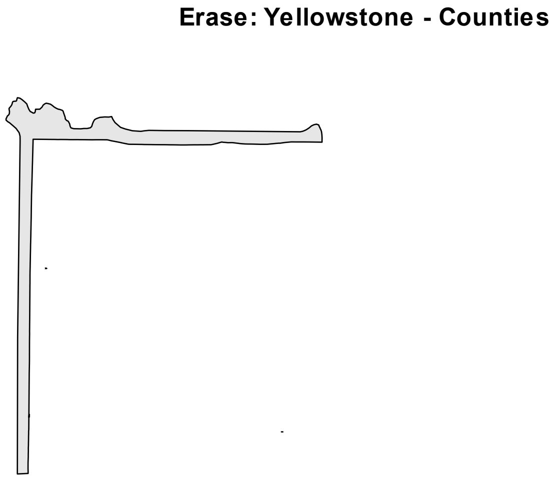
Erase 1


Park + Teton Counties + Yellowstone

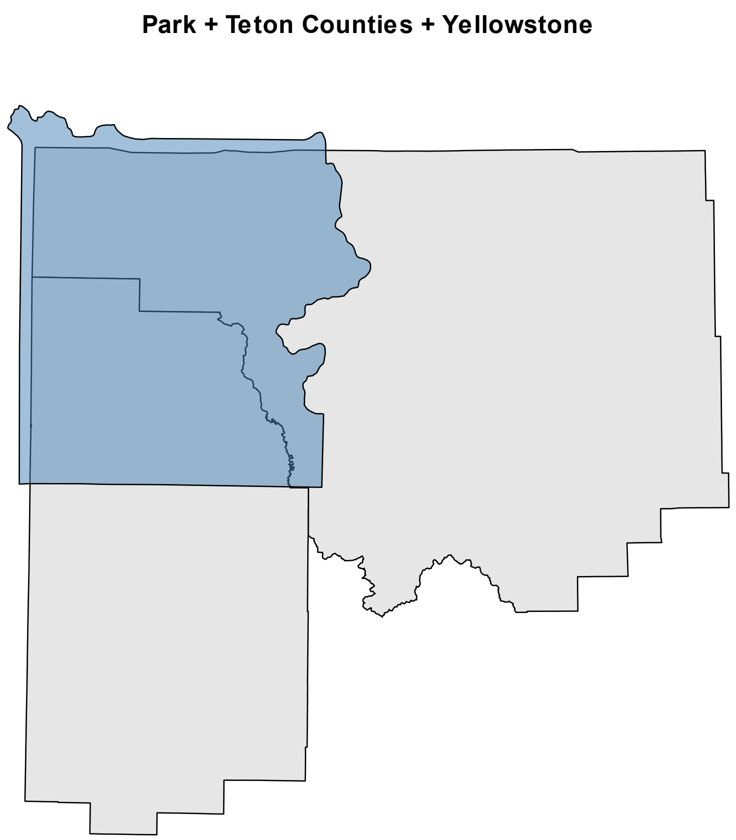

Erase: Counties - Yellowstone

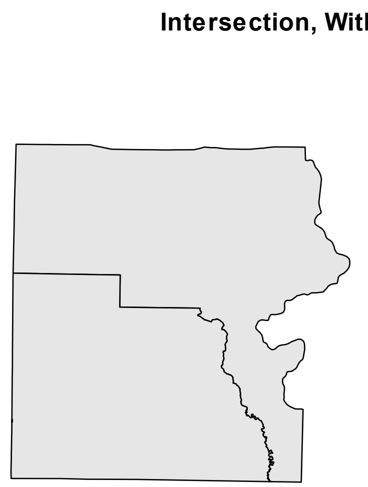
Erase 2


Park + Teton Counties + Yellowstone


Erase: Yellowstone - Counties

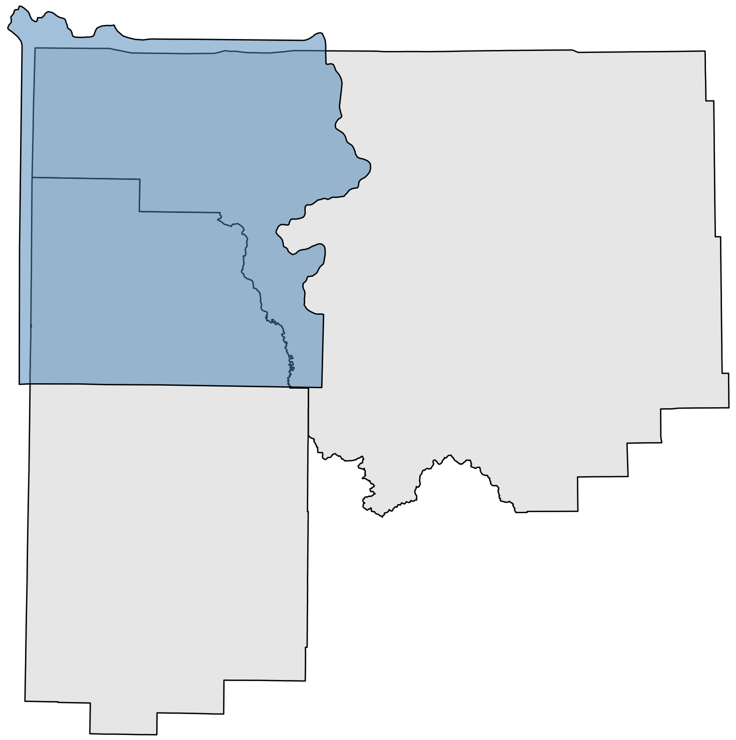
Intersection



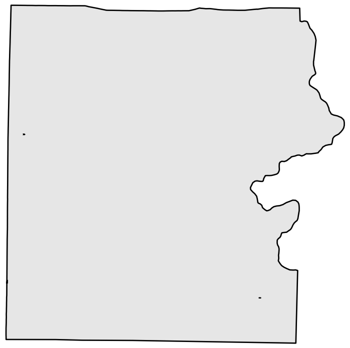


Intersection

Park + Teton Counties + Yellowstone



Intersection, With Dissolve



- Amaya-Gómez, Rafael, Emilio Bastidas-Arteaga, Franck Schoefs, Felipe Muñoz, and Mauricio Sánchez-Silva. 2020. “A Condition-Based Dynamic Segmentation of Large Systems Using a Changepoints Algorithm: A Corroding Pipeline Case.” *Structural Safety* 84 (May): 101912. https://doi.org/10.1016/j.strusafe.2019.101912.