Stockholm Doctoral Course Program in Economics

Topics in Applied Microeconometrics: Using GIS

Lecture 1

Introduction

Masayuki Kudamatsu

IIES, Stockholm University

30 August, 2013

Outline

1. Why GIS in economics?
2. Examples of economics research using satellite images and old maps
3. Spatial data types
4. Coordinate systems

1. Why GIS in economics?

- Expand the set of feasible research questions
  - Use satellite images / old maps (this lecture)
  - Merge different data spatially (L2)

1. Why GIS in economics?

- Expand the set of feasible research questions
  - Use satellite images / old maps (this lecture)
  - Merge different data spatially (L2)
- Come up with more credible identification strategy
  - More covariates to mitigate omitted variable bias (L2, L3, L5)
  - Instruments (L4, L6)
  - RD-design (L7)
  - Estimate the spillover effect on the control group in RCTs (L3)

2. Examples of economics research using satellite images and old maps

- Satellite images
  - Deforestation
  - Nighttime light
  - Wind speed in ocean
- Old maps
  - Roads
  - Ethnic group boundaries

Satellite images: Example 1

Burgess et al. (2012) on deforestation in Indonesia

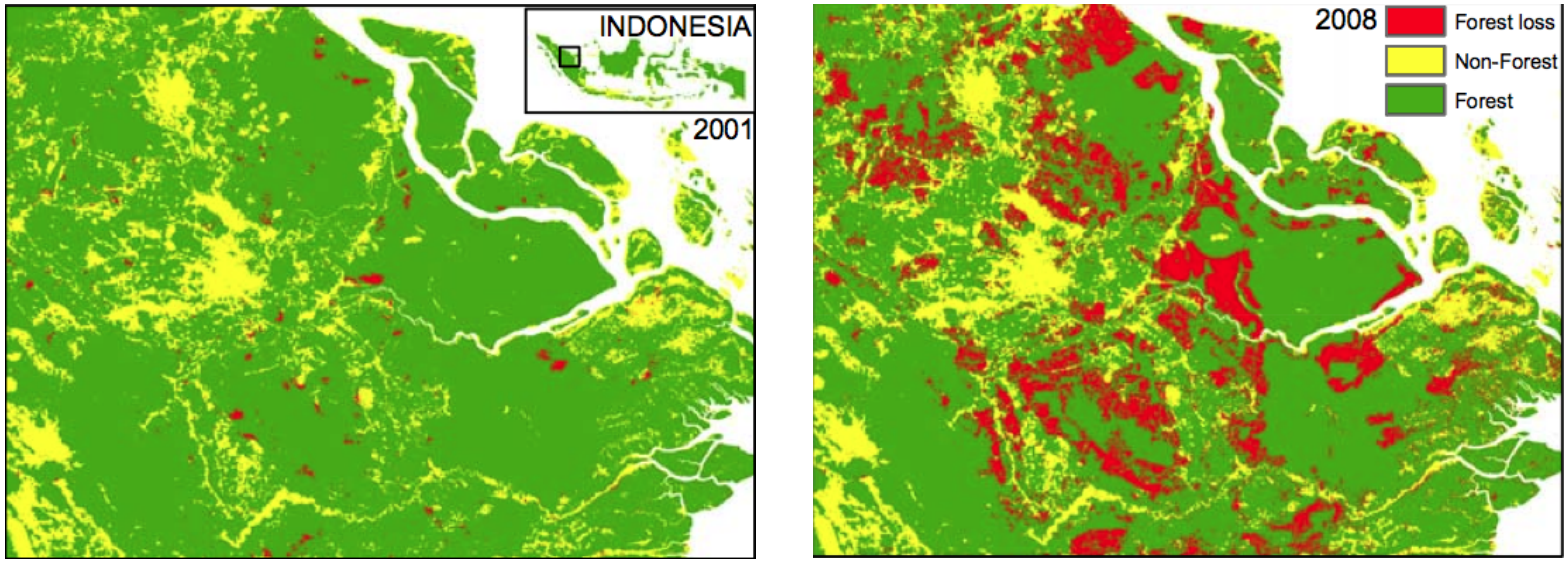
- # of districts in a province ↑

$\Rightarrow$ Each district govt official engages in Cournot competition in selling (illegal) logging permits

$\Rightarrow$ Deforestation in the province ↑

- Cannot rely on official stats of logging

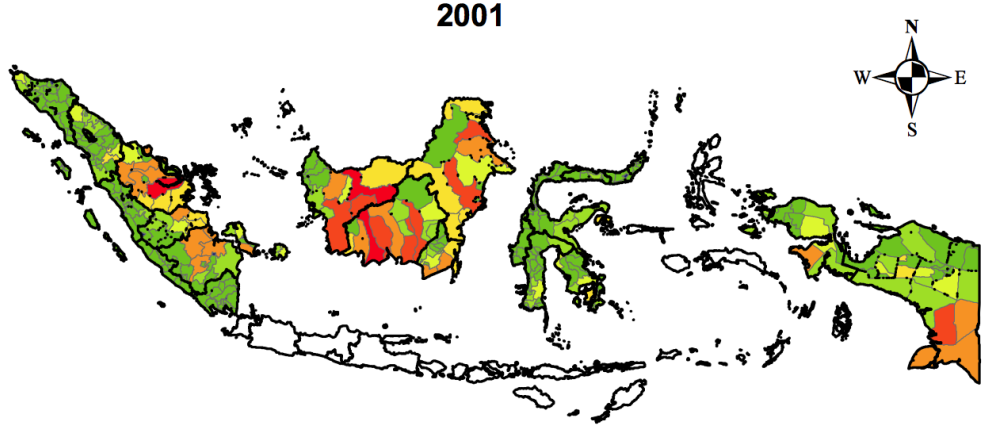
$\Rightarrow$ Use satellite image
  - 250m x 250m pixels images of satellite
  - Measures electromagnetic radiation strength in 36 bands of spectrum
  - Develop algorithm to convert radiation patterns to forest coverage



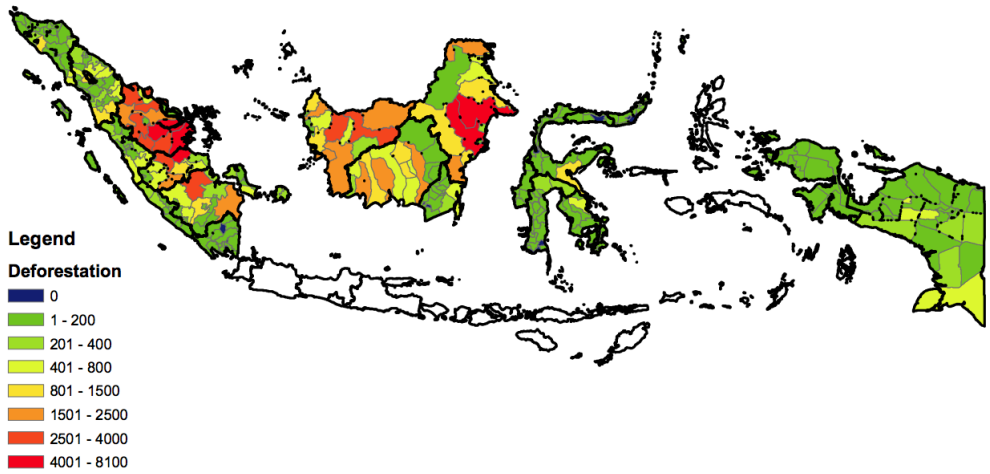
This page: From Figure I of Burgess et al. (2012)

Next page: From Figure II of Burgess et al. (2012)

2001

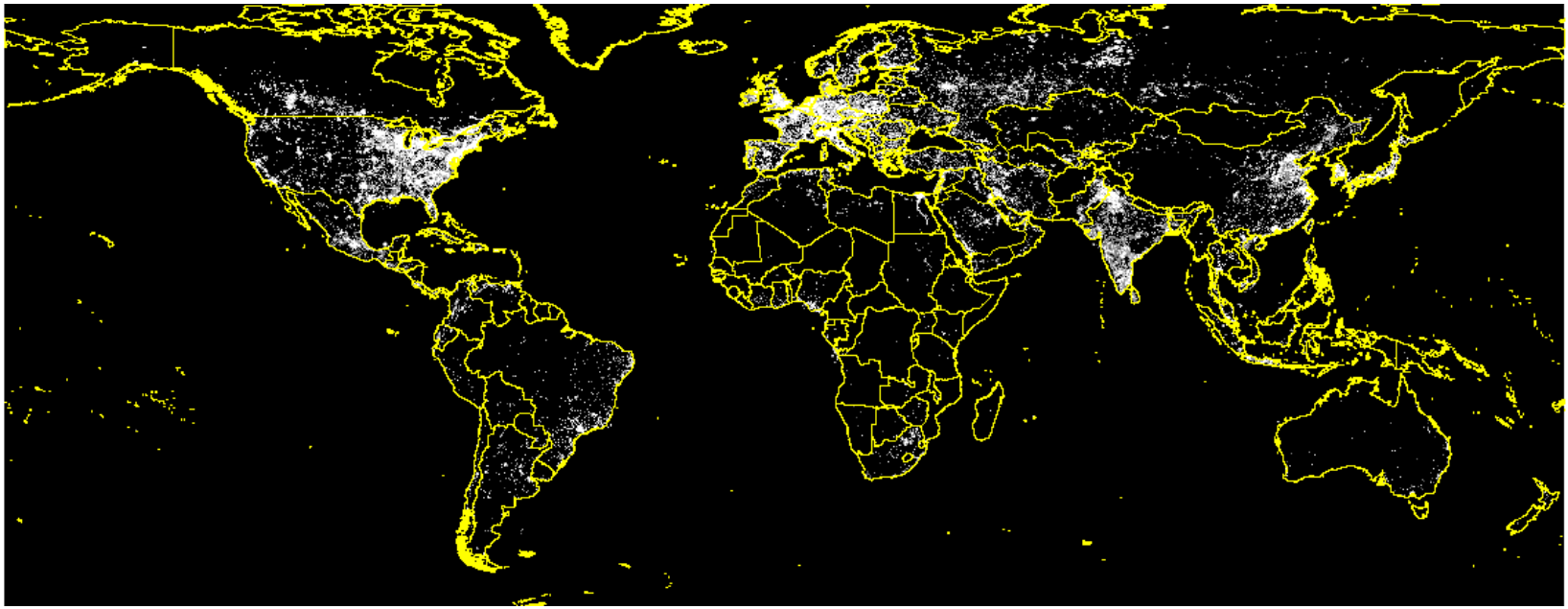

N
W
E
S

2008


Legend
Deforestation
0
1 - 200
201 - 400
401 - 800
801 - 1500
1501 - 2500
2501 - 4000
4001 - 8100

Satellite images: Example 2

Nighttime lights (National Geophysical Data Center 2010) as a measure of development



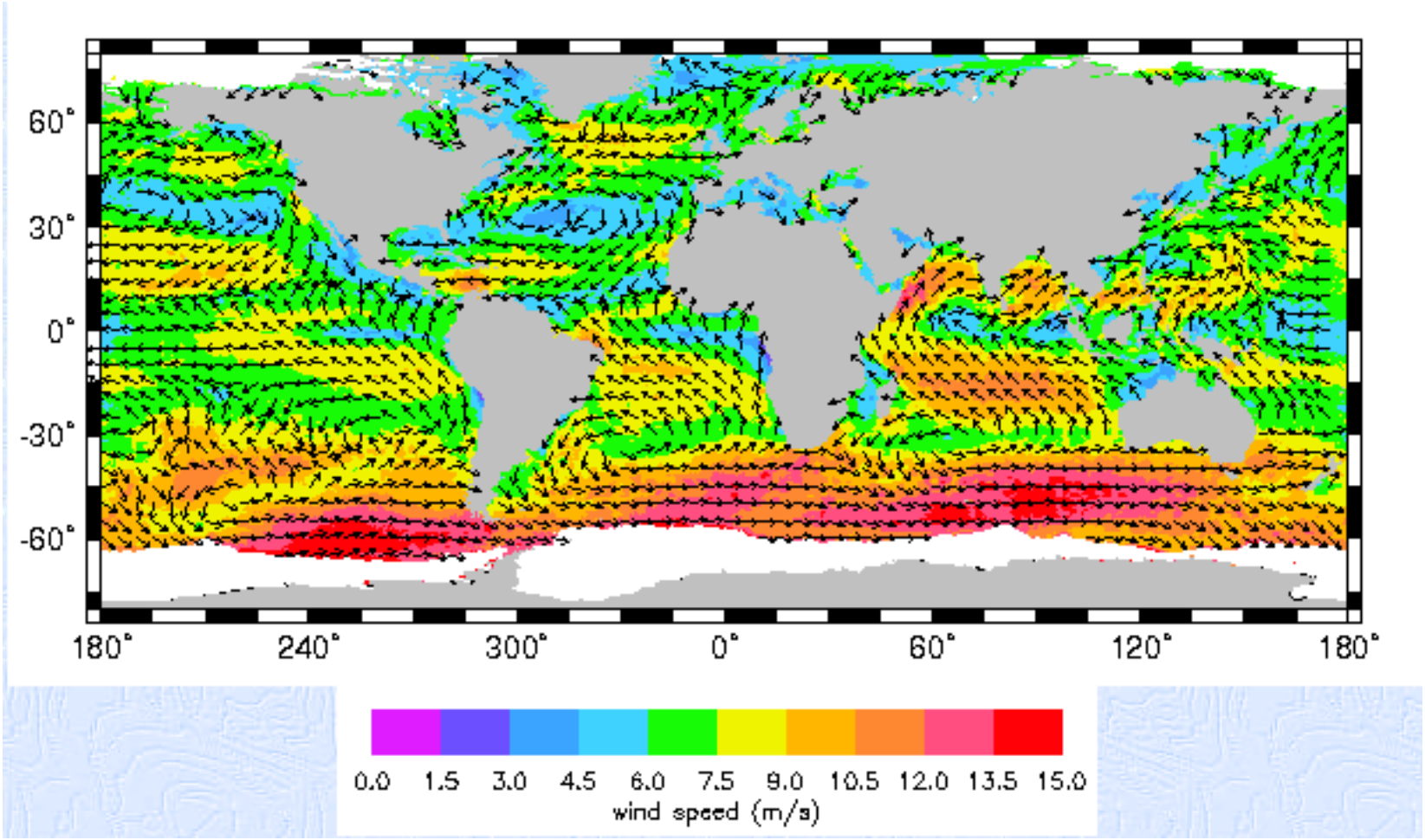
- Henderson et al (2012): estimate real GDP growth
- Michalopoulos & Papaioannou (2012): estimate impact of national institutions on development of African ethnic groups split by the national boundary, by RD design

Satellite images: Example 3

Feyrer & Sacerdote (2009) on wind speed & direction in the oceans

- Estimate the impact of European colonial rules on per capita GDP today for islands
- Wind patterns as instruments for the length of colonization
- Use CERSAT to create instruments

August 2001 by QuikSCAT



www.ifremer.fr/cersat/en/data/tools/browse.htm

Scanned old maps: Example 1

Burgess et al. (2013) on roads in Kenya

- Digitize Michelin maps for Kenya since 1961

Next page: taken from

siteresources.worldbank.org/DEC/Resources/84797-

1257266550602/JedwabR.pdf

Michelin map in 1961

Digitization and Standardization in GIS

Scanned old maps: Example 1

Burgess et al. (2011) on roads in Kenya

- Digitize Michelin maps for Kenya since 1961
- Then track road network expansion over time to see if the president's ethnic group sees more roads built than other groups

Scanned old maps: Example 2

Ethnic group boundaries in Africa

- Drawn by anthropologist George Peter Murdock in his 1959 book
- Used by Nunn (2008), Nunn & Wantchekon (2011) (cf. Lecture 4)
- Used by Michalopoulos & Papaioannou (2012) for RD design

Taken from Figure II in Nunn (2008)

Practical advice

- Very costly (time & money) to create your own data from satellite images
- Digitizing printed maps is also time-consuming but feasible with patience
  - See “From Non-Spatial Data to Spatial Data: Geocoding & Georeferencing in ArcGIS” by Yale Map Collection
- This course at least helps you use these datasets

Prepare for the rest of the lecture

- Download the zipped folder "Lecture1" from T:/economics to C:/TEMP.
- Unzip it and you'll see:
  - 10m-rivers-lake-centerlines.zip
  - F162008.v4.tar
  - g2009_1990_2.zip
  - g2009_2008_0.zip
  - geo_cepii.xls
  - gl_gpwfe_pdens_05_ascii_half.zip

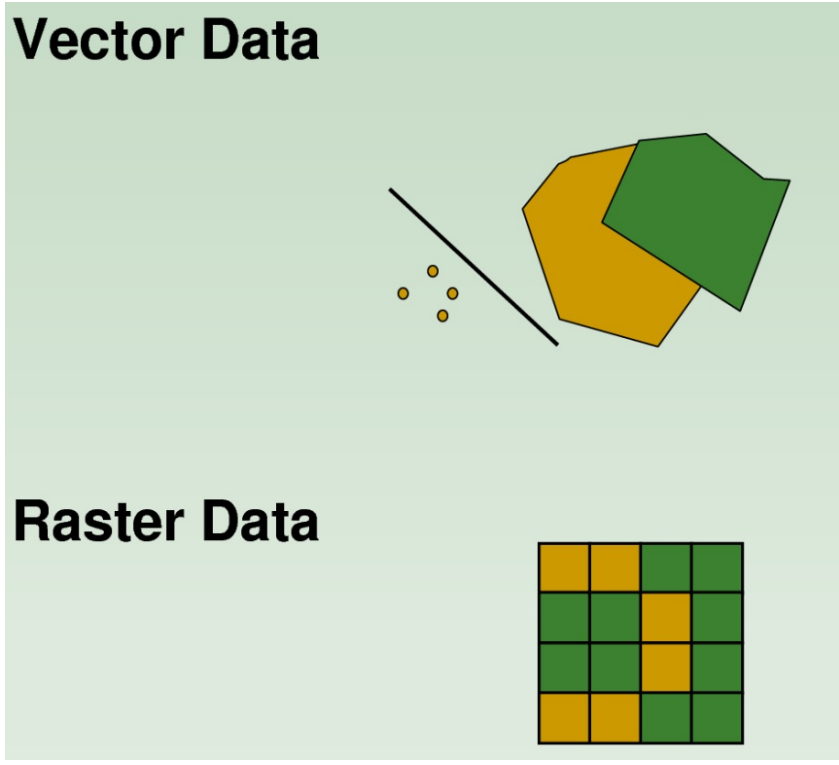
- Uncompress all the .zip files
  - Windows Explorer may hide “.zip” from file names.
- Also uncompress F162008.v4.tar

  This creates three .gz files and three .tfw files.
- Then uncompress .gz files to the same folder as .tfw files are located.
  Otherwise, you cannot read these .tif files in ArcMap. For what .tfw files are, visit:

  gis.ess.washington.edu/data/raster/drg/tfw.html
- Also launch ArcMap 10 (it takes time).

3. Spatial data types

- Spatial data comes in two different types: Vector & Raster



- The way to edit data differs a lot between vector and raster.
- Below we also learn how to read & browse spatial datasets in ArcGIS while learning these different types of spatial data

3.1 Vector data

- Each spatial unit in vector data is called a feature
- Three types of a feature:
  1. *polygon*
  2. *polyline*
  3. *point*
- A set of features of the same type: a feature class
- File format in ArcGIS: *Shapefile* (.shp)

3.1.1 Polygon features

- Represent geographic zones
- Examples:
  - Countries (L4 & L5)
  - Sub-national districts (L6)
  - Areas in which a certain language is spoken (L5)
  - Lakes, Islands, etc.

Exercise 1: Read polygon data

- Data to be read: Global Administrative Unit Layers (GAUL)
  - National boundaries for 2008
    g2009_2008_0/g2008_0.shp
  - Sub-national boundaries for 1990
    g2009_1990_2/g1990_2.shp
- Drag these files in Catalog Window to Data Frame.
  - If you don't see Catalog Window, click "Windows" in the menu bar.
  - If you don't see the data directory, right-click "Folder Connections" and click "Connect Folder..."

Exercise 1 (cont.)

- Ignore the alert message “Unknown Spatial Reference” for the moment. We come back to this issue later in this lecture.

3.1.2 Polyline features

- Represent networks / routes
- Examples:
  - Roads (L7)
  - Rivers (L6)
  - Coastlines (L4)

Exercise 2: Superimpose polylines on polygons

- Data to be read: Natural Earth's Rivers and Lake Centerlines data

  10m_rivers_lake_centerlines.shp

- Read this data (cf. Exercise 1)
- Uncheck the subnational boundary data in Table of Contents
  - If you don't see Table of Contents, click "Windows" in the menu bar.

$\Rightarrow$ Now rivers are shown on national boundaries.

Exercise 2 (cont.)

- Change the color of rivers to blue.
  - Click the symbol (in this case, colored line) just below the data name in Table of Contents
  - Choose the preferred color.

- If you read the river data first and then the national boundary data, the river data will be hidden below the national boundary data.
- In the Table of Contents window (the one on the left), drag the river data and drop it above the national boundary data.
- Then rivers will show up

3.1.3 Point features

- Represent point location
  - Plots (L3)
  - Schools (L3)
  - Surveyed communities (L2)
  - Cities (Exercise 3 below)
  - Centroid of polygon (L4 & L6)
- Can be created from XY data

3.1.3 Point features (cont.)

- XY data: a table in which
  - Each row: point feature
  - Column 1: longitude (x value)
  - Column 2: latitude (y value)
  - Other columns: attributes of point feature (name, statistics, etc.)

Exercise 3: Read XY data

- Data to be read: CEPII Distances Database (geo_cepii.xls)
- Browse the data in Excel
  - Which columns for longitude and latitude?

3.1.3 Point features (cont.)

- XY data format in ArcGIS: tab delimited text file or a worksheet in an Excel file
  - Comma delimited text files sometimes work, sometimes don’t
- For the Stata data file (.dta), use the "outsheet" command
  - Use the "format" command so that ArcGIS won’t round off longitude & latitude values
  - If longitude and latitude are string, first use the "destring" command to make them numerical

If maximal # of decimal digits in XY data is 10, for example, use the following Stata commands:

```
format lon %15.10f
format lan %14.10f
outsheet lon lan using filename.txt, replace
```

- "In %w.df, w is the total output width, including sign and decimal point, and d is the number of digits to appear to the right of the decimal point." (Stata help for format)
- The longitude ranges from -180 to 180; the latitude from -90 to 90.

3.1.3 Point features (cont.)

- XY data can be created on your own
  - During your survey, use a GPS receiver
  - If location names are available, use an online gazetteer
    - www.fallingrain.com/world
    - geonames.nga.mil

3.1.3 Point features (cont.)

- To create a point feature class from an XY data, we need to use geoprocessing tools.
- We do this by using the Model Builder
  - Model Builder helps us write a Python script (cf. L2)

Exercise 3: Read a XY data (cont.)

- Convert the CEPII Distances Database (geo_cepii.xls) into a point feature class in the shapefile format (.shp).
- Geoprocessing tools to be used:
  - Make XY Event Layer
  - Copy Features

How to use Model Builder (1/3)

- Open a Model Builder window by clicking  in the Standard Toolbar.
- In Search Window, type the name of the geoprocessing tool and search.
  - If you don't see Search Window, click "Windows" in the menu bar.
- Drag the tool from Search Window to Model Builder
- Double-click the tool to set inputs, outputs, options etc.

Make XY Event Layer

- XY Table: geo_cepii.xls/geo_cepii$
  - For an Excel file, double click it to choose a worksheet inside
- X Field: lon
- Y Field: lat
- Spatial Reference: WGS 1984 (we learn what this is in Section 4 below)

- To select WGS 1984, click 
- Then click “Select”.
- Navigate the directory structure as follows:
  Geographic Coordinate Systems
  > World
  > WGS 1984.prj

To understand what needs to be filled in for each item...

- Click “Show Help »” on the bottom right. Then click the item you don’t understand. The help document appears on the right column.

Make XY Event Layer (cont.)

- This tool creates a temporary *layer* file out of XY data.
- But the layer file often doesn't properly work with other tools.
- We also want to save the point feature class in the disk.

$\Rightarrow$ Use the Copy Features tool to make it a shapefile data

Copy Features

- Input Features: the output from the Make XY Event Layer
- Output Feature Class: ***.shp
  - It's a good idea to save all the newly created spatial data in a folder different from the one the original data is saved
  - In our case, let's save outputs in D:/temp/outputs

Copy Features (cont.)

- Ignore other items. Rarely used.
- This tool is also useful if you want to keep the original data intact (cf. Exercise 7 below)
  - $\Leftarrow$ Some geoprocessing tools overwrite the input data...

How to use Model Builder (2/3)

To save the model, do the following:

- Click the save icon
- Navigate to the directory in which you will save the model
- Click the toolbox icon (red box at top-right)
- Create a new toolbox (name it, say, lec1.tbx)
- Then click this toolbox
- Type the file name for the model

A model can only be saved inside the toolbox.

- ArcGIS often crashes. It's always a good idea to save the Model frequently while editing it.

How to use Model Builder (3/3)

- To edit an existing model...
  - Locate the model in Catalogue Window
    - If you cannot see a folder you just created, right-click the parent directory and “Refresh”.
  - Right-click the model
  - Click “Edit” (NOT “Open”)

- Now run the Model.
  - Click the triangle icon at top right.
- Read the output point feature class shapefile.
  - Capital cities (and other major cities) around the world should appear as point features.

3.1 Vector data (cont.)

One more thing about vector data...

- Comes with attributes table in the dBASE format (.dbf)
- Attributes table contains fields (ie. variables) which can take a different value for each feature
- To see in ArcMap, right-click the data in Table of Contents and click “Open Attribute Table...”

- Fields can be used to zoom in to a particular set of features

Exercise 4: How to zoom in

Purpose: Zoom in to India

- Method 1: Selection by Attributes
  - In the menu bar, click “Selection > Select By Attributes”
  - Select the shape file (“g2008_0”)
  - Double-click the name of a field based on which the selection is conducted (“ADM0_NAME” in this case)
  - Click “=”
  - Click “Get Unique Value”
  - Double-click “India”
  - In the bottom, now you should see “ADM0_NAME” = ‘India’

Exercise 4: How to zoom in (cont.)

- Method 1 (cont.)
  - Click OK. Now India is selected on the map
  - Right-click the shapefile (“g2008_0”)
  - Click “Selection > Zoom To Selected Features”
- To go back to the whole world, click the Full Extent icon  in the Tools Toolbar.

Now clear the selection by

- In the menu bar, click “Selection > Clear Selected Features”

If you forget doing this and conduct geoprocessing, the outputs will be only for these selected features.

Exercise 4: How to zoom in (cont.)

- Method 2: use the Zoom-In tool
  - Click the Zoom-in Tool icon  in the Tools Toolbar
  - Drag a box that encloses India
- Play around with other tools in the Tools Toolbar

3.2 Raster data

- Divides the earth surface into many “square” cells (or pixels)
- Each cell contains one value
- Often created from satellite images

3.2 Raster data (cont.)

- Examples:
  - Elevation (L6)
  - Suitability for agriculture (L5)
  - Population density (L5)
  - Forest coverage (Burgess et al. 2012)
  - Nighttime light (Michalopoulos & Papaioanno 2012)
- Can be used to assign a new variable to features in the vector data (L5)

3.2 Raster data (cont.)

- File format in ArcGIS: ESRI grid
  - Some other formats (e.g. TIFF) can also be read in ArcGIS, but it may come with some errors
  - Geoprocessing raster datasets is quicker if in ESRI grid.

  $\Rightarrow$ Better to convert to ESRI grid format
- If it comes in the ASCII format (.asc), you need to convert

File name for ESRI grid format

- A file name for ESRI grid format raster cannot be longer than 13 characters.
- No extension needed

Exercise 5: Read tiff raster data

- Data to be read: DMSP-OLS nighttime light

  “F162008.v4b_web.stable_lights.avg_vis.tif”

  - This raster is in the TIFF format. ArcGIS can read it, but the missing value may becomes 255.
  - Use the Copy Raster tool (or the Raster to Other Format tool if many to be converted) to convert it into ESRI grid

Exercise 6: Read ascii raster data

- Data to be read: Gridded World of the Population
  (gl_gpwfe_pdens_05_ascii_half/glds05ag30.asc)
  - Use the Ascii to Raster tool
    - ⇐ This raster data is in the ASCII format
  - Choose “FLOAT” for data type
    - ⇐ This raster data contains population density, which can be decimal.

- When you read the converted population raster data, you may get an error message saying, “Spatial reference is undefined.” We’ll come back to this issue shortly.

4. Coordinate systems

- Earth is a sphere (approximately)
- But we need to represent its surface on a plane
- Various ways to do so. Each way corresponds to a coordinate system.

4. Coordinate systems (cont.)

Why important?

- Properly merge different spatial datasets (cf. the iOS 6 map did this wrong)
- Properly calculate distances and areas

Two types: Geographic / Projected

4.1 Geographic coordinate systems

- Each location is coded by degrees
  e.g. Stockholm
  Latitude: 59 degrees 20 minutes N
  Longitude: 18 degrees 3 minutes E
- Not suitable for calculating distances/areas
  - 1 degree of latitude: 110.6km at equator, 111.7km at the poles
  - 1 degree of longitude: 111.3km at equator, 55.8km at 60 degrees N/S

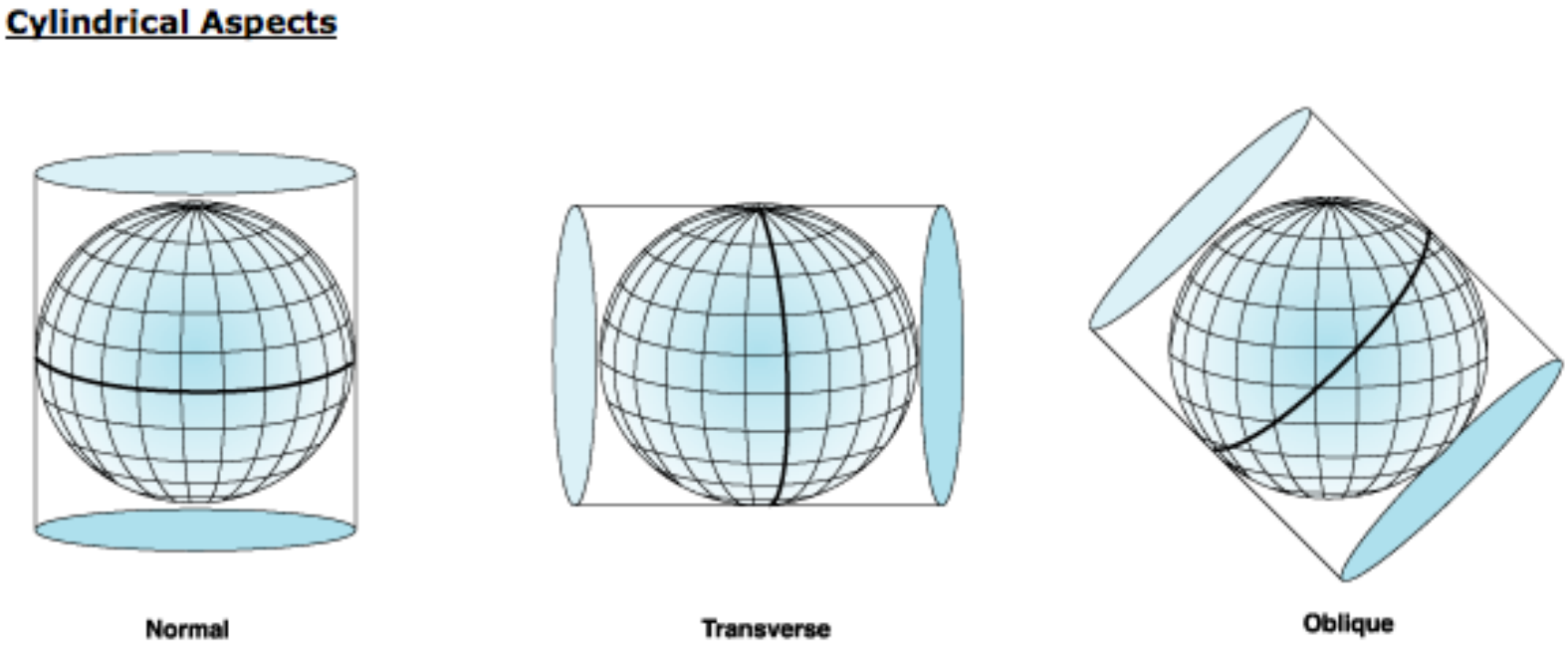
4.1 Geographic coordinate systems (cont.)

- For calculating long distances between two point locations, however, it is useful
  - Formula for great circle distance uses degrees in latitude/longitude
  - Implemented by Stata ado GLOBDIST
- Many spatial datasets adopt *WGS 1984*, the most popular geographic coordinate system

4.2 Projected coordinate systems

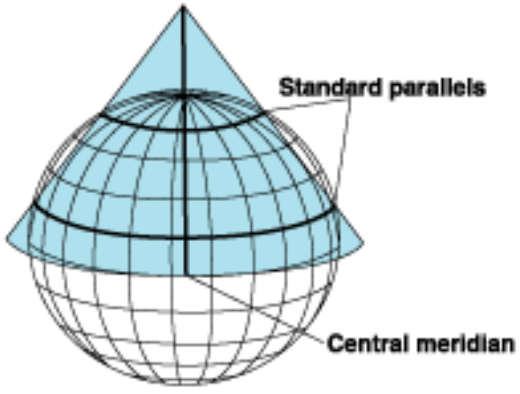
- Earth surface is projected on a flat plane by the “light” from the center of the earth
- Each location: coded in meters from a certain origin on the projected surface
- Various projection methods: cylindrical, conic, arizuminal

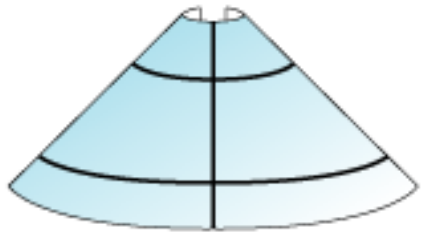
Cylindrical projection

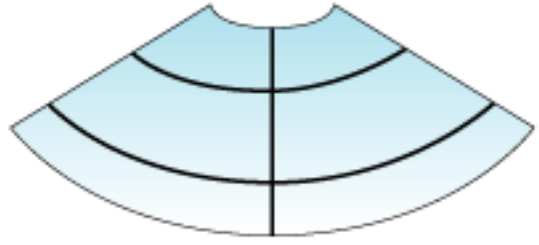


Conic projection

Conic (secant)

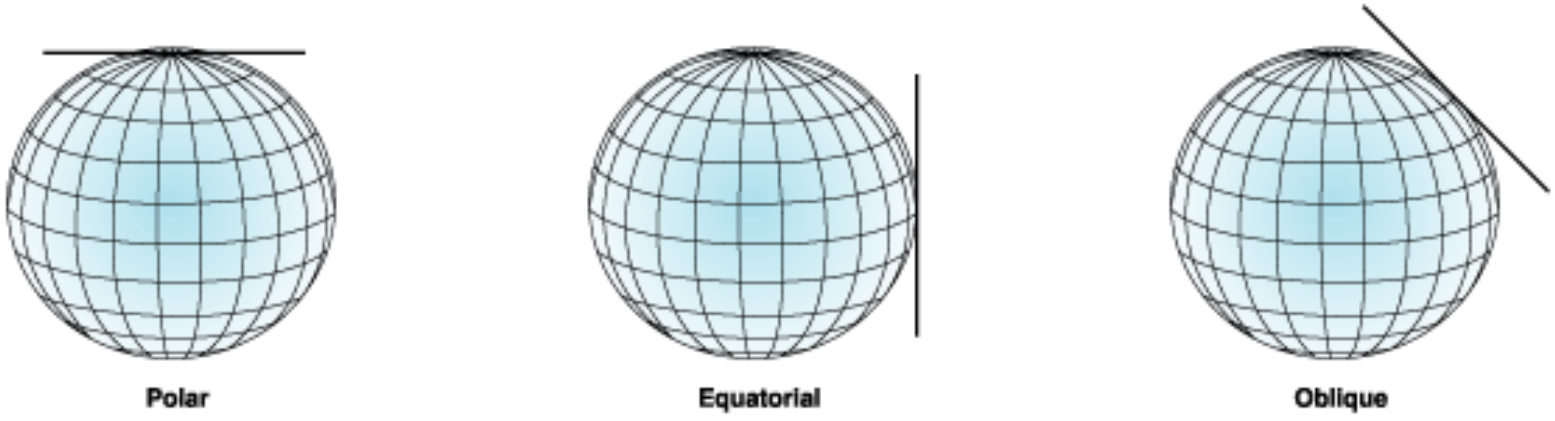






Planar projection

Planar Aspects



4.3 Which coordinate system?

Depends on what you want to calculate.

(There's no definite answer. Here's my suggestion)

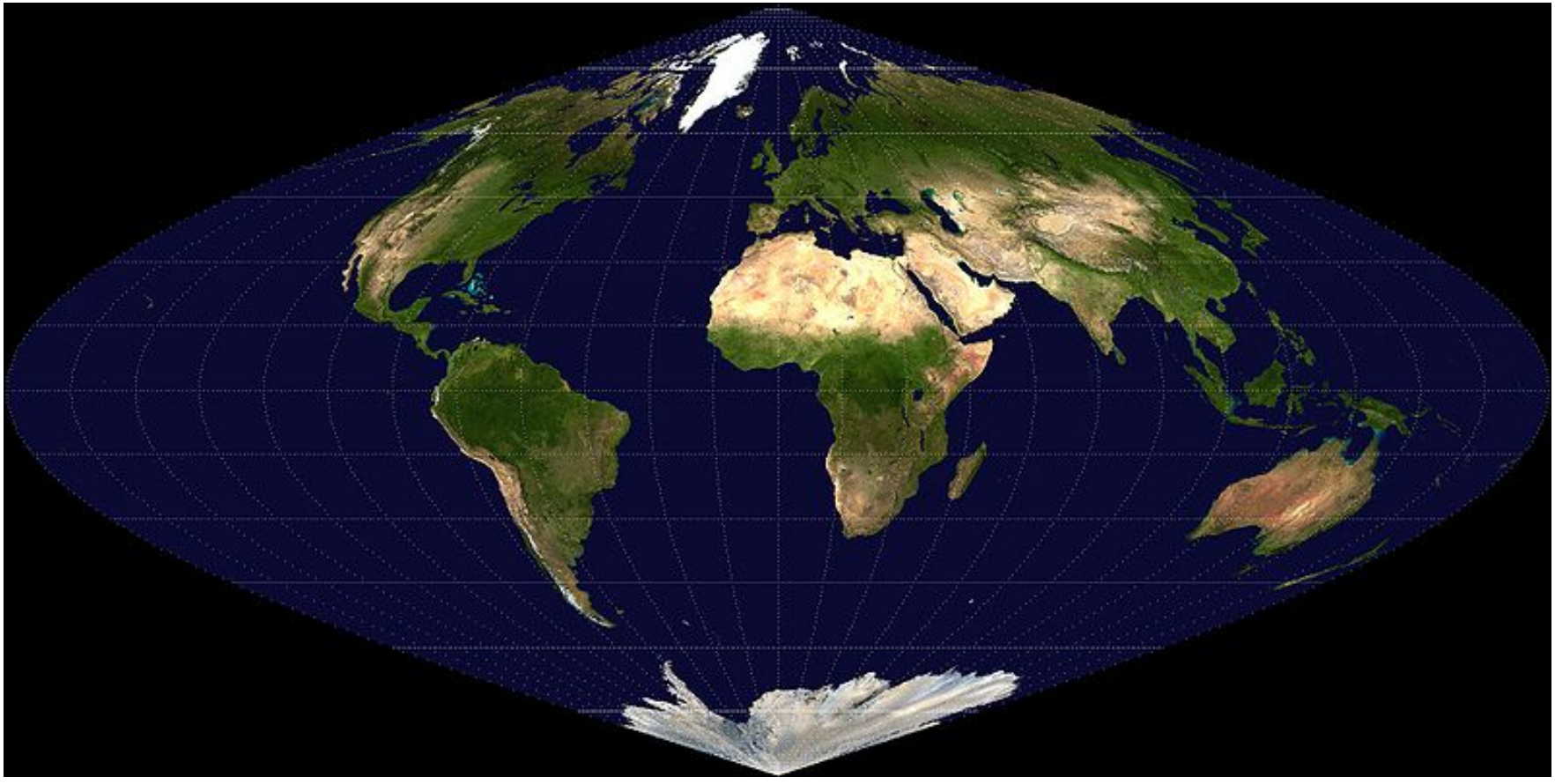
- WGS 1984 (or any other geographic coordinate system)
  - distance between two locations in large study areas (L4)

4.3 Which coordinate system? (cont.)

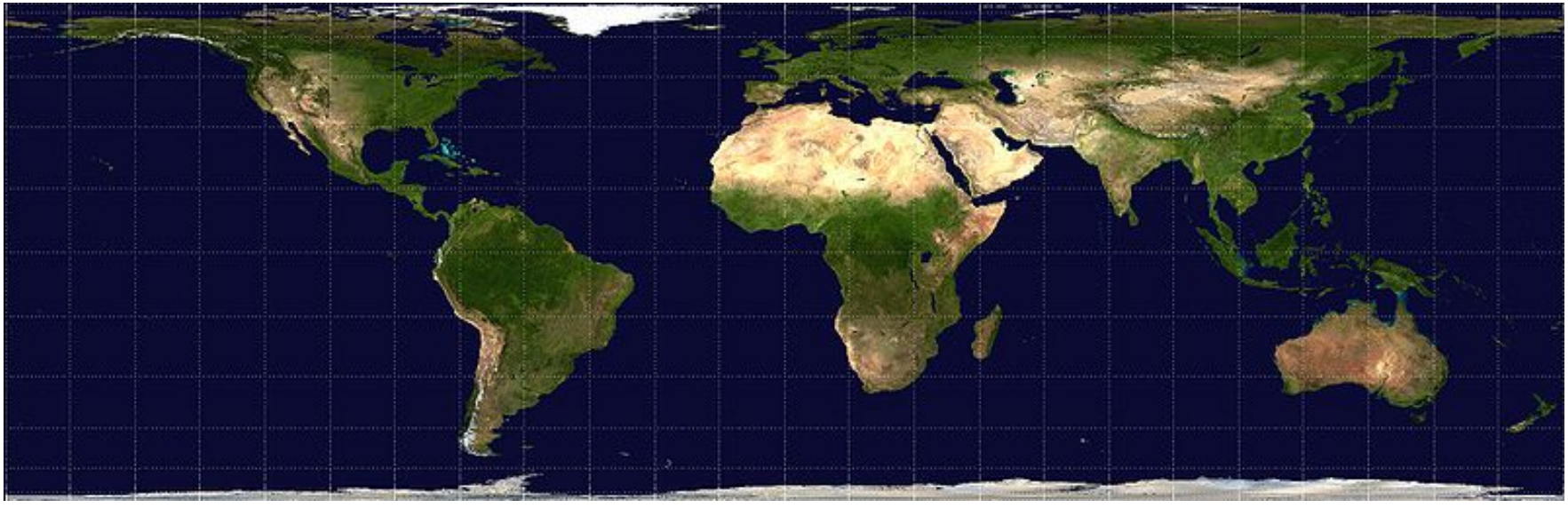
- UTM (L3, L6)
  - length of polyline features
  - distance / area in small regions (within 6 degrees in longitude)
- Any of the “equal area” projections such as Sinusoidal, Lambert Cylindrical, Albers Conic (L4)
  - area of large regions

  cf. Which equal area projection to be used only changes the way the world is shown on a map; all give you the correct area calculation.

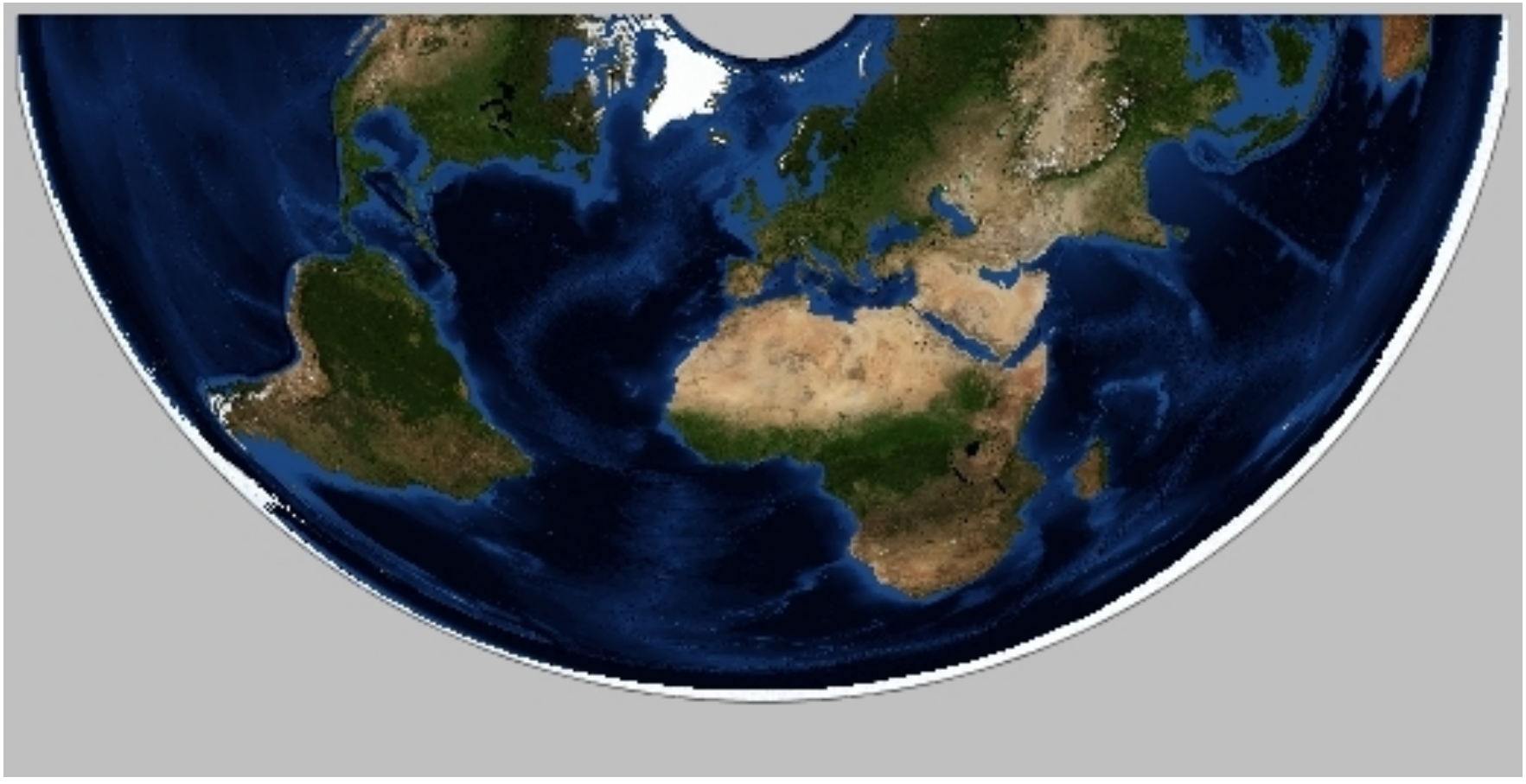
Sinusoidal projection



Lambert Cylindrical Equal Area projection



Alberts Equal Area Conic projection



How to assign coordinate system to spatial data in ArcGIS?

- If undefined, use the Define Projection tool
- If defined and vector data, use the Project tool
- If defined and raster data, use the Project Raster tool (L7)

- Remember when you read the GAUL national / subnational district boundary data and get an alert message “Unknown Spatial Reference”?
- This means the coordinate system is not assigned to these data.

- We can check whether (or what) coordinate system is assigned to each data by:
  - Right-click the data
  - Click “Properties...”
  - Click the “Source” tab
  - Scroll down to “Spatial Reference”
  - If it’s not assigned yet, it says “<Undefined>”

Exercise 7: Assign the projection to GAUL national boundary data

- Spatial data usually comes with a meta data that specify the coordinate system used when the data is created
- For GAUL, the meta data is on the web (Google GAUL FAO).
- The meta data says it's "WGS 1984"

Exercise 7 (cont.)

- Define Projection tool overwrites the file.
- To keep the original file (which is useful for programming), first use Copy Features to make a copy.
- Then use this copy as an input to Define Projection

Exercise 7 (cont.)

- To select WGS 1984, click 
- Then click “Select”. (For ArcMap 10.1, skip this step.)
- Navigate the directory structure as follows:
  Geographic Coordinate Systems
  > World
  > WGS 1984.prj

- It turns out that the population data we converted in exercise 6 above also has no coordinate system assigned.
- See the readme file that comes with the original ascii file. It says that the coordinate system is WGS 1984.
- Follow the same step as in Exercise 7 to assign the coordinate system to the population data.

Exercise 8: Project the GAUL to Sinusoidal

- Now use the Project tool
- Sinusoidal projection is found at Projected Coordinate Systems
  > World
  > Sinusoidal(world).prj

A thing to remember...

- You cannot overlay data with different coordinate systems
- ArcMap cannot display all the data if they are in different coordinate systems
- To read a data with a different coordinate system, open a new map document.

What is a map document?

- It saves the way you overlay, color-code, & symbolize different spatial datasets
- The file extension is “.mxd”
- This file DOES NOT contain spatial data. It just has links to them.
- Make sure setting the relative path to refer to each data

Exercise 9: Save a map document

- Set relative paths as the default
  - In the menu bar, click Customize > ArcMap Options
  - Check “Make relative paths the defaults for new map documents
- Click the save icon in the Standard Toolbar
  - This icon is NOT for saving the data
- Choose the location in which you save the map document
  - best to save in the parent folder for spatial data files

- Now open a new map document.
- Then read the GAUL data in the Sinusoidal projection.

One more thing to remember...

- Browse the spatial data in Windows Explore
- You see many, many files.
  - Shapefile: .shp + .sbx + .sbn + .dbf + .prj
  - Raster: the info folder + the data folder + .aux + .rdd + .prj
  - .dbf: the attributes table (dBASE format)
  - .prj: projection

- Never use Windows Explore to browse, move, copy, or delete spatial data
- Always use the Catalogue Window in ArcMap.

5. What we've learned on ArcGIS

- Convert data format (so data can be read in ArcGIS)
  - XY data
  - Tiff raster
  - Ascii raster
- Assign/change the coordinate system

Do you remember which geoprocessing tools you used for each of these tasks?

Useful references

- The Yale GIS Workshops
- Dell 2009 (econ-www.mit.edu/files/3717)

Burgess, Robin, Matthew Hansen, Benjamin Olken, Peter Potapov, and Stefanie Sieber. 2012. “The Political Economy of Deforestation in the Tropics.” Quarterly Journal of Economics, 127, 1707–1754 (2012).

Robin Burgess, Rémi Jedwab, Edward Miguel, Ameet Morjaria, Gerard Pardo-i-Miquel. 2013. “The Value of Democracy: Evidence from Road Building in Kenya.” http://sticerd.lse.ac.uk/dps/seminarpapers/eopp45.pdf

Feyrer, James, and Bruce Sacerdote. 2009. “Colonialism and Modern Income: Islands as Natural Experiments.” Review of Economics and Statistics 91(2): 245-262.

Henderson, J. Vernon, Adam Storeygard, and David N Weil. 2012. “Measuring Economic Growth from Outer Space.” American Economic Review 102(2): 994–1028.

Michalopoulos, Stelios, and Elias Papaioannou. 2012. “National Institutions and Subnational Development in Africa." Quarterly Journal of Economics, forthcoming. (NBER Working Paper no.18275.)

National Geophysical Data Center. 2010. "Version 4 DMSP-OLS Nighttime Lights Time Series.” National Oceanic and Atmospheric Administration. http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html

Nunn, Nathan. 2008. “The Long-Term Effects of Africa's Slave Trades.” Quarterly Journal of Economics 123(1): 139-176.

Nunn, Nathan, and Leonard Wantchekon. 2011. “The Slave Trade and the Origins of Mistrust in Africa.” American Economic Review 101(7): 3221–3252.