ARCGIS 10 FOR ECONOMICS RESEARCH

LECTURE 1

INTRODUCTION

Masayuki Kudamatsu

3 May, 2016

Press SPACE to proceed.

To go back to the previous slide, press SHIFT+SPACE.

TODAY'S ROAD MAP

1. Why GIS for economics?
2. Satellite images and scanned old maps
3. Polygon, polyline, point, and raster
4. Coordinate systems

1. WHY GIS FOR ECONOMICS?

MAKE MORE RESEARCH QUESTIONS FEASIBLE

Satellite images & old maps (this lecture)

Merge datasets by proximity (L2)

- e.g., weather data with survey data

Estimate the spillover effect on the control group (L3)

MAKE IDENTIFICATION STRATEGY MORE CREDIBLE

Control for more covariates / fixed effects (L2, L3, L5)

Instruments (L4, L6)

RD-design (L7)

2. SATELLITE IMAGES & OLD MAPS

SATELLITE IMAGE EXAMPLE #1

DEFORESTATION IN INDONESIA

(BURGESS ET AL. 2012)

# of districts in a province ↑

⇒ Each district govt official engages in Cournot competition in selling (illegal) logging permits

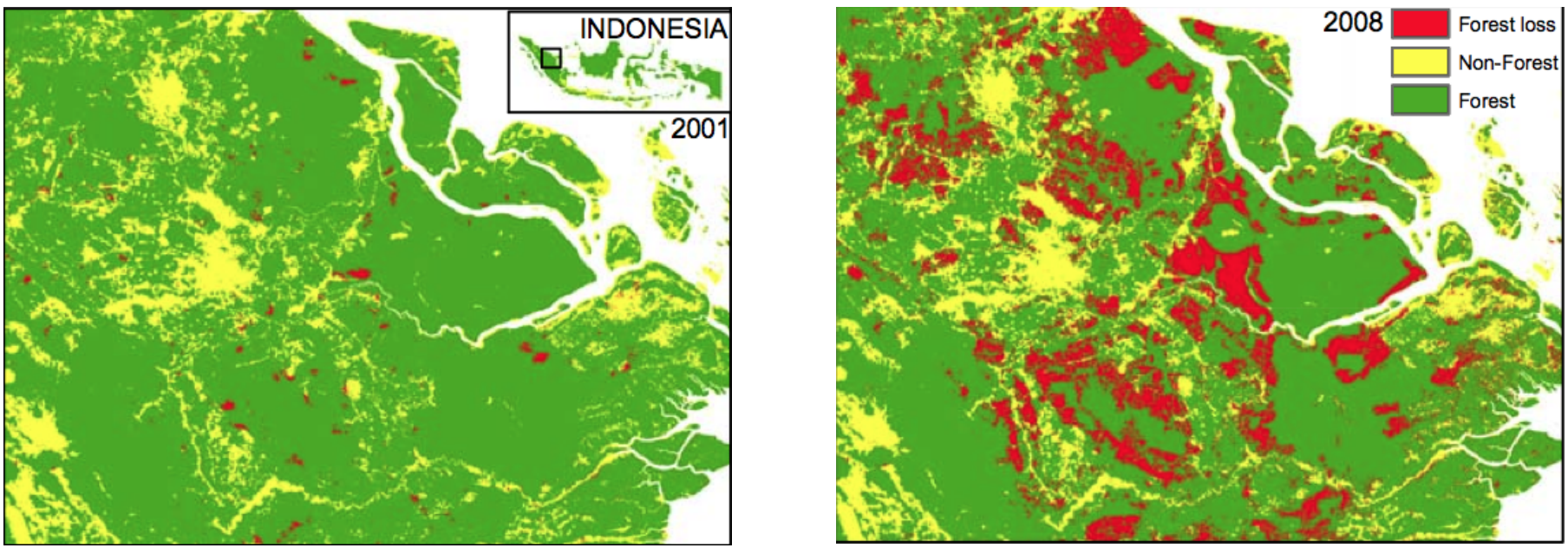
⇒ Deforestation in the province ↑

Cannot rely on official stats of logging

⇒ Use satellite image

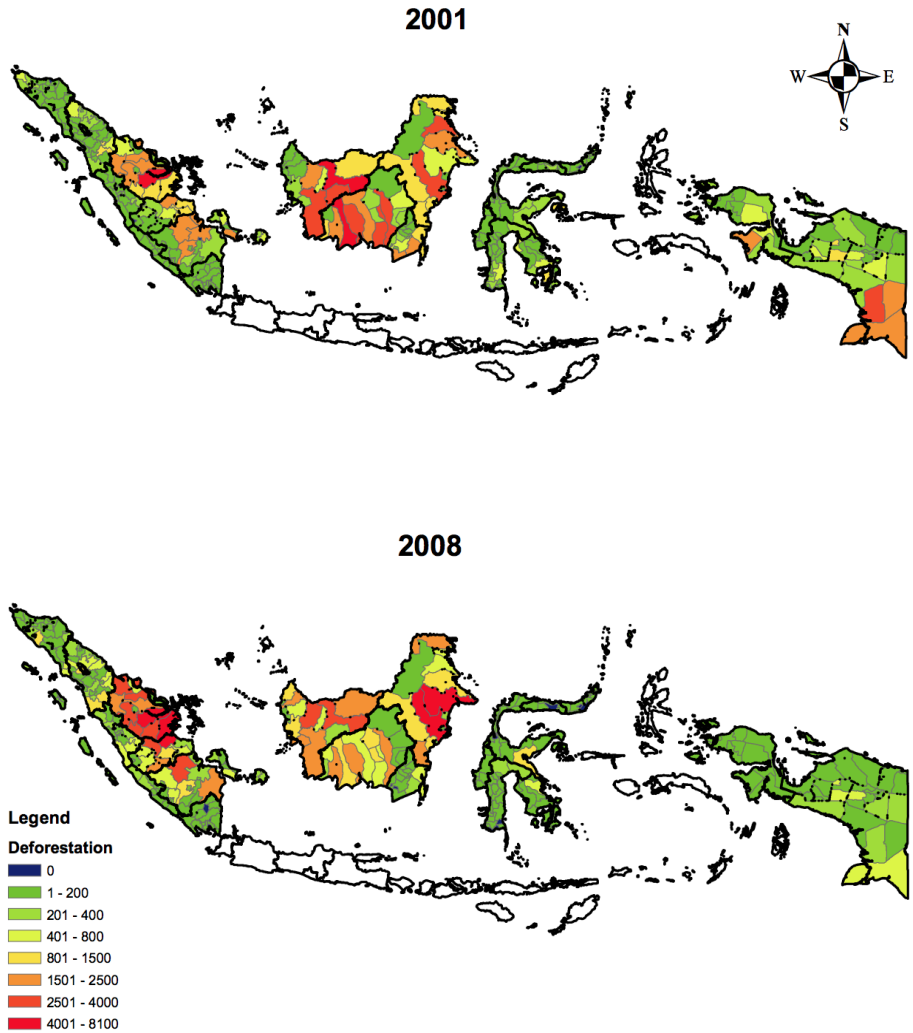
- Spatial resolution: 250m x 250m pixel
- Data: electromagnetic radiation strength in 36 bands of spectrum
- Develop algorithm to convert radiation patterns to forest coverage

PIXEL-LEVEL DATA ON DEFORESTATION



(Figure I of Burgess et al. 2012)

DISTRICT-LEVEL DATA ON DEFORESTATION

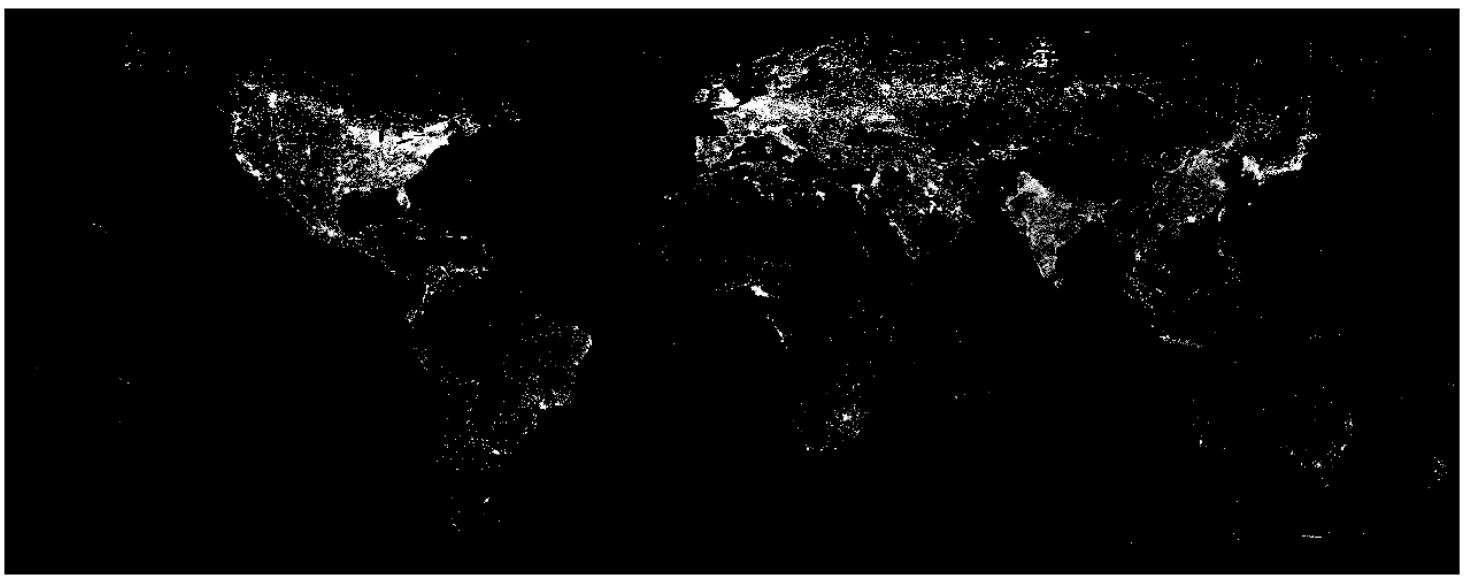


(Figure II of Burgess et al. 2012)

SATELLITE IMAGE EXAMPLE #2

NIGHTTIME LIGHT 1992

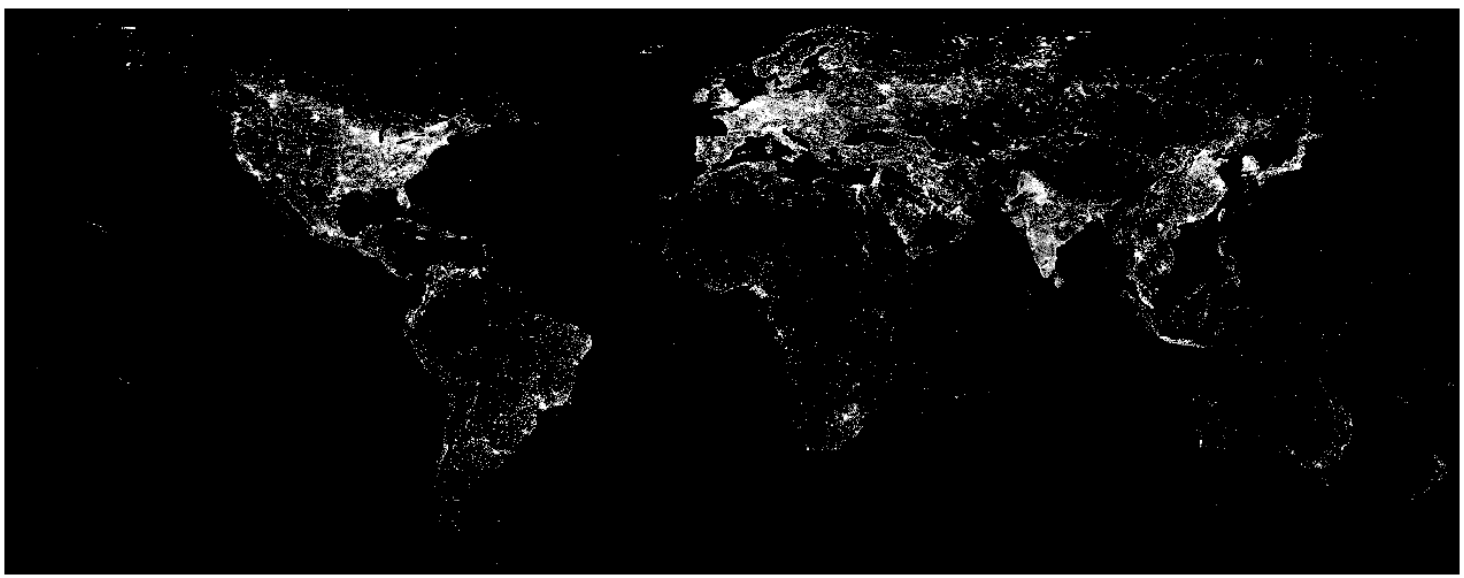
(Source: Version 4 DMSP-OLS Nighttime Lights Time Series)



SATELLITE IMAGE EXAMPLE #2

NIGHTTIME LIGHT 2013

(Source: Version 4 DMSP-OLS Nighttime Lights Time Series)



NIGHTTIME LIGHT IN TOP 5 ECON JOURNALS

Henderson et al (2012): correlate with real GDP growth

Michalopoulos & Papaioannou (2013, 2014), Alesina et al (2016): measure ethnicity-level development in Africa

Hodler & Raschky (2014): Presidents make their home region brighter

OLD MAP EXAMPLE #1

ROAD BUILDING IN KENYA

(BURGESS ET AL. 2015)

Digitize Michelin maps for Kenya since 1961

Track road network expansion over time

See if the president's ethnic group sees more roads built than other groups

DIGITIZING OLD MAPS

Michelin map in 1961

Digitization and Standardization in GIS

(source: Remi Jedwab's presentation slide)

OLD MAP EXAMPLE #2

ETHNIC HOMELANDS IN AFRICA

Drawn by Murdock (1959)

Digitized by Nunn (2008), to match ethnicity-level data on slave trade with country-level data (lecture 4)

Also used by Nunn & Wantchekon (2011), Michalopoulos & Papaioannou (2013, 2014, 2015), Alesina et al (2016)

ETHNIC HOMELAND & COUNTRY BOUDARIES

Figure II of Nunn (2008)

PRACTICAL ADVICE

Satellite images: some are freely available but very costly (time & money) to process

- See "15 Free Satellite Imagery Data Sources" by GIS Geography

Old maps: digitizing is also time-consuming but feasible with patience

- Georeferencing: Yale Map Collection (2009) (pp. 8-10)
- Create vector data: ArcGIS 10: Editing & Creating Your Own Shapefiles (Parts 3-6)

⇒ This course at least helps you **use** these datasets

GIS SOFTWARE

ArcGIS

- Python-friendly
- Buggy; expensive; tricky to create map images; Windows only

QGIS

- Free; easy to create map images; compatible with any OS
- Python-unfriendly
- Tutorial: www.qgistutorials.com

R

- Textbook: Brunsdon & Comber (2015)

PREPARE FOR THE REST OF THIS LECTURE

1. Download the zipped dataset for lecture 1
2. Extract it in Windows Explore, 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

Windows Explorer may hide ".zip" or ".tar" from file names

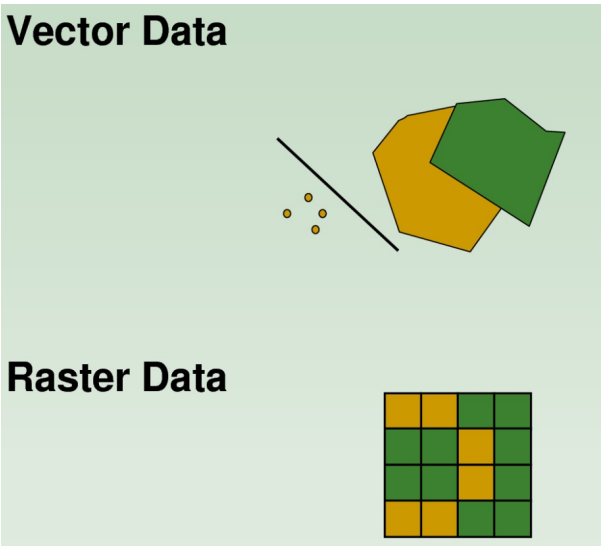
PREPARE FOR THE REST OF THIS LECTURE

3. Extract all the .zip files
    - F162008.v4.tar is also a compressed file, but we will uncompress this file later.
4. Launch ArcMap 10 (it takes time)
5. Take a 10 minute break

3. POLYGON, POLYLINE, POINT, AND RASTER

VECTOR VS RASTER

Spatial data comes in two different formats



How to edit data differs a lot between them

We now learn how to browse spatial datasets in ArcGIS while learning these different formats of spatial data

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: **Shapefile** (.shp)

VECTOR DATA #1

POLYGON FEATURES

Represent geographic zones

- Countries (L4 & L5)
- Sub-national districts (L6)
- Ethnic homelands (L4 & L5)
- Lakes, Islands, etc.

EXERCISE #1

BROWSE POLYGON FEATURES

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 To Folder..."

EXERCISE #1 (CONT.)

Ignore the alert message "Unknown Spatial Reference" for the moment, by clicking OK.

We'll come back to this issue later in this lecture.

VECTOR DATA #2

POLYLINE FEATURES

Represent networks / routes

- Roads (L7)
- Rivers (L6)
- Coastlines (L4)

EXERCISE #2

BROWSE POLYLINE FEATURES

Data to be read: Natural Earth's Rivers and Lake Centerlines data (10m_rivers_lake_centerlines.shp)

Browse 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.

⇒ Now rivers are shown on national boundaries.

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 would 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

If you cannot drag the data in the Table of Contents window, check if the List By Drawing Order icon is selected on top left.

VECTOR DATA #3

POINT FEATURES

Represent point location

- Plots (L3)
- Schools (L3)
- Cities (Exercise #3 below)
- Centroid of polygon (L4 & L6)

Can easily be created from **XY data**

WHAT IS XY DATA?

Each row: point feature

Column 1: longitude (x value)

Column 2: latitude (y value)

Other columns: attributes of point feature

- Locaiton name
- Statistics
- Key
- Foreign key

HOW TO OBTAIN LONGITUDE & LATITUDE?

GPS receiver

- If you conduct your own survey

Online gazetteer

- If location names are available
- Examples:
  - Geonames
  - Global Gazetteer Version 2.3
  - JRC Fuzzy Gazetteer

TO IMPORT XY DATA TO ARCGIS

Data format: Tab-delimited text / Excel worksheet

- Comma delimited text: sometimes work, sometimes don't
- From Stata, use the "outsheet" command
  - ArcGIS may round off longitude & latitude values
  - Use the "format" command to avoid this.

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

Code example for 10 dicimal digits for longitude/latitude

TO IMPORT XY DATA TO ARCGIS (CONT.)

To convert XY data into a point feature class, use:

- Make XY Event Layer
- Copy Features

These are examples of **geo-processing tools**

EXERCISE #3

CONVERT XY DATA INTO POINT FEATURES

Data to be read: CEPII Distances Database (geo_cepii.xls)

Browse the data in Excel

- Which columns for longitude and latitude?

To implement geo-processing tools, we use **Model Builder**

- Model Builder helps us program in Python (L2).

EXERCISE #3 (CONT.)

HOW TO USE MODEL BUILDER (1/3)

1. Click  in the Standard Toolbar.
2. 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.
3. Drag the tool from Search Window to Model Builder
4. Double-click the tool to set inputs, outputs, options etc.

EXERCISE #3 (CONT.)

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

1. Click 
2. Navigate to Geographic Coordinate Systems > World > WGS 1984

EXERCISE #3 (CONT.)

HELP FOR GEO-PROCESSING TOOLS

If you don't know what to fill in on the geo-processing tool window:

1. Click "Show Help >>" on the bottom right
2. Click the input field you don't understand
3. The help document appears on the right column.

EXERCISE #3 (CONT.)

MAKE XY EVENT LAYER (CONT.)

This tool creates a temporary **layer** out of XY data

But the layer often doesn't properly work with other tools

We also want to save the point feature class in the disk

⇒ Always use the Copy Features tool next to convert the layer into a shapefile data

EXERCISE #3 (CONT.)

COPY FEATURES

Input Features: the output from the Make XY Event Layer

- Use the drop-down menu, to specify the input that's already in the Model

Output Feature Class: ***.shp

- Always save newly created spatial data files in a folder different from the one for original input files

Ignore other options. Rarely used.

EXERCISE #3 (CONT.)

COPY FEATURES (CONT.)

This tool is also useful if you want to keep the original data intact (Exercise #6 below)

- Some geo-processing tools overwrite the input data...

EXERCISE #3 (CONT.)

HOW TO USE MODEL BUILDER (2/3)

To save the model:

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

A model can only be saved inside the toolbox.

Save frequently; ArcGIS often crashes.

EXERCISE #3 (CONT.)

HOW TO USE MODEL BUILDER (3/3)

To edit an existing model:

1. Locate the model in Catalogue Window
    - If you cannot see a new folder/file you just created, right-click the parent directory and "Refresh".
2. Right-click the model
3. Click "Edit" (NOT "Open")

EXERCISE #3 (CONT.)

Now run the Model by cliking the triangle icon at top right

Browse the output point feature shapefile

Capital cities (and other major cities) around the world should appear as point features.

One more thing about vector data...

ATTRIBUTE TABLE

Contains **fields** (i.e. variables) which can take a different value for each feature

To browse in ArcMap:

1. right-click the data in Table of Contents
2. click "Open Attribute Table..."

RASTER DATA

Divides the earth surface into many "square" cells (or pixels)

Each cell contains one value

Often created from satellite images

Examples:

- Elevation (L6, L8)
- Suitability for agriculture (L5, L8)
- Population density (L5)
- Forest coverage (this lecture)
- Nighttime light (this lecture, L8)

Can be used to create a new variable for vector data (L5)

RASTER DATA (CONT.)

Most file formats can be browsed in ArcMap

- TIFF (.tif)
- ESRI Grid (no extension)
  - File name cannot be longer than 13 characters
  - File structure is very complicated when browsed in Windows Explore

⇒ I recommend using TIFF format

For the ASCII format (.asc), you need to convert (Excercise #5 below)

EXERCISE #4

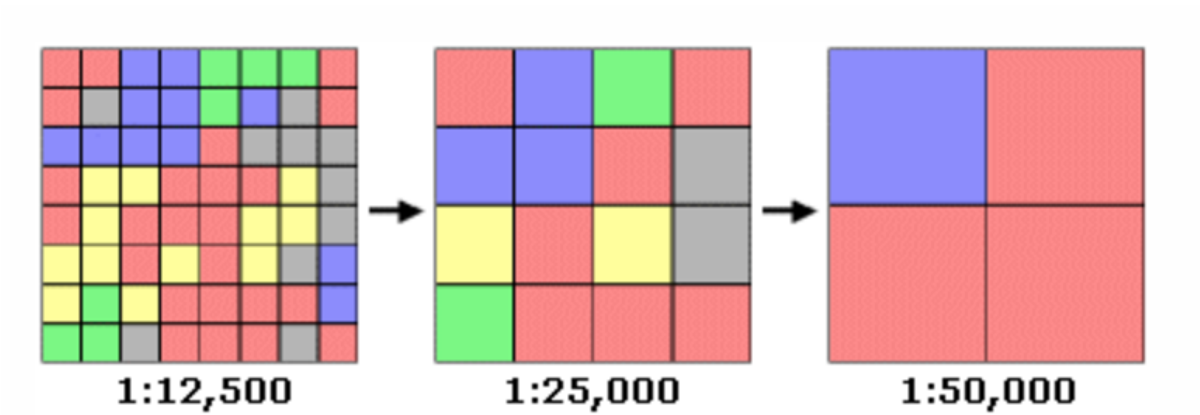
BROWSE TIFF RASTER

Data to be read: DMSP-OLS nighttime light for year 2008

1. Extract F162008.v4.tar
    - Creates three .gz files and three .tfw files.
2. Extract .gz files to **the same folder as .tfw files are located**
    - The .tfw file allows ArcMap to read a TIFF image of the same file name as geo-referenced (see Greenberg 2003 for detail)

EXERCISE #4 (CONT.)

3. In ArcMap, read F162008.v4b_web.stable_lights.avg_vis.tif
4. You'll be asked whether to "build pyramids"
    - If yes, displaying raster data becomes faster



EXERCISE #5

BROWSE ASCII RASTER

Data to be read: population density in 2005 (gl_gpwfe_pdens_05_ascii_half/glds05ag30.asc)

- Source: Gridded World of the Population

Use the **ASCII to Raster** tool

- **Output raster**: add ".tif" to save in the TIFF format
- **Data type**: FLOAT
  - Population density can be decimal

EXERCISE #5 (CONT.)

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.

RASTER IN STATA

Stata can read raster in the ASCII format, with **ras2dta** ado (Muller 2005)

Each cell becomes one row in Stata

To export raster as the ASCII format, use **Raster to ASCII**

4. COORDINATE SYSTEMS

WHAT IS A COORDINATE SYSTEM?

Earth is a sphere (approximately)

Various ways to *two-dimensionally* represent the earth surface

Each way corresponds to a **coordinate system**

- aka. spatial reference / map projection

WHY IMPORTANT?

To merge different spatial datasets accurately

cf. Apple Map did this wrong when it was launched in 2012

To calculate distance and surface area properly

TWO TYPES OF COORDINATE SYSTEMS

1. Geographic
2. Projected

GEOGRAPHIC COORDINATE SYSTEMS

Each location is coded by degrees

e.g., Gothenburg: 57°42'32'' N / 11°58'28'' E

Not suitable for calculating surface area

- 1° in lat: 110.6km at equator / 111.7km at poles
- 1° in lon: 111.3km at equator / 55.8km at 60° N/S

GEOGRAPHIC COORDINATE SYSTEMS (CONT.)

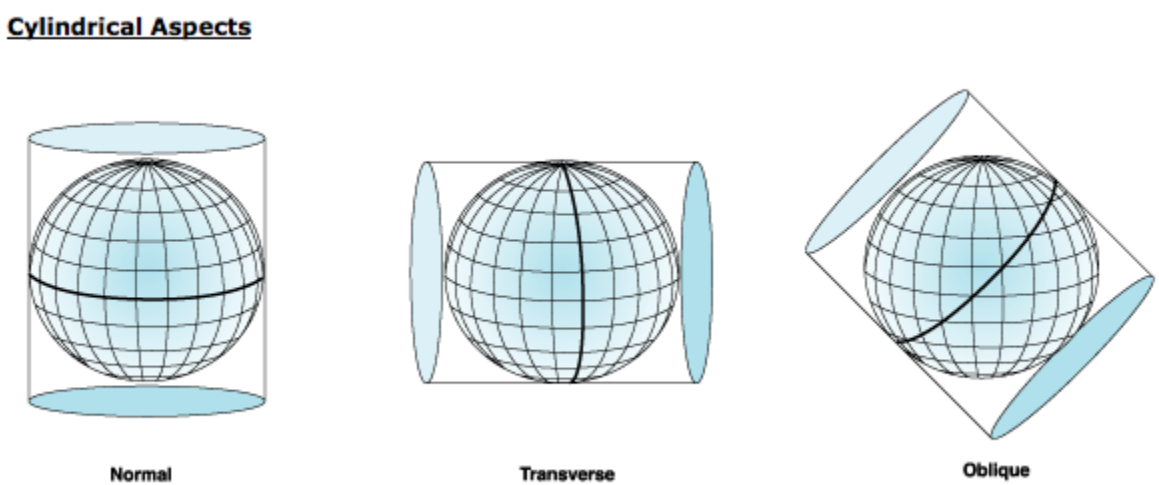
But useful for calculating distance between two locations

- Use great-circle distance formula (L4)
- Can be obtained by Stata ado **globdist**

WGS 1984: most popular

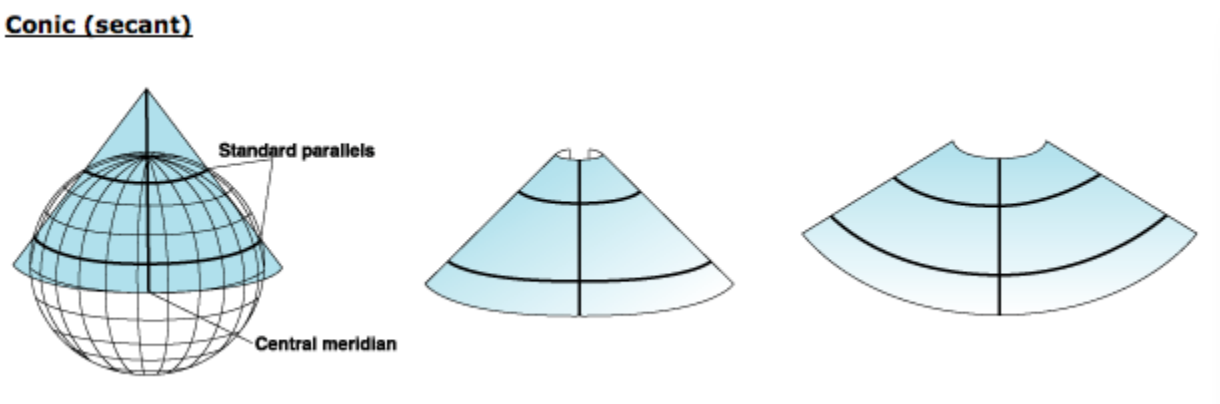
PROJECTED COORDINATE SYSTEMS

Earth surface is projected by the "light" from the center of the earth on *cylinder*:



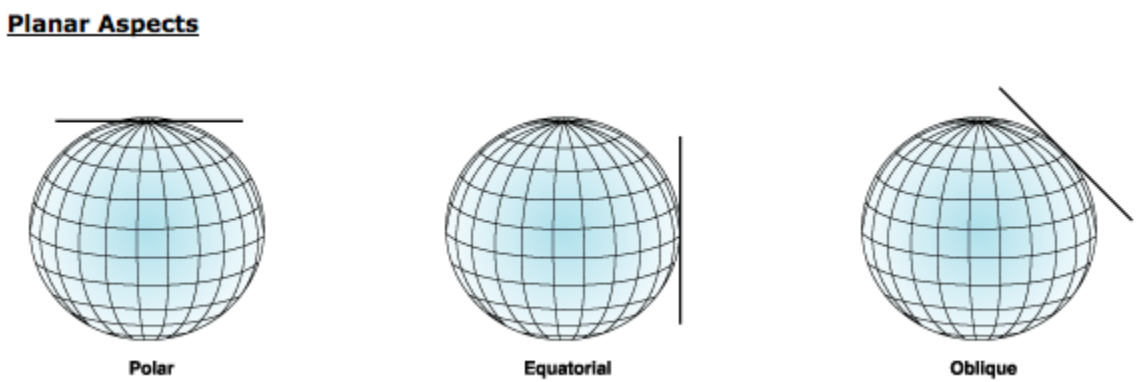
PROJECTED COORDINATE SYSTEMS

Earth surface is projected by the "light" from the center of the earth on *cone*:



PROJECTED COORDINATE SYSTEMS

Earth surface is projected by the "light" from the center of the earth on *plane*:



PROJECTED COORDINATE SYSTEMS (CONT.)

Each location: coded in *meters* from a certain origin on the projected surface

WHICH COORDINATE SYSTEM TO USE?

WGS 1984

- Distance between two locations (L4)

UTM

- Distance / surface area in small regions (L3)
- Length of polyline features (L6)

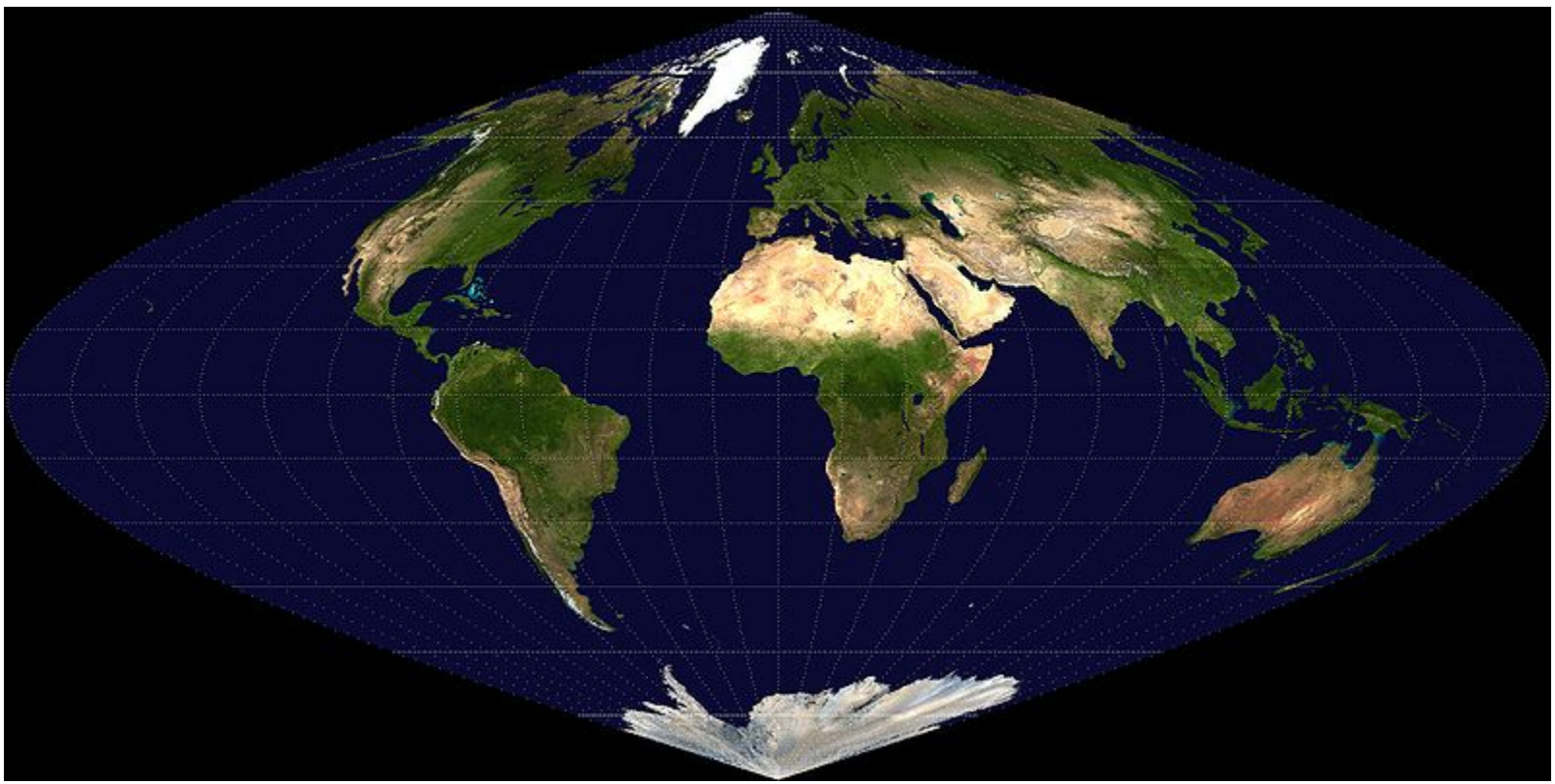
Any equal area projections

- Surface area in large regions (L4)

EQUAL AREA PROJECTIONS

Differ just in how the world is shown

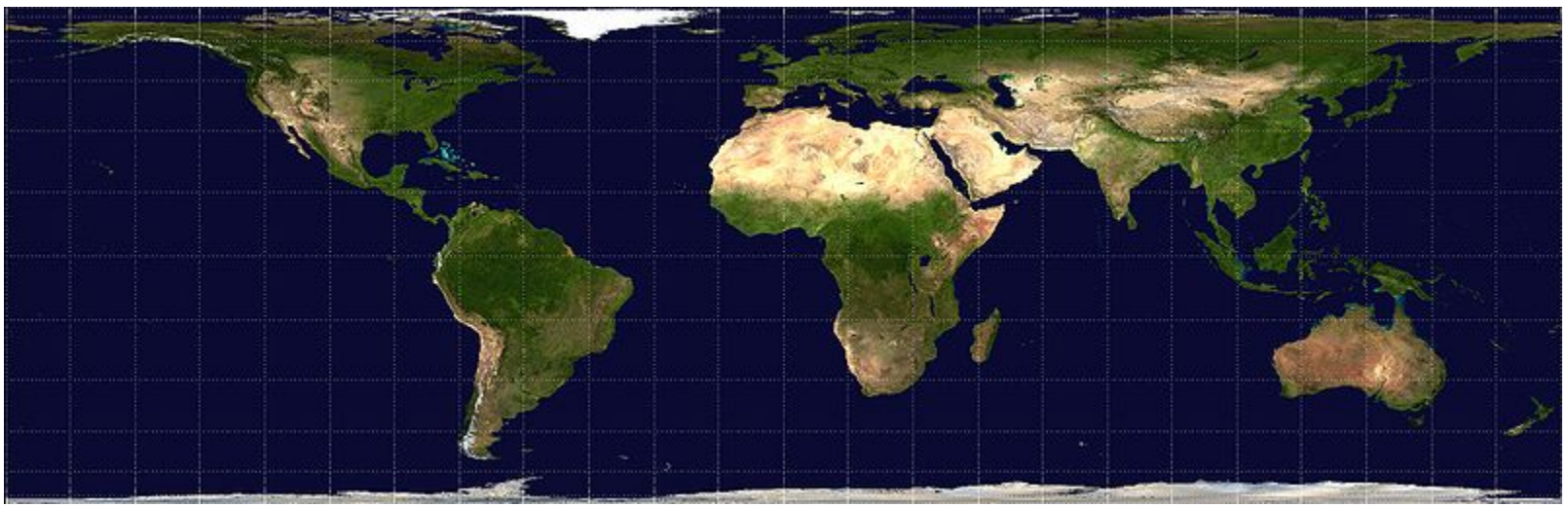
SINUSOIDAL PROJECTION



EQUAL AREA PROJECTIONS

Differ just in how the world is shown

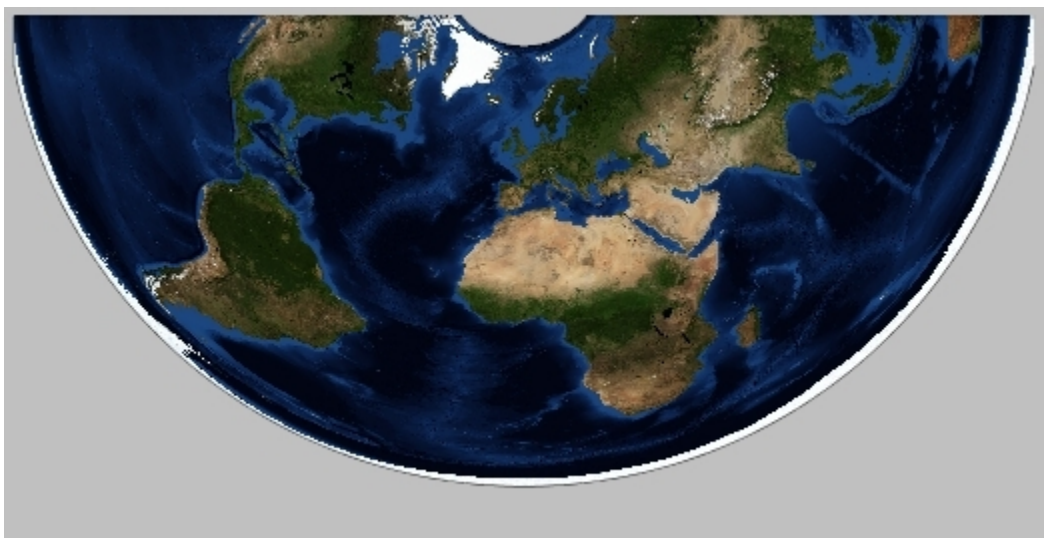
LAMBERT CYLINDRICAL EQUAL AREA PROJECTION



EQUAL AREA PROJECTIONS

Differ just in how the world is shown

ALBERTS CONIC EQUAL AREA PROJECTION



IF YOU WANT TO KNOW MORE:

Map Projections: A Working Manual, by John P. Snyder (U.S. Geological Survey, 1987) (Downloadable for free)

CHECK THE COORDINATE SYSTEM IN ARCMAP

1. Right-click the data in Table of Contents
2. Click "Properties..."
3. Click the "Source" tab
4. Check "Coordinate System" (scroll down, if not shown)

GEO-PROCESSING TOOLS TO ASSIGN A COORDINATE SYSTEM

Project: for vector data

Project Raster: for raster data (L7)

Define Projection: if undefined (for both vector & raster)

- Remember the alert message "Spatial reference is undefined" in Exercise #1?
- It means the coordinate system is not assigned

EXERCISE #6

ASSIGN THE PROJECTION TO GAUL DISTRICT BOUNDARY DATA

Spatial data usually comes with a meta data that specify the coordinate system used when the data is created

GAUL's meta data says it's WGS 1984

EXERCISE #6 (CONT.)

Use Define Projection to assign WGS 1984 (cf. Exercise #3)

This geo-processing tool *overwrites* the input file.

Essential to preserve the original input files for Python programming

⇒ Use Copy Features to create a copy which will be the input for Define Projection

EXERCISE #6 (OPTIONAL)

Remember the alert message "Spatial reference is undefined" in Exercise #5?

The readme file (that comes with the downloaded ASCII raster data) says the coordinate system is WGS 1984

Use Define Projection to assign WGS 1984 just as we did

EXERCISE #7

PROJECT TO SINUSOIDAL

Suppose we are interested in calculating the surface area of districts around the world

Sinusoidal projection allows you to calculate surface area properly

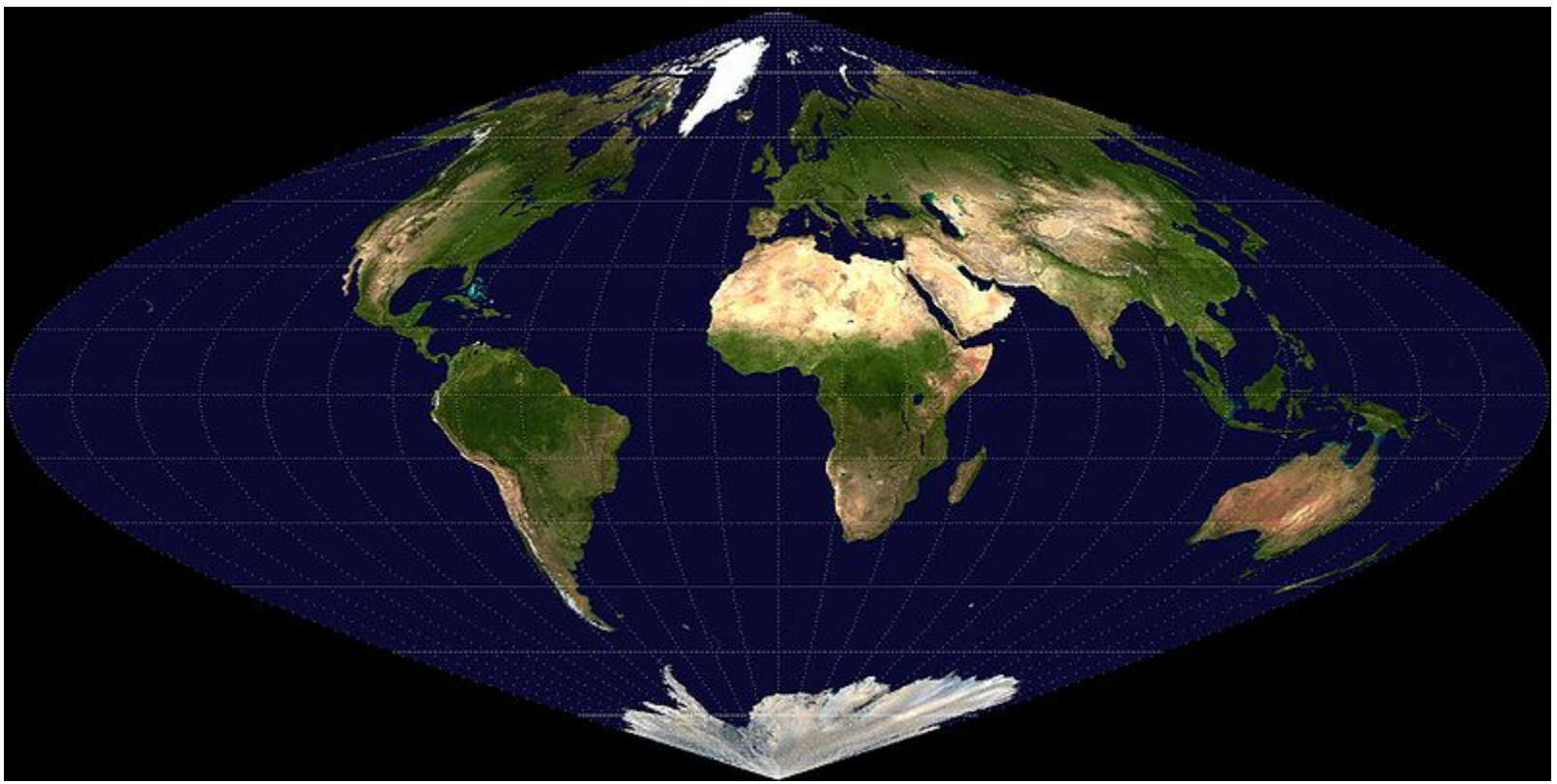
Use Project to change the coordinate system from WGS 1984

Sinusoidal projection is found at:

Projected Coordinate Systems > World > Sinusoidal(world)

EXERCISE #7 (CONT.)

Browse the output. Does it look like this?



BROWSE DATA IN A DIFFERENT PROJECTION

You cannot overlay data with different coordinate systems

ArcMap displays data in the coordinate system of the first data

To browse a data with a different coordinate system, open a new **map document**

WHAT IS MAP DOCUMENT?

Saves the way you overlay / color-code / symbolize different spatial datasets

File extension: .mxd

DOES NOT contain spatial data. It just has links to them

Set the relative path to refer to each data

EXERCISE #8

SAVE A MAP DOCUMENT

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

EXERCISE #8 (CONT.)

BROWSE A DATA IN DIFFERENT PROJECTION

1. Open a new Map Document
2. Drag the GAUL data in Sinusoidal projection from Catalogue Window

ONE MORE THING BEFORE WE CONCLUDE

Browse the spatial data in Windows Explore

You'll see many, many files for one dataset

- Shapefile = .shp + .sbx + .sbn + .shx + .dbf + .prj
- TIFF Raster = .tif + .tfw + .ovr + .aux + .prj

.dbf Attribute table

.prj Projection file

⇒ Use Catalogue Window in ArcMap, not Windows Explore, to move / copy / delete spatial data

WHAT WE'VE LEARNED ON ARCGIS

1. Convert data format
    - XY data
    - ASCII raster
2. Assign / Change the coordinate system

Do you remember which geo-processing tools you used for each of these tasks?

USEFUL REFERENCES FOR GIS

Yale University Library: GIS Workshop Archive

Melissa Dell (2009) "GIS for Applied Economists"

GIS Geography

WHERE TO FIND SPATIAL DATA

FAO Geonetwork

Geocommons

Earth Explorer

Devecondata

Keep an eye on the publication of papers using spatial data