

# ArcGIS for Applied Economists

## Session 4

Mark Westcott

LMU Munich

June 29, 2015

## Session 4

Plan for today:

- Work through Nunn replication
- Introduction to working with Raster files
- Measuring growth using night lights
- Replication of Feyrer & Sacerdote (2009)
- Georeferencing
- Creating/Editing features
- Measuring distances using rasters – e.g. shortest path over land

Also tie up some loose ends: Python, Model Builder, file formats

## Raster file formats

ArcGIS supports a huge number of Raster file formats. When downloading raster data, you are most likely to come across:

- TIF files
- ASCII files (Use the “ASCII to Raster” geoprocessing tool)
- NetCDF files (Use the “MakeCDF Raster” Layer geoprocessing tool)

Climatologists and other natural scientists often distribute data in NetCDF format. The files can contain multiple variables and a time dimension. You can download a NetCDF viewer at

[http://www.giss.nasa.gov/tools/panoply/download\\_win.html](http://www.giss.nasa.gov/tools/panoply/download_win.html)

A note about ASCII (text) files: often the files you download won't be quite ready for importing into ArcGIS, you will need to add a header first. See

<http://resources.arcgis.com/en/help/main/10.2/0012/00120000002s000000.htm> for details.

## Getting used to some raster tools

- Start a new ArcMap document, and add the world's countries to it (ne\_10m\_admin\_0\_countries.shp).
- Let's begin by using the "Create Random Raster (Data Management)" tool to generate some random data. Set the type to INTEGER so we get whole random values. Set the top to 90 and right to 20. Set the cell size to 1. This will create a 90x20 grid starting at (0,0) with each cell one degree wide and high.
- The raster data is automatically shown in a black/white scheme. We can change that using the layer's symbology options.

Note: The raster we have just created, like raster files you obtain from third parties, are in a latitude/longitude grid. This means that the grid cells are not of constant size!

## Raster Geoprocessing tools

Note: there is a ridiculous restriction that means that you are limited to 13 characters when naming raster output!

- Raster Calculator (Spatial Analysis) [examples: Raster + 2, Con(Raster > 5,5, Raster)]
- Reclassify Raster (Spatial Analysis) [examples: (1-2) to 1, (3,4) to 2 etc]
- Clip (Data Management)
- Resample (Data Management)
  - ▶ use 'Nearest' resampling method for discrete data
  - ▶ use 'Bilinear' or 'Cubic' for continuous data
- Project Raster [Switch to Elevation and Cities]
- Zonal Statistics as Table
- Extract Values to Points

## Night lights – motivation

Use night light data to measure economic growth when:

- Official GDP data poor
- We are interested in measuring growth on a finer administrative level than data is available for

Some papers using night lights: Bleakley and Lin (2012), Henderson et al. (2012), Michalopoulos and Papaioannou (2013), Lowe (2014), Pinkovskiy (2013), Storeygard (2012)

There is some scepticism about how good a proxy night lights provide, see for example [http://www.ifw-members.ifw-kiel.de/publications/night-lights-and-regional-gdp/KWP\\_1888.pdf](http://www.ifw-members.ifw-kiel.de/publications/night-lights-and-regional-gdp/KWP_1888.pdf)

## Night lights – data

About night light satellite imagery:

- Available as “Average Visible, Stable Lights, & Cloud Free Coverages” at <http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html>
- Imagery a product of the US Air Force Defence Meteorological Satellite Program
- Satellite orbits the earth 14 times a day
- Raster dataset: each pixel corresponds to 30 arc-seconds in longitude and latitude (about 1 square kilometre at the equator)
- Yearly average of all valid observations for each pixel – i.e. no cloud cover, not summer, dark half of lunar cycle
- Data available from 1992.

## Nightlights – use

Measuring economic output is in principle very simple:

- 1 Load a Shapefile with the areas we are interested in, e.g. countries
- 2 Load in the raster with 'stable lights' from the dataset
- 3 Average nightlight values within each administrative area using the 'Zonal Statistics as Table' tool

But there are two complications we should be aware of:

- 1 Gas flares
- 2 Varying pixel sizes

## Night Lights – gas flares

Gas flares are a man made source of light but not what we have in mind when using night lights as a proxy for GDP. Typically we want to exclude areas with gas flares from our calculations

- Download Shapefiles for your countries of interest from [http://ngdc.noaa.gov/eog/interest/gas\\_flares\\_countries\\_shapefiles.html](http://ngdc.noaa.gov/eog/interest/gas_flares_countries_shapefiles.html).
- Merge them together using the Merge (Data Management)
- Use the Erase (Analysis) geoprocessing tool to create a map of countries of the world with holes in areas of gas flares (but first use the Intersect tool to extract flares over land, otherwise the Erase tool will give an error)

## Night Lights – pixel area

The night lights are recorded in grid of 30 arc-seconds latitude/longitude. As you hopefully know by now, the real area of this 'square' varies, with areas getting larger towards the equator. This means we overweight more northerly areas when calculating the mean cell values.

There are two potential solutions:

- Reproject the raster into an equal area projection using the 'Project Raster (Data Management)' geoprocessing tool. This requires resampling the raster and introduces its own kind of 'distortion', but this shouldn't be systematic
- (or – more complicated) Calculate the shrinkage factor for each cell and divide by this prior to taking the mean. This is equal to  $\cos(lat * 0.0174532925)$  where 0.0174532925 converts from degrees to Radians. However, calculating this factor isn't completely straightforward. See <http://arctgisandpython.blogspot.de/2011/08/latitude-longitude-rasters-in-arctgis-10.html>

## Night Lights – exercise

- Inside this weeks' materials you will find a File Geodatabase called "NightLightsExample.gdb", which contains 2004 and 2009 light data for Europe, the Middle East and Northern Africa.
- For your convenience, it also contains a merged feature class 'Flares' which is the result of merging all the country specific flare shapefiles.
- Use this data to calculate a 'night lights growth table' for countries in Europe and Africa. How does growth in night lights compare to economic growth?

## Replication of Freyer & Sacerdote (2009)

Freyer & Sacerdote are interested in the effect of colonial rule on contemporary economic outcomes. They use wind direction and variability to instrument for length of colonial rule. The idea is that islands with strong easterly winds were more likely to be visited, revisited and colonised.

Data provided:

- “Islands.shp” – shapefile containing Islands in the author’s dataset
- “2008 Wind Data” – contains a NetCDF containing wind data for each month of 2008
- “IslandsData.xlsx” – contains GDP data on each Island

# Replication of Feyrer & Sacerdote (2009)

We want to replicate the following table:

	(1) Log GDP Capita	(2) Log GDP Capita	(3) Log GDP Capita - IV
Number of Centuries a Colony	0.42 (0.076)**	0.491 (0.110)**	0.712 (0.253)**
First Year a Colony			
Final Year A Colony			
Remained A Colony in 2000			
Abs(Latitude)		0.053 (0.012)**	0.054 (0.011)**
Area in millions of sq km		-20.374 (3.894)**	-21.738 (3.970)**
Island is in Pacific		0.752 (0.464)	1.018 (0.559)+
Island is in Atlantic		0.425 (0.395)	0.188 (0.477)
Constant	7.472 (0.205)**	6.033 (0.552)**	5.484 (0.834)**
Observations	81	81	81
R-squared	0.273	0.527	0.498

We need to determine each Island's area, its latitude and whether it is in the Pacific or Atlantic (Indian Ocean is the omitted category). For the IV regressions we need the average Easterly component of the wind and the standard deviation between the months.

## Replication of Feyrer & Sacerdote (2009)

Geoprocessing tools needed:

- Add Geometry Attributes to calculate areas, latitude and longitude for each island
- Make NetCDF Raster layer to import each the NetCDF file for each month's wind data
- Raster Calculator to calculate mean and SD for each grid cell
- Buffer
- Zonal Statistics as Table
- Table to Excel

Also add fields to indicate whether island is in the Pacific or Atlantic Ocean.

## Georeferencing

- Start a new map in ArcGIS and choose a coordinate system to work in (British National Grid)
- Download the John Snow Map from <http://upload.wikimedia.org/wikipedia/commons/2/27/Snow-cholera-map-1.jpg>. This is a raster format, and can be loaded into ArcGIS in the normal way
- Add a Streetmap baselayer
- Enable the Georeferencing Toolbar
- Use 'Fit to Display' on the Georeferencing Dropdown to move the JPG to roughly the right place
- Click 'Add Control Points'
- Click first on the raster and then on the point on the reference (in this case, base layer)
- View Link Table to see all the links you have made. D
- Rectify to save a new Raster (or Update Georeferencing)

Now create polylines for the roads, points for the wells, etc..

## Creating/Editing Features (1)

This is an incredibly brief introduction to creating/editing features. If you need to work with the editor, I recommend you look at the help files at <http://desktop.arcgis.com/en/desktop/latest/manage-data/editing/what-is-editing-.htm>

- (if required) Create a new 'Feature Class' inside a File Geodatabase using the Catalog. Set the coordinate system, the type of feature you want to store (polygon, line, point) and add any fields you want the feature class to include.
- (if required) Alternatively, create a new Shapefile using the Catalog.
- Make sure the Editor toolbar is visible by opening the Customize Menu, selecting Toolbar and then Editor
- Start an edit session by selecting 'Start Editing' from the 'Editor' drop down on the Editor Toolbar, then select the layer you want to edit and press OK.

## Creating/Editing Features (2)

- If your map is displaying in a projection which is different to the projection of the data you will receive a warning that the 'Spatial Reference does not match data frame'. You should change the projection of the data frame or reproject the underlying data
- You can edit existing features by moving around points, lines, vertices of polygons etc.
- To create features make sure that the 'Create Features' window is displaying. Open the 'Editor' drop down on the 'Editor' toolbar and selecting 'Editing Windows' → 'Create Features'. Now select the default feature template from the top half of the Create Features window and a Construction Tool from the bottom.
- You can make changes in the attributes table in an edit session, too.
- Advice: save your edits regularly by selecting 'Save Edits' from the 'Editor' drop down
- End your edit session by selecting 'Stop Editing' from the 'Editor' drop down

## Least Cost Path over Land

- Start with the “ne\_10m” land shapefile (in Session 3 materials, downloaded from NaturalEarthData.com)
- Create uniform Raster using “Polygon to Raster” geoprocessing tool and then reclassify raster
- Select a point feature which represents the ‘from’ point
- Use Cost Distance geoprocessing tool to generate cost and back link rasters. Select the feature with the start point selected as the “Input Raster or feature source data” and the raster we just created as the input cost raster. Make sure you specify an Output back link raster.
- Now select a point feature which represents the ‘to’ point
- Use the Cost Path geoprocessing tool to find the least cost path between the two points. “Feature destination data” should be the feature class with the ‘to’ point selected, the “input cost distance raster” and “input cost backlink raster” should be the to rasters produced in step 4.
- You can use the “Raster to Polyline” to convert the output raster into a line and then “Add Geometry Attribute” to calculate its length.

This is the simplest possible use case. For more advanced uses, see Melissa <sup>18/21</sup>

## Appendix 1: Tips for Analysis

- Create a new File Geodatabase for each project
- Reproject your input data into a common coordinate system
- Give your features/rasters sensible names
- Keep things tidy: delete unnecessary fields, etc

## Appendix 2: File Extensions

- .mxd - Map Document. Describes the layers you currently have loaded, how they are displayed, symbology, etc.
- .gdb (folder) - File Geodatabase (container format), can contain feature or raster data
- .shp - Shapefile, contains one class of feature data
- .dbf - ArcGIS' 'native' database format
- .py - Python Script

## Feedback

It would be great if you could fill out the final feedback form at <http://tiny.cc/GISsurveyFinal> before you leave.

I hope you've enjoyed the course, thank you!