

ArcGIS for Applied Economists

Session 3

Mark Westcott

LMU Munich

June 22, 2015

What's coming up..

Topics for the remainder of the course:

- This week: Geoprocessing on feature data, distances
- Next week: Geoprocessing on raster data, using the Model Builder for loops etc.

Geoprocessing tools

All of ArcGIS' analysis capability is found within its "Geoprocessing tools". There are tools for feature data and tools for raster data.

We will cover some commonly used tools here, but of course it's not possible to cover everything.

Getting ready to work with geoprocessing tools

Enable the Spatial and Network Analyst Extensions:

- Select 'Extensions...' from the 'Customize' menu inside ArcMap
- Put check marks next to the 'Spatial Analyst' and 'Network Analyst' extensions
- Press Close

Turn on the option 'Overwrite the outputs of geoprocessing operations'

- Select 'Geoprocessing options...' from the 'Geoprocessing' dropdown
- Put a tick next to 'Overwrite the outputs of geoprocessing operations' (in the General section)
- Press okay

Getting ready to work with geoprocessing tools - file geodatabases

File geodatabases are a very useful container format – I highly recommend storing your data within them

- Stored on your computer as a folder with a name ending in '.gdb'
- Can contain both raster and feature data
- For example, all the data for the Week 1 Opening Example was stored in a file geodatabase

Geoprocessing tools can store their output inside a file geodatabase; this is the easiest way to keep things organised. Have one Geodatabase for each project you are working on and keep everything in there.

Getting ready to work with geoprocessing tools - creating a File Geodatabase and setting it as the default

Geoprocessing tools will often store their output in the 'default geodatabase'. When you create a new map, ArcGIS sets `Documents\ArcGIS\Default.gdb` as the default geodatabase. So if you are not careful, you will end up with data in here and lose track of it.

For each project, create a geodatabase in the Catalog by right clicking on a folder and selecting 'New File Geodatabase' [important: make sure to get File Geodatabase and not Personal Geodatabase!]

Now set the newly created geodatabase as the default: right click on it and select 'Make Default Geodatabase'

Now save your map. When you reload it the default geodatabase will be remembered.

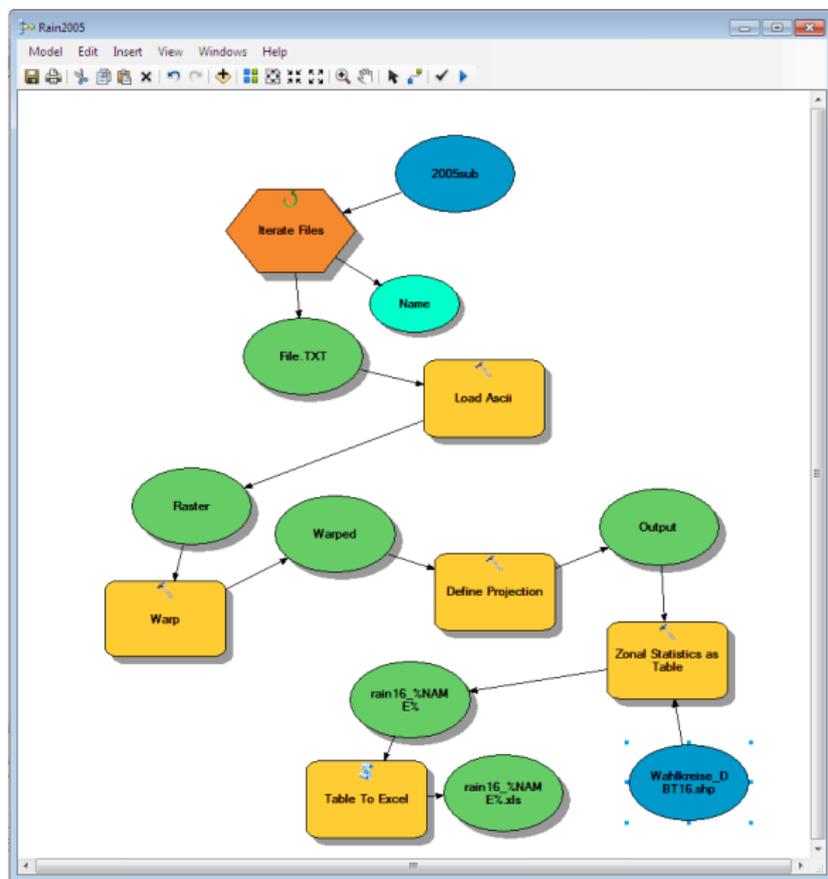
Running geoprocessing tools

Three ways to run the tools:

- Directly (locate tools using the ArcToolbox or search)
- Model Builder [tips at <http://blogs.esri.com/esri/arcgis/2009/02/20/10-things-i-wish-i-had-known-about-model-builder-before-i-started-using-it/>]
- Python

Using the model builder or Python is akin to writing a Stata do file rather than typing into the commands window. Allows you (and others) to reproduce your work, and to run loops etc.

Model Builder



Python

- Python is a general purpose scripting language
- Pretty much all of ArcGIS' capabilities can be called up from within Python
- You might find this easier than the model builder if you have some programming experience.
- Use the python window for individual commands or a Python editor outside of ArcGIS ("IDLE" is installed by default) for working on scripts.
- You can export model builder code to Python by selecting 'Export' from the 'File' Menu inside Python. The ArcGIS Geoprocessing tool reference gives example code for each tool, too.

About geoprocessing tools

- Geoprocessing tools generally take one (or more) data sources as input, and output another data file
- Note that this is different to Stata, where the commands modify the in-memory data!
- Generally, if you have some features selected, the tools will only act on those features.

Running geoprocessing tools

Two ways to find a tool:

- Select 'ArcToolbox' from the 'Geoprocessing' menu and browse to the tool you want, then double click
- Use the 'Search' window

Generally you need to select your input data and where the output data should be stored; this defaults to the geodatabase we set as the default earlier.

Tools run in the background – you will receive a notification in the bottom right of ArcGIS when they are complete.

It is very useful to open the geoprocessing results window (Geoprocessing → Results) when using the geoprocessing tools. This allows you to see what tool is running, error messages, rerun tools etc.

Some basic geoprocessing tools and example uses

- Dissolve tool: create US or 'states' data out of USA county data
- Feature to line tool: create a line of the USA's border
- Near (analysis) tool: find nearest river to each Starbucks location
- Near (analysis) tool: find nearest major city to each Starbucks location
- Buffer tool: draw a 15km circle around each major city
- Buffer tool: buffer around rivers
- Intersect tool: find the points at which Brazos river crosses state boundaries
- Intersect tool: find the border between two states or counties
- Spatial join tool: count the number of Starbucks in each county
- Spatial join tool: find out which county each Starbucks is in

Exercise 1

Load the Cholera data from Session 1

- What proportion of the total deaths were at locations which have Broad Street as their nearest pump? (Hint: first find out the total number of deaths, then the number of those deaths where Broad Street was closest)
- Export an Excel sheet which shows the number of deaths within a 100m of each water pump.
- Show all areas on the map which are within 150m of at least two water pumps.

Making copies of data, adding fields and adding geographic calculations

To make a copy of feature data, either

- Right click on a layer in the Table of Contents and 'Export'
- Use the Copy Features (Data Management) geoprocessing tool.

To add a field: Add Field ('Data Management') Tool

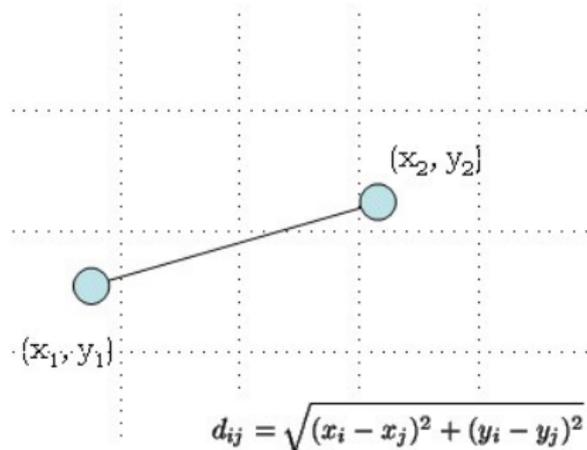
To rename a field: Alter Field ('Data Management') Tool

To delete a field: Delete Field ('Data Management') Tool

To add coordinates field, to calculate area, centroid etc: Add Geometry Attributes ('Data Management') Tool

Calculating Distances

- Question: how to calculate the distance between points? E.g. for creating a distance matrix.



- Problem: By 'flattening' the earth's 3D shape, projected coordinate systems distort distances. These distortions can be significant when working across big areas (e.g. continents).

Calculating Distances Accurately

Solutions:

- Ignore distortions: okay for doing calculations in small regions. Choose a projection that minimises distortion in that region, i.e. one of the UTM projections
- Use an equidistant projection. Equidistant projections preserve distances along given lines or from given points.
- Do accurate calculations based on geographic coordinates. These are 'geodesic' calculations.

Geodesic calculations:

- 'Great circle' – length of shortest curve between two points, treating the earth as a sphere
- Vincenty algorithm – calculates length of shortest curve treating the earth as an ellipsoid. More accurate.

Doing geodesic calculations in ArcGIS

ArcGIS has only very recently gained capability to do geodesic calculations, so it's no longer necessary to use Stata. However, you may still want to.

Here's how:

- `ssc install geodist`
- `geodist lat1 lon1 lat2 lon2, gen(dist)`
- use 'cross' command to help find nearest neighbours. See help geodist.

Otherwise, you can turn on the 'GEODISIC' Method in the 'Near (Analysis)' and Generate Near Table tool.

Creating distance matrices

Earlier we used the 'Near (analysis)' tool to find the nearest Major City to each Starbucks location and the distance from it.

We might, however, want to calculate a full distance matrix. For that we need the 'Generate Near Table' tool. Select 'Major Cities' as both Input Features and Near Features.

Exercise 2

This exercise is meant to bring together a few of the things we have learnt so far. Sorry that it's not so creative, but it's meant to replicate the kind of process you will go through when doing your own work.

- There are some Munich addresses and their lat/long coordinates in `Addresses.xlsx`
- You will find a Shapefile which contains postcode boundaries in `Postcodes.shp`. Unfortunately the projection information is missing from the Shapefile, but according to the website the data is in `GCS_WGS_1984`. Use the appropriate tool to add the projection information to the Shapefile.
- You will also find a Shapefile with all entrances to Munich U-Bahn stations in the materials pack.
- For each address, calculate the distance to the nearest postcode boundary. Rename the field `'distance_to_plz_boundary'`.

Exercise 2 (continued)

- For each address, calculate the distance to the nearest U-Bahn entrance, and rename the field again.
- Use the 'Table to Excel' tool to export the addresses and the distance you calculated to Excel
- Load the data into Stata. Create a new dummy variable called 'closer_to_ubahn' that indicates for each observation if the u-bahn is closer than the postcode boundary. Drop all columns apart from that dummy variable and the OBJECTID field, and save as a new Excel file (don't forget to include column names)
- Use the join function we covered last week to add the 'closer_to_ubahn' field to the display of the existing Addresses shapefile.

The Long-Term Effects of Africa's Slave Trades – Nunn 2008, QJE

Shows negative relationship across countries between number of slaves exported and current economic performance. Instruments number of slaves exported by distances to centres of demand.

Task 1: Calculate area of each country in km^2

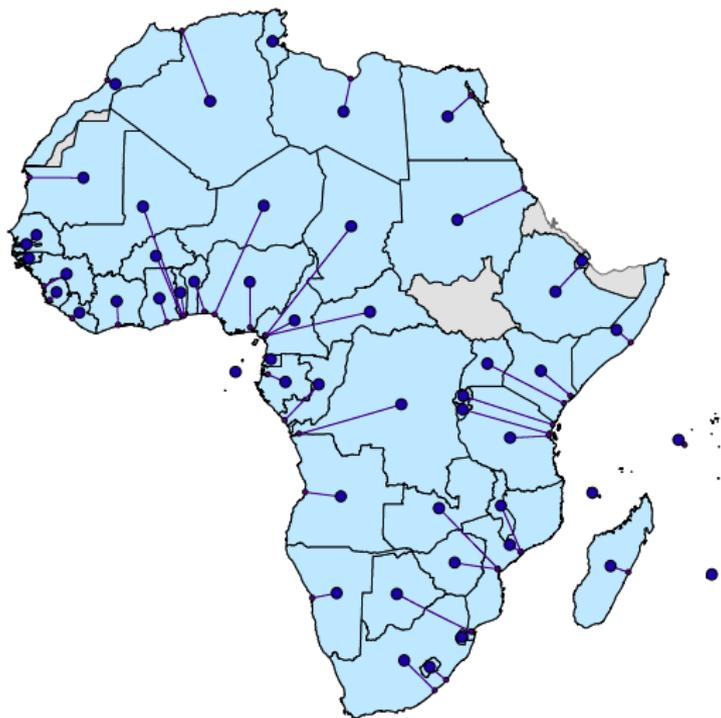
- Need this in order to be able to normalize slaves per square kilometer.

Task 2: Calculate distances from each country to location of slave demand:

- 1 Calculation centroid of each country
- 2 Calculate distance from centroid to coast (and for bonus points reproduce the image on the next slide)
- 3 Calculate distance from point on coast to each of these North American Ports – Virginia, USA; Havana, Cuba; Haiti; Kingston, Jamaica; Dominica; Martinique; Guyana; Salvador, Brazil and Rio de Janeiro, Brazil.

Nunn replication: Distance from centroid to coast

You should try and produce something like this:



Nunn replication: Data sources and tools

In the Session 3 materials ZIP you will find:

- Nunn2008Data.xlsx – contains data on the countries in the study
- ne_10m_admin_0_countries.zip – contains Shapefile for world's countries, downloaded from <http://www.naturalearth.com>
- ne_10m_coastline.zip – contains Shapefile for world's coastline, downloaded from <http://www.naturalearth.com>
- You can find coordinates for the North American ports on wikipedia, using Google Earth, inside ArcGIS, using a Geocoding website, using <http://www.mapcoordinates.net>, etc.

Some of the tools you might need for this are:

- Join (as per last week)
- Project
- Feature to Point
- Near
- Make XY event layer
- XY to line