GY460 project: Overview & Instructions

Assessment: The project is worth 100% of your final mark.
Word limit: 5000 words of text including tables, footnotes, references and any data
                  appendices. Tables should be in a text format, not pasted images. Use the
                  total word count from Microsoft Word or your preferred word processing
                  software.
Due date: See submission link on Moodle

Introduction

The aim of the project is to apply some of the GIS and spatial quantitative methods you have learnt on GY460 to answering a research question which has a 'spatial' dimension. The range of potential topics is very broad, but the defining characteristics for the project are that:

a) the research question should relate to a geographical dimension in the phenomenon you are studying
b) the methods should apply quantitative statistical methods to answering this question (usually regression analysis)
c) the project should involve some use of the GIS and quantitative spatial techniques covered on the course

You can carry out an empirical analysis on a topic of your choice, using data that you have sourced yourself (the '*self-build*' option below). Alternatively, suggested questions are provided in the '*flat-packed*' option below, and there is associated data on moodle.

In both cases, the project should be structured following the general conventional format for empirical academic articles. This should include an introduction, specification of the research question, discussion of theory and brief review of relevant literature, explanation of methods and data, presentation of results, discussion and interpretation of these results and conclusion. (You do not have to stick rigidly to this order or to these headings). An outline of the assessment criteria is provided at the end of this document.

The 'self-build' option

You can choose any topic that satisfies the 'spatial' criteria above, for example related to urban, development, or environmental economics. Please obtain approval for your proposal from the course director (Steve Gibbons) if you follow this path. A browse through the discussion papers in the Centre for Economics Urban and Spatial Programme discussion paper series will give an indication of the kind of topics that are 'spatial' and relevant to this course http://cep.lse.ac.uk/_new/publications/prog.asp?resprog=21. You will also find examples in academic journals in urban, development and environmental economics.

Illustrative examples include: a) assessing to what extent spatial variation in an outcome of interest is attributable to 'sorting' of different people or firms across places, versus effects of places on the average person; b) assessing to what extent we can disentangle spillovers, neighbourhood effects and interactions between neighbouring people or places, from other processes; c) analysing the value of a local amenity or dis-amenity through its effect on house prices; d) evaluating the impact of a policy that has some spatial dimension such as a

transport, infrastructure, or other place-based policy that targets areas rather than individuals; or e) a study of the influence of an area attribute on individual or firm location choices or migration.

Some links to useful data sources are provided on Moodle.

The 'flat-packed' option

Here are two suggested question, which you could explore using data and links provided on moodle.

1. How has the spatial distribution of unemployment/ workless-ness (or some other outcome of interest) changed in London between 2001 and 2011. To what extent are the changes explained purely by residential sorting i.e. where people of different qualifications, ethnicity, age, etc. live within London. Are there spatial concentrations of your outcome variable that are not explained by these patterns of sorting? Use GB census data to answer these questions.

2. Analyse the spatial patterns in the incidence of crime in London in 2011 and/or the changes in crime rates between 2011 and 2013. To what extent to demographic and geographical factors explain these patterns? What evidence is in the data that crime is 'contagious' and tends to spill over from one neighbourhood to another. Use 2011and 2013 data from police.uk linked to 2011 census data to answer this question .

The questions you should be asking include the following. What are the patterns, and what factors explain them? Is there evidence of spatial concentration (autocorrelation)? If so is this due to 'spillovers' (e.g. neighbourhood effects) or due to the spatial distribution of underlying population characteristics like social housing, ethnicity, education? Discuss, in the context of your findings, the limitations of the methods in learning about the causal processes underlying the outcomes you are analysing. This project requires you to construct a regression to explain a dependent variable in terms of a group of explanatory variables. Clearly these regressions will suffer from the spatial econometric problems we have discussed in lectures and seminars and a variety of standard econometric problems. You should discuss these carefully, paying closest attention to the spatial issues. There are limits to what you can achieve given the data and methods available to you. The aim is to get practice in working with these types of data, and to become aware of the limitations of these kinds of techniques, as well as highlight some of their descriptive uses.

Your project should typically include:
i) An initial analysis of the spatial characteristics of the dependent and explanatory variables (e.g. using maps, and/or that spatial statistics)
ii) A regression analysis that accounts for the spatial nature of the data e.g. by including appropriate spatial lags of the explanatory variables.

Software

There is no single piece of integrated software for doing this analysis (unless you are comfortable working in R). You will have to get used to moving data backwards and forwards between various software applications.

There are four distinct activities, which require different software:
1) Organising of the data we've provide in a spreadsheet e.g. Excel, SPPS, STATA, R
2) Visualising and geoprocessing the data using ArcGIS (or QGIS or R).
3) Creating spatial weights matrices, spatially weighted variables, spatial statistics and estimating spatial regressions using ArcGIS, STATA or GeoDa
4) Statistical/econometric analysis using STATA, or R, and/or within GeoDA

We have covered the basics of using ArcGIS and GeoDA in the computer classes. IT services provide some links to online guides to STATA

http://www.lse.ac.uk/intranet/LSEServices/IMT/facilities/software/specialistandteaching.aspx

Data sources for the flat-packed option

For these projects you will need to use the GB Census data for London from 2001 and/or 2011. The census data is readily available on the web via the Nomis service www.nomisweb.co.uk, and links are provided at the end of this document. Shapefiles for the Census boundaries (Middle Layer Super Output Areas, MLSO) for 2001 and 2011 are available from census.edina.ac.uk, and extracts for Greater London are in the project data pack on moodle. You could perform the analysis at various spatial scales, but the shapefiles provided are for MLSOA. These are reasonably consistent from 2001 to 2011 although there are some cases where two MLSOAs in 2001 become one in 2011. I have provided a union of the 2001 and 2011 data which provides a basis for joining the 2001 and 2011 MLSOAs should you wish to (e.g. to look at 2001 to 2011 changes in a census variable within each MLSOA)

You are also provided with crime data (street level) taken from one month (April) in 2011 and 2013 from police.uk. You can augment this with additional data from police.uk if you wish. Details of the crime data are set out below.

You are also provided with house price data (individual transactions) from the Land Registry 'price paid' data set. The full dataset is available on the web https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads.

The files I have provided are in shapefile format, for use in ArcGIS or GeoDA. In the case of point data (crimes, housing sales) the data are also in csv and STATA format. The STATA and csv format files have more variables then the shapefiles, to save storage space.

The source files are in zip file on Moodle. Note there is **NO CENSUS DATA** attached to the MLSOA. You need to get want you want for the analysis yourself from Nomis (see below).

The GB Census data is most easily accessed via the Nomis service, at www.nomisweb.co.uk. (There are other data sources available through Nomis at various levels of geographical aggregation, which you may find useful). I leave you to deduce how to use this web interface.

Census 2011: http://www.nomisweb.co.uk/census/2011

Census 2001: http://www.nomisweb.co.uk/query/select/getdatasetbytheme.asp?theme=26&subgrp=2001+Census

Crime Data:
www.police.uk

House price data: https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads

Boundary data and lookup tables
https://geoportal.statistics.gov.uk/geoportal/catalog/content/filelist.page
and
https://borders.ukdataservice.ac.uk/

The boundary data and police force data for London is also available on moodle.

There is a README file related to each data source that has been provided in the zip file with the data.

Assessment criteria

See separate document on Moodle