NCG608 Data Science Assignment

Introduction

You have been asked to investigate the penetration of domestic broadband networking in Ireland. An early exploration of the available data might start with some maps. What you have learned over the last few weeks, together with some data from the 2016 Census should provide you will that is needed to start.

The goal is to produce maps at *County* and *Electoral Divison* levels. Data can be downloaded from the Central Statistics Office website[1]. Follow the `Census 2016 Small Area Population Statistics (SAPS)` link. You will find CSV files with the data, and also links to boundary files.

The output will be a report detailing the work you undertook to create your maps - document what you did and what was needed to create the maps. Comment on difficulties and challenges.

Data

There will be some work involved in moving from the data download to your finished map.

For the boundary shapefiles, they are only available in geographical coordinates (WGS84). You'll need to transform the boundary data to Irish National Grid or Irish Transverse Mercator. In your report you'll also need a map to show where the counties are in relation to some background mapping - perhaps OpenStreetMap. You'll need to examine the attribute data. The 20m generalised versions should suffice.

For the Census data you'll need the County and Electoral Division (ED) levels. In the CSV files you will find that each county or each ED is represented by a single row in the file. Each column contains the *counts* for a single indicator. There's a glossary[2] which tells you what the column header codes mean (For example `T1_1AGETT` is the count of residents in each county or ED.

The census counts are organised into *themes* and *tables* within each theme. Theme 15 tables 2 and 3 deal with the availability of personal computers and broadband. The counts are of *households*. One of the issues with mapping such data is that the underlying *at risk* population (in this case, households) varies between EDs and between Counties, so you'll need to convert the counts to percentages of households with broadband.

You can use Excel, R, Python, Java, or whatever you want to do the computations.

You'll also need to make a table join to get your data matched to the ED or County boundaries - this will be in QGIS.

Important choices for your maps include the *class intervals*. The QGIS default is to use *equal intervals* (each class represents the same value range in the data. This may or may not be suitable. What happens when you use *quantiles* or *natural breaks*? Comment on the spatial variability of broadband availability.

Submission

You will submit a brief (2000 word *maximum*) report via Moodle. Your report will need an Introduction and a Conclusion. You'll need a section on getting the data, and manipulating the data, as well as a section on creating the maps. The actual maps themselves will be in the *discussion* section with appropriate commentary. What patterns do you observe about the *spatial* distribution of broadband in Ireland?

Submission: 23:59 on Friday 8th December 2017.

Martin Charlton

[1] www.cso.ie/en/census

[2] http://www.cso.ie/en/media/csoie/census/census2016/census2016boundaryfiles/SAPS_2016_Glossary.xlsx

50% for C.A.
50% for final Exam (2 Question)

→ www.cso.ie/en/census

use counties csv from 2016 small counties

We need to write a function to convert
, to — (in no. of people)

* No lecture on 15 Nov