Maynooth University
National University of Ireland Maynooth

# CS621 – SPATIAL DATABASES

## PROJECT REPORT

## FLIGHT PLAN/ROUTE ANALYSES

**Submitted To**

**Dr. Peter Mooney**

**Department of Computer Science**


| | |
|---|---|
| **Name** | Paul Williamson |
| **Student ID** | 18145469 |
| **Course** | HDip – Data Analytics 2018-2019 |

Student: Paul Williamson, 18145469
Course: HDip Data Analytics

**Project:** **County Kildare Flight Plan/Route Analysis**

**Client:** **Mr. Santa Claus, Christmas 2018**

**Purpose:** To use spatial analysis techniques to determine an efficient[1] flight plan for Mr. Claus around Co. Kildare on Christmas Eve. Because Mr. Claus' preferred mode of transport flies over roof-tops he's not restricted to only using roads, rail or waterways networks.

## Project Data Requirements:

- **Geospatial data for Counties in Ireland**
- **Census data for population density of children in Ireland**
- **Geospatial data at population centers[2] (calculated)**
- **OSM standard map (quick map services)**
  - added OSM standard map to QGIS canvas for contextual reasons.

## Methodology:

**Step 1: Download Irish counties shapefile for Ireland**
- add shapefile as vector layer to QGIS panel
- use DB Manager to import shapefile as a table into postgresql

**Step 2: Download 2016 Ireland Census information as small areas shapefile (18600+ polygons)**
- add shapefile as vector layer to QGIS panel
- use DB Manager to import shapefile as a table into postgresql

**Step 3: Develop choropleth map for all County Kildare small areas (740 polygons)**
- use SQL to create new postgresql table – County Kildare-only small areas
- use DB Manager to add new Kildare-only table as new vector layer to QGIS panel
- modify the new layer symbology properties (graduated, column, mode e.g. Jenks)

**Step 4: Get all small area (polygon) centroids.**
- use postgis ST_Centroids to compute the centroids of all Co. Kildare small area polygons
- add new centroids table as centroids layer to QGIS canvas (DB Manager)
- confirm layer properties retained using QGIS "info" tool 

**Step 5: Calculate scores to apply to each route segment**

(1) find starting point – assuming Santa's travelling N to S, pick most northerly centroid.
(2) using SQL find "nearest neighbour" (NN) for current point based on following formula:

$$\text{NN score} = \frac{\text{population density of small area polygon containing next centroid (pd)}}{\text{distance from current centroid to next centroid (dist)}}$$

1 Efficient route: defined here as a route which starts at the most northerly point/area (latitude) and moves progressively southward, each next step determined by a formula.
2 Population centers: Initially the resolution was going to be per residential building but this changed to small area.

## Methodology contd.

(3) SQL 'insert' current point into a route table
(4) delete all rows from original table which use the current point (either start or end)
(5) check original table length > 0 i.e. route incomplete (y/n)
(6) the new current point is the NN point that maximises the NN score above
(7) repeat from (2) above.

**Step 6: Analyse route**

- is this route the minimum distance/cost route
- does this route minimise intersections/backtracking
- does the route exhibit a pattern in accordance with the choropleth map/KDE heatmap

## Shapefile considerations:

- Initially I had planned to use a buildings shapefile for Co. Kildare. However, the most up to date buildings shapefile I could find for Co. Kildare turned out to be incomplete with a lot of missing data e.g. new residential areas not included, etc.

- As a work-around, I used the Ireland small areas shapefile, freely available from the C.S.O. website. The Irish map is divided into 18641 small areas. The C.S.O. data conveniently contains all the small area geometries and the population statistics (from Census 2016) for each area. I imported this shapefile as a new table into the postgresql database. There I filtered the data using a select query to extract the Co. Kildare only data which contained 740 rows. This formed the main dataset table on which all later operations were performed.

- I also used this Co. Kildare only data to develop the choropleth map indicating the various population counts in those small areas.

- Using the Ireland small areas shapefile, also allowed me to address the issue of the Modifiable Areal Unit Problem (MAUP), since the geometry of the small areas was designed in such a way as to minimise population variance between the areas. I further normalised the population counts using the actual area of each areal unit (small area) to obtain a measure of population density. Addressing this was an important consideration for me because I later relied on the population density with point-to-point distance calculations (ST_Distance) between the small areas, to determine the particular scores for the route segments.

- I used the Ireland County Boundaries shapefile to “clip” the original Ireland buildings shapefile with QGIS clipping utility (Vector > Geoprocessing Tools > Clip). This allowed me to extract Co. Kildare buildings data which I used to produce the KDE heatmap.

- While developing the above scores algorithm I used the QGIS Select Features by Polygon tool to select a subset of small areas (i.e. Newbridge town) which I used for test purposes during development. I imported this subset as a table into the postgresql database where I was able to use a “Create Table As (Select...” query to further extract an even smaller subset of the Newbridge small areas. This reduced dataset proved a lot easier to manipulate and also the results of any analysis here were easier to interpret during development.

## QGIS & Other considerations:

- The QGIS tool (Vector Geometry::Centroids) can be used to generate centroids of the polygons of interest (e.g. Co. Kildare small areas). However it produces a vector consisting of multi-points. When I tried to manipulate the resulting table of points in the postgresql database, I realised that the multi-points required a bit more processing before I could compute a simple route table with them.

- The workaround I used was to create the small areas centroids table (using ST_Centroids) in the postgresql database, which produced a table of ST_Point geometries. I used this new centroids table to compute the ordered route table in postgresql which I was then able to add as a new layer to the QGIS canvas. Once in QGIS, I used the Vector Creation::Points to Path tool to actually create a new vector layer with a polyline connecting all the points. (The ST_Makeline postgis function produced a linestring table (1r1c) which could not be added to the QGIS canvas correctly.)

## Webmap from QGIS:

file:///C:/Users/PAUL/AppData/Local/Temp/qgis2web/qgis2web_2018_12_29-07_40_00_888217/index.html#10/53.1555/-6.8145

## Analysis:

The subset of small areas I chose to develop the algorithm consisted of 6 small areas around Newbridge. (30 possible routes).

The Newbridge town small areas subset I used for testing the algorithm consisted of 84 small areas. (6972 possible routes which took approximately 10 seconds to compute in postgresql.

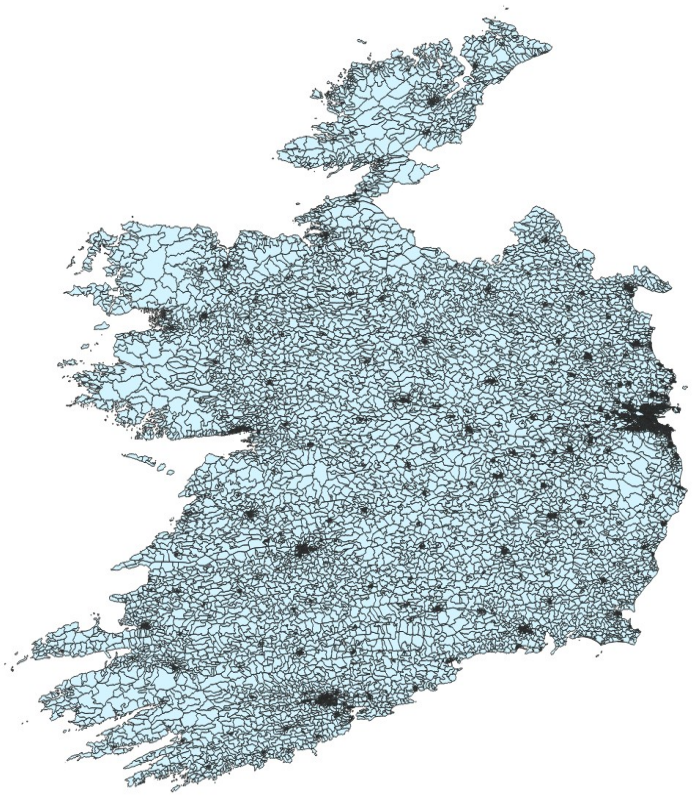
The main Co. Kildare small areas dataset for this project consisted of 740 small areas. (546,860 possible routes which took approximately 33 minutes to compute in postgresql).

The Co. Kildare buildings (type="houses") dataset which I used to create the KDE (heatmap) consisted of 8464 "houses". I filtered this dataset using both the clipping tool in QGIS (county boundary) and a select query in postgresql database to build a subset table.

## Conclusions:

## Small Areas Maps

**Fig. 1.1** The Small Area Boundaries were created using resources including the National boundary dataset and consistent sub-divisions of Electoral divisions (ED).



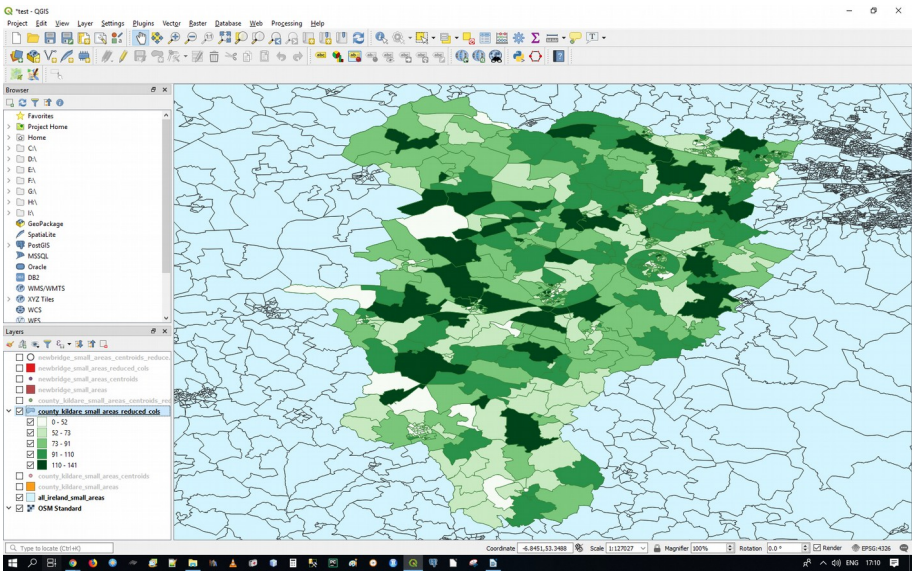
They were created, where possible, not to cross natural features. Each area is a defined area with a minimum number of GeoDirectory building address points. The defined area was initially created with an average of approximately 90 residential address points (with a minimum of 65). The final map was generated using two bespoke algorithms which incorporated the ED and Townland boundaries, ortho-photography, large scale vector data and GeoDirectory data.

Before the 2011 census small area definitions were split in relation to motorways and dual carriageways. After the 2011 census some boundaries were merged and others divided to maintain privacy of the residential area occupants.

MAUP:

Using the Small Area Boundaries defined above, allowed me to address the Modifiable Areal Unit Problem in this project in that the areas were specifically designed to minimise population variance between areas.

**Fig 1.2** Ireland County Boundaries: these were used to “clip” the County Kildare small areas data from the entire Ireland small areas dataset. I later used a Postgresql select query to create a sub-subset table of the County Kildare small areas data (Newbridge areas) for purposes of testing (see fig 2.2).



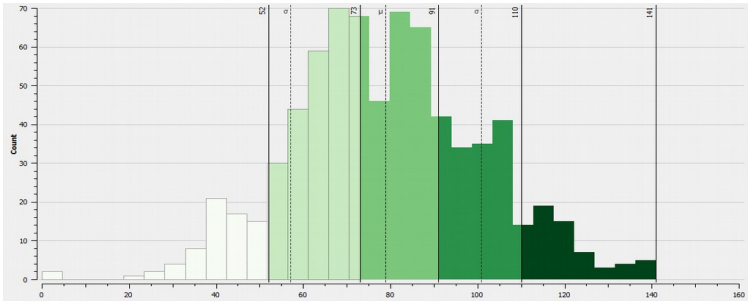



**Fig. 2.1 & 2.2** County Kildare Small Areas with QGIS generated Choropleth Map with Histogram indicating population statistics for children in county Kildare.

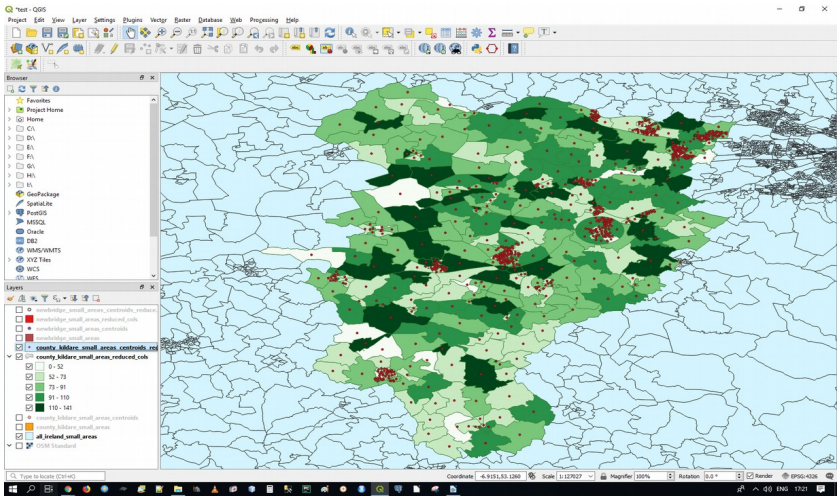


**Fig. 2.3** County Kildare Small Areas centroids.

Initially this vector was generated using the QGIS toolbox (Vector Geometry::Centroids) but this generated a Multipoints vector.

Later, I used the ST_Centroids postgis function to generate single point geometries for these centroids.

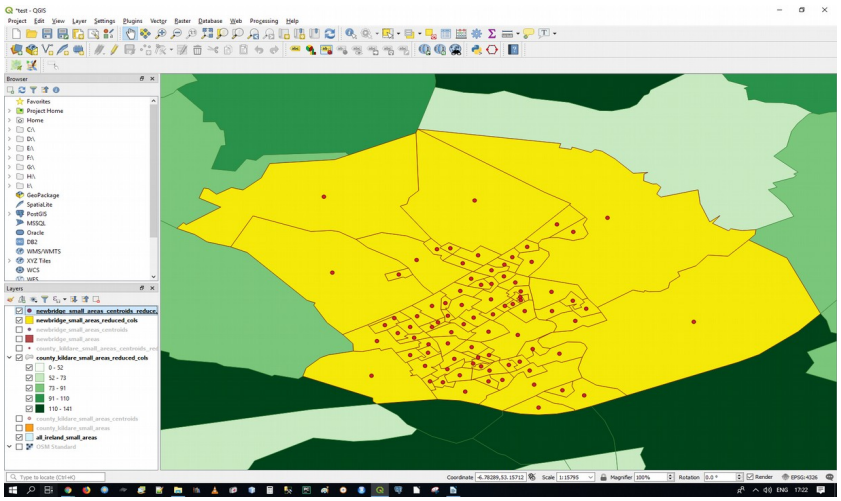


**Fig. 2.4** Subset of County Kildare Centroids used for developing algorithm for calculating route. This subset table was created in QGIS using Select Features by Polygon tool. This was reduced further in Postgresql using a “Create Table As (Select...” query. The centroids shown are from the postgis ST_Centroids function.

## Geospatial Data sources:

http://www.geofabrik.de/data/shapefiles.html
http://download.geofabrik.de/
http://download.geofabrik.de/europe.html
http://download.geofabrik.de/europe/ireland-and-northern-ireland.html

Documentation for the layers in the Geofabrik shape files is available at following link:

http://download.geofabrik.de/osm-data-in-gis-formats-free.pdf

## Other Useful links:

https://openrouteservice.org/