The penetration of domestic Broadband Network in Ireland

SUBMITTED BY

Amol Nanaware (19250328)

UNDER THE GUIDANCE OF

Prof. Martin Charlton

MSC. DATA SCIENCE AND ANALYTICS

INTRODUCTION

This report is on investigation of the penetration of domestic broadband network in Ireland for which the data has been obtained from 2016-census website[1]. The goal is to produce maps at County and Electoral Division levels and to detail out the work that has been undertaken to create the maps, and to explain all the difficulties and challenges faced. It includes the spatial data mapping and analysis of the broadband distribution throughout the Ireland.

1. Requirement gathering and data analysis

The data includes many kinds of aspects such as, age factor, the occupancy of the houses the facilities with the population, etc. We will investigate data access and how to manipulate them based on our requirement.

1.1 Getting the data

On the CSO website under Census 2016 Small Area Population Statistics, we will be downloading the shapefiles, from the table, of the Counties and Electoral Division under 'Generalised 20M' column. The page also includes 'SAPS 2016 Glossary' file which contains information on the columns present in the data files. We will download the complete set of SAPS from the link provided below the table. From this table we will need only two files named, 'SAPS2016_CTY31' and 'SAPS2016_ED3409'.

So, we are having total four data files,

- SAPS2016_CTY31.CSV(Data file)
- SAPS2016_ED3409.CSV(Data file)
- CTY31_gen20_ING.SHP
- Electoral_Divisions__CSO_Generalised_20M.SHP

[Note: A shapefile is a simple, nontopological format for storing the geometric location and attribute information of geographic features.]

As there was GUID issue between the downloaded County shape file and the data from CSV, we will be using the county shapefile provided by Prof. Martin Charlton for further investigation.

1.2 Data Manipulation, challenges and difficulties

After referring the glossary file, we found out that the broadband data is included in the column named 'T15_3_b' for both the county and electoral division data files. But the problem here is that each column contains the count for the corresponding row data. In county data file the count of broadband is with respect to the county area and similarly for the ED. So, to create a map of these, we will need the common measuring parameter. The table also contains the total count column named 'T15_3_t', so, now we can calculate the percentage from these parameters.

1.2.1 Case 1

For data manipulation we will be using PostgreSQL and Excel. If we look at the numeric data in the data files, it contains commas which will give error while calculations. So, to solve this problem follow the following steps,

- Open data files in Excel
- Select all cells (ctrl + A)
- Select Number Format for the data as 'General'

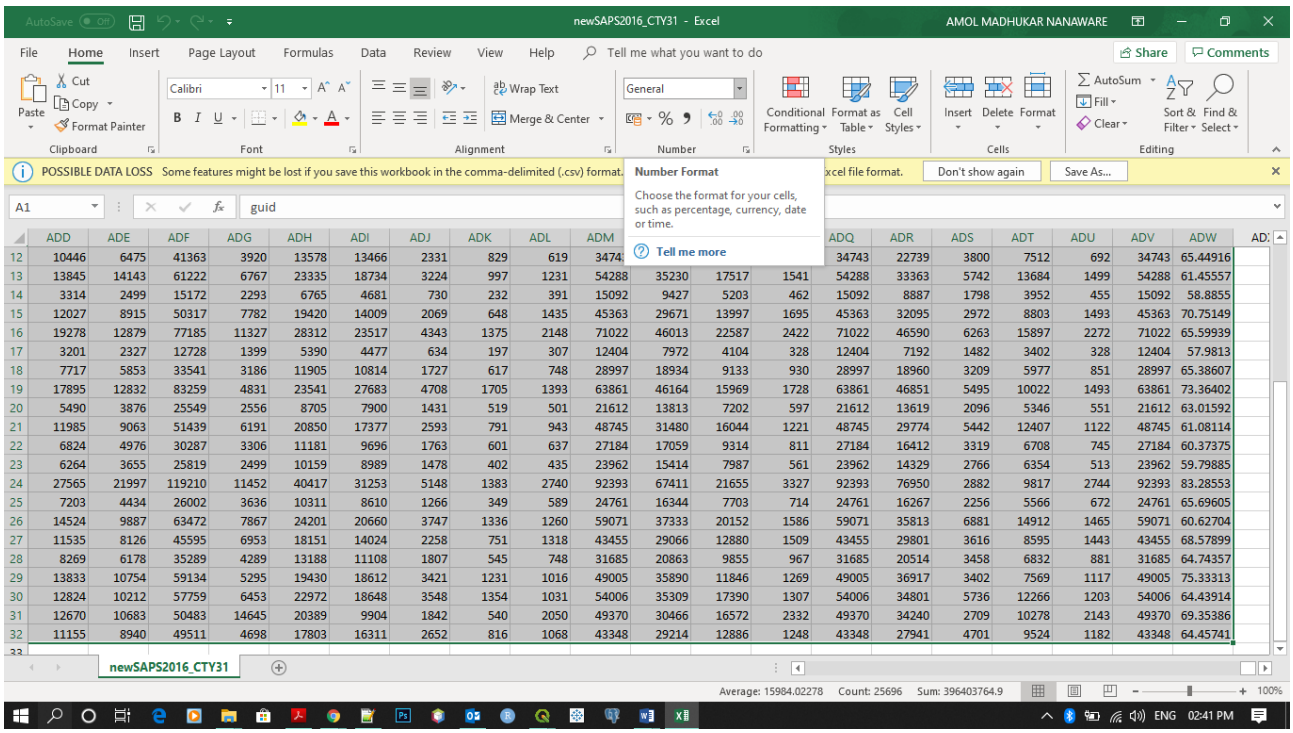


Figure 1.2.1.1

Now, we can see that all the numeric data present in file does not contains commas.

1.2.2 Case 2

To create the common percentage column, we will be using the PostgreSQL (including postGIS). Please follow the link [2], to create the database connection between QGIS and PostgreSQL and for more details.

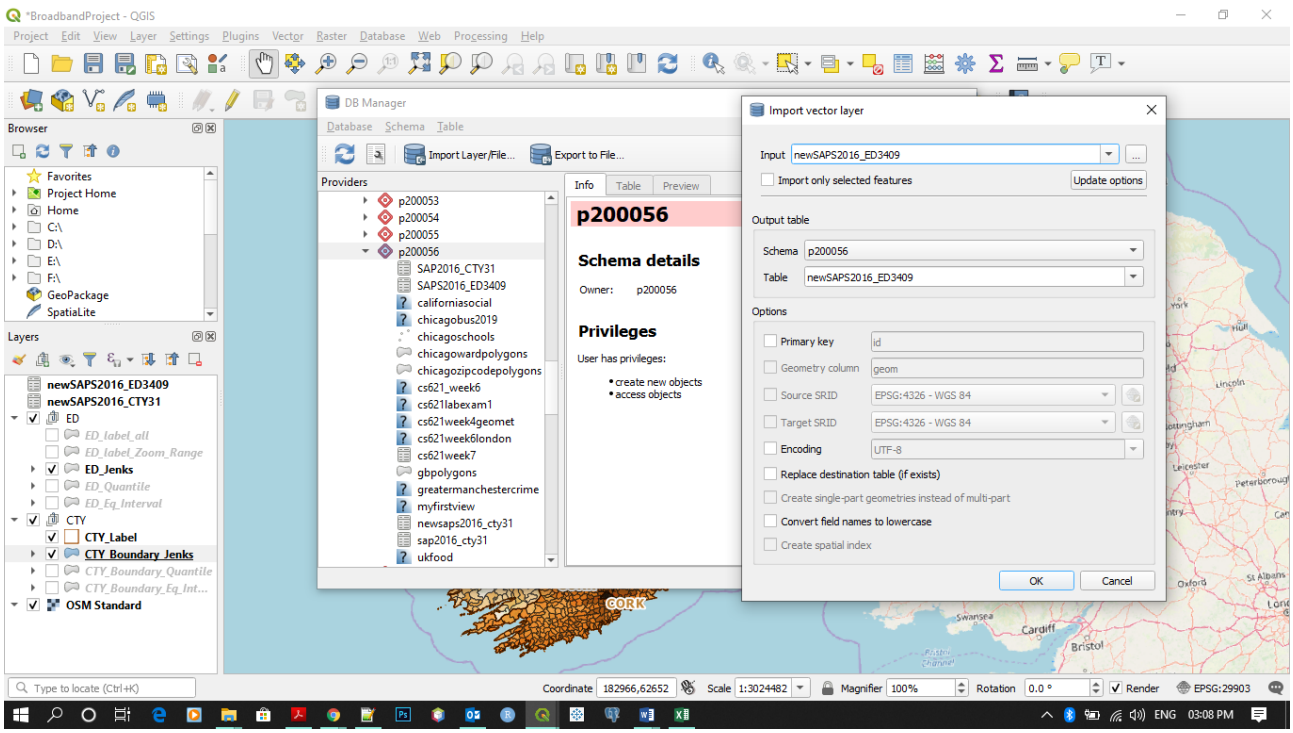


Figure 1.2.2.1

Add the csv data file and shapefile in QGIS as a layer by using the 'Layer' menu option. Once the connection is established open the DB Manager in QGIS and open your schema in postGIS from left side panel, please refer fig.1.2.2.1. Now click on 'Import Layer' button and import the csv in postGIS as shown in the figure.

After this go to pgAdmin and open Query tool and write all the commands shown in the fig.1.2.2.2 and execute.

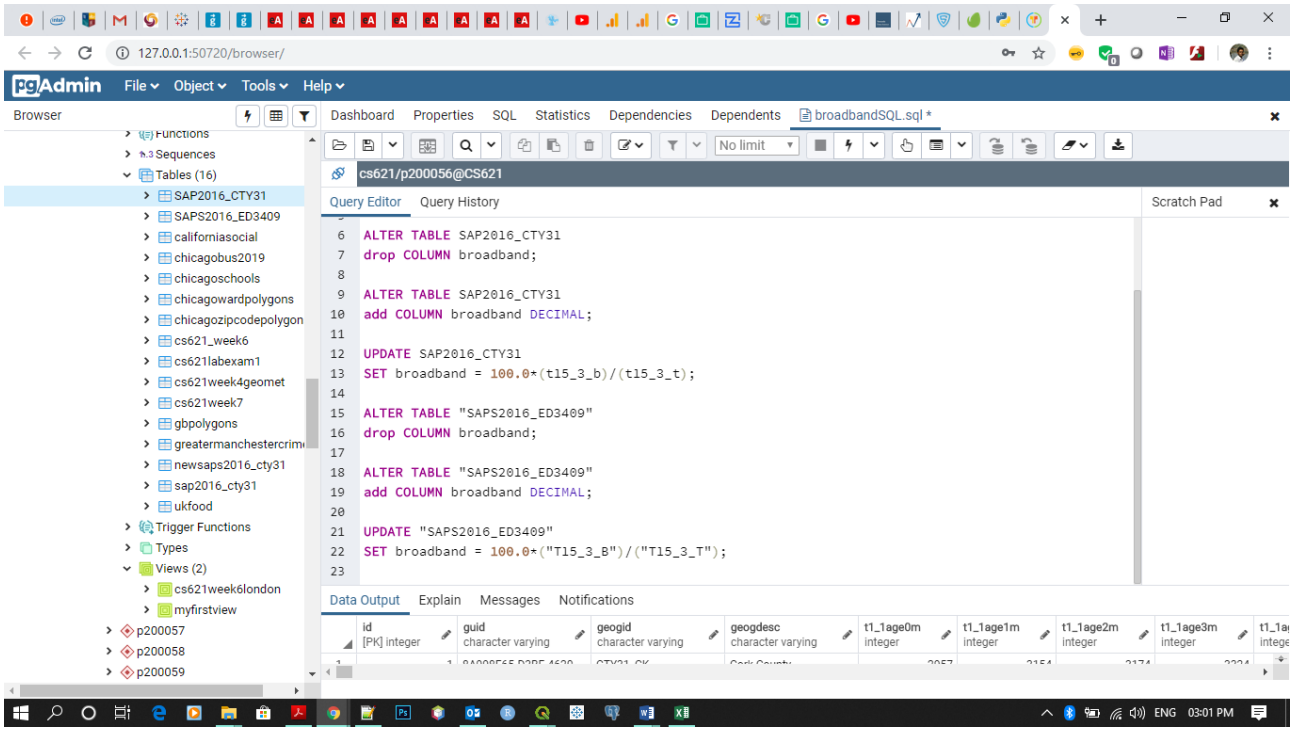


Figure 1.2.2.2

With this new column will get created for broadband percentage. Also, to map all these percentages we need to do the table join of shapefile and csv data file, we will discuss this is 'Creating maps' section. Once we do all above process for both the csv files then our data is ready for creating map.

2. Creating Maps

To create maps for the data that we have processed please follow the following steps,

- Open QGIS and load all the data and shapefiles as layers. Click on Layer menu option then Add Layer, then click on Add vector layer and load the shapefile and click on add button. Do the same for the other shapefile. To add the csv file, do the same but instead of vector layer click on Add delimited text layer.
- Right click on the shapefile from the Layer panel and click properties. Join the table by using Join option from the left panel and add labels from Label option. Do the same for the other shapefile as well.
- Please follow the steps shown in the figure 2.1 for creating the gradient pattern

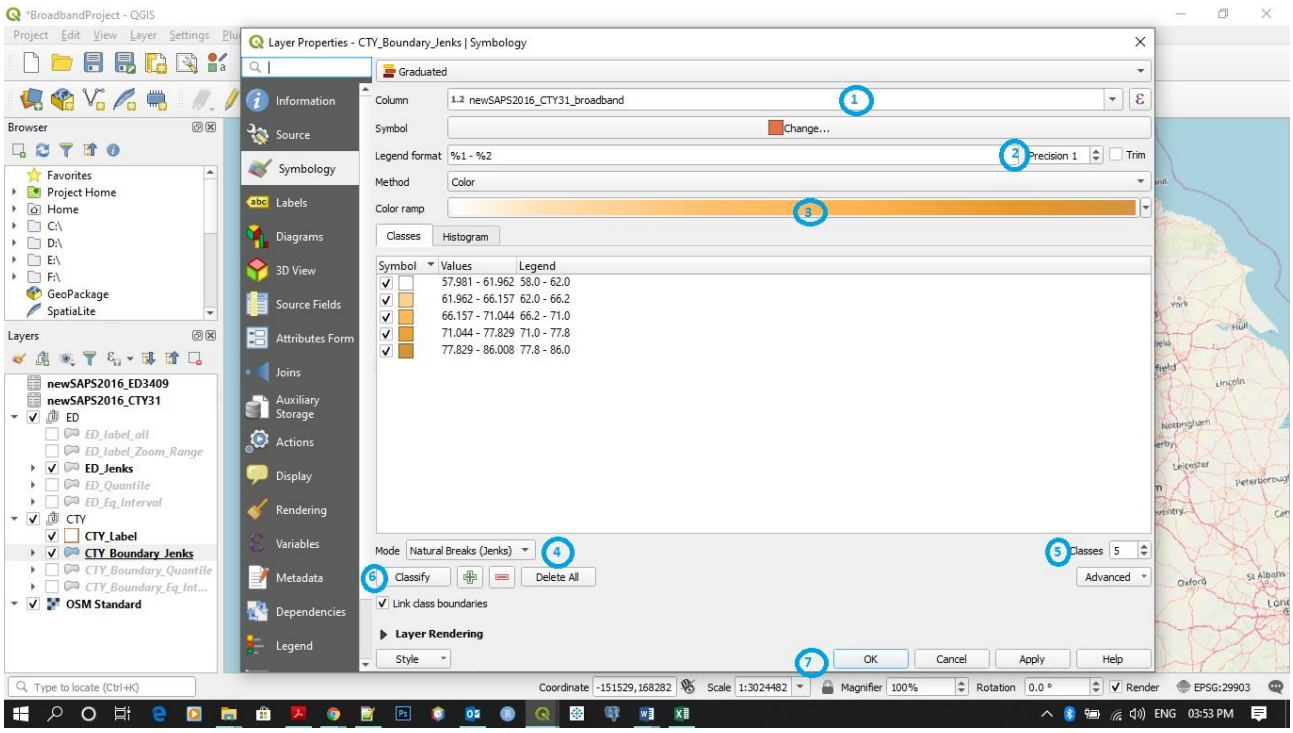
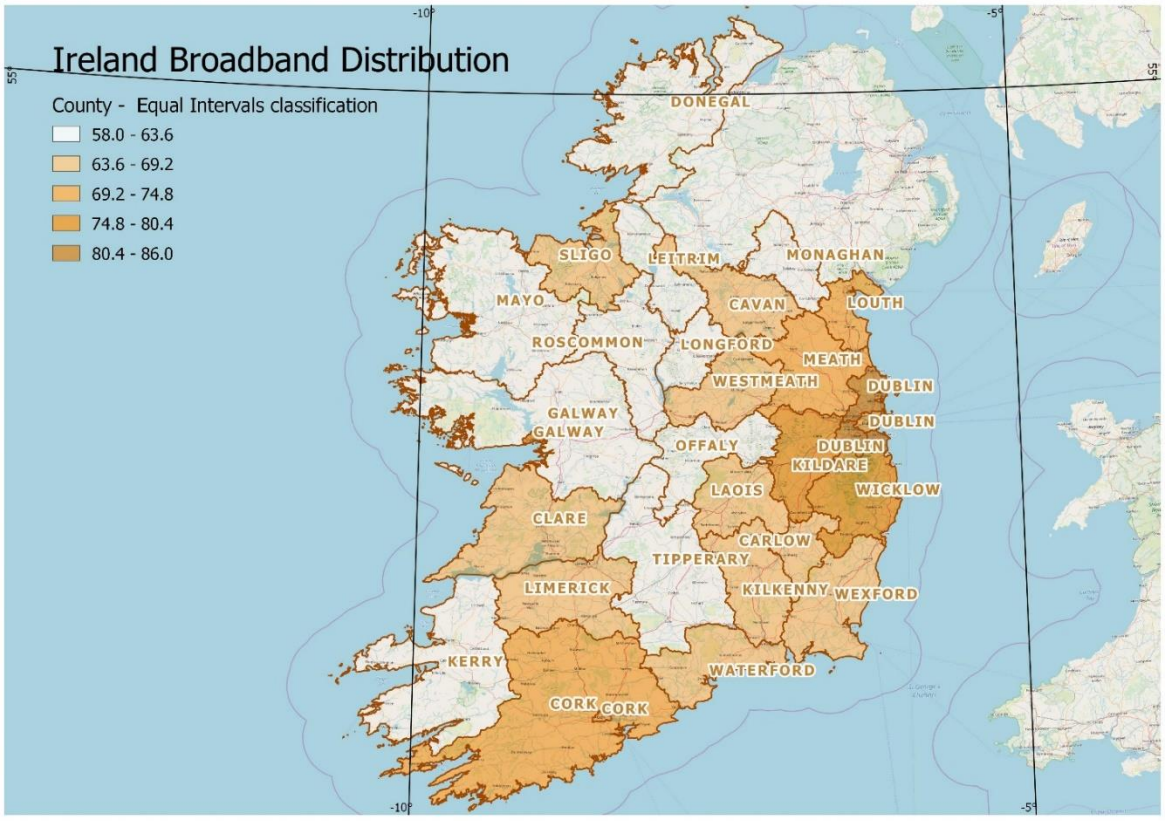


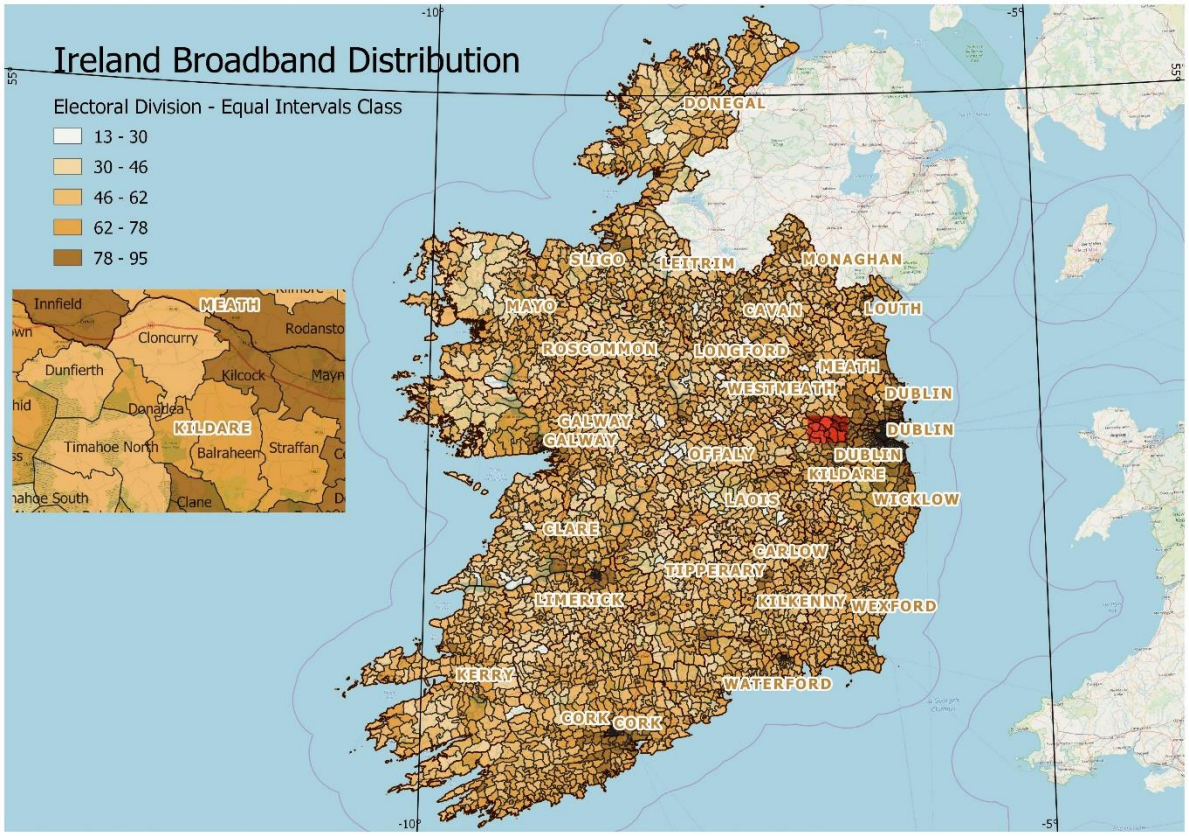
Fig.2.1

- Create 3 duplicate layer each for county and ED shapefile layers. For each of them change the mode from the properties fig.2.1 to Equal Interval, Natural breaks and Quantile respectively.
- In the end we will have 6 map layers in total. Please follow the link[3] provided in reference for more details.
- Create print layout for each of them and export it as image files

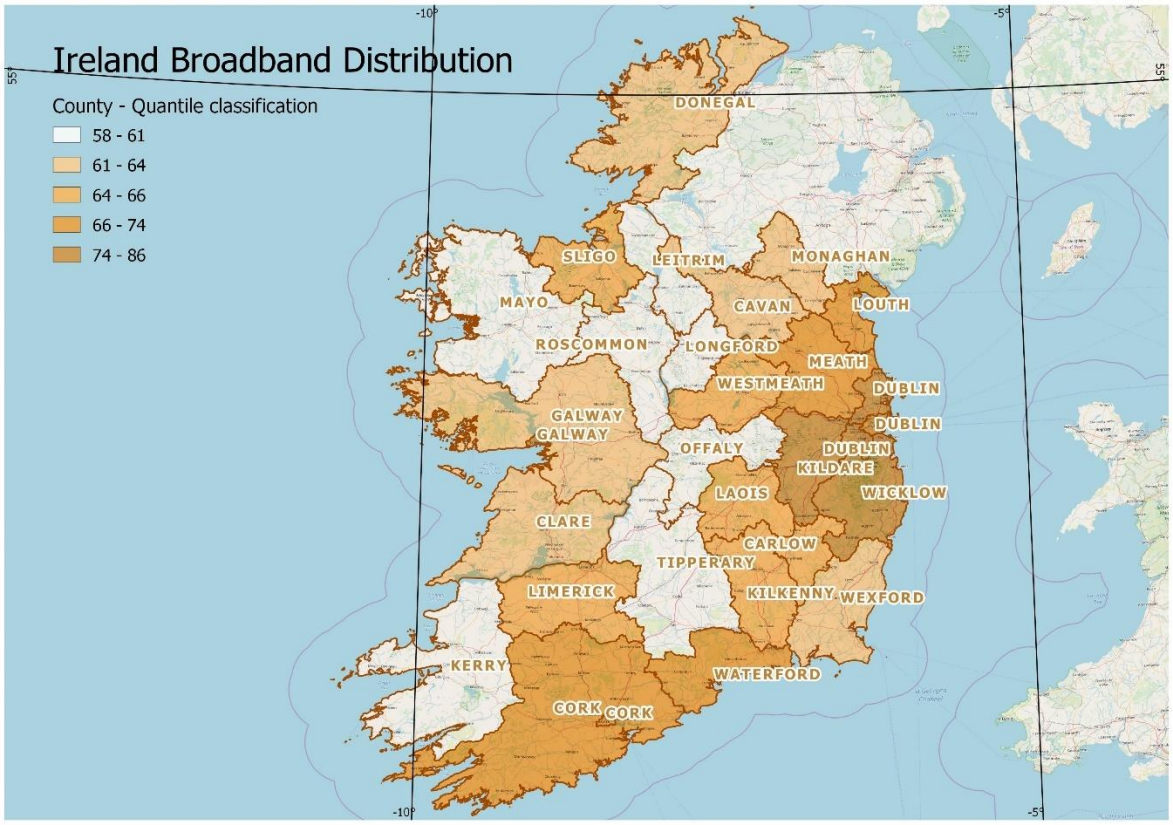
- County_Equal_Interval_map



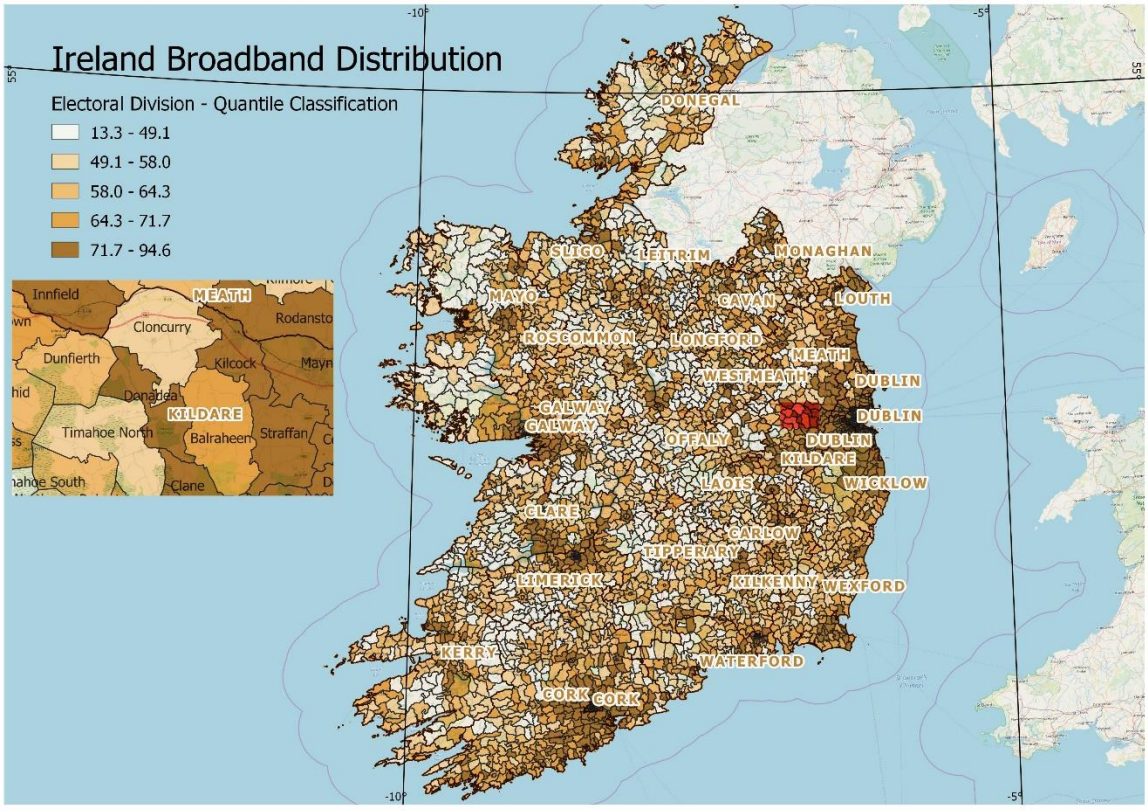
- ED_Equal_Interval_map



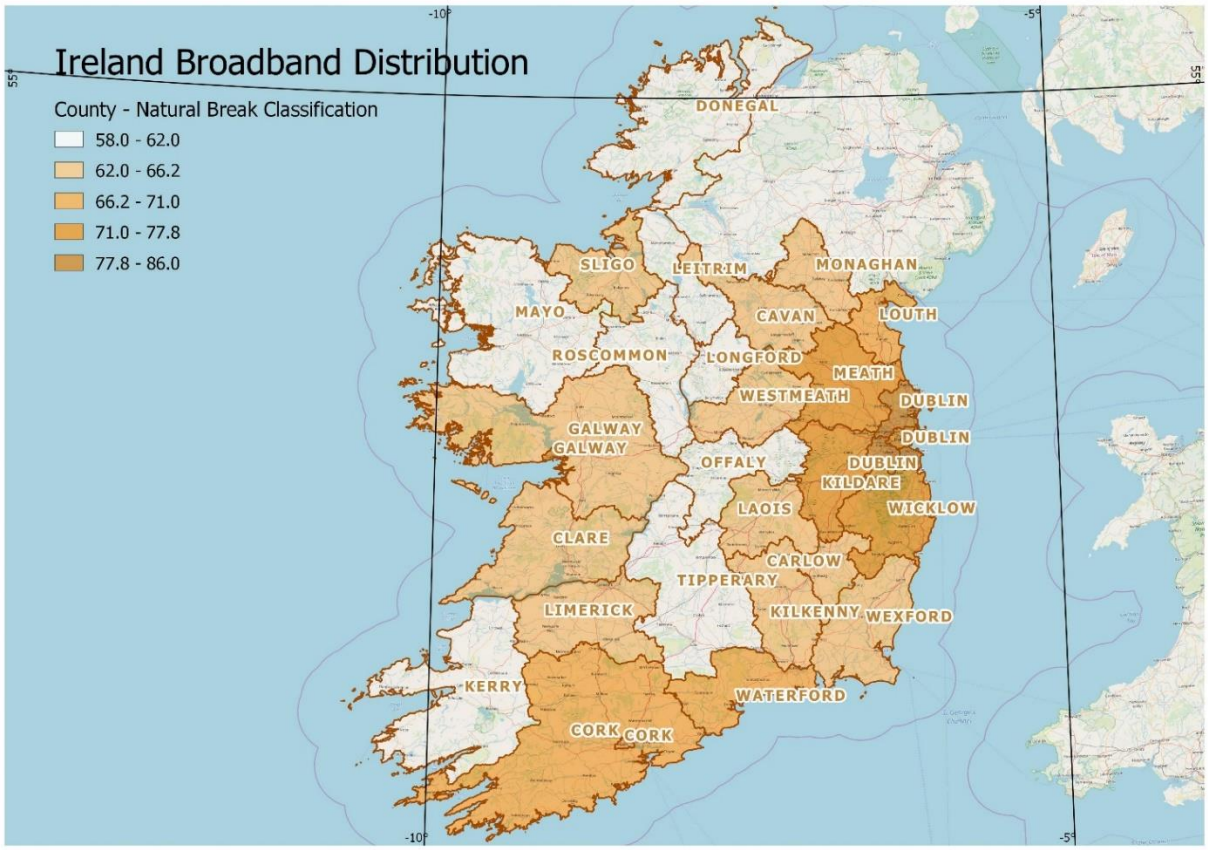
- County_Quantile



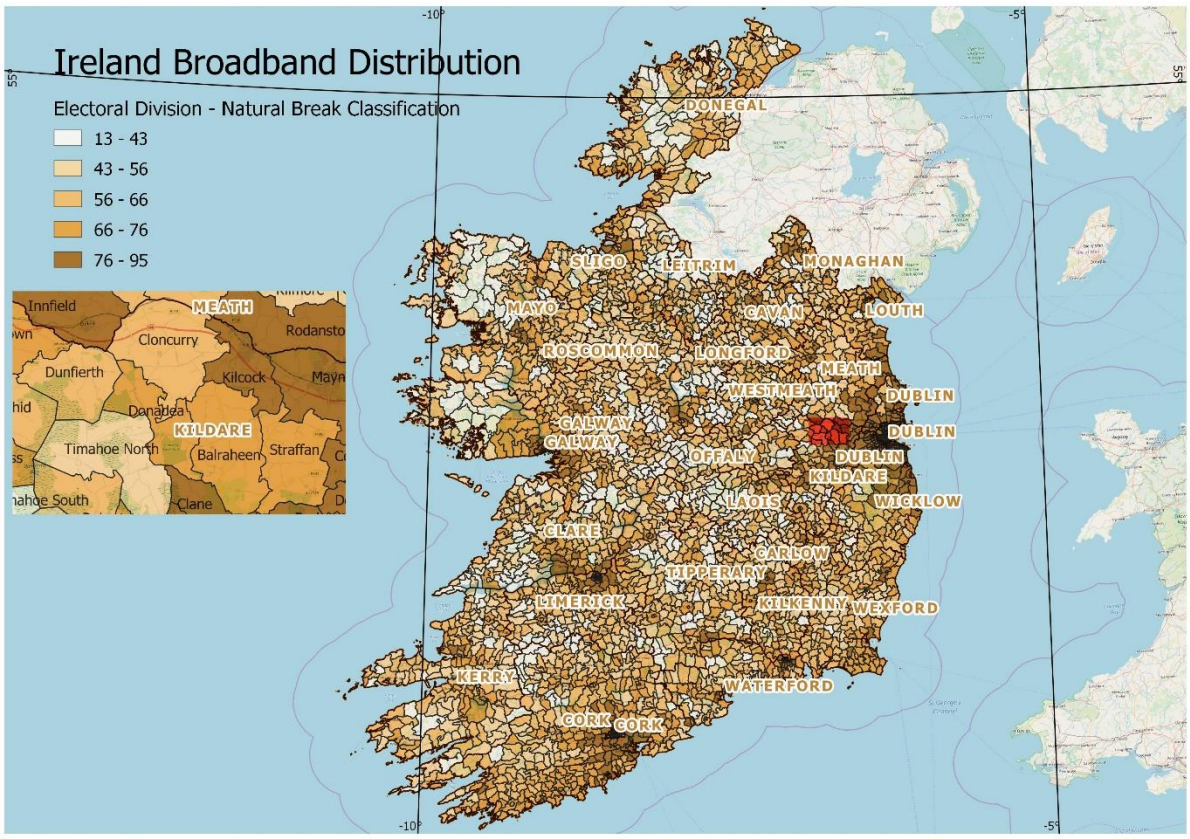
- ED_Quantile



- County_Natural_Break



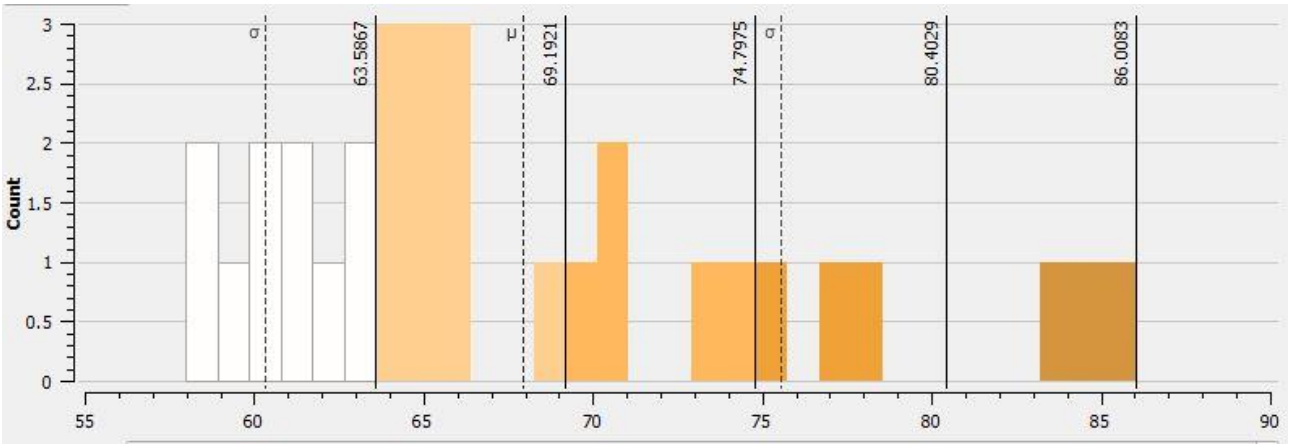
- ED_Natural_Break



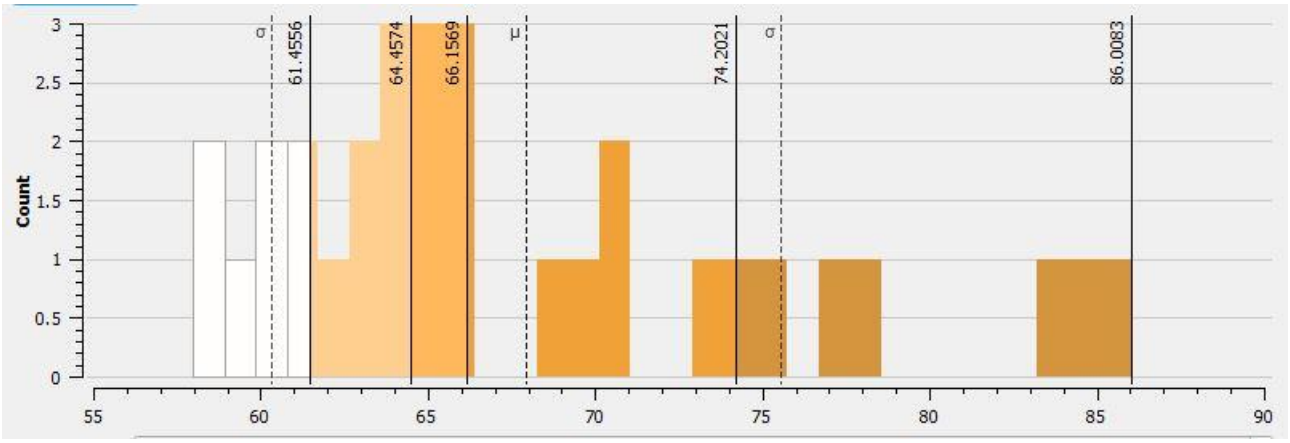
3. Discussion

The maps shown above for all three types of mode of classification is done based on the min and max values of the broadband percentage for each range. For more details we will investigate the histogram plots of the same.

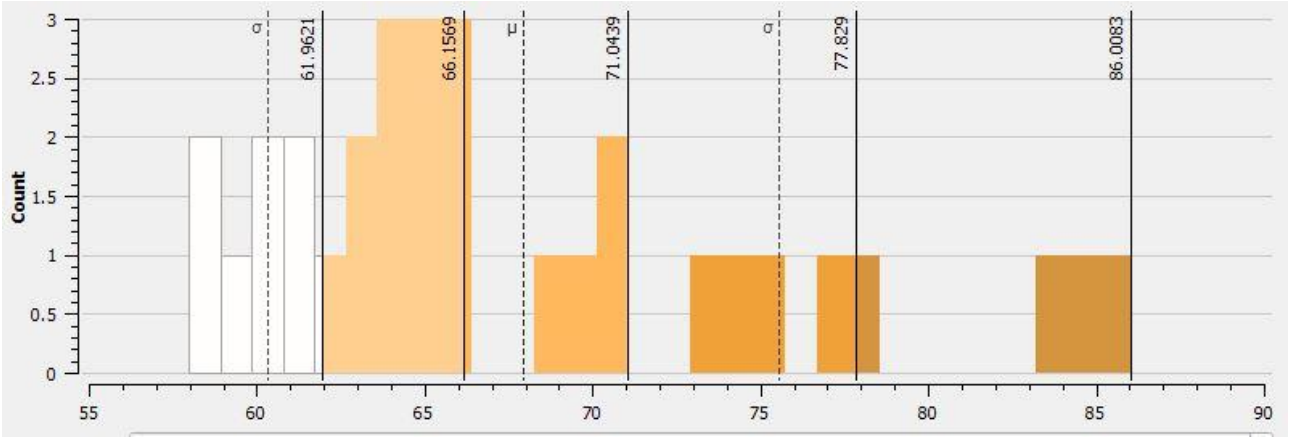
- County_Equal_Interval



- County_Quantile

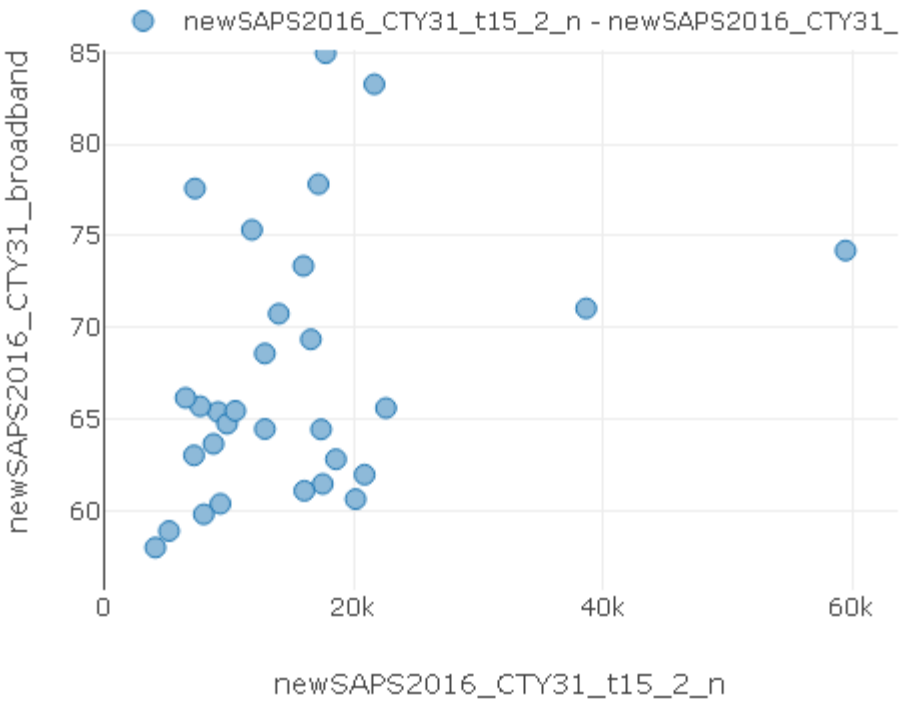


- County_Natural_Break

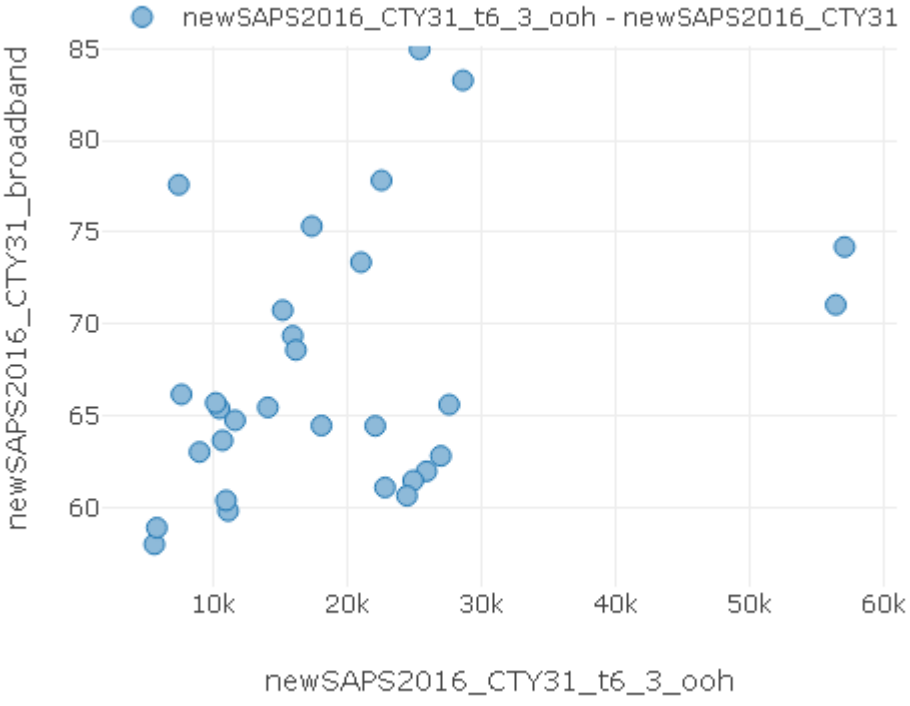


- By analysing data, we found out that there are few factors which affects the broadband distribution and plotting broadband values against various aspects we get the following results

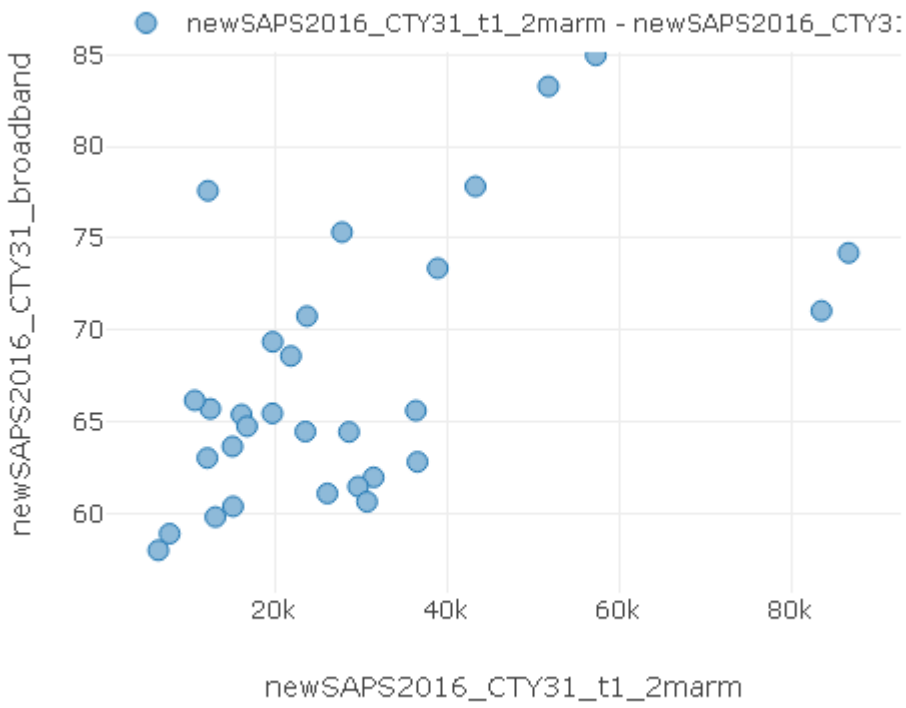
- Broadband vs Houses with Personal computer



- Broadband vs Owner occupied residents



- Broadband vs Married population



4. Conclusion

Based on the above data we can say that there are more usage/distribution of broadband network in the urban areas. From the scatter plots we can say that the relationship between married population and broadband distribution is linear and mostly owner-occupied houses has more usage of the network. Hence, we can say that broadband-network-distribution has linearity to the density of the population with personal computers.

REFERENCES

[1] https://www.cso.ie/en/census/census2016reports/census2016smallareapopulationstatistics/

[2] http://www.postgresqltutorial.com/

[3] https://www.qgistutorials.com/en/docs/3/making_a_map.html

[4] Full project is uploaded in GitHub: https://github.com/NanawareAmol/GIS_Broadband_Network_Distribution_in_Ireland