**Data Analysis and Visualisation**

**Module: Tools for Data Analytics**

**Student: Galina Lopez 10333429**

**Q1.** Using MS Excel, use Pivot tables and Pivot charts to provide the following analysis:

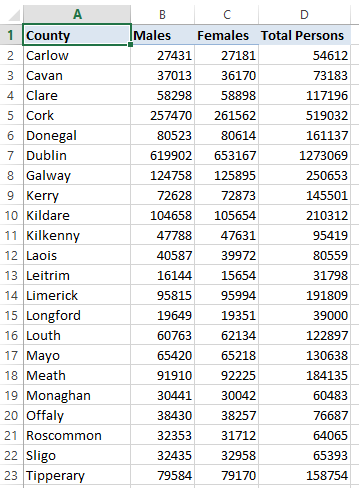
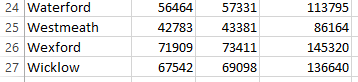
a. A chart which breaks down the population by gender.

b. A table which shows the total population by province.

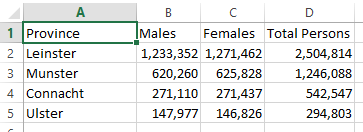
c. Use a pivot table to find out what % of the population does each county account for in 2011.

1. First we needed to extract, load and cleanse the census data in MS Excel into two tables.

First table of population: males, females, total persons for each of the 26 counties.

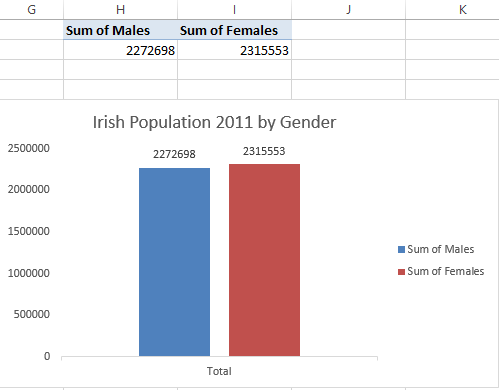
 

Second table of population: male, female, total persons for each of the 4 provinces.

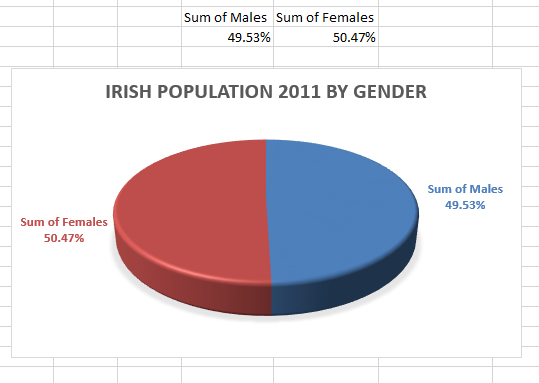


a. Pivot table and chart which breaks down the population by gender.

By sum of totals: males and females.

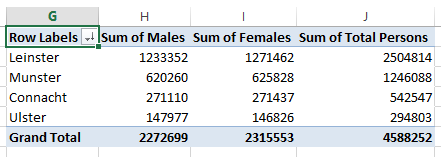


By percentages from the grand total: males and females.



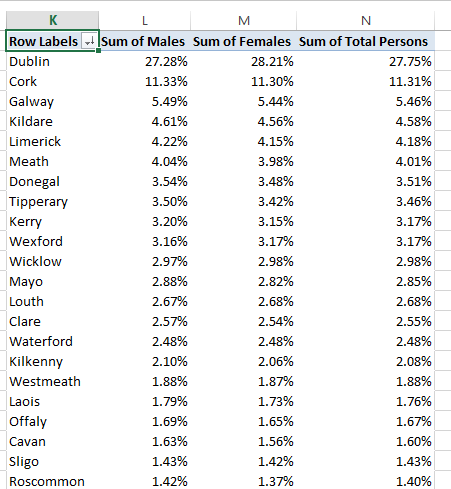
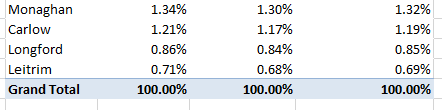
b. Pivot table which shows the total population by province.

Results sorted from largest to smallest to get better insight.



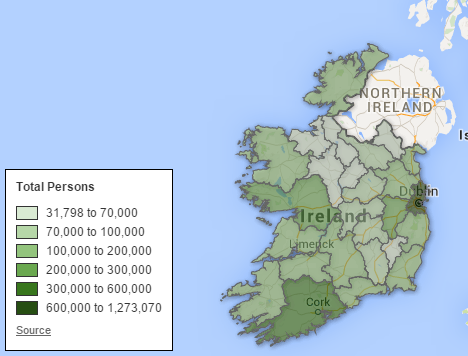
c. Pivot table to find out what % of the population does each county account for in 2011.

Results sorted from largest to smallest to get better insight.

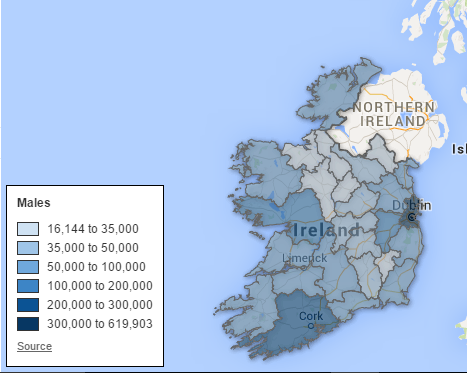
 

2. Using Google Fusion, create a heatmaps which visualises the population data per county.

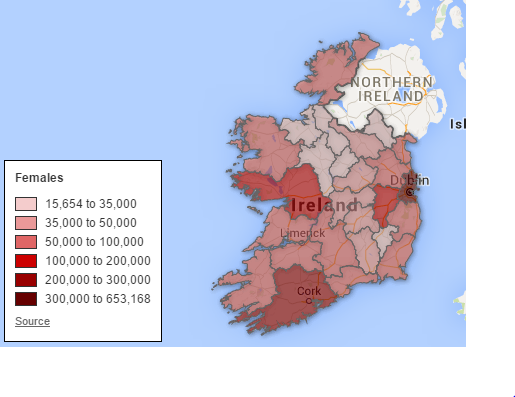
Total People per County



Total Males per County



Total Females per County



Link to the above Google Fusion Tables

https://www.google.com/fusiontables/DataSource?docid=1OGrbb2fV\_FEcWcvMEylEuhJ0XG6IjLoRc3XGZLZa

3. Write a report:

a. An overview of what benefits data analytics and data visualisation could bring to an organisation of your choice.

b. Your analysis of the population data.

c. A written description of

. How you achieved the heatmap.

. What information could be gleamed from the heatmap.

. What other ideas/concepts could be represented in the heatmap.

**3. Report**

**a. Overview of what benefits data analytics and data visualization bring to an organisation.**

Data Analytics is the discovery and communication of meaningful patterns in data. Especially valuable in areas rich with recorded information, analytics relies on the simultaneous application of statistics, computer programming and operations research to quantify performance. Analytics often favours data visualization to communicate insight. Data visualization refers to the techniques used to communicate data or information by encoding it as visual objects e.g., points, lines or bars contained in graphics. The goal is to communicate information clearly and efficiently to users. Well-crafted data visualization helps uncover trends, realize insights, explore sources, and tell stories. Data visualization is closely related to information graphics, information visualization, scientific visualization, exploratory data analysis and statistical graphics.

Firms may commonly apply analytics and data visualization to business data, to describe, predict, and improve business performance. Analytics is a multidimensional discipline. There is extensive use of mathematics and statistics, the use of descriptive techniques and predictive models to gain valuable knowledge from data—data analysis. Data visualization then communicates the insights from data used to recommend action or to guide decision making rooted in business context.

Marketing has evolved from a creative process into a highly data-driven process. Organizations like the Virgin Group, use analytics to determine the outcomes of campaigns or efforts and to guide decisions for investment and consumer targeting. Demographic studies, customer segmentation, conjoint analysis and other techniques allow Virgin marketers to use large amounts of consumer purchase, survey and panel data to understand and communicate marketing strategy. Data visualization aids this communication.

Web analytics allows marketers in Virgin to collect session-level information about interactions on their website. Google Analytics is an example of a popular free analytics tool that marketers use for this purpose. Those interactions provide the web analytics information systems with the information to track the referrer, search keywords, IP address, and activities of the visitor. With this information, a Virgin marketer can improve the marketing campaigns, site creative content, and information architecture.

Web analytics and optimization of web sites and online campaigns now frequently work hand in hand with the more traditional marketing analysis techniques. These tools and techniques support both strategic marketing decisions such as how much overall to spend on marketing and how to allocate budgets across a portfolio of brands and the marketing mix and more tactical campaign support in terms of targeting the best potential customer with the optimal message in the most cost effective medium at the ideal time. Through data visualization then these insights into a rather sparse and complex data set, communicate its key-aspects in a more intuitive way.  As a picture is worth a thousand words, data displayed graphically allows for an easier comprehension of the information. Proper visualization provides a different approach to show potential connections, relationships, etc. which are not as obvious in non-visualized quantitative data.

Virgin recognizes that being a successful, data-driven company requires skilled developers and analysts. They grasp how to use data to tell a meaningful story or picture that resonates both intellectually and emotionally with an audience. Virgin marketers are responsible for this story/picture; as such, they're often the bridge between the data and those who need to learn something from it, managers or executives that make decisions based on its analysis. They can tailor the story to that audience and effectively use data visualization to complement the narrative. They know that data is powerful and with a good story, it's unforgettable.

A good data visualization does a few things. It stands on its own; if taken out of context, the reader should still be able to understand what a chart is saying because the visualization tells the story. It should also be easy to understand. And while too much interaction can distract, the visualization should incorporate some layered data so the curious can explore.

Virgin marketers are responsible for messaging; as such, they are the bridge between the data and those who need to learn something from it, like managers or executives that make decisions based on its analysis and need to be grabbed in the first few minutes otherwise they lose interest. By rethinking the way they use data and understanding their audience, they create meaningful stories or pictures through data visualisation that influence and engage their audience on both an emotional and logical level. Using the right colours, charts, plots on their vivid dashboards their message gets through their audience much clearer making data analytics and data visualisation indispensable in the decision making process of the Virgin Group organisation.

**b. Analysis of the population data.**

We analysed the Irish populate census data from 2011. We extracted, loaded, and cleaned this data in MS Excel into two tables. The first one showed the Irish population 2011 figures by male, female and total persons for each of the 26 counties. The second one showed the figures for each of the 4 provinces.

We then used a pivot table and charts to show the breakdown of the population by gender in 2011. We got that Ireland had in 2011 more females than males. Females figure is 2,315,553 and males is 2,272,698. In the corresponding bar chart we noticed the difference was slight as the bars are closed in height and looking at the data confirms this. Their percentage over the total population, females 50.47 % and males 49.53% further proves this finding. A pie chart visualises this fact.

We then did a pivot table to show the total population by province. We sorted the result from largest to smallest to get a better insight. We got Leinster as the highest populated province in the country with 2,504,814 followed by Munster with 1,246,088 then Connacht with 542,547 and finally Ulster as the lowest populated province in the country with 294,803.

This finding make sense as Leinster is home to counties like Dublin where most of the Irish population lives in major cities like Dublin city which have great infrastructure hence attracting those big numbers. Munster have counties with the next highest population levels like Cork with cities like Cork city to account for being next in population number. Connacht has counties like Galway with cities like Galway city. While Ulster does not.

Another interesting find was that Leinster has double the population of Munster which in turn has more than double the population of Connacht which is nearly double the population of Ulster. So there is massive population differences between provinces, especially Leinster and Ulster. Leinster’s being 8.5 times higher than Ulster’s. Leinster accounts for 54.59% of the total population, Munster 27.16%, Connacht 11.82% and Ulster 6.43% for a ratio of 8.5: 4.2: 1.8: 1.

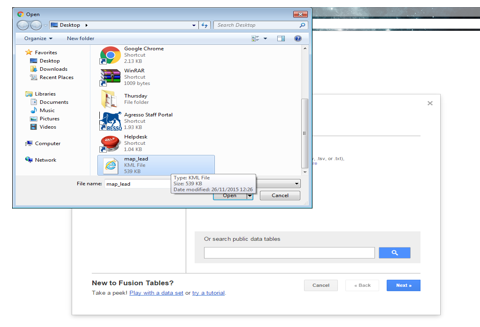
We finally did a pivot table to find out what percentage of the population does each county account for in 2011. We once again sorted the result from largest to smallest to gain insight into our result. Not surprisingly we got County Dublin at the top with 27.75% of the grand total followed by county Cork with 11.31% and Galway with 5.46%. At the bottom we had Leitrim with a very low 0.69% and Longford with 0.85%. These results support our previous province results. They particularly illustrate the massive gap in population levels between the counties at the top of the table and the counties at the bottom. Top counties have major cities on them and great infrastructure to account for their high population levels while the bottom ones don’t. Counties in the top half of the table belonged mainly to Leinster and Muster provinces while counties in the bottom half belonged mainly to Connacht and Ulster, further confirming our previous finding results.

**c. To achieve the heatmap:**

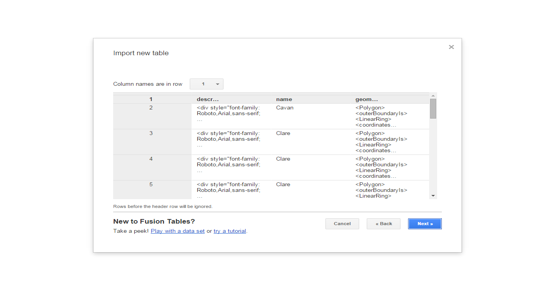
We log in to Google Drive. Select Google Fusion – Add New File.



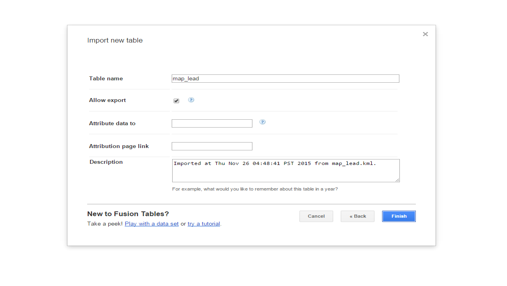
Choose File and browse to the kml File: map\_lead.kml and select it.



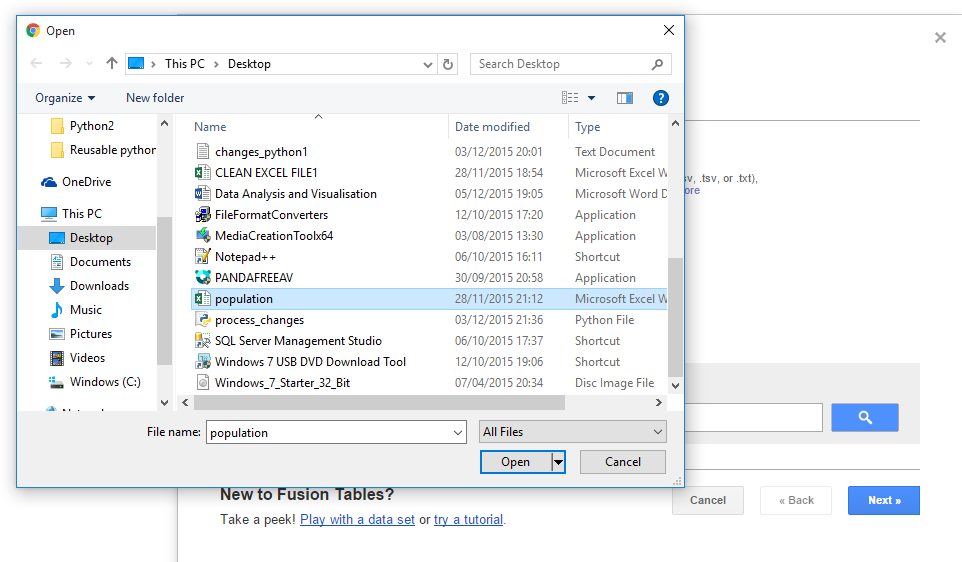
Click Open > Next



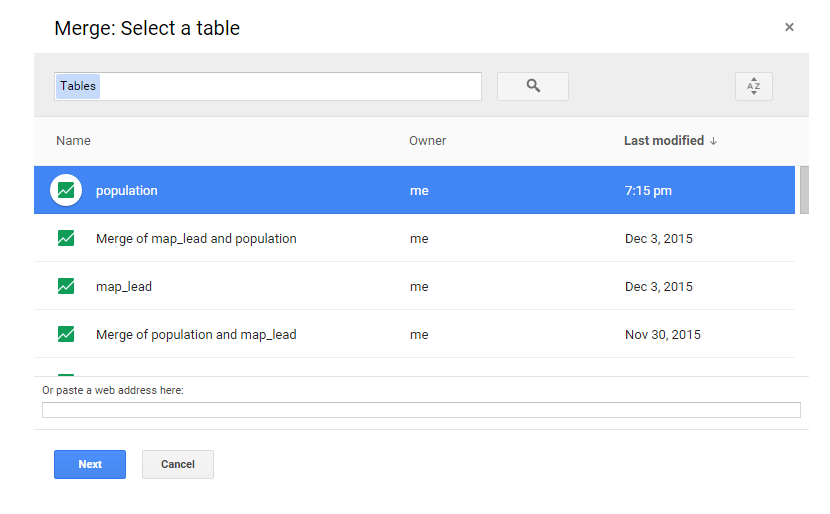
Click Finish



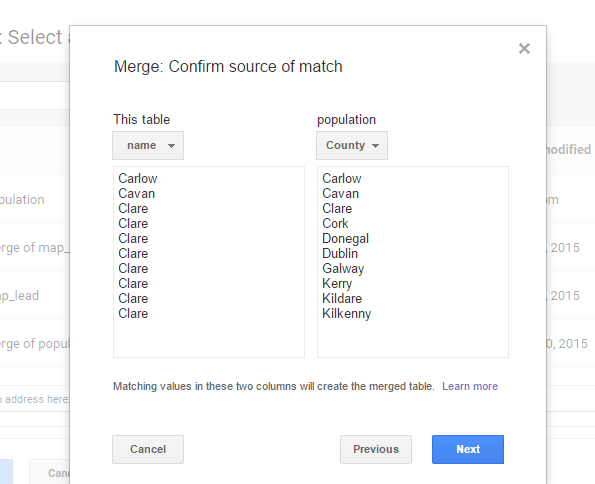
The file is now loaded in to Fusion Drive. Repeat the same steps with the Clean Excel File: population.xlsx.



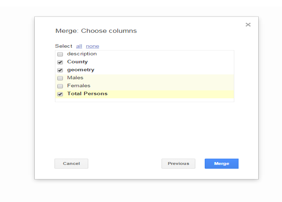
When the excel file loads Click File > Merge. Select the map\_lead\_klm file from list of loaded files. Click > Next.



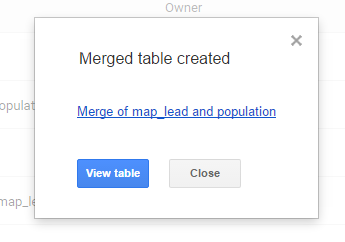
Select County fields in both tables to Merge on County fields. Click >Next.



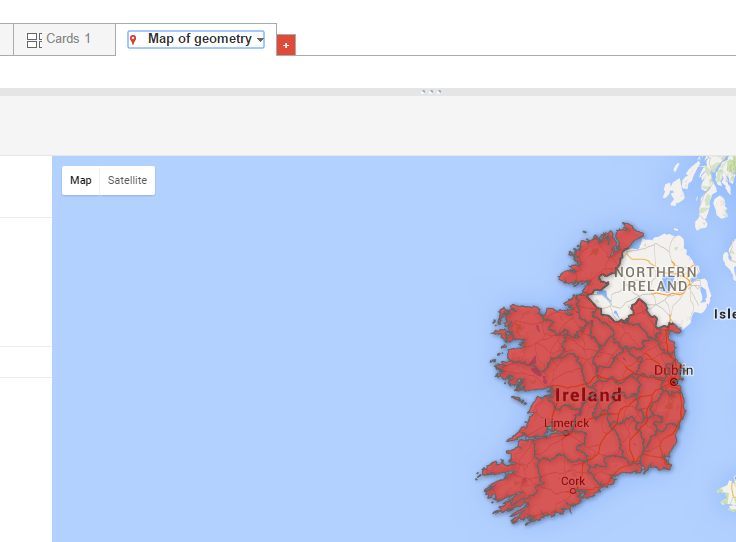
Deselect description, Males, Females Click > Merge.



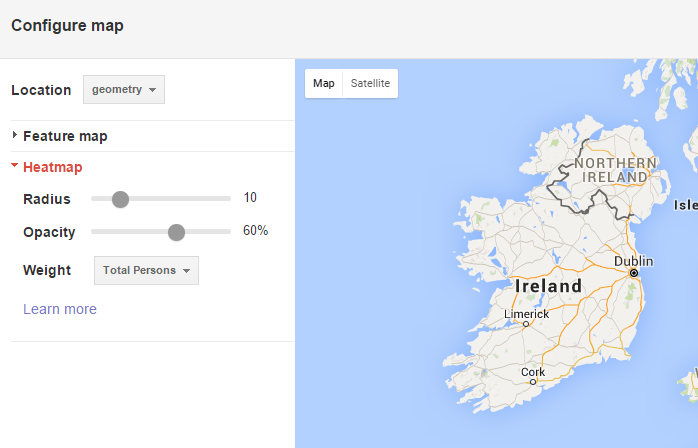
Click View Table.



Click > Map of geometry. All Counties are shown now.

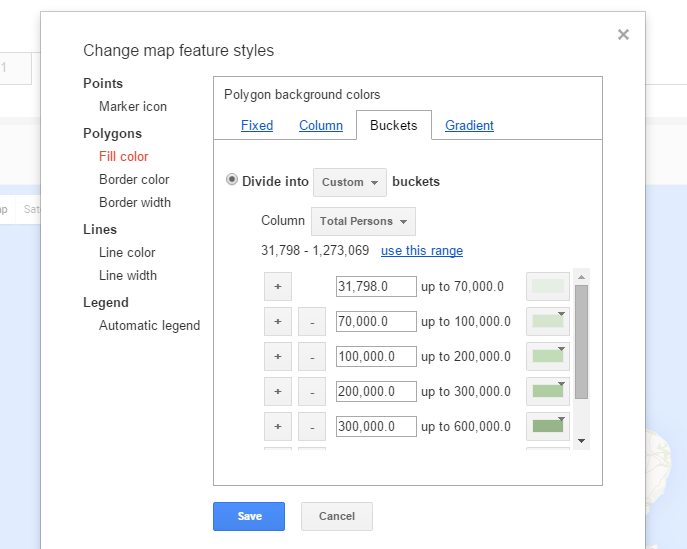


Click Heatmap. In the Weight field select Total Persons and adjust the Radius and Opacity controls.

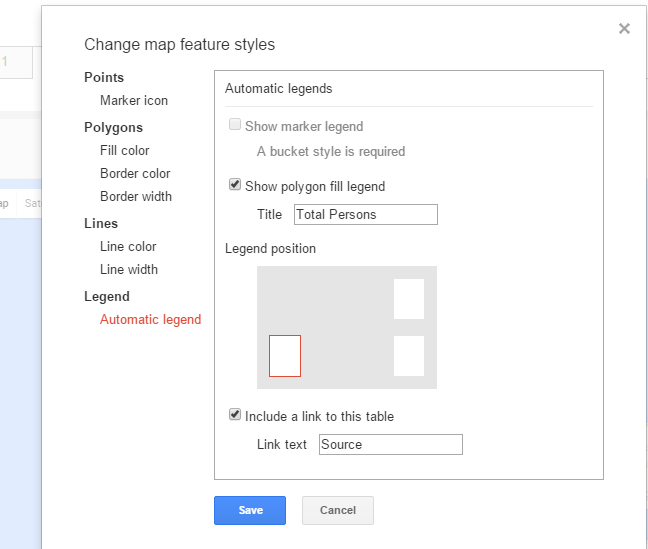


Click Feature Map to view results.

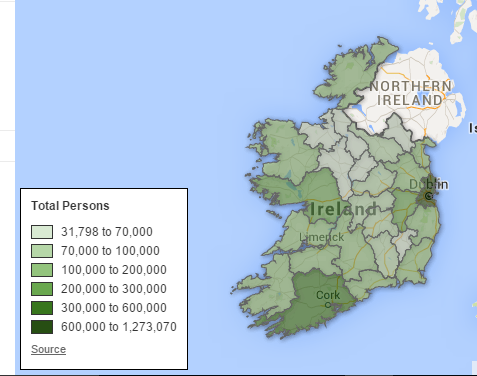
In the Feature Map Click Change feature styles… In the Change map feature styles select Polygon > Fill colour. In the Polygon background colours select the Buckets tab. Click the Divide into … Buckets and select number of buckets you need. I chose 6 to divide my data meaningfully. Choose colour to best represent the contrast of your data. I chose different intensities of the same colour to represent my data as it makes it easier to read and understand. Darker shades for higher population areas and lighter shades for lower population areas. Use the population range given. Divide your buckets into meaningful increments of it.



Now select Legend > Automatic legend. Tick the Show polygon fill legend box. Select desired legend position and a legend will be automatically added to your map.



Click > Save

`````

Follow above steps for Males and Females to get their corresponding Heatmaps.

**d. What information could be gleamed from the heatmap?**

The heatmap uses intensity of colour to show population levels by county. The darker the colour the higher the population level and vice versa.

The heatmap clearly highlights by the intensity of its colour, darkest green, Dublin as the county with the highest total population level in Ireland. It then points out with the next shade of dark green to Cork in the south and Galway in the west as counties with the following highest population levels in the country respectively. These findings back up our previous analysis of the population data in excel. Dublin, Cork and Galway are counties with the main cities on them. They have better infrastructure compared to their surrounding counties as well. Dublin County has the capital city, main airport, port, main colleges and hospitals. Cork has an airport, main college and hospital etc. Galway also has main college and is closer to America. Their superior road and overall infrastructure will attract companies to set up themselves in those counties therefore creating more jobs. So people will naturally gravitate towards these counties. In the east to Dublin, in the south the Cork and in the west to Galway as they offer more jobs, opportunities, amenities for their population.

The heatmap also highlighted in the lightest shade of green, a corridor with the lowest population level in the country, under 70,000, comprising Sligo, Leitrim, Roscommon and Longford. The eastern counties on this corridor, Leitrim and Longford, being the lowest populated counties in the country with population figures below 40,000, were showing in the map almost white. Monahan nearby also showed low population as well as Carlow still in the Midlands with figures below 70,000 as well. The next corridor of low population level, less than 100,000, was also highlighted in the midlands with the next lightest shade of green, comprising the counties of Westmeath, Offaly, Laois and Kilkenny.

In general, the midland counties experience lower population levels compared to the rest of the country as population tends to gravitate towards the counties with the better infrastructure, major cities, amenities and job opportunities.

Dublin, Cork and Galway have the highest population levels showing in the map with the darker green. While Leitrim and Longford have the lowest showing with the lightest green shade.

**e. What other ideas/concepts could be represented in the heat map?**

With this project we have done primary research and now we can do some secondary research of the census data. The census questionnaire also gathers extra information of its subjects, like marital status, employment status among other things. So we could perform further analysis on this data. Specific examples like the number of employed or unemployed males and females, the married or unmarried percentages of the population, single parent families or two parent families etc. All this data gathered in the census we need to extract, clean and load into excel to merge to the kml file so we can create new heatmaps from which to gather new insights into this data as part of our secondary research. This secondary research by being more specific will be of great value to specific government departments or organisations/companies or anyone depending on this specific data for future decision making.

We could also get the data from previous census and visualise population growth or decline by counties over the years. We could use filters of the data sets to visualise population levels by specific counties or use other filters to get population figures by gender as well as geographical geometry. We could merge our census data sets with maps of province geographical boundaries and get visual insights into province population levels as well.

All this information could help us find out population trends over the years and predict likely areas of population growth or decline for the future. This information could be very useful to government in their planning strategies, budgeting and allocation of resources.

**References:**

Google website, Virgin Group.

Wikipedia, Data Analytics and Data Visualisation.