

TECHNICAL REPORT

National College of Ireland

BSc(Hons) in Computing(Data Analytics)

2016/2017

Identifying new Health Centre Locations in
Leinster, Ireland



Student: Declan Barnes
Email: Declan.Barnes@student.ncirl.ie
Student No: x13114352
Supervisor: Dr. Paul Stynes

Declaration Cover Sheet for Project Submission

SECTION 1 *Student to complete*

Name: Declan Barnes
Student ID: x13114352
Supervisor: Dr.Paul Stynes

SECTION 2 **Confirmation of Authorship**

The acceptance of your work is subject to your signature on the following declaration:
I confirm that I have read the College statement on plagiarism (summarised overleaf and printed in full in the Student Handbook) and that the work I have submitted for assessment is entirely my own work.

Signature: **Declan Barnes**

Date: **10/05/2017**

NB. If it is suspected that your assignment contains the work of others falsely represented as your own, it will be referred to the College's Disciplinary Committee. Should the Committee be satisfied that plagiarism has occurred this is likely to lead to your failing the module and possibly to your being suspended or expelled from college.

Complete the sections above and attach it to the front of one of the copies of your assignment,

Table of Contents

IDENTIFYING NEW HEALTH CENTRE LOCATIONS IN LEINSTER, IRELAND	1
EXECUTIVE SUMMARY	5
1 INTRODUCTION	6
1.1 BACKGROUND	6
1.1.1 <i>Research</i>	7
1.2 AIMS	7
1.3 TECHNOLOGIES	8
1.4 STRUCTURE	9
2 SYSTEM	10
2.1 FUNCTIONAL REQUIREMENTS	10
2.1.1 <i>Output Requirements</i>	10
2.1.2 <i>User requirements</i>	10
2.1.3 <i>Data requirements</i>	10
2.1.4 <i>Use Case Diagram- Overall System</i>	11
2.1.5 <i>R1: Identify Health Centre locations</i>	12
2.1.6 <i>R2: Identify locations of people with specific disabilities</i>	13
2.2 NON-FUNCTIONAL REQUIREMENTS	15
2.2.1 <i>Performance/Response time requirement</i>	15
2.2.2 <i>Availability requirement</i>	15
2.2.3 <i>Recover requirement</i>	15
2.2.4 <i>Robustness requirement</i>	15
2.2.5 <i>Security requirement</i>	15
2.2.6 <i>Reliability requirement</i>	15
2.2.7 <i>Maintainability requirement</i>	15
2.2.8 <i>Extendibility requirement</i>	16
2.2.9 <i>Reusability requirement</i>	16
2.2.10 <i>Resource utilization requirement</i>	16
2.3 DESIGN AND ARCHITECTURE	16
2.4 SYSTEM EVOLUTION	17
2.5 IMPLEMENTATION	18
2.5.1 <i>Identifying new HSE Support Centre Locations</i>	18
2.5.2 <i>Disability Application</i>	24
2.6 GRAPHICAL USER INTERFACE (GUI) LAYOUT	28
2.7 TESTING	29
2.7.1 <i>Integration Testing:</i>	29
2.7.2 <i>Usability Testing (with 2 users):</i>	29
2.7.3 <i>Testing Summary:</i>	30
2.8 EVALUATION	30
3 CONCLUSIONS	31
4 FURTHER DEVELOPMENT OR RESEARCH	32
5 MOTIVATIONAL SECTION	33
6 BIBLIOGRAPHY	34

7	APPENDIX.....	35
7.1	PROJECT PROPOSAL	35
7.2	PROJECT PLAN	42
7.3	MONTHLY JOURNALS	46
7.4	USABILITY TESTING-USER FORMS	53

Executive Summary

This report describes the analysis of datasets and shapefiles to identify HSE support centre locations based on the location and population number in Leinster, Ireland. This analysis is useful for executives in the Health Service Executive. It provides them with planning details to optimise where their resources should be used across Leinster to provide better coverage to people with a disability. This report used various data analysis processes to complete the tasks that were set out. One of these processes was to use the KDD(Knowledge discovery and databases) in order to properly conduct the analysis on the health centre and disabilities datasets being used, such as data cleaning and data transformation. The KDD was followed as it is used substantially in data analytics to gain knowledge of data. The proposed system identifies new HSE support centres locations using the current location of HSE support centres and the population density in each small area of Leinster, Users will also be able to see a breakdown of disabilities such as: the types of disabilities(physical, hearing loss, blindness etc.) the locations of these disabilities shown using a graphical representation in the form of a Java Desktop application.

1 Introduction

A major concern in the HSE is identifying suitable locations for support centres in the country where there is a lack of disability support services due to budget cuts. For example, John Dolan, CEO of DFI stated on the disability federation website (John Dolan, 2015, www.DisabilityFederation.ie) that cuts of €159 million since 2008 to disability services has left a gaping hole in the support services across the country. Disability services such as speech therapy for those who have the likes of down syndrome, physical therapy for those with physical disabilities, learning support for those with intellectual disabilities have all been affected. This report aims to identify the locations of new support centres in Leinster, Ireland, so that more services for people with a disability become available especially in rural areas where the lack of services is worse (NDA, www.NDI.ie). These new health centres will be placed using the population density in the small areas of Leinster that do not have a HSE support centre, the new locations where the centres will be of the most value to those in need of disability support services. Also, as part of the above goals, is the showing relevant information on disabilities in Leinster such as showing the amount of male and females in Leinster who suffer from some form of disability, showing the different age groups suffering with disabilities, and the types of disabilities (Intellectual, physical, hearing loss, blindness, cognitive etc.). This analysis will be shown in a java application.

1.1 Background

It is important to try to understand or work out where there are limitations in the services provided to disabled people across Leinster, Ireland and what's the best locations for the new HSE support centres and where they will have the most benefit. This report will show the limitations in a constructive way that will help improve the quality of the services that disabled people rely on daily for their everyday needs. It would be beneficial overall to show where new support centres should be positioned in Leinster as it is unacceptable to expect disabled people in rural areas specifically, to travel a significant distance to get to their nearest support centre, while these same support centres could be failing to provide adequate services/coverage to the people who need them by being out of reach. On the disability-federation website it states the following in regards to disabilities (DFI, 2015, www.DFI.ie): People in Ireland that are living with a disability and parents who have special-needs children are at a much higher risk of suffering depression than those who do not. Identifying new support centre locations would be beneficial to disabled people and the parents of special needs children as these support centres could provide very important care and or support to these people to help relieve the pressure that they live with daily by creating more health centres that are much better placed to provide the necessary coverage across the country that they need. Another interesting fact that was noticed through research was the attitudes towards people of a disability. On the national disability authority, it stated that negative attitudes towards disabled people had become a significant barrier to the inclusion of the disabled into various aspects of society and therefore preventing people with a disability from reaching their full potential. The National Disability Authority has conducted a series of surveys at regular intervals on the public's attitude towards people living with a disability (2001,2006,2011). The NDA's surveys showed that attitude towards

disabled people improved in the years 2001 and 2006, but the year 2011 showed a decrease in attitudes.

1.1.1 Research

The research conducted involved searching various online sites such as the HSE, Disability Ireland and Disability.ie to see if there is anything similar on the market. After searching online for an extended period, it became apparent that while there was similar analysis done on disabilities in Ireland by government run organisation such as the HSE, there was no analysis implemented that would identify the locations of where new HSE support centres could be placed in the Leinster. After researching for several more days, it was clear that the identification of where new HSE support centres could be placed regarding this project had not been done before by other people or organisations specifically in relation to Leinster, Ireland.

1.2 Aims

The **aim** of identifying new locations for HSE support centres in Leinster, Ireland was to acquire the necessary information regarding where the current support centres were located and the population amount throughout Leinster. In order to fulfil the aim there was a need to investigate several datasets such as the Health Centre dataset and small areas shapefile acquired from the Central Statistics Office. The Health centre dataset contained key information such as the description of the support centre, its latitude and longitude and the small area shapefile contained the population amount of each small area of Leinster. The analysis aspect will also need data to display the findings, this data will be like the following: how many people are living with a disability in Ireland, what type of disability do they have, what sex and age are they. The information that will be used will come from the datasets mentioned above that were acquired from public resources such as the central statistics office.

The **main objective** will consider where current support centres are located and what the population amount is in each small area of Leinster to identify where new centres could be built i.e. in areas where there is no current HSE support centre and the population in those areas is over a required amount and then it will be shown in a graphical representation where the new centres could be placed in relation to this information.

The information that will be shown will be used to represent the key analysis of the locations, for example: how many people live in that area and what services do nearby support centres provide.

1.3 Technologies

The technologies that will be used are MySQL, the R programming language, and java. MySQL will be used to store the datasets containing all the variables needed to complete the research that is being set out. The R programming language will be used to create scripts that will identify the health centre locations. R will also produce the necessary analysis by pulling the data from the MySQL database and performing data manipulation and data cleaning. Finally, data visualisation will be used to visualize the findings with charts through a java application.

RStudio:

RStudio is a free and open-source integrated development environment (IDE) for R, a programming language for statistical computing and graphics. RStudio also allows the use of packages, each package helps the user perform certain tasks. Some of the packages that will be used for this project are:

- **RMySQL:** Database interface and MySQL driver for R.
- **TidyR:** Used for data manipulation and cleaning.
- **Stringr & Magrittr:** Data cleaning.
- **SP:** Provides classes and methods for spatial data
- **Rdgal:** Provides access to projection and transformation operations.
- **Plyr:** Also, used for data manipulation, usually used in conjunction with TidyR.
- **GoogleVis:** Data Visualisation.
- **Sqlf:** Provides an easy way to perform SQL selects on R data frames.
- **ggplot2:** Used for Visualizing data.
- **Leaflet:** Used to create interactive maps in R.

MySQL:

MySQL will be used for the storage of all the datasets. The reason for using MySQL is due to the author's familiarity with it. It also has a relatively straight forward interaction with R and RStudio through R's RMySQL package which will make interacting with my datasets friendlier.

NetBeans:

NetBeans is an open source IDE used to develop desktop, mobile and web applications with multiple programming languages such as Java, C++, HTML5. NetBeans will be used to create the java application portion of this project; this application will be created in Java and it will be a three-tier architecture system.

1.4 Structure

The structure is as follows:

Chapter 1: Describes the introduction, background/research, aims, technologies and structure.

Chapter 2: Describes the system, which contains the pages related to the following: functional requirements, output requirements, data requirements, user requirements, use cases and all the non-functional requirements that apply.

Chapter 3: Describes the conclusion in relation to this research paper.

Chapter 4: Further development or research describes how the research and project could be expanded in the future, if required.

Chapter 5: References- contains all the references used for this research paper.

Chapter 6: Appendix, the appendix contains the old project proposal, project plan, the monthly journals, and other material used.

2 System

2.1 Functional Requirements

This report should identify the locations of where new HSE support centres should be placed in Leinster, Ireland to provide better support services to people with disabilities. The java application should allow a user to search for a specific disability and see a heat map showing the amount of that disability across the Leinster.

2.1.1 Output Requirements

2.1.1.1 Identify new HSE Support Centre Locations

As part of this report, new support centre locations will be identified. By identifying where these centres could be placed there is hope that they will increase the quality of service to those with disabilities around Leinster by making more of these support centres readily available and reducing the amount of travel the disabled have to do to access these support centre services. These new centre locations will be displayed using RStudio and it will be displayed as an interactive spatial map using the leaflet package.

2.1.1.1 Descriptive Analysis (java Application)

Descriptive statistics, and the summarization of the data with visuals i.e. graphs: heatmaps etc. Explore and analyze disabilities in Leinster, Ireland using a number of variables such as: gender and type of disability. A java desktop application will be used to output the analysis.

2.1.2 User requirements

The User requirements for this report is that the user should be able to look at all data associated with identifying HSE support centre locations and the information shown should be easy to understand, while the java application itself will need to be clean and well presented for the user to easily understand and use the application to find the information they are looking for.

2.1.3 Data requirements

The following data requirements for this report are subject to change throughout the project. For the java application, the data requirements are that the datasets acquired contain relevant information that will allow a decent depth of analysis to be performed, these would include: disability type, age-group, and sex. For the identification of potential HSE support locations a dataset containing the latitude and longitude of the current health centres will be needed as well as a shapefile of Leinster's small areas and further information such as: population. The following is an example of the variables needed in the datasets that will be used for the project:

Data:

Sex: Male & female,

Age Group: 0-14,15-44, 25-44, 45-64, 65+,

Types of disability:

Blindness or a serious vision impairment, Deafness or a serious hearing impairment, A condition that substantially limits one or more basic physical activities, an intellectual disability, Difficulty in learning, remembering or concentrating, Psychological or emotional condition, Other disability, including chronic illness, Difficulty in dressing, bathing or getting around the home, Difficulty in participating in other activities.

HSE Support Centre locations:

Locations: Latitude/Longitude, small-area shapefile

Population: Amount by small areas

2.1.4 Use Case Diagram- Overall System

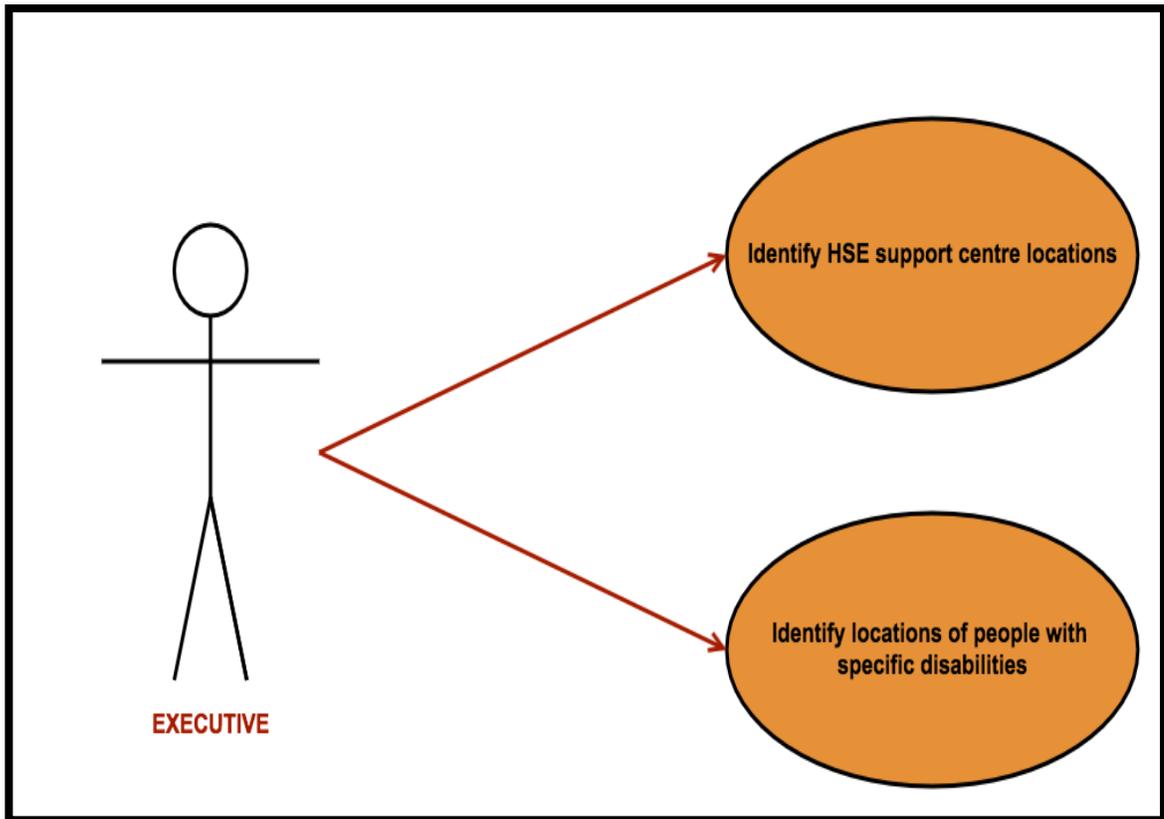


Figure 1: Overall System

2.1.5 R1: Identify Health Centre locations

2.1.5.1 Description

In this requirement, the locations of where HSE support centres should be placed in the future are decided using the small areas of Leinster that do not have a current support centre and have a population in those areas over a specific amount. The <executive> will be able to see where throughout Leinster, these new support centres could be located, so they can make an informed decision of where to move HSE resources and funds in Leinster and build new support centres in those areas to provide more adequate services to people with disabilities

2.1.5.2 Use Case

Scope

The scope of this use case is identifying the locations of HSE support centres in Leinster, Ireland that provide services to the disabled.

Use Case Diagram

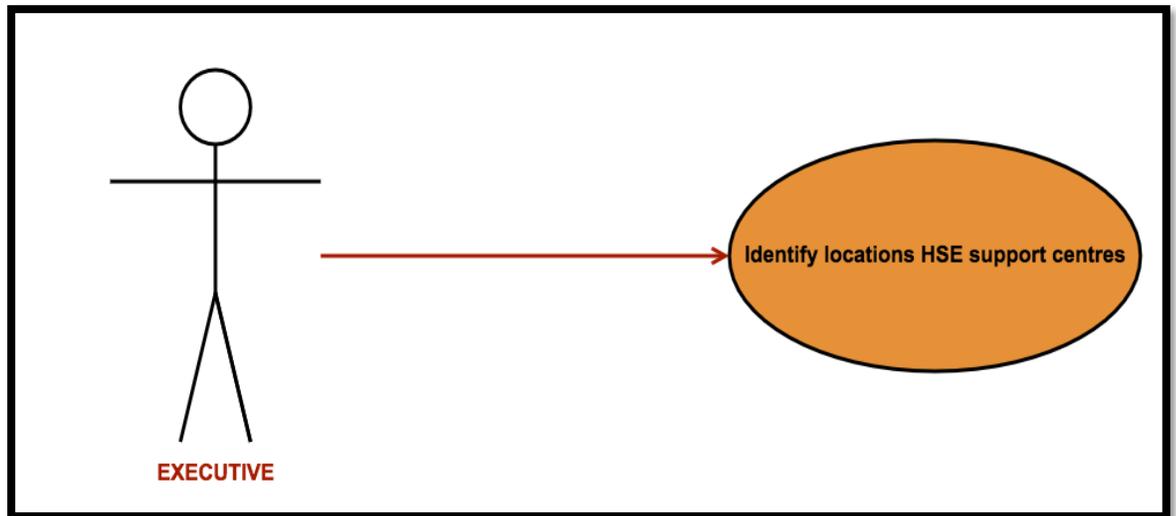


Figure 2: Identify Health Centre Locations

Flow Description

Precondition

None. This is the first requirement; therefore, it can begin straight away.

Activation

This use case starts when the <executive> generates a map that identifies the possible locations of new support centres.

Main flow

1. The <executive> accesses and runs the system to generate the spatial map.

2. The <executive> analyses the spatial map where the current support centres, population density and new locations for HSE support centres are shown.

Exceptional flow

E1: < Support centre location Error >

1. The system cannot generate spatial map.
2. The <executive> seeks support to fix the error.
3. The use case continues at position 5 of the main flow

Termination

This use case will be aborted if there is no data to be found, and if there is a problem with the database.

Post condition

The <executive> can use the information shown to move resources in the country to provide better services to the disabled.

2.1.6 R2: Identify locations of people with specific disabilities

2.1.6.1 Description & Priority

This requirement will involve users such as an executive in the HSE being able to access a java application to get a breakdown of disabilities in Ireland. The executive will be able to search for a specific type of disability and a graphical representation will be displayed to show the executive the key information associated with that disability including the locations of the chosen disability.

2.1.6.2 Use Case

Scope

The scope of this use case is to identify the location of people with specific types of disabilities.

Use Case Diagram

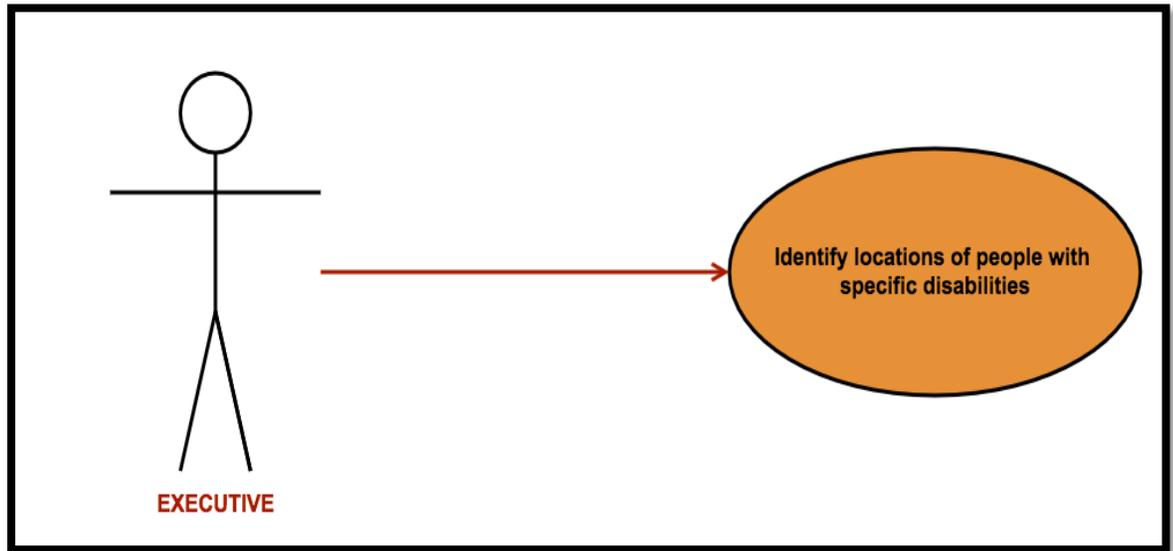


Figure 3: Identify Locations of People with Specific Disability

Flow Description

Precondition

Adequate data will need to be acquired before this requirement can be started and the java application should be completed. The data must have the locations and type of disabilities.

Activation

This use case starts when the <user> has opened the java application.

Main flow

1. The <executive> opens java application.
2. The <executive> selects the disability they want to search.
3. The <executive> views a heat map displaying the disability searched.

Exceptional flow

E1: <Disability not found>

1. The <executive> searches for a specific disability, but cannot find it on the application.
2. The <executive> searches for a different disability.
3. The use case continues at position 2 of the main flow.

Termination

This use case will be aborted if the <executive> cannot find the disability that they are searching for. Or if the application does not provide the necessary results.

Post condition

The <executive> closes the application.

2.2 Non-Functional Requirements

2.2.1 Performance/Response time requirement

Even though performance and response time is an important aspect of many systems, it does not apply to this report as the volume of data being used as part of the data analysis does not need to be analysed at a fast speed or with any real urgency. Therefore, the speed of performance and response time is not an important aspect to this report.

2.2.2 Availability requirement

The data will remain available to the system throughout.

2.2.3 Recover requirement

In the event of a hardware or software failure the data and database being used for this system will need to be backed up on a regular basis. The data folders containing the datasets and database schema's will be backed up on GitHub using version control. The data folder will also be copied and saved to multiple locations such as google drive and Dropbox for extra safety.

2.2.4 Robustness requirement

The robustness requirement does not apply to this system.

2.2.5 Security requirement

The data has been acquired from several websites such as the Central Statistics Office which is a public site, the datasets that have been downloaded from the CSO site are all publically available and require no special permissions to use them. The system itself can only be accessed on a personal laptop including the MySQL database which is protected with a username and password. The database is also hosted using gearhost.com which is also being protected by a username and password.

2.2.6 Reliability requirement

The data that is being used for this system came from the Central Statistics Office which compiles this data on a yearly basis. Therefore, the system is dependent on the accuracy and availability of data produced by the CSO.

2.2.7 Maintainability requirement

This system does not necessarily need to be maintained once it is created. It may need to be updated with new data in the future, but for this project the data that will be used has already been acquired.

2.2.8 Extensibility requirement

At the time of this requirement specification, there is no plan to further the study at this, but the methods and processes that are being used could be used again in the future.

2.2.9 Reusability requirement

This system currently has no requirement for reusability.

2.2.10 Resource utilization requirement

There is a need for a laptop, and cloud storage for backups. Some of the software that will be used for this system still needs to be installed, they are: RStudio, MySQL Workbench, Gearhost, Google Drive, Dropbox and NetBeans.

2.3 Design and Architecture

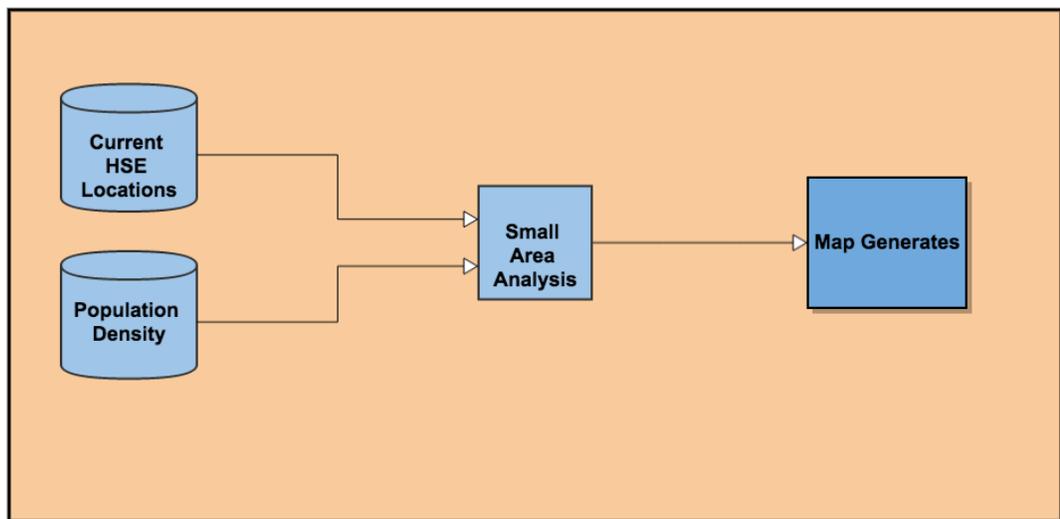


Figure 4: System Architecture

This section describes the system architecture and shows the components that make up the project and show how they interact with each other and how the extraction of the data works. The components that make up this project include an input source which is the datasets mentioned previously, that will be acquired from various data sources such as websites etc. There is a MySQL database component that is used to store the datasets as tables and RStudios is used to connect to the database in order to access the data and conduct all the analysis required for the project. After all the analysis is complete the datasets are visualized through a java based application by running the R-scripts against the data in the database. All the components listed above make up the key aspects of the system architecture.

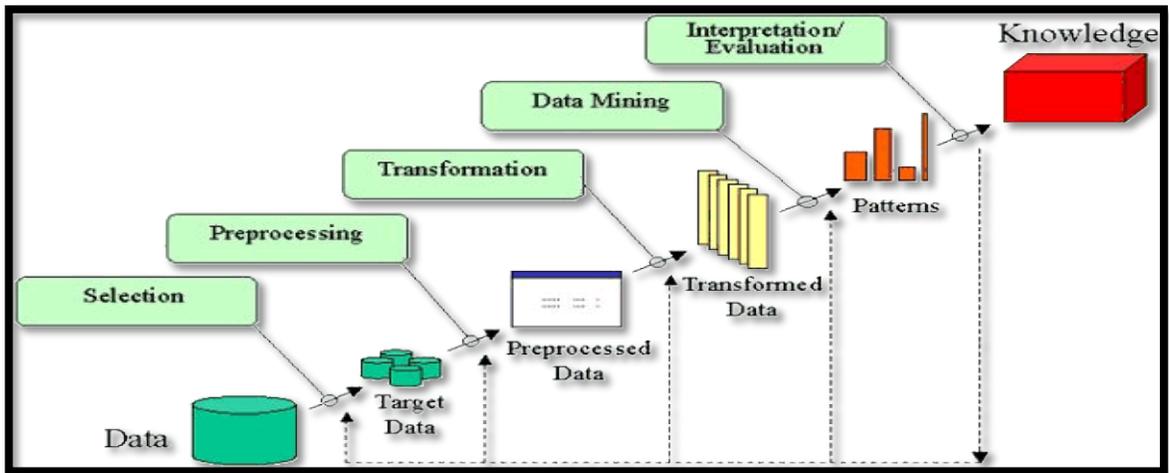


Figure 5: KDD

The data analysis aspect of this project, which will be used with the java application will follow the KDD (Knowledge discovery and databases) process. The KDD will be used to map out the steps to be implemented for the analysis of disabilities. The steps are:

Data Cleaning: In this step, the data will be cleaned by removing inconsistent data. **Data Integration:** In this step, multiple datasets can be combined to get the result needed. **Data Selection:** In this step, the data in the database is retrieved for analysis tasks to be completed. **Data Transformation:** The required data is transformed using aggregation. **Data Mining:** In this step, patterns are extracted. **Pattern Evaluation:** the patterns are evaluated. And **Knowledge Representation:** the knowledge is represented.

2.4 System Evolution

This system could evolve in the future to include further analysis if more datasets are acquired. Machine Learning could also be added, for example a prediction of learning disabilities in school children. Another way this system could evolve is by expanding the geographic focus past Leinster and have it identify where health centres should be placed in other regions of Ireland and other countries or specific regions i.e. Europe, North America etc. while also conducting analysis of disabilities in those very same locations. By expanding the focus, it will allow further research into support centre locations outside of Leinster, Ireland, comparing of disability statistics from several different countries or regions to analyze the difference if there is one in the amount of disabilities affecting men, women and children of various types disabilities. The java application could also be expanded later to include more functionality.

2.5 Implementation

In this section, the key portions of the code are shown as well as screenshots to show how this report was implemented. The code that will be shown will be both R and Java, which were the two programming languages used in this project.

2.5.1 Identifying new HSE Support Centre Locations

Setting up the Environment:

```
1 #Libraries
2 library(rgdal)
3 library(sp)
4 library(rgeos)
5 library(maptools)
6 library(ggplot2)
7 library(plyr)
8 library(leaflet)
9
10
11 #Download, Unzip and Read in shapefile
12 download.file("http://census.cso.ie/censusasp/saps/boundaries/Census2011_Small_Areas_generalised20m.zip", "Census2011_Small_Areas_generalised20m.zip")
13 unzip("Census2011_Small_Areas_generalised20m.zip")
14 data <- readOGR(".", "Census2011_Small_Areas_generalised20m")
15
```

Figure 8: R code for setting up the environment

Firstly, the R packages/libraries that will be needed, need to be installed these packages are ‘rgdal’, ‘sp’, ‘rgeos’, ‘maptools’, ‘ggplot2’, ‘plyr’ and ‘leaflet’. These packages are needed to allow us to use various functionalities in R, such as using the (sp) library to download, unzip and read in the small areas shapefile that will be needed later, as shown above in **figure 8**. Once all the necessary libraries were downloaded and the shapefile was successfully read into RStudio, it was now possible to move to the next part.

Sub-setting the shapefile:

```
15
16 #Subset population shapefile to show only Leinster
17 newdata <- subset(data, data$COUNTYNAME=='D<fa>n Laoghaire-Rathdown' | data$COUNTYNAME=='South Dublin' | data$COUNTYNAME=='Dublin City'
18 | data$COUNTYNAME=='Carlow County' | data$COUNTYNAME=='Dublin County' |
19 | data$COUNTYNAME=='Kildare County' | data$COUNTYNAME=='Kilkenny County' |
20 | data$COUNTYNAME=='Laois County' | data$COUNTYNAME=='Longford County' |
21 | data$COUNTYNAME=='Louth County' | data$COUNTYNAME=='Meath County' |
22 | data$COUNTYNAME=='Offaly County' | data$COUNTYNAME=='Westmeath County' |
23 | data$COUNTYNAME=='Wexford County' | data$COUNTYNAME=='Wicklow County' |
24 | data$COUNTYNAME=='Fingal')
25
26 shapeData <- spTransform(newdata, CRS("+proj=longlat +datum=WGS84 +no_defs"))
27
```

Figure 9: R code for sub-setting the shapefile to Leinster

For the next step of Identifying new locations for possible HSE Support Centres in Leinster, the R subset function was used to subset the small areas all-Ireland shapefile down to show

just the small areas within the region of Leinster, Ireland. The subset function was used by first selecting the dataset that was to be subsetted (the “data” variable, which stores the shapefile) and selecting all the areas within the shapefile that are within Leinster by specifying the locations name in the shapefile itself. For example: `data$COUNTYNAME == ‘Carlow County’` will select the rows in the shapefile where the column COUNTYNAME is equal to ‘Carlow County’ etc. This was done for all the areas that are in Leinster, and was then stored to a variable call ‘newdata’ as shown in **Figure 9**. The ‘newdata’ variable now only contains the information on the small areas within Leinster and nowhere else. Finally, the `spTransform` function of the (sp) library was used to set the spatial coordinates to longitude and latitude as shown on line 26 in **Figure 9** and is stored to the ‘shapeData’ variable.

Read in healthcentres.csv and subset to Leinster:

```

28 #Read in centre dataset
29 centres <- read.csv(url("https://data.gov.ie/dataset/729f95ee-62ff-4a7c-a14d-e19f9d8ea622/resource/31a7e92f-f46c-4637-9a0e-564fdb1ea816/download/healthcentres.csv"))
30
31 #convert data into a data-frame
32 data.frame(centres)
33
34
35
36 #####
37 #subset healthcentre data to show only Leinster
38 leinster <- subset(centres, County == "Wicklow" | County == "Wexford" | County == "Dublin" | County == "Kildare"
39 | County == "Meath" | County == "Carlow" | County == "Kilkenny"
40 | County == "Laois" | County == "Offaly" | County == "Westmeath"
41 | County == "Longford" | County == "Louth")
42

```

Figure 10: R code for reading in centre data and sub-setting.

In **Figure 10** above, the csv file that contains the information on the HSE support centres that are currently in Ireland is read into RStudio, this information will be needed later as the locations of the current support centres will be need to be known. The information contained within the healthcentres.csv file is the latitude and longitude of each centre, the address, and the services provided by each of the centres. As shown in **Figure 10**, The csv is read in and stored to the variable ‘centres’, then it is converted to a data frame to make it more manageable, once this is complete the centres data frame is subsetted similarly to the shapefile earlier by sub-setting it to show only the current support centres that are located within Leinster, Ireland and then storing it to the variable called ‘leinster’. With both the shapefile and the current support centre locations successfully subsetted to Leinster, it is now possible to move onto the more challenging aspect of creating a spatial map and using it to plot the current support centres and the population of all the small areas of Leinster.

Creating Leaflet map and plotting data:

```
67 #Create Leaflet map
68 m <- leaflet() %>% addTiles() %>%
69   setView(lng = -7.1050, lat=53.2000,zoom=8) %>%
70
71   addPolygons(data=shapeData,weight=1, popup = map_popup2, fillOpacity = 0.5,
72     color = ~colorQuantile("YlOrRd", shapeData$TOTAL2011)(TOTAL2011)) %>%
73
74   addPolylines(data = shapeData,color = "black", weight = .5)%>%
75
76   addCircles(data = leinster, lng = ~long_xcord,
77     lat = ~lat_ycord, popup = map_popup,
78     radius = 1218.69, fillOpacity = 1,
79     color = 'black', weight = 2, label=leinster$Service.name, fillColor = "blue")%>%
80
81   addLegend(pal = pal, values = shapeData$TOTAL2011, opacity = 0.7, title = "Leinster's-Population Density",
82     position = "bottomright")%>%
83
84   print(m)
85
86
```

Figure 11: R code for Leaflet map and plotting of data

In **Figure 11** above, the spatial map is created using the leaflet package which is used in the R programming language to create interactive maps. The first task of creating the spatial map was to set the view of the map to overlook the area of Leinster. This was accomplished by finding the latitude and longitude of Leinster and setting the ‘setView’ function of the leaflet map to those co-ordinates with a zoom level of 8 as shown below in **Figure 12**:

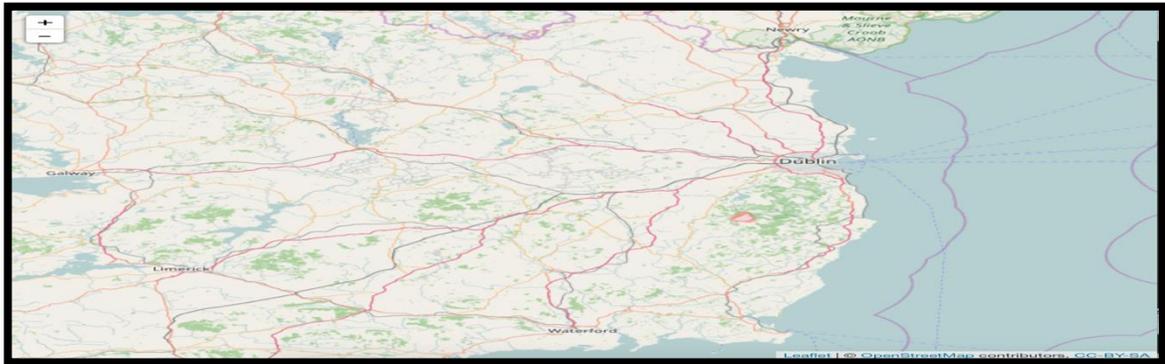


Figure 12: Spatial map-Leinster

The second task was to plot the shapefile of the small areas of Leinster showing the population density in each of those areas. The population of each area would be important as it is one of the criteria that will be checked later to justify a location as a potential location for a new HSE support centre i.e. if the population in the area is over a certain amount, then there is enough people in that area for a new support centre to provide services to. The shapefile was added to the spatial map using the ‘addPolygons’ function of the leaflet package as shown on line 71 in **Figure 11** above. The ‘addPolygons’ function set the fill opacity, and the colour quantile using the TOTAL2011 column in the shapeData and the addPolylines function on line 74 in Figure 11 simply adds a border around each polygon/small area. The TOTAL2011 column was the total population amount in each area. The results of plotting the shapefile is below in **Figure 13**.

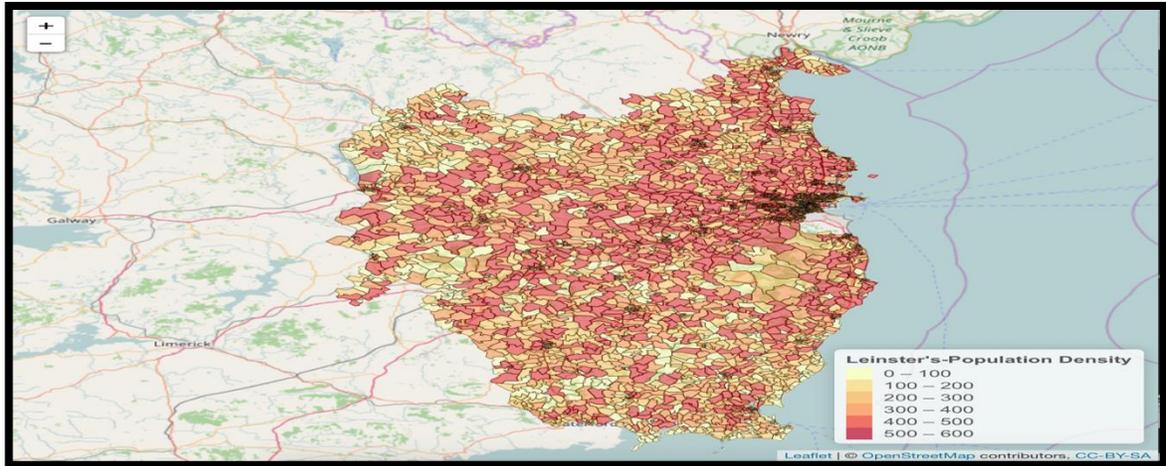


Figure 13: Spatial map-showing population density

The third task involved adding the ‘centre’ data to the leaflet map, which contained all the locations of the current HSE support centres in Leinster. The data was added to the map using leaflets addCircles function, which takes the latitude and longitude of the support centres and applies them to them map as circles, marking their location. The addCircles function also allows the design of the circles to be modified as well as adding a popup when the user clicks on one of the points. This popup shows all the important information about that specific support centre such as the address and the services it provides. The spatial map with the current support centres plotted is shown below in **Figure 14**.

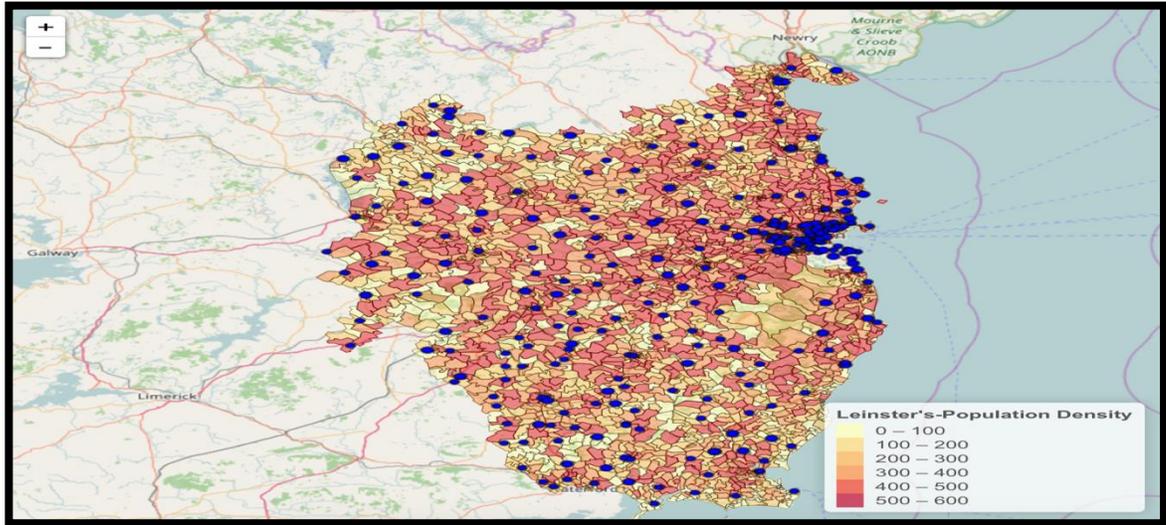


Figure 14: Spatial map- with current HSE support centres

```

44 #Create a pop ups for the map
45 map_popup <- paste0("<strong>HSE Support Centre: </strong>",
46                   leinster$Service.name,
47                   "<br><strong>Services Provided: </strong>",
48                   leinster$Service.name,
49                   "<br><strong>Address: </strong>",
50                   leinster$Address)
51
52 map_popup2 <- paste0("<strong> Area Name: </strong>",
53                   shapeData$EDNAME,
54                   "<br><strong>Population: </strong>",
55                   shapeData$TOTAL2011,
56                   "<br><strong>Suitable New Site: No </strong>")
57
58 heat <- colorNumeric(
59   c("Yellow", "Red"),
60   domain = shapeData$TOTAL2011)
61
62 bins <- c(0, 100, 200, 300, 400, 500, 600)
63 pal <- colorBin("YlOrRd", domain = shapeData$TOTAL2011, bins = bins)
64

```

Figure 15: R code for Spatial map-popups

In **Figure 15** above is the R code used to create the popups for the spatial map. ‘map_pop’ provides the popup information when a user clicks on one of the current support centre points on the map, this popup will show information about that support centre i.e. the name of the centre, the services it provides and their address. ‘map_popup2’ provides the popup information for the small areas on the map, this popup will show the areas name and the population amount in that specific area.

```

92
93 ##find polygons with a health centre in it##
94 coordinates(leinster) <- ~ long_xcord + lat_ycord
95
96 # Set the projection of the SpatialPointsDataFrame using the projection of the shapefile
97 proj4string(leinster) <- proj4string(shapeData)
98 p <- over(leinster, shapeData, returnlist = FALSE)
99
100 #get the polygons with no health centres.
101 polys_notint <- shapeData[!(shapeData$GEOGID %in% unlist(p$GEOGID)), ]
102
103 shapeData.sub <- polys_notint[polys_notint$TOTAL2011 > 350,]
104
105 map_popup3 <- paste0("<strong> Area Name: </strong>",
106                   shapeData.sub$EDNAME,
107                   "<br><strong>Population: </strong>",
108                   shapeData.sub$TOTAL2011,
109                   "<br><strong>Suitable New Site: Yes </strong>")
110
111
112 #Map Potential HSE Support Centre sites
113 m <- addPolygons(m, data=shapeData.sub,weight=1, popup = map_popup3, fillOpacity = 0.5,
114                color = "green") %>%
115
116 addPolylines(data = shapeData.sub,color = "black", weight = 1)%>%
117
118 addLegend("bottomright", colors= "green", labels=" New HSE centre Locations", title="HSE Support Centres")%>%
119 print(m)

```

Figure 16: R code for finding new support centre locations

In **Figure 16** above, is the R code that was implemented to identify new HSE support centre locations. On line 97 the proj4string function was used to set the projections of the ‘leinster’ data frame to be the same as the projections of the shapeData data frame so they can be overlaid on top of each other and be mapped together. Next, as another part of the criteria to identify suitable locations for future HSE support centres, the small areas on the leaflet map that do not have a support centre need to be identified. This was accomplished using the R code on line 101 in **Figure 16** above, this code is getting the small areas from the overlay done on line 98 that do not have a point (support centre) contained within them and is storing those areas to a new variable called ‘polys_notint’. The next step (line 103) is to

find which of those areas stored in 'polys_notint' that have a population that is over 350, this is how new potential locations for HSE support centres will be identified. The final step is to plot this new information to the spatial map created above and to show the areas that meet the criteria for a new HSE support centre by changing their colour to green, so they can be identified on the map. The result is shown below in **Figure 17**.

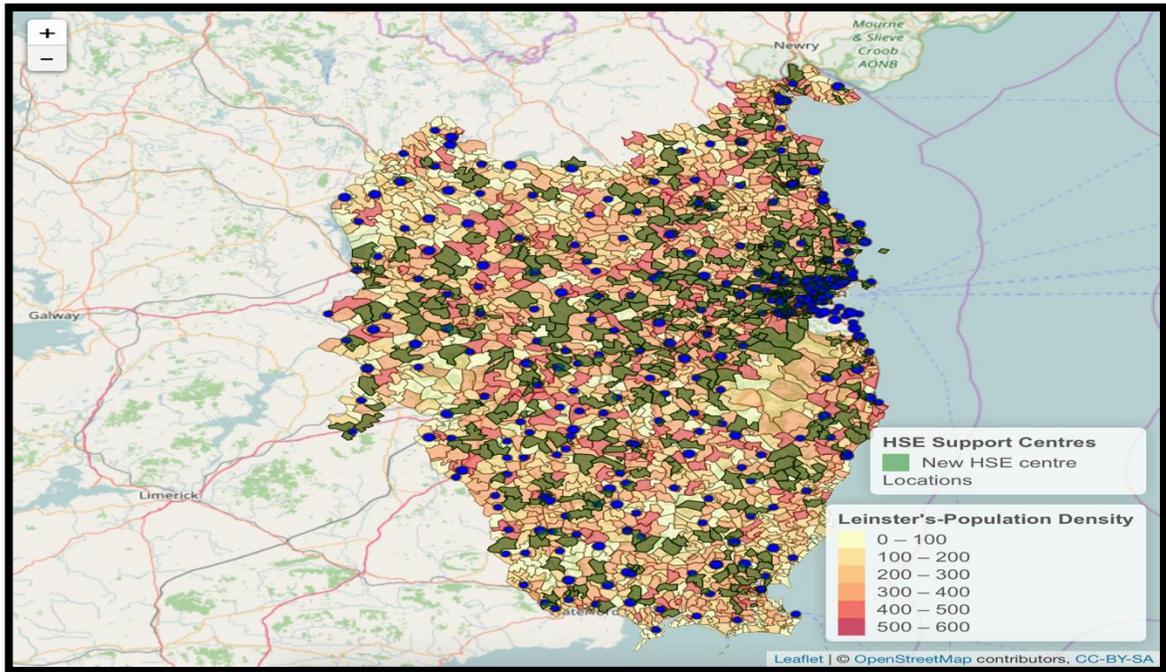


Figure 17: Spatial map- showing potential HSE support centre sites

In **Figure 17** above, the small areas/polygons that meet the criteria for been a suitable location for a new HSE support centre in the future are coloured green so they easily stand out to the viewer. These green polygons symbolise having no support centre at the moment and that they have a population amount greater than 350, therefore they are suitable HSE support centre locations for the HSE to look at in the future.

2.5.2 Disability Application

Creating the applications GUI's:

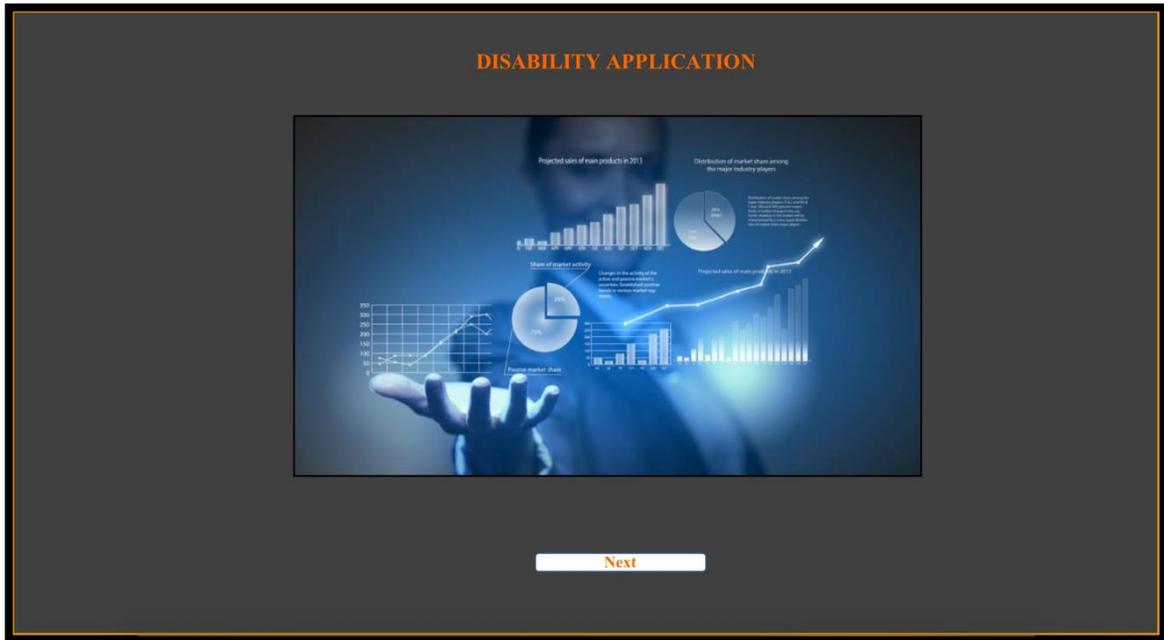


Figure 18: mainGUI.java

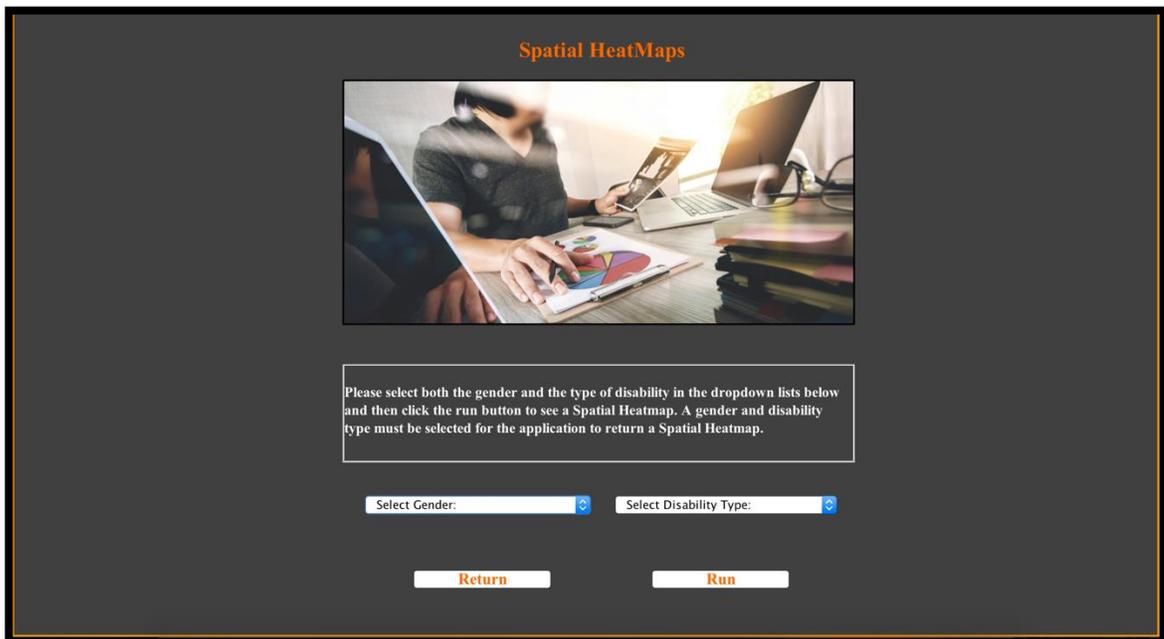


Figure 19: heatmapGUI.java

Regarding the Disability Application that was created using Java and R, the first task was to create the graphical user interfaces(GUI's) as shown above. Creating the GUI's was

straightforward in NetBeans, labels were selected and used to store various text such as headings and titles whereas a scroll pane was used in `heatmapsGUI.java` for a descriptive box. Buttons were added to allow the user to move back and forth between the GUI's as they see fit. A menu was not needed as there was a limited number of GUI's to navigate, making a menu bar unnecessary. Images were added into the Disability Application to add a better-looking design to the GUI's instead of having just a bare design. The images themselves were stored within the java project itself and added into a label for display.

Set up Java and R Interface:

```
//Set up Rcaller
RCaller caller = new RCaller();
RCode code = new RCode();

// Tell java where r is stored on computer
caller.setRscriptExecutable("/Library/Frameworks/R.framework/Versions/3.3/Resources/Rscript");
code.clear();
caller.setRCode(code);
```

Figure 20: RCaller-setup

A key aspect of the Disability Application was the interfacing of the Java and R programming languages, after extensive research into the many possible ways of accomplishing this, two possibilities were narrowed down, they were: JRI and RCaller. JRI (Java/R interface) allows java applications to run R code as single threads by loading the R dynamic library into Java and then provides an API from Java to R functionality. While adding JRI to this project at an early stage, it was found that the JRI was overcomplicated for bridging the Java language with R due to having to set up numerous and various file paths to where R and other tools were stored on my laptop. It became more apparent as the application progressed that JRI would not be ideal for handling and displaying plots, this was a problem as the main functionality of the Disability Application was displaying spatial heatmaps of the various types of disabilities throughout Leinster, Ireland. After further research, another way of bridging java and R together was found: RCaller. RCaller was another tool for interfacing Java and R programming languages together in a Java application, a good thing about RCaller was that its approach to bridging the two languages together was a lot more simplistic than JRI's approach. RCaller was also very efficient at handling plots since its 2.07 update, which allows the plotting of R graphs in a java application to happen with just five lines of code. Therefore, after experimenting with both JRI and RCaller, it was decided that RCaller would be provide a simpler and more efficient interface for java and R regarding the Disability Application. **Figure 20** above, shows how RCaller was set up within the Disability Application. The first step to getting RCaller set up and running within the application was to import the `rcaller.jar` file into the project in NetBeans, then was it was the setting up of RCaller and RCode which would allow us to use RCaller's various methods. Once this was complete all that was then needed was to tell java where R was stored on the laptop by setting the file path of the `Rscript` as shown in **Figure 20**. Now that this was done, RCaller was fully integrated into the Disability Application and all that needed to be done now was to start using it to create the main functionality of the application.

Create Spatial Heatmap of Disabilities:

```
code.addRCode("data3 <- data3[,-1]");
code.addRCode("data3['id'] <- NA");
code.addRCode("data3$id <- c('C01','C02','C03','C04','C05','C06','C07','C08','C09','C10','C11','C12','C13','C14','C15')");
code.addRCode("data3<- data3 %>% select(id, everything())");
code.addRCode("data3$id = as.character(data3$id)");
```

Figure 21: Example of RCaller's *addRcode* method ()

Now that RCaller was set up and running, it was time to start creating the main functionalities of the application. This was started not in NetBeans but in RStudio the IDE been used for R programming in this final year project, the reason for starting in RStudio was to create working R scripts that would return the required spatial heatmaps that would later be shown in the Disability Application. In each R script, R code was being used to first select the R libraries that were needed, and then read in the shapefile of all the counties in Leinster, Ireland. The next few lines of R code were then used to connect to the MySQL database where the disabilities dataset was being stored, and import the data about a certain disability and gender into the R workspace using the RMySQL Package to run a MySQL select query on the dataset in question i.e. (Select * **From disabilitydb.disabledsurvey where (Category = 'Difficulty in learning, remembering or concentrating' AND Sex = 'Male');**). Once the required data was imported successfully, it was time to prepare both the shapefile and the disability data to be utilised for a heatmap. To create a heatmap in R, both the shapefile and disability dataset would need to be joined together. But as both files are not of the same type: one a shapefile and the other a data frame, the R Function Fortify was used to convert the shapefile into a data frame so it and the disability data frame could be merged. With the merge done, all that was left to do was to use the ggplot2 function to create and display the heatmap using the newly merged data frame. The ggplot2 function allows the selecting of the data that you want to use for the heatmap as well as handling the design of the heatmap itself, such as background colour, titles, subtitles etc. All the steps discussed above had to be done for every disability 3 times: once for Males, once for Females and once for Both Genders to have heatmaps for each possible selection that the user could choose from on the applications dropdown list. Now that all the R scripts were done it was time to implement them into the Disability Application. It was felt that it would be better create and get working R scripts that would create heatmaps of disabilities within Leinster, Ireland first and then simply apply that same working R code into the application by using RCaller's *addRcode* () method to add and run R code and return the output, which in this case would be a heatmap depending on the disability type and gender selected by the user of the application. A small example of the *addRcode* () method in use is shown above in **Figure 21**. Now that everything was ready and in order, the next step was to create all the java classes that when called would run R code for a specific gender and disability i.e. when blindnessMale.java is run it would run R code and return the spatial heatmap showing the number of males with blindness throughout Leinster. Altogether there were 18 java classes created to add all the disabilities within the dataset that was being utilised.

For the user to select the gender and type of disability that they wanted to see, two dropdown lists would be utilised, one for the selecting of the gender and another to select the disability. For the selection to work an if statement was created behind the run button that would return the result. This if statement would work as follows: if gender == Male and disability Type == blindness, then run the blindnessMale.java class to return the heatmap for males with blindness in Leinster. Shown below in **Figure 22** is the Disability Application showing a heatmap based on the selection of a user.

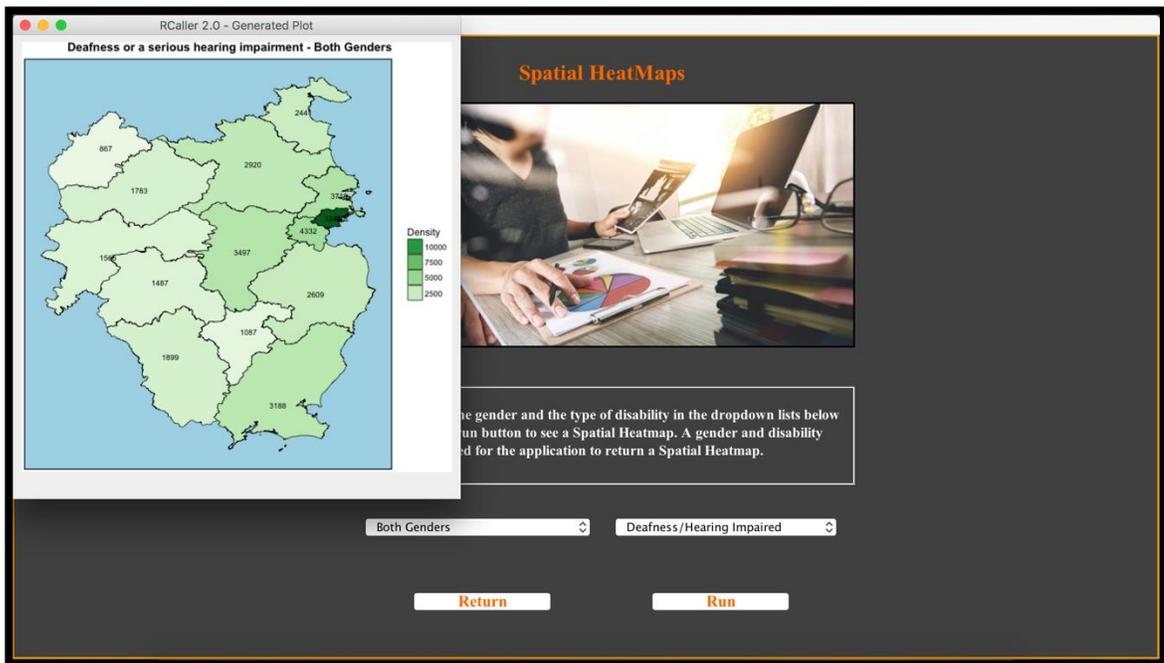


Figure 22: End Result-Spatial Heatmap of Deafness disabilities in Leinster (Both Genders)

2.6 Graphical User Interface (GUI) Layout

The following is the mock-ups of how the java application GUI's.

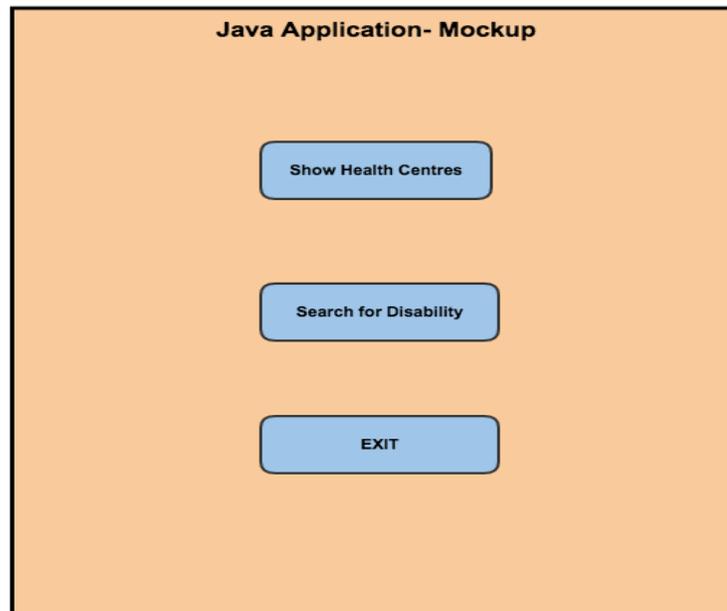


Figure 23: Menu

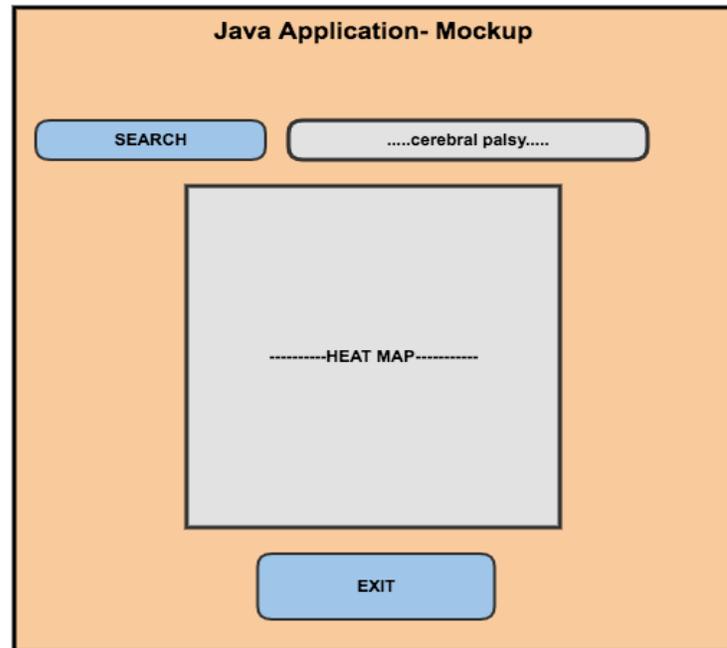


Figure 24: Disability Search

2.7 Testing

2.7.1 Integration Testing:

To check that all important parts of this project are working correctly, both the R scripts and Disability Application will have testing performed on them to make sure the vital parts of each script and the application are performing the way they should be. The first testing that was done was the testing of all R scripts that this project is using, this included testing over 20 R scripts in total. In R programming, the *testthat* library was used to test various R functions and code in each of the R scripts being tested. The *testthat* library allows tests to be made up in a separate script, all that needs to be done is to use R's *source ()* function to show R to each of the R scripts that are to be tested, this is done by placing each scripts file path into its own *source ()* command as shown in **Figure 25** below.

```
25 source("/Users/declanbarnes/Desktop/DeclanBarnesx13114352-FinalYearProject/DisabilityApplication-R scripts/Total Persons-Female.R")
26
27 ###TESTING###
28
29 #Testing that plot returns object
30 test_that("Plot returns ggplot object",{
31   p <- plot_fun(plotData)
32   expect_is(p,"ggplot")
33   print("TRUE, ggplot object is being returned")
34 }
35 })
36
37 #Test the correct data is being used
38 test_that("Plot uses correct data", {
39   p <- plot_fun(plotData)
40   expect_that(plotData, equals(p$data))
41   print("TRUE, plot is using the correct data")
42 })
43 #Test 'data' is a SpatialPolygonsDataFrame
44 test_that("Test that data is a SpatialPolygonsDataFrame",{
45   d <- data
46   expect_is(d,"SpatialPolygonsDataFrame")
47   print("Variable: data, is a SpatialPolygonsDataFrame")
48 }
49 })
50 #Test that 'con' is a MySQL connection
51 test_that("Test that con is a MySQL Connection",{
52   c <- con
53   expect_is(c,"MySQLConnection")
54   print("con is a connection for MySQL")
55 }
56 })
```

Figure 25: Testing.R

The only tests that needed to be done with the R scripts was to test that the various graphs were in fact using the correct data, and that certain variables were of the correct type such as a data frame, or a SpatialPolygonDataFrame etc. This was done to make sure that the graphs such as the heatmaps for the Disability Application would be produced properly without problems and that they were being plotted using the right data such as: the disability type, amount, and gender. The *testthat* library was very straight forward to use, as the tests written using the library only had to be 4 to 5 lines of R code and its functions such as the *expect_is ()*, was very easy to use regarding testing if data is a data frame etc.

2.7.2 Usability Testing (with 2 users):

The Disability Application was tested using Usability testing. Usability testing was used instead of JUnit testing as there was not much to test on the java side of the application as most of the work being conducted in order to set up and plot a spatial heatmap of disabilities

in Leinster was done on the R side of this project, therefore it was decided to instead use Usability testing involving two volunteers. The Usability test involved getting two users to run the application, each time the users ran the Disability Application, they had to choose a different gender and disability type. Several steps were followed during the Usability tests, these steps were: step 1: whenever a user asked a question, it would be written down as an error. Step 2: The users must not be told how the application works, they would only be giving a scenario. These steps were followed throughout all tests that were conducted in regards to the Usability testing. The overall feedback from the Usability tests was very positive. Both users were able to quickly open and use the Disability Application without any problems, they fully understood how to move from GUI to GUI and get the output that they wanted in regards to the specific gender and disability that they wanted to see in the form of a heatmap. After the test were concluded the users were asked if the application could/should be made easier to use, both agreed that the current way the application was setup was perfectly fine and already easy to navigate and use. The forms used during the Usability tests for both users are displayed in the appendix.

2.7.3 Testing Summary:

As shown above testing was conducted in two different ways, both integration testing and Usability testing in order to ensure that everything that has being completed in regards to this report is functioning properly and without errors. Overall the testing of all R-scripts and the Disability Application using both testing methods stated above in sections 2.7.2 Usability Testing and 2.7.1 Integration Testing worked out well and would have allowed possible errors to be rectified if found, which in the case of this report no errors were found. All aspect of the system is functioning properly.

2.8 Evaluation

Throughout this report, both the Identifying of new HSE support centre locations as well as the development of the Disability Application have been continuously evaluated. Any errors or mistakes were fixed and progression was quickly made afterwards. The tasks of identifying new locations for HSE support centres and creating an application for show heatmaps was accomplished. A user can view various disability types by gender through a spatial heatmap and the HSE can find new suitable locations for future support centres for those with disabilities. The results found with the Disability Application heatmaps show that disabilities involving chronic illnesses are higher among females than males in most areas of Leinster. Whereas both genders have similar amount when it comes to learning disabilities. When it comes to the total amount of disabilities in Leinster, females are shown to have a higher amount of disabilities than males. In regards to identifying new HSE support centre locations, several potential areas meet the conditions to justify a new support centre being placed i.e. they do not have a support centre currently and the population in those areas is over the required amount, this gives the HSE a significant amount of options if or when they decide to place more support centres in Leinster, Ireland.

3 Conclusions

This report was carried out in order to identify new HSE support centres locations in Leinster and to create a java application that would allow a user to view a heatmap of various disabilities based on gender. The task of identifying new support centre locations was met by plotting the locations of the current support centres and then analysing the population density to see if the population was greater than a certain amount in each of the small areas of Leinster. If the small areas did have a population density greater than the specific amount and didn't have a current support centre then they were marked green to identify them as suitable locations for future support centres. Now that this task is complete, several areas are easily identified for the HSE to look at if they plan to build new support centres sometime in the future to ensure that HSE resources are used well.

The task of allowing a user to generate and view a spatial heatmap of disabilities in Leinster based on gender was met by creating a java desktop application. The application integrates with the R programming language to run analysis on disabilities types and genders within Leinster and then plots a heatmap based on that data. The heatmap shows the users the amount of people with a specific disability based on the gender and disability type chosen from a drop-down menu. The results from the application show that there are more females with disabilities in Leinster, Ireland than males. For example, there are more females who suffer from disabilities that fall into the Chronic illness category than males and females have a higher amount than males regarding total disabilities. These results would allow the user such as a HSE executive to see that more resources may need to be more focused on females than males due to a higher amount of disabilities among that gender. The advantages of this report are that it includes the identification of new potential sites for HSE support centres in Leinster, Ireland, which after research conducted by the author at the beginning of this report has not been done before in Ireland. This gives the report a unique aspect to it. The disadvantages of this project are that while these new sites for HSE support centres can be identified, it does not necessarily mean that they will be built. The HSE support centres in Ireland are funded by the Irish government, which means there may be a lack of resources and funds to build new support centres in the country. With this lack of funding and resources, identifying where new support centres should be built may not be realistic in Ireland until the government decides to invest more funding into this area of the HSE. The opportunities that this project provides is that if the Irish government does decide to spend more government resources to build new HSE health centres then this project could give them a head start by identifying the best possible locations to place the health centres, saving valuable time and resources. The limits to this report is that the author had to learn by doing as the author had not learnt specific portions of was to be used to create the project.

4 Further development or research

The results of this report could lead to more expansive disability analysis and provide more focus on how badly funded the disability services in Ireland are. With more resources, development and research this report could expand to provide a better analysis on support centres and support services to disabled people in the country.

With further research, machine learning could become an even bigger asset to the HSE in Ireland. Machine learning could help predict certain disabilities in children, which will help the HSE plan more effectively regarding their resources throughout the country.

5 Motivational Section

The motivation for this research is to address the challenges faced by people with disabilities having to travel long and difficult journeys to reach vital support services.

This research was undertaken for personal reasons due to the fact that the author of this paper was born with a birth defect (born with one hand), While he does not personally consider himself disabled due to the fact that he is not hindered in any way by having one hand as it doesn't stop or prevent him from doing anything that he wants to do, but it did get him thinking about other people around the country who have various disabilities that prevent them from doing certain activities or require them to use certain services that are provided by health centres around the country.

While researching the support services in the country, it became apparent the lack of support services in the country so, it was decided to implement a Data Analysis project identifying where new support centres could be placed in Leinster, Ireland so more services could be provided.

6 Bibliography

- Authority, N. (2015). [online] Nda.ie. Available at:
<http://nda.ie/Publications/Transport/Transport-Publications/Provision-of-Transport-Services.html> [Accessed 9 Dec. 2016].
- Cran.r-project. (2016). [online] Available at: <https://cran.r-project.org/web/packages/sqldf/sqldf.pdf> [Accessed 9 Dec. 2016].
- Cso.ie. (2008). [online] Available at:
<http://www.cso.ie/en/newsandevents/pressreleases/2008pressreleases/nationaldisabilitysurvey2006-firstresults/> [Accessed 9 Dec. 2016].
- Data.gov.ie. (2016). [online] Available at: <https://data.gov.ie/data> [Accessed 9 Dec. 2016].
- Disability-federation.ie. (2016). [online] Available at: <http://www.disability-federation.ie/index.php?uniqueID=29> [Accessed 9 Dec. 2016].
- DBD, U. (2017). KDD Process/Overview. [online] [Www2.cs.uregina.ca](http://www2.cs.uregina.ca). Available at: http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html [Accessed 8 May 2017].
- Educationinireland.com. (2017). Profile of National College of Ireland for International Students - Education in Ireland. [online] Available at: <http://www.educationinireland.com/en/Where-can-I-study-/View-all-Private-Higher-Education-Institutions/National-College-of-Ireland-Profile.html> [Accessed 8 May 2017].
- Unctad.org. (2017). unctad.org | Analysis. [online] Available at: <http://unctad.org/en/Pages/analysis.aspx> [Accessed 8 May 2017].
- User, S. (2017). Company profile - Ampercom. [online] [Ampercom.com](http://www.ampercom.com). Available at: <http://www.ampercom.com/en/> [Accessed 8 May 2017].
- Disability.ie. (2017). Disability.ie. [online] Available at: <http://www.disability.ie/> [Accessed 9 May 2017].
- Ohri, A. (2017). Navigating the Choices in R and Cloud Computing.
- Trap.ncirl.ie. (2017). Trap.ncirl.ie. [online] Available at: <http://trap.ncirl.ie/701/1/stephenreid.pdf?&session-id=a4e24db53f57f4ed02ec6be9c63d03d4> [Accessed 9 May 2017].

7 Appendix

7.1 *Project Proposal*

Project Proposal

Identifying new Health Centre Locations in Ireland

Declan Barnes, X13114352, Declan.Barnes@student.ncirl.ie

BSc (Hons) in Computing

Data Analytics

16/10/2016

Objectives

(Max. 1 Page)

The **first part** of this project is to acquire the necessary information needed to complete the project. This information will come in the form of several datasets that will be acquired from public resources such as the central statistics office. Each dataset will contain key information that I will need for my final year project in relation to disabilities in Ireland i.e. how many people are living with a disability in Ireland, what type of disability do they have, what sex and age are they etc.

The **second objective** of this project is to analyse disabilities in Ireland. My aim of doing this is by using several datasets that I have acquired through extensive online research as stated above. The data sets in question will show the statistics of disabled people across Ireland such as: Sex, Age Group, and type of disability. This information will be used to create a dashboard representing key analysis of the disabilities, for example: how many disabled people require homecare in Ireland and how many people with disabilities are not working due to their disability.

The **third objective** of this project will be to map all the health centres in Ireland showing their locations and proximity to each other. These health centres will show us how much of Ireland is ideally covered by health centres and where there is a lack of centres in certain areas (i.e. rural areas) of the country.

The **fourth objective** of this project is to use machine learning to predict where new health centres should be placed in Ireland to provide better services and coverage to disabled people in the country. This objective will take the above-mentioned information (in the second objective) into account of where current health centres are located and predict where new centres should be built i.e. in areas where the nearest health centre is too far for disabled people to travel to, and then it will show on a map where the new centre should be placed in relation to this information.

The **above objectives** will be done using MySQL, the R programming language, and Tableau. MySQL will be used to store the datasets that I will be using for my project, the R programming language will be used pull the data from the MySQL database and perform data manipulation, data cleaning on it, as well as being used for the machine learning aspect of the project. Finally, Tableau will be used to visualize my findings with its professional looking charts.

Background

(Max. 2 Pages)

There are a significant number of people in Ireland that are living with some form of disability and they are people of all ages and backgrounds. I thought that it would be of

interest to various people and organisations to breakdown or analyse disabilities in Ireland for my final year project.

The reason I chose to do this project was for personal reasons due to the fact that I was born with a birth defect (I was born with one hand), While I don't personally consider myself disabled due to the fact that I am not hindered in any way by having one hand as it doesn't stop or prevent me from doing anything that I want to do, but it did get me thinking about other people around the country who have various disabilities that prevent them from doing certain activities or require them to use certain services that are provided by health centres around the country. For my final year project, I wanted to investigate further into this by analysing disabilities across the country in detail.

I wanted to show how many health centres there are and where in the country they are located so I could see if new health centres could be built to provide better coverage of services to disabled people in Ireland by helping to predict where these new health centres should be placed in the country.

While researching information for my project I searched online to see if other people, companies etc. had done something like what I am trying to do in relation to my final year project, after searching for a significant time I was unable to find anything in relation to predicting where new health centres should be built. I could find government run websites that contained statistics about disabilities in Ireland but they didn't breakdown the statistics the way I am going to show a better picture of disabilities in Ireland.

I believe it is important to try to understand or work out where there are limitations in the services provided to disabled people across Ireland, by doing this project I hope to show these limitations in a constructive way that will help improve the quality of the services that disabled people rely on daily for their everyday needs. I think it would be beneficial overall to show where new health centres should be positioned in Ireland as it is unacceptable to expect disabled people in rural areas specifically, to travel a significant distance to get to their nearest health centre, while these same health centres could be failing to provide adequate services/coverage to the people who need them by being out of reach.

While I was deciding whether to base my project on the analysis of disabilities in Ireland I looked at several informative websites to get a better picture of what was going to be involved with doing this project, for example I researched the disability-federation website to come up with ideas of what type of information would be useful for me to show in my project.

On the disability-federation website it states the following in regards to disabilities: It states that people in Ireland that are living with a disability and parents who have special needs children are at a much higher risk of suffering depression than those who do not. I realised after reading this that my idea of predicting new health centres would be a beneficial idea to those disabled people and the parents of special needs children as these

health centres could provide very important care and or support to these people to help relieve the pressure that they live with daily by creating more health centres that are much better placed to provide the necessary coverage across the country that they need.

Another interesting fact that I noticed through my research was the attitudes towards people of a disability. On the national disability authority, it said that negative attitudes towards disabled people had become a significant barrier to the inclusion of the disabled into various aspects of society and therefor preventing people with a disability from reaching their full potential.

The National Disability Authority has conducted a series of surveys at regular intervals on the public's attitude towards people living with a disability (2001,2006,2011). The NDA's surveys showed that attitude towards disabled people improved in the years 2001 and 2006, but the year 2011 showed a decrease in attitudes.

I thought that the information stated above about the attitudes toward the disabled was interesting, so much in fact that as part of my dashboard I plan to include sentimental analysis of twitter to show the positive and negative tweets in relation to disabilities.

Technical Approach

Brief description of the approach to be followed (Max. 1 Page), Research, literature review, requirements capture, implementation etc....

Research:

I carried out a great deal of research before starting my project. There were a lot of different factors to be figured out. My main priority was to create a project that was both interesting and enjoyable so I conducted a lot of research into the data analysis of various topics to give me an idea of what I needed to do for my final year project.

During my research, I focused on specific topics such as storing datasets in a MySQL database and if it was possible to pull or access the data from MySQL using RStudio and the R programming language. I also conducted research on data manipulation, data cleaning, data visualization and machine learning. These topics are what I will be using as part of my project so I felt it was best to research into them and find how I could use them in relation to my project.

The aim of researching all these topics was to give me the confidence that I need to start and complete my final year project.

Literature review:

As part of my approach to my project I reviewed several websites and books that I thought would be of benefit to me. These websites and books are listed below:

- Machine Learning with R - Brett Lantz
- <https://www.r-bloggers.com/>
- https://cran.r-project.org/doc/contrib/Zhao_R_and_data_mining.pdf

Requirements capture:

- Acquire needed datasets,
- Dashboard representing the following:
 - -Disability analysis represented on charts,
 - -Predicting location of future health centres.
 - Predictive analysis of equipment needed at each health centre location.

Implementation:

I will implement my project using the R programming language which I am currently learning in my final year at NCI. I will use R to pull my data from the MySQL database then perform data cleaning and data manipulation to get the analysis results that I am aiming to show.

On the dashboard, I plan to also show the sentimental analysis of twitter in the R language. This analysis is intended to show where tweets about disabilities are positive or negative in anyway and if they are, how or why are they negative so I can show a better picture of why some people have a negative attitude towards people living with a disability.

I also intend to conduct a word cloud in the R language representing the keywords associated with disabilities, I thought the result might be interesting to see and therefore well worth while investing the time in my project to include it on the dashboard.

I will then use Tableau to import my R scripts and produce a dashboard representing my data. I hope to have the dashboard completed by the end of semester 1.

For semester 2 I will use the R programming language to conduct the machine learning aspect of my project to predict where new health centres should be placed in the country as well as conducting the predictive analysis.

Special resources required

I will use the following resources to help me create and complete my project:

Machine Learning with R - Brett Lantz:

This is a book that was recommended to me for figuring out how I should go about the Machine learning aspect of my project. I will read this near semester 2 when I intend to start and integrate machine learning into my project.

R-Bloggers:

This is a very popular website containing useful information in regards to anything related to the R programming language. I will use this to help me learn R and how to use it for my project.

Zhao R and data mining:

This is a PDF of a well-known book that give insight into how to manage data mining and how to use best practices when working with data. I will rely on this to ensure that I am doing everything correctly when I'm working with my data.

MacBook Pro Retina:

I will be using my own laptop to do my project instead of the college computers as it will allow me to work from home more easily.

RStudio:

Since I will be using R as my programming language, I will therefore be using RStudio. RStudio is a great IDE for R. Its interface is very user friendly and easy to learn.

Tableau:

I will be using Tableau to show my data as their charts are more professional looking. I think Tableau will be better suited for presentations due to their attractive charts.

Technical Details

Implementation language and principal libraries

The language that I aim to use for my final year project is the R programming language. R is a statistical programming language that has powerful packages for nearly everything that you would want to do with data.

The IDE that I will be using is the standard IDE for R which is RStudio. The R language has great support and a massive community that releases new packages and tutorials on a routine basis.

I am currently learning R on Pluralsight and in college as part of my Data applications module, which will give me the necessary skills that I need for the completion of my project.

My project will make good use of multiple Packages in R such as:

- **RMySQL:** database interface and MySQL driver for R.
- **TidyR:** Used for data manipulation and cleaning.
- **Stringr&Magrittr:** Data cleaning
- **Dplyr:** Also, used for data manipulation, usually used in conjunction with TidyR.
- **GoogleVis:** Data Visualisation.
- **Sqlf:** Provides an easy way to perform SQL selects on R data frames.
- **ggplot2:** Used for Visualizing data.

MySQL:

I will be using MySQL for the storage of my datasets. The reason for using MySQL is that I am very familiar with it as I have been using it for the last 3 years in college, it also has a relatively straight forward interaction with R and RStudio through R's RMySQL package which will make interacting with my datasets friendlier.

Evaluation

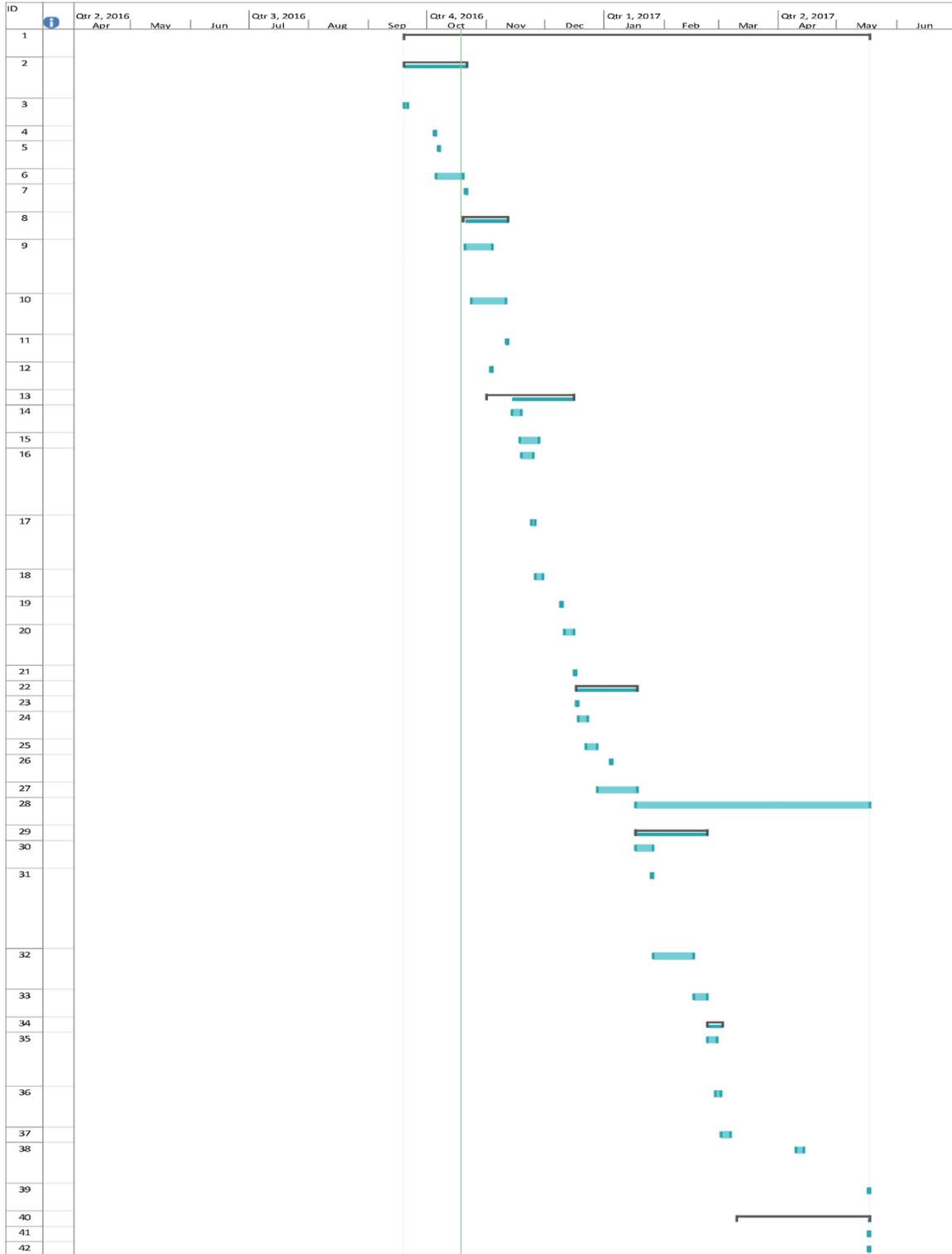
I will evaluate my data analysis project using testing by getting volunteers to look at my dashboard and the machine learning aspect to my project by checking that everything is displayed properly and professionally.

I will also ask for input from certain lecturers and my supervisor in NCI to see if my project is done properly and functioning as it should or if they have any input on how to improve the project further.

I aim to meet with my supervisor on a routine basis to make sure my project is going in the right direction while getting my supervisor to test what I have done up to each meeting to ensure accuracy and consistency.

Declan Barnes
16/10/2016

7.2 Project Plan



Task Name	Duration	Start	Finish
Project-Disability Analysis	173 days	Mon 19/09/16	Wed 17/05/17
Project Proposal & Project Plan Documents	25 days	Mon 19/09/16	Fri 21/10/16
Project Brainstorming	2 days	Mon 19/09/16	Tue 20/09/16
Project Pitch	1 day	Wed 05/10/16	Wed 05/10/16
Reflective Journal(Sept)	1 day	Fri 07/10/16	Fri 07/10/16
Project Proposal	10 days	Thu 06/10/16	Wed 19/10/16
Documentation Review & Upload	1 day	Fri 21/10/16	Fri 21/10/16
Requirement Specification	17 days	Thu 20/10/16	Fri 11/11/16
Research for Disability Datasets on CSO website	10 days	Fri 21/10/16	Thu 03/11/16
Requirement Specification Document	14 days	Mon 24/10/16	Thu 10/11/16
Documentation Review & Upload	1 day	Fri 11/11/16	Fri 11/11/16
Reflective Journal(Oct)	1 day	Thu 03/11/16	Thu 03/11/16
Project Analysis & Design	33 days	Tue 01/11/16	Thu 15/12/16
Set up Data Warehouse	5 days	Mon 14/11/16	Fri 18/11/16
Data Cleaning	7 days	Fri 18/11/16	Sun 27/11/16
Filter statistics on male/female and type of disabilities	5 days	Sat 19/11/16	Thu 24/11/16
filter on age groups and type of disabilities	2 days	Thu 24/11/16	Fri 25/11/16

Create Prototype(Web Application)	3 days	Sat 26/11/16	Tue 29/11/16
Reflective Journal(Nov)	1 day	Fri 09/12/16	Fri 09/12/16
Prepare for Mid-Point Presentation	5 days	Sun 11/12/16	Thu 15/12/16
Mid-Point Presentation	1 day	Fri 16/12/16	Fri 16/12/16
Post Mid-Point	23 days	Sat 17/12/16	Tue 17/01/17
Review Feedback from panel and makes changes	1 day	Sat 17/12/16	Sat 17/12/16
Finish Web Application	5 days	Sun 18/12/16	Thu 22/12/16
Christmas Break	4 days	Thu 22/12/16	Tue 27/12/16
Reflective Journal(Dec)	1 day	Wed 04/01/17	Wed 04/01/17
Study Break for exams	15 days	Wed 28/12/16	Tue 17/01/17
Project Final Documentation	87 days	Tue 17/01/17	Wed 17/05/17
Model Finalization	27 days	Tue 17/01/17	Wed 22/02/17
Research Machine Learning(Prediction)	7 days	Tue 17/01/17	Wed 25/01/17
Meet with Supervisor about implementing the machine learning aspect	1 day	Wed 25/01/17	Wed 25/01/17
Implement Machine Learning	15 days	Thu 26/01/17	Wed 15/02/17
Finish Machine Learning	5 days	Thu 16/02/17	Wed 22/02/17
Testing Phase	6 days	Thu 23/02/17	Thu 02/03/17

Test the Web Application & Machine Learning	3 days	Thu 23/02/17	Mon 27/02/17
Meet with Supervisor on progress	3 days	Mon 27/02/17	Wed 01/03/17
Final testing phase	3 days	Thu 02/03/17	Mon 06/03/17
Prepare Showcase Materials	4 days	Mon 10/04/17	Thu 13/04/17
Upload final code submission	1 day	Wed 17/05/17	Wed 17/05/17
Technical Report, Findings	49 days	Fri 10/03/17	Wed 17/05/17
Review	1 day	Wed 10/05/17	Wed 10/05/17
Upload Final Documentation	1 day	Wed 10/05/17	Wed 10/05/17

7.3 Monthly Journals

Reflective Journal

Student Name: Declan Barnes

Programme (BSc in Computing (Data Analytics)):

Month: September 2016

My Achievements

This month, I was able to research my idea for my final year project, while researching I was able to acquire datasets that I could use for my project. I also attended my first meeting with Michael Bradford on the 30th of September. This was a great opportunity to see if my idea was good enough for a final year project.

My contributions to the projects included:

- > My first meeting in regards to my project,
- > Acquiring datasets for my project.

My Reflection

I felt, it worked well to find suitable datasets that could be used for my project, as the getting hold of datasets that you can use is one of the most important parts to a data analytics project. I also felt that meeting with my supervisor before the project pitch was a good idea, this allowed me to get an idea on what I should add to my project to improve it.

Intended Changes

Next month, I will try to get started on step 1 of my project, as I want to give myself plenty of time to work on the project as I anticipate that I may encounter issues later on towards the final part of my project. Therefore I want to have enough time to sort out any of the issues that arise.

Meetings

Date of Meeting: 30th September 2016

Items discussed: My project idea, was it a good idea, what I could do to improve it.

Action items: Michael Bradford told me that it was great that I had already acquired datasets to use and that instead of just doing a dashboard, I should also add a machine learning aspect to my project as the dashboard alone would not be enough.

?

?

Reflective Journal

Student Name: Declan Barnes-x13114352

Programme (BSc in Computing (Data Analytics))

Month: October 2016

My Achievements

This month, I was able to set up my MySQL database and acquire more datasets for my project. I also set up an account with Tableau, which I will use for data visualization. I also started to think about what I want to show for my mid-point presentation. I also had my first meeting with my supervisor (Paul Stynes).

My contributions to the projects included:

- Meeting with my supervisor and setting a schedule for weekly meetings with him.
- Setting up the database that will house my datasets.
- Planning out what to show for the mid-point presentation.
- Creating an account with Tableau.
- Acquiring more datasets for my project.

My Reflection

I felt it worked well to find suitable more datasets that could be of use for my project. I also thought about what kind of information I want to display on my dashboard. By doing this I have set myself further goals that I need to do by the end of next month. I felt that October was overall a productive month for my project.

Intended Changes

Next month, I will try to get the more work done on my dashboard and start preparing for a mid-point presentation.

Supervisor Meetings

Date of Meeting: 31st October 2016

Items discussed: My project idea, Requirement specification.

Action items: Paul Stynes told me that the next deliverable due was the requirement specification and that I should try to work out the use cases for my project by the next meeting.

?

?

Reflective Journal

Student Name: Declan Barnes

Programme: BSc in Computing (Data Analytics)

Month: November 2016

My Achievements

This month, I wrote up and completed the requirements specification for my final year project, and began to prepare for the technical report towards the end of the month. I continued to research my machine learning algorithms but was unable to do more due to time constraints with CA's and projects.

My contributions to the projects included:

- Uploaded Requirements Specification
- Continued to meet with my supervisor.

My Reflection

I felt, it worked well to continue to meet with my supervisor (Paul Stynes) as I had to write up the requirements specification and get ready for the technical report. By meeting with Paul, it gave me a good amount of insight into what was expected.

Intended Changes

I had to change parts of my requirements specification to the way Paul Stynes suggested. This involved my trimming down my use cases. I was also told by Paul to write in 3rd person and not to refer to myself within the technical report.

Meetings

Date of Meeting: 14th & 21st November 2016

Items Discussed: Requirements Specification, Building a Machine Learning App.

Action Items: Paul Stynes gave me advice on how to write up my requirements specification and he also said that for my Machine Learning portion to my project, I could build an app that would show the output.

?

?

Reflective Journal

Student Name: Declan Barnes

Programme: BSc in Computing (Data Analytics)

Month: December 2016

My Achievements

This month, I prepared for the mid-point presentation and created my prototype

My contributions to the projects included:

- Prepare prototype
- Prepare mid-point presentation.

My Reflection

I worked on creating my prototype to show during my mid-point presentation as well as preparing my tech report. I thought I was well on track to do well in the mid-point. After having the mid-point, I am very pleased with the results.

Intended Changes

I had great feedback from my supervisor (Paul Stynes) and Damian McNamara in relation to how I should proceed, I will research more into the ideas they gave me for the web application aspect of my project.

Meetings

Date of Meeting: 13th & 16th December 2016

Items discussed: Tech Report and Mid-Point Presentation.

Action Items: Paul Stynes gave me advice on what I should do for my prototype (A 3 tier-Architecture web application) and he also gave me advice on my tech report.

Reflective Journal

Student Name: Declan Barnes

Programme: BSc in Computing (Data Analytics)

Month: January 2017

My Achievements

This month, I was able to do further work on the web application aspect of my project while continuing my research into which machine learning approaches would suit the identification of new HSE health centre locations in Ireland. I came to the conclusion that optimisation would be the best way to go about this part of my project.

My contributions to the projects included:

- Conducted further work on the web application,
- Narrowed down the machine learning approach to optimisation.

My Reflection

I felt that I got a decent amount of work done for the month of January even though I had 4 exams to study and sit for two weeks in the month. I managed to narrow down the best machine learning approach to identifying new HSE support centres, which allowed me to focus on specific machine learning algorithms that are used for optimisation. By completing this I saved myself from wasting further time, so I can now move to the next stage of this portion of my project which will involve applying the machine learning approach and getting it working.

Intended Changes

Next month, I will try to get step 1 of my project working, which involves implementing the machine learning aspect of my project. I will also continue to work on the web application in order to improve it further.

Meetings

Date of Meeting: I didn't have any meetings with my supervisor in January as I had exams to study and sit. Also Semester 2 didn't start until the final week of January and my supervisor set up new meetings for February onwards.

Reflective Journal

Student Name: Declan Barnes

Programme: BSc in Computing (Data Analytics)

Month: February 2017

My Achievements

This month, I was able to do further work on the Machine Learning aspect of my project, I managed to plot the current HSE support centres on a GIS map with a distance radius around each centre location, this will be needed for the next stage of the project. I have also done more work on my web application, I fixed up the user interface and figured out how to run R code from within the application itself.

My contributions to the projects included:

- Conducted further work on the web application,
- Mapped current HSE support centre locations on a GIS map using the R package (Leaflet).
- I added a distance radius around each of the plotted HSE support centres, this will be needed in the next stage of the machine learning part of my project.

My Reflection

I felt that I got a good amount of work done for the month of February. I am very happy with the work that I managed to get done in regards to identifying new HSE support centre locations. I am also happy with how the web application is turning out and I have a clear idea about what I need to do next. All in all I am making very good progress in my final year project and I am optimistic about the overall outcome of the project.

Intended Changes

Next month, I will try to add the population density to the GIS map for the machine learning part of my project, I will need to know where the people are situated in order to properly identify the optimal locations for the support centres. I will also try to get some R code running through the web application that will produce adequate output.

Meetings

Date of Meeting: 16th, 20th and 27th of February

I met with my supervisor (Paul Stynes) on the 16th, 20th and 27th of February in order to get advice on my project and to make sure I was moving in the right direction. The feedback that I got back from Paul was very helpful and enabled me to move forward in my project.

Reflective Journal

Student Name: Declan Barnes

Programme: BSc in Computing (Data Analytics)

Month: March 2017

My Achievements

This month, I attempted to improve my Java application by adding the integration into the application. I researched the various ways of doing this and settled on using a caller, a caller provided me with exactly what I needed to progress.

My contributions to the projects included:

- Conducted further work on the application.
- Researched a caller (a Java and a bridge).
- Applied a caller to the Java application.

My Reflection

I felt that I got a good amount of work done for the month of March. I am very happy with the work that I managed to get done in regards to the web application, this will allow me to progress further with my project.

Intended Changes

Next month, I will try to add several plots such as heatmaps to the Java application and see if there is any further work to be done to increase the functionality of my Java application.

Meetings

Date of Meeting: 13th and 27th of March

I met with my supervisor (Paul Stynes) on the 13th and 27th of March in order to get advice on my project and to make sure I was moving in the right direction. The feedback that I got back from Paul was very helpful and enabled me to move forward in my project. I got more advice on how best to go about my Java application.

7.4 Usability Testing-User Forms

User 1	Date of test: 04/05/2017
Purpose of Test	The purpose of the test was: The Disability Application is functioning properly without problems. The User can use the application with relative ease.
Test Environment	The environment was: Hardware used: MacBook Pro (Retina, 13-inch)-macOS Sierra and R version 3.3.1 (2016-06-21).
Test Steps	User 1 starts application User1 navigates to heatmap GUI User 1 selects gender and disability type from drop-down menu User 1 clicks 'run' button User 1 views and interprets spatial heatmap results.
App Performance	When User 1 ran, and used the application, the application performed as it should have with no errors or mistakes of any kind.
# of Errors/Mistakes	0
User 1 Feedback	User 1 mentioned that the application was easy to use and that they had no problems understanding the results of the heatmap.
Actions Needed	N/A
Resolution	N/A

Date of Test	Date of Test : 06/05/2017
Purpose of Test	<p>The purpose of the test was:</p> <p>The Disability Application is functioning properly without problems.</p> <p>The User is able to use the application with relative ease.</p>
Test Environment	<p>The environment was:</p> <p>Hardware used: MacBook Pro (Retina, 13-inch)-macOS Sierra and R version 3.3.1 (2016-06-21).</p>
Test Steps	<p>User 2 starts application</p> <p>User 2 navigates to heatmap GUI</p> <p>User 2 selects gender and disability type from drop-down menu</p> <p>User 2 clicks 'run' button</p> <p>User 2 views and interprets spatial heatmap results.</p>
App Performance	<p>When User 2 ran and used the application, no errors or problems occurred from the application. The application performed well.</p>
# of Errors/Mistakes	0
User 1 Feedback	<p>User 2 found the application easy to navigate and use, they said they had no issues when using the application. User 2 also said they understood the results of the spatial heatmap.</p>
Actions Needed	N/A
Resolution	N/A