

National College of Ireland

BSc (Honours) in Computing - Evening

Data Analytics

2020/2021

Jonathan Sammons

17102561

X17102561@student.ncirl.ie

Analysis of Salary Scales and Higher Education
Attainment in the Republic of Ireland

Contents

Executive Summary	2
1.0 Introduction	2
1.1. Background	2
1.2. Aims.....	2
1.3. Technology.....	2
1.4. Structure	4
2.0 Data	4
3.0 Methodology.....	7
4.0 Analysis	15
5.0 Results.....	15
6.0 Conclusions	17
7.0 Further Development or Research	18
8.0 References	18
9.0 Appendices.....	18
9.1. Project Plan	19
10.0 Objectives.....	20
11.0 Background	21
12.0 Technical Approach.....	22
13.0 Special Resources Required	22
14.0 Project Plan	22
15.0 Technical Details	22
16.0 Evaluation	23
16.1. Reflective Journals	23
Introduction	24
October 10 th , 2020	25
November 30 th , 2020	26
February 28 th , 2021	27
March 23 rd , 2021.....	29
April 19 th , 2021	30
16.2. Link to the interactive map.....	30

Executive Summary

The purpose of this report was to research how beneficial higher levels of education are when you want to pursue a higher salary. As a student myself, I was curious as to how much of an increase of salary I could expect if I decided to pursue a master's degree after this degree. The report shows how I went about creating an interactive map of the Republic of Ireland, including the technologies used and an examination of the data. After building the map and interacting with the data, I found that you tend to lose from your salary by moving from an Advanced Certificate to a Higher Certificate. Unsurprisingly, Dublin tends to pay the most for each level of degree that you have, up until you obtain a Ph.D, where the highest median income is actually in Laois. I also looked at the breakdown of degrees between men and women, and women tended to have a larger percentage of obtaining degrees in each county up until the level of a Ph.D, where men overtake women in nearly all counties. From the data collected, I can see that for, personally, pursuing a postgraduate degree should bring roughly a 19% increase in salary if I decided to pursue it, since I am based in Kildare. The question then is if I feel that the time spent is worth that 20%, to which, at the moment, I think I'll take a break and enjoy being finished with college.

1.0 Introduction

1.1. Background

As a student pursuing my honours degree, the question was in my mind, as I am sure it is in my fellow students minds, if I should immediately start working towards a postgraduate degree after I finish my honours degree, or should I take a break? As I pondered this question, I wasn't sure what kind of financial gain I could expect if I decided to move on to the next level of education – would it be a 5% or 10% gain; or could it be something much higher like 40%? I decided that this was an important factor in determining the desirability of pursuing this higher level of education.

1.2. Aims

When undertaking the project, the initial goal was to find out how much of an increase in salary I could expect if I decided to move on to a postgraduate degree. I decided to expand upon this and take a look at each level of higher education available in the Republic of Ireland, as I thought that this data would be interesting to evaluate and could possibly serve some purpose to providing feedback to the government as to the current state of education in the country.

I decided that, while the financial information is interesting, what could be even more interesting to build on top of that is the breakdown of males versus females in each county in the Republic, along with the number of people in each county that actually have each level of each degree.

1.3. Technology

There are quite a few technologies being used in my project, the main being RStudio with my R script containing all of my code. Most of the rest of the technologies used are

frameworks from within RStudio. Here's a breakdown of each framework and what it is used to achieve:

- **Shiny**: Shiny is a framework that allows the user to build interactive web applications inside of RStudio. It allows you to utilize HTML, CSS, and JavaScript all from within RStudio. I had to learn this very verbose framework from the ground-up for this project so that I could create the interactive maps (Shiny, 2021).
- **Shinydashboard**: This framework allows you to create dashboards on top of your Shiny application to add additional functionality (Create Dashboards with 'Shiny' [R package shinydashboard version 0.7.1], 2021).
- **Leaflet**: The Leaflet framework allows you to create and customize interactive maps in RStudio or with the Shiny package. I use this framework to create my base layer map of the world that my shape file of the counties in the Republic of Ireland sits on (Leaflet, 2021).
- **Rgeos**: rgeos is used to perform topology operations on geometries, which would be the shape files that I use for all of the counties in the Republic of Ireland (Rgeos, 2020).
- **Dplyr**: This framework is used a small bit in my project for selecting data between datasets when I was creating the data frames that I wanted by using the 'select' keyword. The framework offers lots of different ways to manipulate data (A Grammar of Data Manipulation, n.d.).
- **Maptools**: This is another framework used to interact and manipulate geometric data, similar to rgeos (Maptools, 2021).
- **Rgdal**: This framework is used in conjunction with the sp package below. Rgdal allows you to provide bindings to the 'Geospatial' Data Abstraction Library and perform projection and transformation operations (Bindings for the 'Geospatial' Data Abstraction Library [R package rgdal version 1.5-23], n.d.).
- **Sp**: Sp is used in conjunction with rgdal and it provides access to classes and methods for spatial data, such as plotting data as maps or spatial selection and retrieving coordinates (Classes and Methods for Spatial Data [R package sp version 1.4-5], 2021).

- **Sf:** Sf is used with spatial data and represents simple features as records in data frames or tibbles with a geometry list-column. It helps to read and write data from spatial information (Simple Features for R, n.d.).

In addition to these frameworks in RStudio, I also used shinyapps.io. Shinyapps.io is a service within RStudio that allows you to publish your Shiny projects to the web using RStudio's servers. Fortunately, they allow up to 5 projects to be live at a time with up to 25 hours of interactivity allowed per month before you have to pay. I used this service to make my project live and accessible by anyone on the web, which saved a lot of headache that I thought I would originally have by using Amazon Web Services or a similar web-hosting platform.

In addition to the above technologies, I also used Microsoft Excel to obtain and view my original data, as it was provided in CSV format from the Central Statistics Office of Ireland, and I performed some data manipulation in Excel as well. Lastly, I used Google Chrome as my browser of choice when browsing the internet to obtain data or find information at any point when I was stuck in creating this project.

I had originally also planned on using ArcGIS as well as the Google Maps API, however these initial plans fell through. ArcGIS was a very expensive tool to obtain, so I had to look for alternatives to using it. I was actually using the Google Maps API in early iterations of my project to create my base map, but I found it too difficult to get a map that I was happy with as the map colors and types provided by the API was quite limited for what I was looking for. After doing research, I came across Leaflet and decided that it was a preferable solution.

1.4. Structure

The rest of this documentation will take a look at my data and the processing that I performed on it as well as how I analysed the data. I'll cover why I performed certain actions as well as discuss problems that I ran into, and explain my final analysis and results. I'll finish this off by commenting on further work that I would like to carry out in the future to further expand on this project, as well as appending previous work that was completed leading up to this finished paper.

2.0 Data

The data that I used for my project was all sourced from the Central Statistics Office of Ireland. In total, I ended up pulling 4 datasets from the Central Statistics Office and combining them together to build my project. I'll describe each dataset and their variables below, so that you have an understanding of them.

Median earned income per person by highest level of education achieved and county of work, 2016 – This was the first dataset that I acquired and, fortunately, gave me a great jumping off point for my project. The CSO had already prepared the data in the exact format that I was looking for, so it was easy to pull into my project.

Show Table: Table 5.2: Median earned income per person by highest level of education achieved and county of work, 2016

Table 5.2: Median earned income per person by highest level of education achieved and county of work, 2016

County of work	€	€	€	€	€	€	€	€	€	€	€
	No education	Primary	Lower secondary	Upper secondary	Technical	Advanced certificate	Higher certificate	Ordinary degree	Honours degree	Postgraduate	Ph.D
State	14,684	15,165	19,809	19,978	20,717	26,940	26,506	32,193	37,830	46,916	60,912
Carlow	19,092	18,906	22,521	18,556	19,552	27,345	23,057	29,136	34,326	47,277	61,091
Dublin City	20,038	23,996	27,266	22,143	25,549	32,569	33,841	40,336	44,664	53,757	66,250
South Dublin	23,018	24,511	26,957	26,888	26,302	35,223	32,617	38,623	44,479	51,660	74,566
Fingal	21,966	22,891	26,207	26,251	24,701	31,626	30,780	35,676	42,927	49,811	73,647

As you can see from the above example image, the dataset had each county in the Republic of Ireland and the median of the income earned for the level of education that the people in that county had achieved. I was able to take this data and drop off the data that I didn't need (that being everything below higher education), and upload it to RStudio with doing almost nothing to it. The data can be viewed on the CSO's website through [here](#).

Population Aged 15 Years and Over 2011 to 2016 – This dataset was the next that I was able to locate to help with my analysis. This dataset allowed me to select the year of 2016 to match the median earned income per person dataset, and it provided me with a breakdown of the number of men and women in each age group, by county, along with the highest level of education that they attained.

Population Aged 15 Years and Over 2011 to 2016 (Number) by Highest Level of Ed, Sex, Age at which Full Time Education Ceased, County and City and Census Year			
			2016
	No formal education		
	Male		
		18 years	
		State	811
		Carlow	3

Unfortunately, the structure of the data from this file was quite jarring and not at all easy to navigate, so I had to perform restructuring on it to make it useable for my needs and so that it matched my other datasets. I'll cover this in more depth in the next section, as this was a multi-step process. This dataset can be obtained from the CSO [here](#). That said, the data that the dataset contains is the level of education that was attained along with the gender, age, and the number in each county. This proceeds from the age of 18, up to 25+.

Census2011_Province_generalised20m – This dataset that I obtained from the CSO is the shapefile that I was able to use to overlay the counties of the Republic of Ireland on top of the map I generated using Leaflet. This shapefile contains 21 variables that I'll outline below:

- **NUTS1**: An identifier for the NUTS1NAME. Since NUTS1NAME is all Ireland, the identifier here is IE0 for all values.
- **NUTS1NAME**: This is the country in question. The value for all datapoints here is Ireland.
- **NUTS2**: This is a unique identifier for the NUTS2NAME. The values found here depend on the regions in NUTS2NAME, being either IE01 or IE02, depending on if it's Border, Midland, and Western (IE01) or Southern and Eastern (IE02).

- **NUTS2NAME**: This is used with NUTS2, it's either Border, Midland, and Western or Southern and Eastern, depending on the county location.
- **NUTS3**: This is a unique identifier for NUTS3NAME, it is either IE011, IE012, IE021, IE022, IE023, IE024, or IE025, depending on the value in the NUTS3NAME.
- **NUTS3NAME**: This describes the location of the county as being Border, South-East, Dublin, Mid-East, Midland, South-West, or Mid-West.
- **COUNTY**: This is a unique number for each county, ranging from 1 to 34.
- **COUNTYNAME**: This is the name of the county for the row.
- **GEOGID**: This is a unique ID – it's the same as the COUNTY variable, just with a C in front of the number.
- **MALE2011**: The number of Male's in the county in 2011.
- **FEMALE2011**: The number of Female's in the county in 2011.
- **TOTAL2011**: The number of people in the county in 2011.
- **PPOCC2011**: Permanent private housing units.
- **UNOCC2011**: Total unoccupied dwellings (vacant dwellings + temporarily absent dwellings).
- **VACANT2011**: Total vacant dwellings (vacant houses + vacant flats + holiday homes).
- **HS2011**: Total housing stock.
- **PCVAC2011**: Vacancy rate.
- **LAND_AREA**: Land area of the county.
- **TOTAL_AREA**: Total area of the county (including sea).
- **CREATEDATE**: The date that the dataset was created (27-03-2012).
- **Geometry**: The geometry of the county. This is the most important part of the data for my needs as this gives the shape of each county that I can use to overlay on my map.

The page from the CSO that I used to obtain this data can be found [here](#), while the specific dataset is the [Census2011 Province generalised20m.zip](#) file from the available files. The fact that this data is from 2011 is okay, as I'm not using any of the datapoints that would have really changed since then, like the number of males or females in the county – I used my next dataset for that. The most important datapoint that I really needed from this was the geometry.

VACANT2011	HS2011	PCVAC2011	LAND_AREA	TOTAL_AREA	CREATEDATE	geometry
7328	33044	22.2	1791.17522	1837.48425	27-03-2012	list(list(c(170576.515, 170674.58, 170695, 170800.463, 17089...
7277	33711	21.6	1855.88871	1932.05547	27-03-2012	list(list(c(204973.6421, 205332.9274, 205549.1104, 205796.9...
23996	83918	28.6	4764.35784	4862.72065	27-03-2012	list(list(c(146275.913, 146256.582, 146179.322, 146215.798, ...
3398	25056	13.6	1273.44242	1294.71893	27-03-2012	list(list(c(263301.0136, 263539.5217, 263696.7101, 263695.9...

Population 2016 – The final dataset that I brought into my project was a fairly simple and straightforward dataset showing the number of people in each county, by gender.

County	Males	Females
Carlow	28465	28467
Dublin City	272270	282284
Dun-Laoghaire Rathdown	104584	113434
Fingal	145240	150780
South Dublin	136277	142490

3.0 Methodology

There was quite an extensive process in preparing the data so that I could combine everything together and generate the map – not to mention adding the ability to interact with it so that the user can choose to see the data that they'd like to see. The first obstacle was getting the data ready so that I could combine them together. I decided that the best way to do this was to combine each data frame by the county name, so that the data would combine across the data frames based on that. The only issue was that there was slight variations in the names of some of the counties, so I had to correct this in Excel first. The first set of data that I loaded in was my shapefile, as I figured that I needed everything to match up to it since it held the actual geometry of each county. After loading it into RStudio, I created a new variable containing only the parts of the original that I wanted to keep – GEOGID, NUTS1, NUTS1NAME, NUTS2, NUTS2NAME, NUTS3, NUTS3NAME, COUNTY, COUNTYNAME, and geometry. I then proceeded to set the county names to what I would maintain through all of the datasets so that they would be the same, basing this off of the county names in my first dataset – median earned income per person.

```
#Updating the county names to match to the median_earned_income csv
county_data[2, 9] = "Limerick"
county_data[6, 9] = "Waterford"
county_data[8, 9] = "Galway"
county_data[9, 9] = "Leitrim"
county_data[10, 9] = "Mayo"
county_data[11, 9] = "Roscommon"
county_data[12, 9] = "Sligo"
county_data[13, 9] = "Cavan"
county_data[14, 9] = "Donegal"
county_data[15, 9] = "Monaghan"
county_data[16, 9] = "Carlow"
county_data[20, 9] = "Dun-Laoghaire Rathdown"
county_data[21, 9] = "Kildare"
county_data[22, 9] = "Kilkenny"
county_data[23, 9] = "Laois"
county_data[24, 9] = "Longford"
county_data[25, 9] = "Louth"
county_data[26, 9] = "Meath"
county_data[27, 9] = "Offaly"
county_data[28, 9] = "Westmeath"
county_data[29, 9] = "Wexford"
county_data[30, 9] = "Wicklow"
county_data[31, 9] = "Clare"
county_data[34, 9] = "Kerry"
```

I also updated the variable names of my new dataset, changing COUNTY to COUNTYNUMBER and COUNTYNAME to County, as I felt this made the data easier to understand.

I then proceeded to upload my median earned income dataset to RStudio and grab only the data that I wanted from it as well, dropping anything that wasn't the County or higher education. The next obstacle was getting the gender information cleaned up which was a multi-step process as I mentioned before. I started by removing the education that wasn't needed from the CSV file. After that, I manually combined data in the CSV following the tidydata process, so that I could combine it easily in RStudio.

Education Level Attained	Gender	Age	County	Occurrences
Technical	Male	18 years	State	19803
Technical	Male	18 years	Carlow	277
Technical	Male	18 years	Dublin City	1649
Technical	Male	18 years	Dun Laoghaire-Rathdown	531
Technical	Male	18 years	Fingal	1128
Technical	Male	18 years	South Dublin	1034
Technical	Male	18 years	Kildare	899
Technical	Male	18 years	Kilkenny	493

After getting to this point with the data, I uploaded it into RStudio so that I could use it to merge the data into a manageable data frame. The code is commented out and I have removed it from my code since it was a once-off operation, but here's the code before I removed it:

```
##Cleaning up the gender_data by summarising and merging it
#genderDF = education_by_gender %>%
# group_by(County, Education.Level.Attained, Gender) %>%
# summarise(list(Occurrences = sum(Occurrences)))
#Removing unnecessary row
#genderDF = genderDF[-c(193),]
#genderDF = genderDF[-c(1),]
#genderDF = genderDF[-c(364:377),]

#genderDF = apply(genderDF, 2, as.character)
#write.csv(genderDF, "C:\\Users\\Sammons Family\\Desktop\\Co1")
```

This resulted in a clean dataset that I was able to then make a few more manual adjustments to in Excel with column headers, resulting in data that I could use and merge with my other datasets:

County	Technical Male	Technical Female	Advanced certificate Male	Advanced certificate Female
Carlow	788	1101	966	487
Cavan	1342	1787	1361	769
Clare	1499	2052	1964	1038
Cork City	1482	1811	1221	807

Once I had the gender data cleaned and ready, I just uploaded my population data into RStudio and merged all of the data frames by County, pulling all relevant information for each county into their row. This resulted in a data frame that I called all_data that contained all of the data from my median earned income, gender data, and population data. I proceeded to clean up my data from here, setting the salary character values to numeric so that I could perform operations on it later.

```

# Cleaning up and converting salary data to numeric
#Technical
all_data$Technical = gsub(" ", "", all_data$Technical)
all_data$Technical = gsub(",", "", all_data$Technical)
technical = as.numeric(all_data$Technical)
#Advanced certificate
all_data$Advanced.certificate = gsub(" ", "", all_data$Advanced.certificate)
all_data$Advanced.certificate = gsub(",", "", all_data$Advanced.certificate)
advanced_certificate = as.numeric(all_data$Advanced.certificate)
#Higher certificate
all_data$Higher.certificate = gsub(" ", "", all_data$Higher.certificate)
all_data$Higher.certificate = gsub(",", "", all_data$Higher.certificate)
higher_certificate = as.numeric(all_data$Higher.certificate)
#Ordinary degree
all_data$Ordinary.degree = gsub(" ", "", all_data$Ordinary.degree)
all_data$Ordinary.degree = gsub(",", "", all_data$Ordinary.degree)
ordinary_degree = as.numeric(all_data$Ordinary.degree)
#Honours degree
all_data$Honours.degree = gsub(" ", "", all_data$Honours.degree)
all_data$Honours.degree = gsub(",", "", all_data$Honours.degree)
honours_degree = as.numeric(all_data$Honours.degree)
#Postgraduate
all_data$Postgraduate = gsub(" ", "", all_data$Postgraduate)
all_data$Postgraduate = gsub(",", "", all_data$Postgraduate)
postgraduate = as.numeric(all_data$Postgraduate)
#Ph.D
all_data$Ph.D = gsub(" ", "", all_data$Ph.D)
all_data$Ph.D = gsub(",", "", all_data$Ph.D)
ph_d = as.numeric(all_data$Ph.D)

```

At this point, I was ready to try and merge in my geometry data with the rest of my data, however shape files are a special data type, so I had to set it as spatial data. I also ran into the problem where the coordinate data provided by the CSO was in a different coordinate type than what RStudio could understand, so I had to update the coordinate type to datum WGS84:

```

convertedData = as_spatial(county_data)
convertedData$id = rownames(convertedData)

# Must convert the datum so that it's readable by leaflet
convertedData = spTransform(convertedData, CRS("+proj=longlat +datum=WGS84 +no_defs"))
oo = merge(convertedData, all_data, by="County")

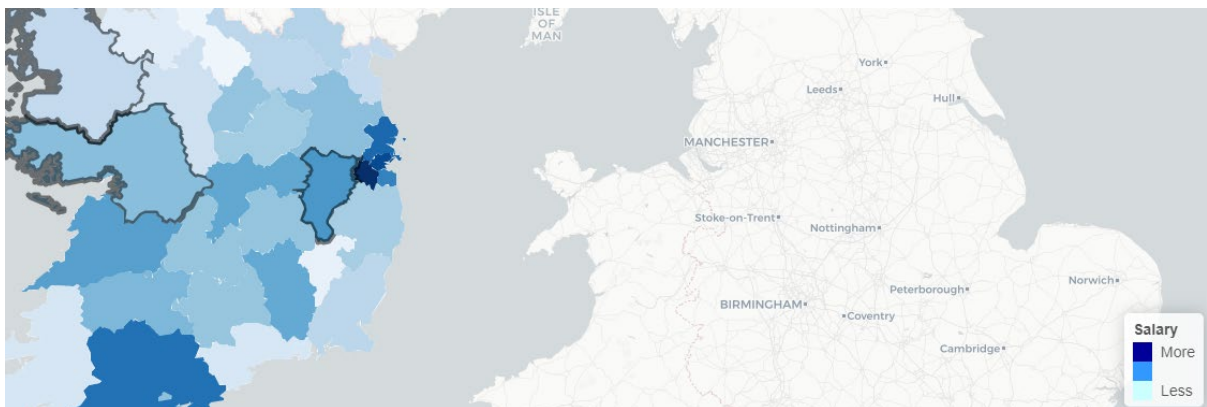
```

Finally, I had my data frame that contained all of my datasets, oo. I then proceeded to update some more values to numeric so that I could perform mathematical operations on them. I also setup a colour scheme on my map so that each county was represented by a shade of blue, from light to dark, depending on where on the scale of values it sat. This is all pulling from the salary scale, so the higher the salary in a county, the darker blue that county will be for the chosen level of education:

```

# Setting up the color scheme based off of
#Technical
pal_technical = colorNumeric(
  palette = "Blues",
  domain = oo$technical)
#Advanced certificate
pal_advanced_certificate = colorNumeric(
  palette = "Blues",
  domain = oo$advanced_certificate)
#Higher certificate
pal_higher_certificate = colorNumeric(
  palette = "Blues",
  domain = oo$higher_certificate)
#Ordinary degree
pal_ordinary_degree = colorNumeric(
  palette = "Blues",
  domain = oo$ordinary_degree)
#Honours degree
pal_honours_degree = colorNumeric(
  palette = "Blues",
  domain = oo$honours_degree)
#Postgraduate
pal_postgraduate = colorNumeric(
  palette = "Blues",
  domain = oo$postgraduate)
#Ph.D
pal_phd = colorNumeric(
  palette = "Blues",
  domain = oo$ph_d)

```



At this point, I'm finally ready to start working with the Shiny framework and building the user interface that will be interactive for anyone that uses the tool. There are quite a few moving parts with the whole user interface, so I'll break it down into chunks, starting with the initial page when it's loaded.

```

##Shiny UI
css = HTML("
  .leaflet-top, .leaflet-bottom {
    z-index: unset !important;
  }

  .leaflet-touch .leaflet-control-layers, .leaflet-touch .leaflet-bar {
    z-index: 1000000000 !important;
  }
")
ui = fluidPage(
  tags$head(tags$style('.selectize-dropdown {z-index: 10000}')),
  absolutePanel(top = 10, left = 500, fixed = TRUE,
    tags$div(style = "opacity: 0.90; background: #FFFEE; padding: 4px; ",
      htmloutput("header")
    ),
  selectInput("education", "Level of education:",
    list("Technical", "Advanced Certificate", "Higher Certificate", "Ordinary
  ),
  textOutput("result"),
  leafletOutput("mymap"),
  absolutePanel(bottom = 100, left = 10, fixed = FALSE,
    tags$div(style = "opacity: 0.90; background: #FFFEE; padding: 4px; ",
      helpText("Please click on a county for more information."),
      htmloutput("text")
    )
  ),
  plotOutput("genderPie"),
  plotOutput("popPie")
)

```

To start with, I have some CSS on the page to prevent a drop down from appearing behind the map generated by leaflet. Prior to adding this CSS, some of the drop-down options were obscured by the map which made navigation difficult. After that, I have a header displayed which is just an explanation as to what the user is looking at and a drop down that allows the user to choose the level of education that they would like to take a look at. The list is cut off in the screenshot due to length, but the complete list is Technical, Advanced Certificate, Higher Certificate, Ordinary Degree, Honours Degree, Postgraduate, and Ph.D.

After that, I just have the UI showing the map that's generated as well as text below the map for information on the county that has been clicked on. At the bottom, I have two pie charts that are also generated depending on the county, showing the number of males versus females in the county have this degree and the amount of the population in the chosen county with the degree.

That covers the user interface part of the shiny framework – the next part is the server side, or the actions that are taken based on how the user interacts with the map. This is where the bulk of the processes (and code) are. I have a separate function for each level of education, here's an example of the Technical level of education:

```

##Shiny server
server = function(input, output, session){
  output$header = renderUI({
    header_str = paste("Salary Scales by County in the Republic of Ireland")
    HTML(paste("<h2>",header_str,"</h2>"))
  })

  output$mymap = renderLeaflet({
    #####
    ##Technical Map##
    #####
    if(input$education == "Technical"){
      # Using leaflet to plot the shape file on a map for interactivity
      leaflet(data=oo,
              options = leafletOptions(zoomControl = TRUE,
                                       minZoom = 6, maxZoom = 10,
                                       dragging = TRUE)) %>%

        addTiles() %>%
        addProviderTiles(providers$CartoDB.Positron) %>%
        # setting the view to Ireland
        setview(lat=53, lng=-8, zoom=7) %>%
        #Adding the polygon information for the counties
        addPolygons(fillColor = ~pal_technical(oo$Technical),
                   color = NA,
                   fillOpacity = 1,
                   highlight = highlightOptions(weight = 2,
                                                color = "black",
                                                fillOpacity = .1,
                                                bringToFront = FALSE),

                   label = ~County,
                   layerId = ~County) %>%
        # Adding a legend to the map
        addLegend(data = convertedData,
                  colors = c("#000099", "#3399ff", "#ccffff"),
                  labels = c("More", " ", "Less"),
                  opacity = 1.0,
                  title = "Salary",
                  position = "bottomright")
    }
  })
}

```

The server renders what the user is seeing, so as you can see at the top of this code, it's passing the header that's displayed at the top of the page that explains what the user is seeing on the web page. After that, the leaflet map is being rendered and there are quite a few functions that control how the map is rendered and what it looks like. I'm basing the map being generated by the education level being selected in the drop down by the user, and changing this in if/else if statements. Leaflet is then generating the map based on the data frame oo and there's a few settings configuring how the map looks upon loading. I have it set to allow a zoom of 6 up to 10 and they can drag around the map. The colour scheme of the map generated by leaflet is being controlled by the 'providers\$CartoDB.Positron' line.

I had to manually find a latitude and longitude for Ireland and set it so that the map generated starts there, hence the lat = 53 and lng = -8, and a zoom of 7 is how the map starts. The addPolygons section is adding in the geometry data from my oo spatial data frame so that each county is on top of the leaflet map. I have the fillColor set to the different shades of blue that I configured prior to the Shiny framework section and the counties are highlighted with a border after scrolling over them. Next, the name of the county appears when you mouse over it thanks to the ~County for label and the layerId is used for passing

information when a county is clicked, which we'll look at after this. Lastly, the legend is added to the map with the information displayed in it.

Next, I have the listener for click events setup which pulls through the data on the county once it's clicked on. There are a lot of variables being setup and math taking place in this part, so I'll break it down into a few sections.

```
observe({
  if(input$education == "Technical"){
    click = input$mymap_shape_click
    sub = oo[oo$county==input$mymap_shape_click$id, c("County", "Technical", "Advanced.certificate",
    #County
    ob_county = sub$county
    #Median Salary
    ob_salary = sub$Technical
    #Number of males with this degree
    ob_male = sub$Technical.Male
    #Number of females with this degree
    ob_female = sub$Technical.Female
    #Number of males in the county
    ob_male_pop = sub$Males
    #Number of females in the county
    ob_female_pop = sub$Females
    #Total population of the county
    ob_total_pop = ob_male_pop + ob_female_pop
    #Total number of people with this degree in the county
    ob_total_edu = ob_male + ob_female
    #Percentage of the county with this degree in the county
    ob_total_pc_pop = round((ob_total_edu/ob_total_pop)*100)
    #Percentage of males in the population with this degree in the county
    ob_male_pc_pop = round((ob_male/ob_total_pop)*100)
    #Percentage of females in the population with this degree in the county
    ob_female_pc_pop = round((ob_female/ob_total_pop)*100)

    ##Salary change data
    ob_adv_cert = sub$Advanced.certificate
    ob_high_cert = sub$Higher.certificate
    ob_ord_deg = sub$Ordinary.degree
    ob_hon_deg = sub$Honours.degree
    ob_post = sub$Postgraduate
    ob_ph_d = sub$Ph.D
```

First, I have this running as an if/else if section once more, looking at the drop-down selection like the previous code. In the image above, it's the section for the Technical education level. The first thing it's doing is checking for a click in the map and then creating a subset of the data from the oo data frame. I specify the data that I want pulled through here, and for each one I'm pulling through the county along with the salary data for each degree level and gender information. You can see that I'm also getting the percentage of the number of people in the county with the degree by taking the total number with the level of education and dividing by the total population, then multiplying by 100. I repeat this for the number of males and females as well.

```

ob_tech_to_adv_cert = round(((ob_adv_cert-ob_salary)/ob_salary)*100)
ob_adv_cert_to_high_cert = round(((ob_high_cert-ob_adv_cert)/ob_adv_cert)*100)
ob_high_cert_to_ord_deg = round(((ob_ord_deg-ob_high_cert)/ob_high_cert)*100)
ob_ord_deg_to_hon_deg = round(((ob_hon_deg-ob_ord_deg)/ob_ord_deg)*100)
ob_hon_deg_to_post = round(((ob_post-ob_hon_deg)/ob_hon_deg)*100)
ob_post_to_ph_d = round(((ob_ph_d-ob_post)/ob_post)*100)

##Plot data
ob_pie = c(ob_male, ob_female)
ob_pop_pie = c(ob_total_pop, ob_total_edu)
if(is.null(click))
  return()
else
  output$text = renderUI({
    county_str = paste("<b>Selected county: </b>", ob_county)
    county_pop_str = paste("<b>County population: </b>", ob_total_pop)
    salary_str = paste("<b>Median salary for current education level: </b>", ob_salary)
    male_str = paste("<b>Males with this level of degree in 2016: </b>", ob_male)
    female_str = paste("<b>Females with this level of degree in 2016: </b>", ob_female)
    change_str = paste("<b>Salary % change with each education level: </b>", "<br/>",
      "Technical to Advanced Certificate: ", ob_tech_to_adv_cert, "%", "<br/>",
      "Advanced Cert to Higher Cert: ", ob_adv_cert_to_high_cert, "%", "<br/>",
      "Higher Cert to Ordinary Degree: ", ob_high_cert_to_ord_deg, "%", "<br/>",
      "Ordinary Degree to Honours Degree: ", ob_ord_deg_to_hon_deg, "%", "<br/>",
      "Honours Degree to Postgraduate: ", ob_hon_deg_to_post, "%", "<br/>",
      "Postgraduate to Ph.D: ", ob_post_to_ph_d, "%")
    HTML(paste(county_str, county_pop_str, salary_str, male_str, female_str, change_str, sep = '<br/>'))
  })

```

Continuing down the code, I next have the math happening for finding the change in salary going from one level of education to the next. As an example, you can see that I'm taking the salary for advanced certificate and subtracting the salary of technical, then dividing by the technical salary and multiplying by 100 to get the percentage change of salary for the county when moving from a Technical to Advanced Certificate.

Next, I'm getting values set for the pie charts that are generated, then I have a small if/else statement that checks if the map was clicked on or not. After that, I have the actual data being rendered on the page using HTML. You can see that the information being passed is the County, County Population, Median salary for the current education level selected, Males with that type of degree in 2016, Females with that type of degree in 2016, and then the percentage that your salary changes for each level of degree.

```

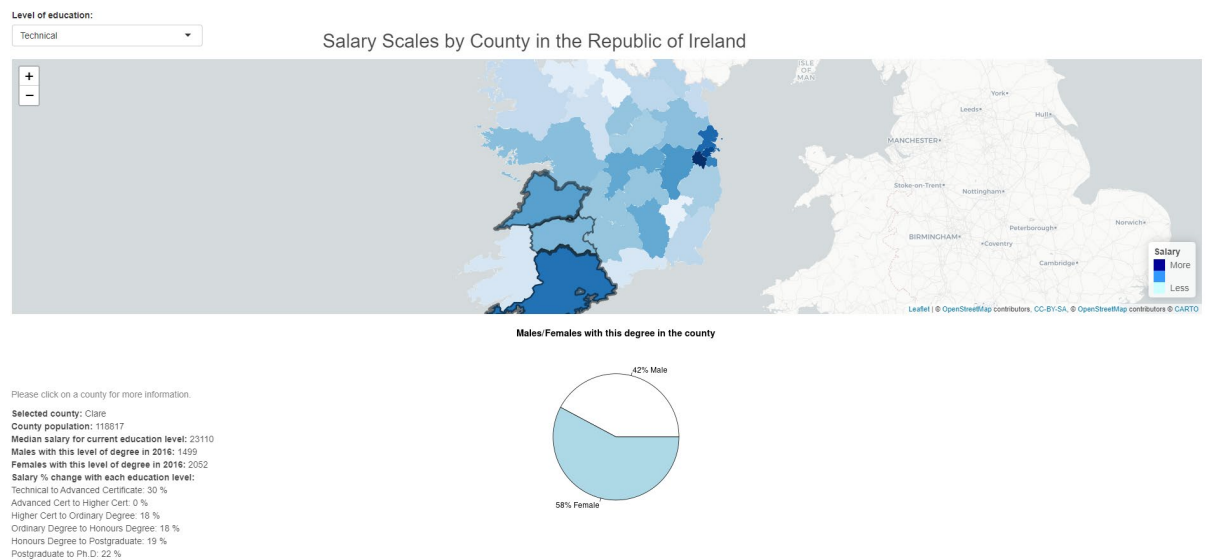
output$genderPie = renderPlot({
  total = ob_male + ob_female
  female_pct = round((ob_female)/sum(total)*100)
  male_pct = round((ob_male)/sum(total)*100)
  female_lbl = paste(female_pct) # add percents to label
  female_lbl = paste(female_pct, "% Female", sep="") # add % to label
  male_lbl = paste(male_pct) # add percents to label
  male_lbl = paste(male_pct, "% Male", sep="") # add % to label
  pie(ob_pie, labels = c(male_lbl, female_lbl), main = "Males/Females with this degree in the county")
})
output$popPie = renderPlot({
  lbl = paste(ob_total_pc_pop) # add percents to label
  lbl = paste(ob_total_pc_pop, "% with this degree", sep="") # add % to label
  pie(ob_pop_pie, labels = lbl, main = "Percentage of people with this degree in the county")
})
}

```

Lastly, I just have the pie charts being generated and added to the page below the above information. You can see that I have to setup two labels that I then combine into a single pie chart for the difference between men and women in the county with the degree. Below that is the pie chart that is generated showing the total number of people in the county with the degree. That's everything for my code at that point, it continues on for the rest of the levels of education, just changing the data slightly while keeping the structure the same.

4.0 Analysis

The end result of the above code results in a working map of the Republic of Ireland where the user can interact with each county for interesting information in regards to education and salary.



The has full reign to choose the county that they're interested in learning more information about in regards to the data and make decisions based off of it. As I mentioned previously, the map offers and easy, at-glance approach to knowing which counties are more beneficial to live in with a certain level of education, thanks to the easy to understand colouring scheme that is applied. The tool that I've created is very simple to use and I believe that the information is quite easy to understand as well, which I believe is incredibly crucial when it comes to presenting data. The ability for the user to interact with the data makes it, I believe, more enjoyable and interesting, rather than looking at a table of information.

5.0 Results

At the outset of my journey to create this application, I was curious as to whether I should continue pursuing my degree or not. I wasn't sure what kind of change was expected in my salary if I decided to go from an Honours degree to a Postgraduate, but thanks to this tool that I've created, I now know that I could probably expect around a 19% increase in salary if I went from an honours degree to a postgraduate degree:

Selected county: Kildare
County population: 222504
Median salary for current education level: 41178
Males with this level of degree in 2016: 5598
Females with this level of degree in 2016: 6789
Salary % change with each education level:
Technical to Advanced Certificate: 33 %
Advanced Cert to Higher Cert: -6 %
Higher Cert to Ordinary Degree: 30 %
Ordinary Degree to Honours Degree: 7 %
Honours Degree to Postgraduate: 19 %
Postgraduate to Ph.D: 37 %

Having this information at hand, I feel that I think I'll take some time off and enjoy being finished with college before I look to return and further my education more. There's a lot of other interesting information that I was able to find from this project as well. In almost every county, it is detrimental to go from an Advanced Certificate to a Higher Certificate. As you can see from the above screenshot, you could expect a drop in salary for a job that would require the higher certificate. I believe this is because a lot of advanced certificates are specialized jobs like being a mechanic or plumber that actually pay quite well, while a higher certificate isn't always for such a specialized role.

From the data in my map, I've also found that in almost every county and level of education, there are more women than men with a degree in the chosen level of education, with Leitrim having the highest difference with postgraduate degrees showing that women had 68% of them:



But as I mentioned, there were two levels of education where women were behind men in almost every county – Advanced Certificate and Ph.D. I believe that the reason that women are behind men for advanced certificates is because, as I mentioned, advanced certificates are usually given for jobs like being a mechanic or plumbing, both being jobs that are typically performed by men. As for the change in data when it comes to a Ph.D, my guess is that a lot of women are focusing on having a family at the point in their lives where they could dedicate the time to that level of education so not as many are able to attain it.



Figure 1 Ph.D in Laois

Males/Females with this degree in the county

Please click on a county for more information.

Selected county: Laois
County population: 84697
Median salary for current education level: 28565
Males with this level of degree in 2016: 1308
Females with this level of degree in 2016: 652
Salary % change with each education level:
Technical to Advanced Certificate: 30 %
Advanced Cert to Higher Cert: 2 %
Higher Cert to Ordinary Degree: 22 %
Ordinary Degree to Honours Degree: 13 %
Honours Degree to Postgraduate: 21 %
Postgraduate to Ph.D: 60 %

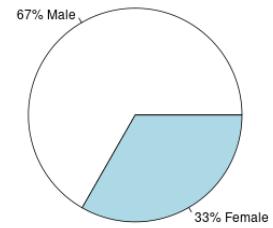
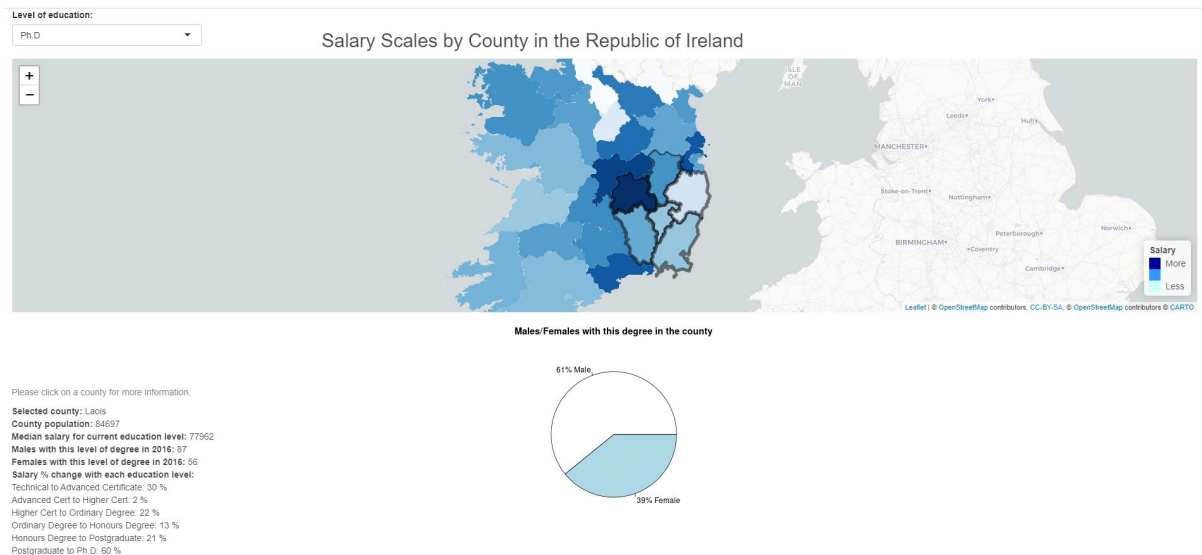


Figure 2 Advanced Certificate in Laois

Lastly, I also found it interesting that the highest salary with a Ph.D wasn't in Dublin as I would have expected, but was instead in county Laois with a median salary of 77,962:



6.0 Conclusions

This project was extremely interesting and challenging to work on. I just learned R with RStudio in this final semester of college and I had to learn a very large framework (Shiny) on top of it to create this project. I also know that this project is quite different to what I imagine a lot of Data Analytics projects will be like, but I think that's really part of its strength. I feel that a lot of the time, data is just put in your face in graphs, plots, charts, and tables and it all starts to run together. I feel that my project really allows the user to explore the data and have some fun with it, while providing interesting education at the same time. That said, I feel like the strengths of my project are also closely related to some of its weaknesses.

For instance, while the data is really interesting to interact with, it's hard to get a full picture of the gender issues or how much of the population have attained each level of education easily, as you have to click through each county to find this information. I feel that, in that regard, some tables and plots or graphs could greatly benefit making the data understandable. What I do know is that this has helped answer questions for me as to what

I plan on doing going forward, which is to take a break from education, at least for a little while, and focus on work and my family. Lastly, I hope that you've found this project interesting and I encourage you to try it out for yourself [here](#)!

7.0 Further Development or Research

With more time and resources available, I would certainly like to expand on the data and how it's presented as I mentioned above. I think being able to present some graphs or plots to get across some of the harder to see data, like between men and women when it comes to education level attainment, would definitely benefit my tool. I also know that the user interface could be polished up to look much cleaner and better – design has never been my strong suit. I think it could also be interesting to take my concept and apply it to the next census to see how the data changes and see if there are any trends. Adding predictive analysis could also be quite interesting and show what could be expected to happen in each county.

8.0 References

Shiny.rstudio.com. 2021. *Shiny*. [online] Available at: <https://shiny.rstudio.com/>

Cran.r-project.org. 2021. *Create Dashboards with 'Shiny' [R package shinydashboard version 0.7.1]*. [online] Available at: <https://cran.r-project.org/web/packages/shinydashboard/index.html>

Cran.r-project.org. 2021. *Leaflet* [online] Available at: <https://cran.r-project.org/web/packages/leaflet/leaflet.pdf>

Cran.rstudio.com. 2020. *Rgeos* [online] Available at: <https://cran.rstudio.com/web/packages/rgeos/rgeos.pdf>

Dplyr.tidyverse.org. n.d. *A Grammar of Data Manipulation*. [online] Available at: <https://dplyr.tidyverse.org/>

Cran.r-project.org. 2021. *Maptools* [online] Available at: <https://cran.r-project.org/web/packages/maptools/maptools.pdf>

Cran.r-project.org. n.d. *Bindings for the 'Geospatial' Data Abstraction Library [R package rgdal version 1.5-23]*. [online] Available at: <https://cran.r-project.org/web/packages/rgdal/index.html>

Cran.r-project.org. 2021. *Classes and Methods for Spatial Data [R package sp version 1.4-5]*. [online] Available at: <https://cran.r-project.org/web/packages/sp/index.html>

R-spatial.github.io. n.d. *Simple Features for R* [online] Available at: <https://r-spatial.github.io/sf/>

9.0 Appendices

This section should contain information that is supplementary to the main body of the report.



National College of Ireland

Project Proposal

Financial Benefits of Further Education

08/11/2020

BSc (Honours) in Computing - Evening
Data Analytics

2020/2021

Jonathan Sammons

X17102561

X17102561@student.ncirl.ie

10.0 Objectives

The research that I'm performing is to analyse the data available for the level of education achieved and how it impacts the median earned income broken down by county. I plan to take this data and find the variance between each level of education achieved to see how much of an increase or a decrease can be expected as one obtains higher levels of education in each county in Ireland. I plan to further break this data down by bringing in the gender of those that have obtained each level to see what percentage of men and women are obtaining this higher level of education in each county.

By researching this information, we can answer a few questions:

- What is the financial incentive to return for a higher level of education?
- Is there a point where the return (financially) is very low near pointless to return for higher levels of education?
- At what rate are men versus women completing each level of education?

Answering the above questions will provide us with insight to other questions that we can ask, such as why a certain county may have such a high rate of PhD students, or why in a certain county there is a very low rate of men completing higher levels of education instead of women.

11.0 Background

As a current student myself pursuing this Honours Degree, the thought of if I should go for a Masters or PhD in my field is already on my mind. While I do enjoy learning these subjects, it is a very hard balance to strike with home life, work, and college all together. I often wonder what the potential financial benefits could be for continuing my education past this Honours degree once I have it and if the jump in salary is as high as I would be expecting.

Having a question that impacts myself directly is what has caused me to choose and research this project specifically. I initially wanted to make this even more granular, and find out about the potential change in salary for someone who pursues higher education directly after secondary school versus someone who attends college later in life – if it would be more beneficial to get on the job experience first or not? Unfortunately, the data that I had available at the time wouldn't allow me to dive into this question.

Since I couldn't research my first option, I decided to keep the general over-arching idea of my project, but take the age bracket out of it since that was the granularity that I was having trouble finding data for. Fortunately, I was able to locate data that told me the story that I was looking for – I just need to analyse it now.

While I do think that this is a good question to ask on its own, I felt that brining in some more information could be very beneficial. I decided that it would be very interesting to find out if there is a more or less even split between men and women attaining their higher levels of education in Ireland, or if there might be some interesting data to potentially tell us a story and maybe even raise concern if one gender is obtaining higher education at levels higher than would be expected than the other.

I believe that researching and providing this information can be very important for any potential students looking to either go to college for the first time or return to college in the future but are unsure if they're making the correct call. This information could also potentially give the Irish government needed data to investigate education gaps in the country and what may be causing the gaps, hopefully offering solutions in time to those that would be interested in seeking higher education but are hitting a barrier along the way.

I plan on making this data easily accessible and readable by plotting it on a map of Ireland with different toggles so that you can find the information that you're looking for, broken down by each county. Making the data as easy to understand as possible is key in spreading the story that it tells.

12.0 Technical Approach

For this project, I have obtained my data from the Central Statistics Office of the Republic of Ireland and this is the data that I will be analysing. I plan on using R Studio to clean, compute, and display my data by using the Shiny library. Shiny is a framework that will allow me to create a web application using the data and the tools that R Studio provides. By using the Shiny framework, it will allow me to make my data available on a website so that individuals can easily and comprehensively interact with it.

I'm also looking at using the Google Maps API to get a map of Ireland broken down by county for the display of the data. If I can get this same functionality from R Studio, then I may forgo leveraging the Google Maps API, however I need to further research this at this time. I plan on hosting this site using Amazon Web Services so that it's available to the public.

Lastly, I plan to use ArcGIS to assist in plotting the data points on the interactive map that is displayed.

13.0 Special Resources Required

A list of special resources that I will need for this project are:

ArcGIS

Google Maps API

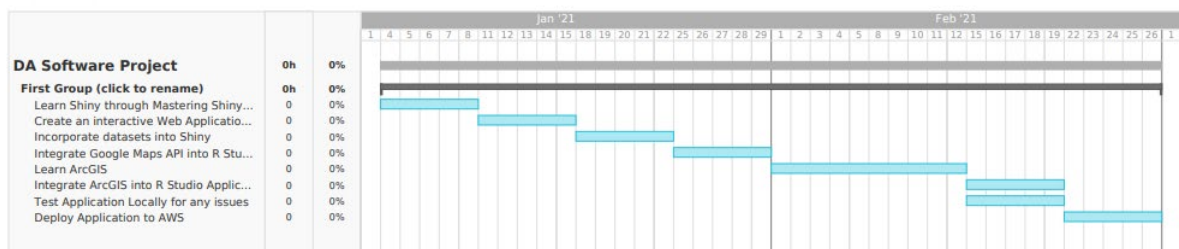
Mastering Shiny – A guide on how to utilize the Shiny framework

A website hosted by Amazon Web Services

14.0 Project Plan



Created with Free Edition



15.0 Technical Details

R Studio

ArcGIS

Google Maps API

Shiny Framework

16.0 Evaluation

I will be evaluating that the system is working correctly by testing it once it's complete to ensure that there are no errors with the data points, interactions on the map, and that the website is reachable.

16.1. Reflective Journals

Thoughts on Accomplishment

4th Year Reflective Journal

Jonathan Sammons x17102561

Introduction

This is a journal of my thoughts as I work on my final year of college. My name is Jonathan Sammons and I've just started my final year of my four-year degree with the National College of Ireland. I'm extremely proud of what I've accomplished, not only in the previous 3 years of college, but in the years since I've moved to Ireland at the end of 2014. So much has changed in my life, such as moving countries, getting married, buying a home, starting college, and having a little baby girl, and to think that none of it would have happened without the initial leap of faith that it took to uproot and move.

I think back sometimes on what my life was like in America, where I was and what I had done; and it still leaves me in disbelief at times, where I am now. I'm extremely lucky to have everything that I have in my life, I realize that – but I know that I've also worked hard to get where I am. I won't dwell on this section too much as it makes me uncomfortable to talk about myself like this, but I'm happy to be where I am and I'm excited to finish my degree since higher education isn't something that has been achieved much in my family.

October 10th, 2020

I have just begun my reflection journal, which, had it not been mandatory for this class, I know I would not have done since it's not normally something that I would do. Now that I'm typing though, I don't find it too bad. I just started back to classes last week so it's been a busy time between normal life, work, and classes. I feel bad that Joanna (my wife) has her hands full three days a week while I'm in class with Maebh (our daughter who is 13 months old). I hope she knows how much I appreciate everything she does and has done while I finish my education.

I'm excited to be back in college because I do find it enjoyable to learn, but I'm also a bit nervous about the projects in AI that seems quite challenging and the Software Project class, where I will be working on my final project all year long. Since I'm in the Data Analysis stream, it needs to be on Data Analysis which I didn't expect before hearing of it. I was a bit bummed out initially, but the more I thought about it the more interested I am in the project. I think it's quite interesting to gather data and analyse it to see what information I can get, but I'm a bit nervous about ensuring that my project is good enough and I hope that I'm not wandering off into the wrong direction.

I'm currently thinking about getting the data on people who are mature students (> 25 years of age) and finding out what their salary is before and after finishing higher education in Ireland. I would like to further break this down by county, and see what it might look like across each county of Ireland. It could be quite cool to get an interactive map through Google maps or something similar where this data can be displayed and you can click through the different data sets. This is my initial idea, I just need (and I hope!) that the data sets that I need are available and I can find them. I just need to remember to **avoid procrastination at all costs!**

November 30th, 2020

It's been a very busy month, but unfortunately not for my Data Analytics project. I had a test for my prob and stat class, a CA due for Data Application Development that was a bit of a doozy, a CA for Strategic Management that I knocked out, a CA for my Web API class that I finished and I finally made progress on the CA for the AI class where we have to make a chess game that plays against you. I've gotten the game working fully, so that you could play an entire game with no issues, however I still need to implement the AI side of the project. Because of all of this work, I haven't had a chance to make much progress on this project, unfortunately as I'm already spending as much of my workday as I can working on class work and evenings are either taken up with lectures or trying to help Joanna around the house and help taking care of Maebh as soon as work finishes. I'll be very happy to have this last year of college finished, just so I have time to breath once more.

All of that said, I did meet with my advisor and my project has been approved with a caveat that another data set should be introduced, so I decided to find data on genders by county with degrees that I can add to my analysis. I've also decided that I'm going to use the Shiny framework in R Studio to build my own webpage for the application that should add a nice complexity to the project. I only hope that next semester is quieter so that I have more time to focus on the project, as right now it feels like I'm managing deadlines instead of actually learning anything.

February 28th, 2021

Unfortunately, this is my first entry in a couple of months. At the end of December and into January, there was a lot of work that I was doing in regards to final projects and TABA's that I was working on, and not much happening on the front of my final years' software project. It wasn't until earlier this month that I really started to get my hands dirty with the project once more. The progress that I've been making on it has been exciting, but there's still plenty of work ahead of me. I have been learning the Shiny framework in R which is used to create interactive web applications for the R language. I started with the basics to understand the different syntax and flows that is used with the framework. I also tried looking into ArcGIS as I was planning on using it for the interaction, however it looks like that will no longer be needed – also it's quite an expensive software to get my hands on.

So far with Shiny, I've created a few interactive applications on my localhost testing out the various functionalities like picklists, sliders, and text entry and how you use the information entered to influence what the server returns to you. This will be important in selecting different levels of education or income in the final project when the user wants to find specific information.

I've also setup an account with Google for the Google API and I've used this to generate a map of Ireland. What I'm currently trying to do is take the map of Ireland and overlay the boundaries of the different counties, since that's what my project will focus on. After I get the boundaries of the different counties working, I'll just need to add the information for each county based on the selected input by the user and use this information to draw conclusions from the data. Once I have this working, I can start looking into getting the application hosted online so that it's available to others to use.

Fortunately, I believe that Shiny actually has an option to make an application live through their platform, so I may not need to use AWS to host the application, which will save a lot of a headache.

March 23rd, 2021

I've been working as much as I can on the project, but it's been tough with other projects in other classes. I have made some good progress though. I had to ditch the Google Maps API but found that I could use Leaflet to generate maps and I've got that working now. Getting the shape file to apply was a battle but I eventually got that working as well. I'm currently trying to get the map to update when I click on it with the name of the county but so far I'm having no luck. I think I also need to get some data for the population of each county so I need to check the CSO website to see if I can find a file that contains this information for each county like I'm currently doing. I really need to keep working on this project but between work, school, and family life, it's very tough to find enough time to dedicate to everything.

April 19th, 2021

I've managed to work on this project on and off over the last few weeks and made some great progress since my last post. I have the ability to click on a county and get the county data to load, showing the name of the county, median salary, number of males and females with the degree, and the change in salary for the next level of education. The map also has a nice colour coding to it showing how a higher salary is a darker blue than a lesser salary, being a lighter blue. I ran into a couple of really frustrating problems that I managed to get fixed. The first being that the coordinates provided in the shape file wouldn't work and I had to convert them to another type of longitude and latitude that RStudio can actually understand. The second major problem that I was running into was merging my data on the levels of education with my geometry data. Because the geometry data is in a spatial dataframe, I couldn't merge the two like I normally do, but I was able to find a way to merge them eventually and I have everything working now. It's a sigh of relief at this point and I feel like it's great progress. I think I'd just like to add in some sort of graph or chart to show the difference between the number of men and women in each county with the degree for an additional layer of visual data since it can be hard to tell the difference by just looking at the numbers.

16.2. [Link to the interactive map](https://salary-scales-by-county-in-the-republic-of-ireland.shinyapps.io/RepublicOfIrelandSalaryScales/)
<https://salary-scales-by-county-in-the-republic-of-ireland.shinyapps.io/RepublicOfIrelandSalaryScales/>