Dublin Institute of Technology
Institiúid Teicneolaíochta Átha Cliath

An Exploration of Levels of Education as Predictors of Voter Turnout: A Case study of Dublin City Electoral Results

DT265 – Data Analytics

Higher Diploma in Computing

Robert Porter (D13124708)

Project Supervisor – Damian Gordon

School of Computing

Dublin Institute of Technology

23rd June 2014

School of Computing
Scoil na Ríomhaireachta

Abstract

This project sought to explore in detail the relationship between levels of education with regard voter turnout within Dublin City Electoral Districts. Research methods and data analysis skills that were learned through the completion of DT265 Higher Diploma in Data Analytics were applied to the project. Analysis of CSO data along with information supplied by the Dublin Returning Officer enabled the construction of various predictive models that enable manipulation of variable statistics in order to reproduce election results and counterfactually modify values of educational attainment in order to examine hypothetical scenarios. Possible future research is also briefly touched upon by broadening the scope of the project to the national level along with other possible data analysis techniques that could be potentially applied.

Acknowledgements

I would very much like to thank my supervisor Damian Gordon for the advice, enthusiasm and helpful suggestions that helped shepherd this project to its conclusion. I would also like to thank my fellow class mates for the feedback and which always proved a good sounding board for what was practically achievable with the project.

I would also like to express sincere gratitude to Sean Donnelly from *Elections Ireland*, who was very helpful in clarifying the finer detail regarding the disparate nature of the electoral system at the local level and for allowing me to corroborate election data that I had compiled with his own.

Lastly, I would like to thank my family and my partner Anna for their support and understanding over the past number of months of examinations and research.

Table of Contents

Glossary of Terms 5

1. Introduction 6

1.1 Project Background 6

1.2 Project Description 6

1.3 Research Methods 7

1.4 Project Aims 8

1.5 Project Scope 8

1.6 Dissertation Roadmap 8

2. Literature Review Introduction 9

2.1 Literature Review 9

Conclusion 10

3. Design 11

3.1 Introduction 11

Suitability 11

3.2 Dataset 12

Pre-processing 13

3.3 R 15

3.4 Introduction 15

3.5 Importing Data and Preliminary investigation 15

4. Experiment 17

4.1 Introduction 17

Correlations – 2009 Local Election Data 17

Correlations 2011 General Election Data 19

Multi-Linear Regression Model – Turnout Prediction 21

Correlations between Populations and Electorates 22

5. Results and Evaluation - Introduction 24

5.1 Results and Evaluation 24

6. Conclusions and Future Work 29

6.1 Introduction 29

6.2 Conclusions 29

6.3 Future Work 30

Appendix 31

Bibliography 33

Glossary of Terms

Constituency: Dáil Éireann is constituted by 166 TDs, representing 43 parliamentary multi-seat constituencies. Each constituency must have at least 3 members. This district magnitude (No. Seats) varies on the basis of boundary and population with a given constituency. The most recent Census from 2011 indicates a ratio of one TD for every 27,640 persons in the country.

Electoral Divisions: These are the smallest administrative areas for which population statistics are published. There are 3,440 electoral divisions in the State. Electoral divisions are referred to by their established statutory names. In some cases, these names differ from addresses and place names currently used. Electoral Divisions (EDs) are also the smallest legally defined administrative areas in the State for which Small Area Population Statistics (SAPS) are published from the Census. There are 32 EDs with low population, which for reasons of confidentiality have been amalgamated into neighbouring EDs giving a total of 3,409 EDs

First Order Elections: Are considered to be the most significant national elections. These are defined as parliamentary elections in the case of parliamentary systems and presidential, in presidential systems such as the United States.

Local Electoral Areas: Under local government legislation, the Minister for the Environment, Community and Local Government is responsible for dividing each county and city into electoral areas (also referred to as local electoral areas) for the purposes of local elections. Generally, a number of electoral divisions are grouped to form an electoral area.

Polling District: Polling districts are geographical areas created by the division of Electoral Districts into smaller parts.

Polling Station: A Polling Station is a building or area in which polling stations are selected by the Returning Officer.

Returning Officer: Under the Electoral Act 1992, the duties of a Returning Officer are defined as requiring all things necessary for effectually conducting a Dáil election in his or her constituency, in accordance with the Act, to ascertain and declare the results of the election and to furnish to the Clerk of the Dáil a return of the persons elected for the constituency. For local elections the position is held by the administrative head of the local council. The returning officer for presidential elections and referendums is a senior official in the franchise section of the Department of the Environment.

Second Order Elections: Are generally characterised as having lower levels of Turnout in national elections. Here they represent both Local and European elections, which are held on the same day.

Voter Turnout: Voter turnout is the number of eligible voters who cast a ballot in an election. Turnout percentage is calculated by dividing this number by the total registered voters.

1. Introduction

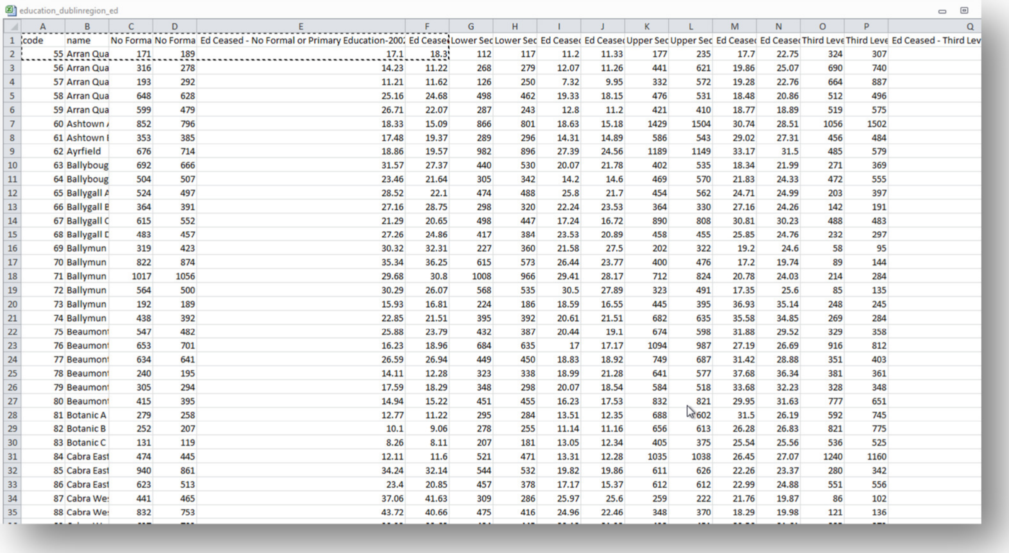
1.1 Project Background

Voter turnout is essential for the efficient operation of any democracy. Evidence has shown that there has been a decline in voter turnout in many western countries in the past number of decades. Low turnout is considered undesirable by political scientists and there is much debate regarding the factors that affect it. To date, there remain various arguments that appear to remain inconclusive for the causes of low turnout.

Robert Putnam's research points to the decline of civic engagement and participation within the political systems in his book *Bowling Alone* as one of the potential reasons for the decline in turnout. Other notable researchers such as Arend Lijphart (1997) have commentated that *"unequal participation spells unequal influence"*. With this sentiment in mind the basis of the research is to highlight potential areas where deficiencies can be addressed within both the educational and democratic systems.

1.2 Project Description

Educational data-sets of educational attainment based on Electoral Divisions in Dublin were compiled availing of open-data sources on dublinked.ie and the Central Statistics Office. The data offers a breakdown of various levels of educational attainment. Datasets for 2002 and 2006 are available for all Electoral Divisions for all of the Dublin area while census data for the entire country is available for the 2011 census. The data consists of multiple Excel spreadsheets with extensive entries seen below. No data omissions have been observed. Datasets can be found at the following URL: http://dublinked.com/datastore/datasets/dataset-247.php



education_dublinregion_ed

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	code	name	No Forma	No Forma	Ed Ceased - No Formal or Primary Education-200	Ed Cease	Lower Sec	Lower Sec	Ed Ceasec	Ed Ceasec	Upper Sec	Upper Sec	Ed Ceasec	Ed Ceasec	Third Leve	Third Leve	Ed Ceased - Third Lev
2	55	Arran Qua	171	189	17.1	18.3	112	117	11.2	11.33	177	235	17.7	22.75	324	307	
3	56	Arran Qua	316	278	14.23	11.22	268	279	12.07	11.26	441	621	19.86	25.07	690	740	
4	57	Arran Qua	193	292	11.21	11.62	126	250	7.32	9.95	332	572	19.28	22.76	664	887	
5	58	Arran Qua	648	628	25.16	24.68	498	462	19.33	18.15	476	531	18.48	20.86	512	496	
6	59	Arran Qua	599	479	26.71	22.07	287	243	12.8	11.2	421	410	18.77	18.89	519	575	
7	60	Ashtown /	852	796	18.33	15.09	866	801	18.63	15.18	1429	1504	30.74	28.51	1056	1502	
8	61	Ashtown I	353	385	17.48	19.37	289	296	14.31	14.89	586	543	29.02	27.31	456	484	
9	62	Ayrfield	676	714	18.86	19.57	982	896	27.39	24.56	1189	1149	33.17	31.5	485	579	
10	63	Ballyboug	692	666	31.57	27.37	440	530	20.07	21.78	402	535	18.34	21.99	271	369	
11	64	Ballyboug	504	507	23.46	21.64	305	342	14.2	14.6	469	570	21.83	24.33	472	555	
12	65	Ballygall A	524	497	28.52	22.1	474	488	25.8	21.7	454	562	24.71	24.99	203	397	
13	66	Ballygall B	364	391	27.16	28.75	298	320	22.24	23.53	364	330	27.16	24.26	142	191	
14	67	Ballygall C	615	552	21.29	20.65	498	447	17.24	16.72	890	808	30.81	30.23	488	483	
15	68	Ballygall D	483	457	27.26	24.86	417	384	23.53	20.89	458	455	25.85	24.76	232	297	
16	69	Ballymun	319	423	30.32	32.31	227	360	21.58	27.5	202	322	19.2	24.6	58	95	
17	70	Ballymun	822	874	35.34	36.25	615	573	26.44	23.77	400	476	17.2	19.74	89	144	
18	71	Ballymun	1017	1056	29.68	30.8	1008	966	29.41	28.17	712	824	20.78	24.03	214	284	
19	72	Ballymun	564	500	30.29	26.07	568	535	30.5	27.89	323	491	17.35	25.6	85	135	
20	73	Ballymun	192	189	15.93	16.81	224	186	18.59	16.55	445	395	36.93	35.14	248	245	
21	74	Ballymun	438	392	22.85	21.51	395	392	20.61	21.51	682	635	35.58	34.85	269	284	
22	75	Beaumont	547	482	25.88	23.79	432	387	20.44	19.1	674	598	31.88	29.52	329	358	
23	76	Beaumont	653	701	16.23	18.96	684	635	17	17.17	1094	987	27.19	26.69	916	812	
24	77	Beaumont	634	641	26.59	26.94	449	450	18.83	18.92	749	687	31.42	28.88	351	403	
25	78	Beaumont	240	195	14.11	12.28	323	338	18.99	21.28	641	577	37.68	36.34	381	361	
26	79	Beaumont	305	294	17.59	18.29	348	298	20.07	18.54	584	518	33.68	32.23	328	348	
27	80	Beaumont	415	395	14.94	15.22	451	455	16.23	17.53	832	821	29.95	31.63	777	651	
28	81	Botanic A	279	258	12.77	11.22	295	284	13.51	12.35	688	602	31.5	26.19	592	745	
29	82	Botanic B	252	207	10.1	9.06	278	255	11.14	11.16	656	613	26.28	26.83	821	775	
30	83	Botanic C	131	119	8.26	8.11	207	181	13.05	12.34	405	375	25.54	25.56	536	525	
31	84	Cabra East	474	445	12.11	11.6	521	471	13.31	12.28	1035	1038	26.45	27.07	1240	1160	
32	85	Cabra East	940	861	34.24	32.14	544	532	19.82	19.86	611	626	22.26	23.37	280	342	
33	86	Cabra East	623	513	23.4	20.85	457	378	17.17	15.37	612	612	22.99	24.88	551	556	
34	87	Cabra We:	441	465	37.06	41.63	309	286	25.97	25.6	259	222	21.76	19.87	86	102	
35	88	Cabra We:	832	753	43.72	40.66	475	416	24.96	22.46	348	370	18.29	19.98	121	136	

These Educational statistics are cross-referenced with electoral turnout data from the Dublin City Returning Officer. The electoral data is sourced from the Elections Ireland website and dublincountyreturningofficer.com. Data from the Returning Officers website is presented on various webpages for the various Local Electoral Areas. This data was scrapped and compiled into CSV format using Excel. A subset of the data that can be obtained is from the following url:http://www.dublincityreturningofficer.com/2011_general_election/poll_scheme_voter_turnout/dublin-north-west.php

Preliminary investigation using R is used to ascertain whether:

- There is any correlation in both first order and second order elections
- Correlations or relationships exist between the electorate and/or turnout
- There is information that can be revealed regarding the electoral register and recorded populations within Dublin City electoral districts.

In addition to examining the academic literature regarding the impact of education and voter turnout, various multi-linear regression models approximating to Ordinary Least Squares using R are constructed in order to estimate the composite make up of Dublin electoral districts in terms of education. The objective of these models is to attempt to predict voter turnout.

1.3 Research Methods

The exploratory research approach I undertook required researching how the census is conducted in Ireland, how it is devised and how the scope of the data available can be of use to researchers. Further research into the electoral system required attempting to locate data from various agencies such as numerous Dublin councils and research the law pertaining role and procedures of agencies and staff involved in the electoral process.

Other research necessitated determining which methods in relation to data pre-processing were to be undertaken in order to achieve a dataset that would best enable exploratory research into the topic. This research required learning methods of conditional formatting in excel and the attendant functions in order to test and validate the amalgamation of the data into a superset.

Research was also undertaken into the established literature concerning elections and education, both domestic and international. This was undertaken to frame the examination of the data in context with established findings.

Lastly, data analysis research was undertaken in order to determine which methods were best suited in order to assess the data and evaluate it. Coding through R, an approach was taken to research employing linear regression models as a means to investigate and examine the data.

1.4 Project Aims

The aim of the project is to examine electoral and educational attainment data to ascertain whether there is any correlation with low voter turnout and where possible spoiled ballot papers within Dublin electoral divisions and constituencies in the Dublin. The objective is to confirm or dismiss the role of educational attainment within the democratic process in terms of voter turnout in Dublin City. Information that maybe gained may point to disadvantaged areas which require additional educational resources to address potential deficits in terms of civic engagement with the electoral process.

1.5 Project Scope

The scope of the project is to construct multiple predictive models in order to determine whether there are correlations between levels of education and voter turnout. An investigation of possible correlation in both first order and second order elections i.e. local elections and general elections, will be undertaken. Researching the academic literature regarding the impact of education and voter turnout will help regarding framing findings and conclusions in their appropriate context. Lastly, I propose to attempt classification of Dublin EDs in terms of geographical location, both North and South.

1.6 Dissertation Roadmap

The road map for this dissertation begins with a limited literature review of the political science literature regarding education and voter turnout. Assumptions and conclusions will accompany this review along with a brief historical snapshot of some of the existing research into the topic domestically.

At the next juncture, information is supplied regarding the contours and composition of the data that was selected for the project and how this was manipulated and tested in order to arrive at a workable solution in order to gain insights from the data.

The dissertation then briefly describes how RStudio was employed to gain the initial insights into the candidate variables for prediction and then move on to experimentation. Experimentation and the results therein are further explained in their statistical context.

Lastly, results and observations regarding the objective of the project are discussed and conclusions offered. Further research relating to the topic is also considered.

2. Literature Review Introduction

Across the academic literature regarding voter turnout and education it is generally accepted that effective citizen participation affected by the educational system. Authors have noted that "education is the key ingredient of any relationship between socioeconomic status and voting turnout" (Lewis-Beck et al. 2008, pp: 102). In terms of Ireland, much work has examined the combined socio-economic, demographic and educational factors at play, however many of these observations are informed by other extraneous variables such as rural urban divides and types of voter abstention

2.1 Literature Review

University of Wisconsin political scientist Barry Burden (2009) in his investigation of American electoral turnout states that apart "from physical characteristics such as race, sex, and age, formal educational attainment is one of the most clearly exogenous factors that predict electoral behaviour". This position has been extensively researched and many studies have reinforced the strong relationship between educational attainment and political participation (Miller 1992; Miller and Shanks 1996; Rosenstone and Hansen 1993; Verba et al. 1995). Though the premise has been underpinned that educational attainment bares an influence on political participation, this statement is framed in the context of a dilemma as although education is supposed to predict whether individuals will vote, over time rising levels of education in the US did not increase aggregate turnout (Brody, 1978). Burden asserts that the effect of education in terms of voter turnout may not have been constant and that this explains why voter turnout has not risen on par with levels of education.

In terms of Irish research, it has been observed in previous studies that education is a not leading determinate of voter turnout in first order elections. Despite the counterfactual intuition that education provides knowledge to process the various political issues at play along with the functional knowhow in terms of registering to vote and making sense of ballot papers and electoral systems, no clear indicator has been observed whereby educational attainment is seen to significantly boost turnout at the national level. Research conducted by Lyons and Sinnott (2002) using ASES (Asia Europe Survey) and Eurobarometer survey data reported that the level of turnout in the 1997 general election among those with primary education or less, was according to the ASES Survey 89 percent, while the sample as a whole was less than this figure at 85 percent. The authors report however that lower educational attainment seemed to have a bearing moreover on European Elections whereby those with primary or lower levels of education manifested a lower turnout than the sample. Speculating on the reason why, it was determined that the cues and signals that provide incentives to vote in national elections are to a larger extent absent at the European election level.

Lyons and Sinnott's (2003) study showed that where education does seem to have had a perceivable impact upon turnout, that it has been in association with another factor; namely, the rural urban divide. Lyons and Sinnott describe this where the relationship between turnout and education pertaining to the 1981 general election differed strongly. In rural areas there was little or no relationship between the percentage with secondary education and turnout, while in urban areas the relationship was positive.

A dominant feature of Irish electoral behaviour is that an estimated 37 percent of registered voters do not in fact vote. Though this number may seem sizable it should be noted that this is the percentage based upon the actual Electoral Register and not the eligible population of voting age, meaning that the actual percentage of non-voting eligible voters may be greater. Another strong feature of the electoral register is that there is a significant difference in turnout between those for whom residency and registration coincide and those for whom it does not. Marsh *et al.*, put the figure at 29 percent. Further to their analysis, the high percentage of non-voters is split between those abstaining on circumstantial grounds and those who are disaffected by the political system. There are four factors that affect turnout. The first two distinguish between voter *facilitation* and voter *mobilisation*. Facilitation is essentially the provision to make it easier to vote, whereas mobilization refers to anything that makes a voter want to vote. Increasing facilitation is said to address the primary cause for voter abstention in terms of circumstantial inhibition, while the impetus in terms of mobilisation of the electorate is said to lower abstention in terms of disaffection. The latter two factors are obtained by Cross-classifying facilitation and mobilisation in terms of a further distinction between institutional (political and civil society organisations) and individual level organisation (Marsh *et al.,* 2008). Through the lens of this model it has been observed that there appears to be a higher level of voluntary abstention among young people, suggesting that issues of mobilisation are more at play.

Conclusion

The effect of education on voter turnout has been broadly described as a composite mix of the ability to overcome a number of inhibitory thresholds. In terms of getting to grips and understanding the system, education has provides people with skills in order to understand issues at play in current affairs, thus requiring " more abstract thought than does everyday activity" (Delli Carpini and Keeter 1996). With regard to its simplest effect, it allows for the skills necessary to overcome voter registration impediments to voting (Highton 2004; Powell 1986; Timpone 1998; Verba *et al.* 1995; Wolfinger and Rosenstone 1980). Lastly educational institutions can be seen to allow the formation of social networks where the greater levels of education inculcate a sense of civic duty and expose them to elite recruitment efforts (Campbell *et al.*, 1960; Rosenstone and Hansen 1993; Wolfinger and Rosenstone 1980).

In the case of Ireland the institutional evidence suggested by Sinnott *et al.* (Sinnott, 2008), indicates that Ireland is characterized by lower levels of institutional mobilisation at the sub–national level of governance i.e., local and European elections. While the effect of low levels of educational attainment have been considered to affect European election turnout negatively, as the mobilisation effects are not as strong in terms of stimulus, that European elections coincided with local elections, any assumption that the same lack of stimulus applies to local politics must surely not hold given that local issues must resonate as a corollary. It maybe speculated that the disincentive for voter engagement at local elections may be partly due to the fact that local government in Ireland is generally considered to be week by international comparisons due to "very limited budgetary powers" (Sinnott, 2008).

Based on the literature reviewed in this chapter, the following chapter will explain the key design choices in this research, including, most importantly, the design of the datasets.

3. Design

3.1 Introduction

In this chapter the design of the dataset and the design of the experiments will be discussed. The design of the dataset will include a discussion of the suitability of the dataset, and a detailed description of the dataset including how data was merged, checked, summed, aggregated and compiled.

Suitability

The suitability of combining the census data with the electoral data is ideal in that the census data is compiled on exactly the same geographical unit as the electoral unit. By aggregating the EDs it is possible to scale up the geographical magnitude thus allowing for analysis of societal characteristics vis-à-vis electoral outcomes.

It should be noted however that there is one central caveat when attempting to correlate census data with the specific electoral data for a given geographical unit. The census data reports data in relation to where people are residing on a particular night, while the electoral data is concerned with the Electoral Register. There is thus a potential degree of mismatch in terms of persons who are residing in a particular Electoral District and registered in another.

Another factor that should be born in mind is that the higher the percentage of the electorate divided by the population, it may be assumed that the more people over 18 are included i.e., for those districts and areas where the percentage is high this may infer an aging population. As can be seen in the table below, the electorate as a percentage of the population has gone from 71% in 1985 to 81% in 2004. It dropped back to 78% in 2009 and is down to 73% in 2014 with only 66% in Dublin (indicating a potentially younger population than the rest of the country).

Local Election Year	No. LEAs	Seats	Population	Electorate	E/P
2009	171	883	4,239,848	3,297,426	77.77%
2004	180	883	3,917,203	3,166,033	80.82%
1999	180	883	3,626,087	2,872,305	79.21%
1991	177	883	3,525,719	2,536,967	71.96%
1985	177	883	3,443,405	2,444,247	70.98%

Given that the last census was three years ago, it may be assumed that population has probably continued to trend upwards during that time.

3.2 Dataset

The dataset is primarily numeric in nature consisting of 161 tuples and 113 variables. A derivative variable was created to distinguish North and South Dublin city, also giving the data a categorical element. Of the numeric data most variables are integer, while candidate variables such as voter turnout percentage are continuous.

There are several concepts and types of entities that require understanding in order to accurately describe the hierarchy of geographical areas that relate to defined units within the data. Common to most peoples understanding are geographical areas namely as constituencies. Within the Republic of Ireland there are currently 43 Parliamentary constituencies. These constituencies are defined by the Electoral Boundary Commission and have a constitutional prerequisite of having not more than a population of 30,000 per TD. Currently there is one sitting TD for every 27,640 people on the basis of 2011 census data. At the local level, matters are confused somewhat in that counties and cities are divided into Local Electoral Areas (LEAs). Dublin City Council is currently constituted by 11 LEAs:

Each LEA is weighted by the number of councillors that are returned to the council. These LEAs have different characteristics such as the population, number of registered electors, and percentage of turnout. There were a total of 52 seats available to Dublin City Council in 2009. Local Electoral Areas are comprised of further sub-units called Electoral Divisions (EDs). These electoral divisions can vary in number when comprising an ELA, this can be seen in the table below. EDs differ from each other and are characterized by varying populations. They number a total of 162 individual instances that makeup the 11 LEAs, these constitute Dublin City Council.

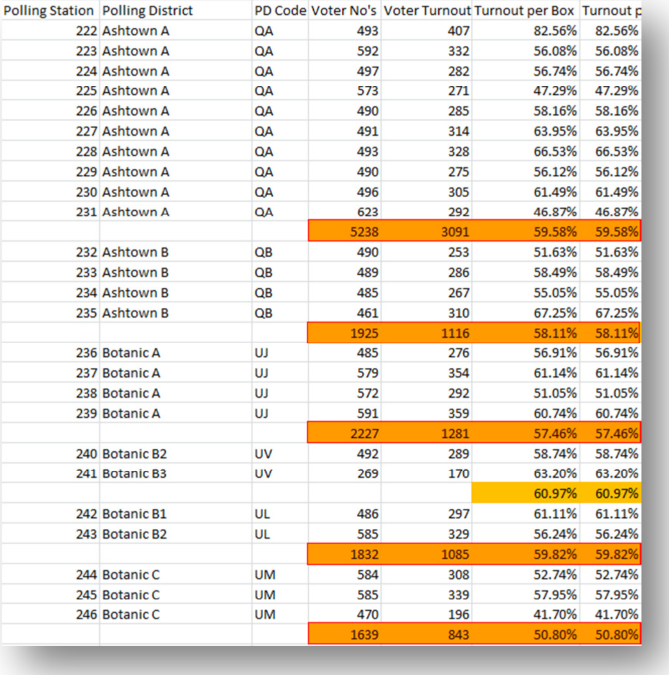
Dublin City Council			Population		
Constituency	**Local Electoral Area**	**EDs**	**2006**	**2011**	**% Change**
Dublin NC	Artane-Whitehall	16	47,095	46,773	-0.68%
Dublin SC	Ballyfermot-Drimnagh	13	37,398	38,602	+3.22%
Dublin NW	Ballymun-Finglas	17	50,957	52,053	+2.15%
Dublin Central	Cabra-Glasnevin	13	44,618	46,483	+4.18%
Dublin NC	Clontarf	15	48,934	48,857	-0.16%
Dublin SC	Crumlin-Kimmage	15	39,246	39,335	+0.23%
Dublin NE	Donaghmede	12	41,793	44,951	+7.56%
Dublin Central	North Inner City	19	60,056	67,309	+12.08%
Dublin SE	Pembroke-Rathmines	16	60,277	60,622	+0.57%
Dublin SE	SE Inner City	11	40,028	43,211	+7.95%
Dublin SC	SW Inner City	15	35,809	39,416	+10.07%

Totals	**162**	294,529	307,495	+4.40%

As can be seen from the above table, population figures are available for each Local Electoral Area. This is derived by cross referencing the population and summing for each ED in the 2011 census.

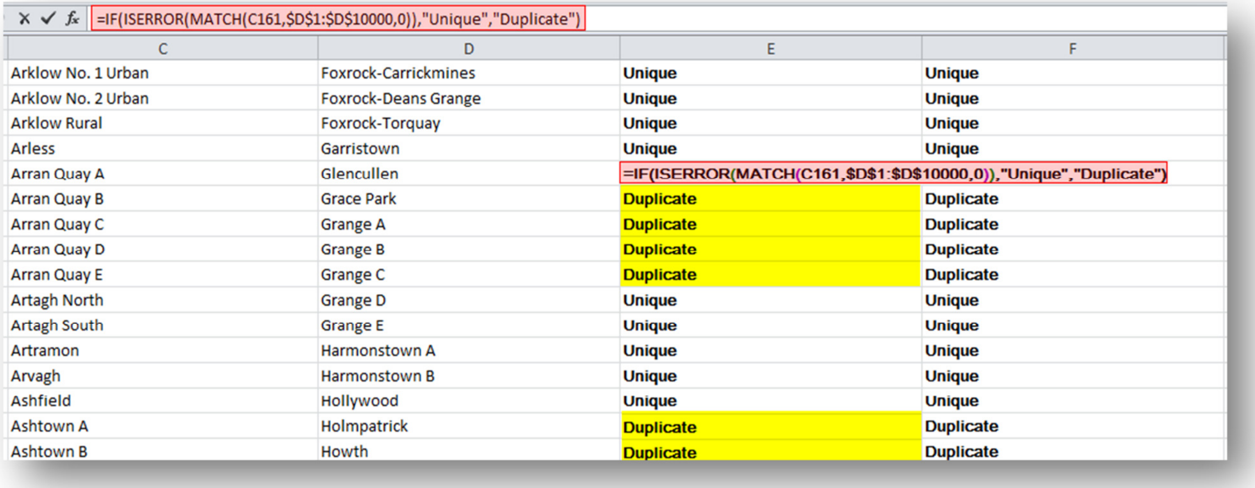
Pre-processing

Pre-processing the data required several stages to complete. In terms of the electoral data, the information supplied by the Dublin City Returning Officer required summing the voter numbers and turnout per box in each Polling District. The Polling Districts were then totalled to arrive at the correct figure for each Electoral District. In many cases similarly named Electoral Districts can be found in different LEAs across the county, thus this can be quite challenging in order to keep track of certain districts. An example of where this occurred can be cited whereby the *Crumlin* EDs A, B and E are located in the LEA of *Ballyfermot and Drimnagh*, while *Crumlin* C and D are located in the *Crumlin and Kimmage* LEA. Restoring all of the correct Polling Districts electoral data into EDs was carried out using Excel. This process was achieved by creating a separate sheet and colour coding each distinctive LEA and could then be sorted using conditional formatting functions in Excel.



Polling Station	Polling District	PD Code	Voter No's	Voter Turnout	Turnout per Box	Turnout p
222	Ashtown A	QA	493	407	82.56%	82.56%
223	Ashtown A	QA	592	332	56.08%	56.08%
224	Ashtown A	QA	497	282	56.74%	56.74%
225	Ashtown A	QA	573	271	47.29%	47.29%
226	Ashtown A	QA	490	285	58.16%	58.16%
227	Ashtown A	QA	491	314	63.95%	63.95%
228	Ashtown A	QA	493	328	66.53%	66.53%
229	Ashtown A	QA	490	275	56.12%	56.12%
230	Ashtown A	QA	496	305	61.49%	61.49%
231	Ashtown A	QA	623	292	46.87%	46.87%
			5238	3091	59.58%	59.58%
232	Ashtown B	QB	490	253	51.63%	51.63%
233	Ashtown B	QB	489	286	58.49%	58.49%
234	Ashtown B	QB	485	267	55.05%	55.05%
235	Ashtown B	QB	461	310	67.25%	67.25%
			1925	1116	58.11%	58.11%
236	Botanic A	UJ	485	276	56.91%	56.91%
237	Botanic A	UJ	579	354	61.14%	61.14%
238	Botanic A	UJ	572	292	51.05%	51.05%
239	Botanic A	UJ	591	359	60.74%	60.74%
			2227	1281	57.46%	57.46%
240	Botanic B2	UV	492	289	58.74%	58.74%
241	Botanic B3	UV	269	170	63.20%	63.20%
					60.97%	60.97%
242	Botanic B1	UL	486	297	61.11%	61.11%
243	Botanic B2	UL	585	329	56.24%	56.24%
			1832	1085	59.82%	59.82%
244	Botanic C	UM	584	308	52.74%	52.74%
245	Botanic C	UM	585	339	57.95%	57.95%
246	Botanic C	UM	470	196	41.70%	41.70%
			1639	843	50.80%	50.80%

The next stage of pre-processing required matching all of the 162 Electoral Districts relating to Dublin City Council election results to their corresponding instances in the National Census, 2011. All of the Dublin EDs for all local authorities were identified, sorted and extracted from the Census data. The National Census contained information relating to 332 County Dublin EDs. The Dublin City ED data that was reconstituted from Polling Districts was inserted into a column beside the total number of all Dublin local authority EDs. A matching function was the used in Excel to indicate whether an instance one column was unique or a duplicate in the other. All of the duplicate tuples were then kept via sorting and the unique instances discarded, thus providing a complete set of electoral results for each Dublin City ED and its complimentary data pertaining to the 2011 census.



=IF(ISERROR(MATCH(C161,$D$1:$D$10000,0)),"Unique","Duplicate")

C	D	E	F
Arklow No. 1 Urban	Foxrock-Carrickmines	Unique	Unique
Arklow No. 2 Urban	Foxrock-Deans Grange	Unique	Unique
Arklow Rural	Foxrock-Torquay	Unique	Unique
Arless	Garristown	Unique	Unique
Arran Quay A	Glencullen	=IF(ISERROR(MATCH(C161,$D$1:$D$10000,0)),"Unique","Duplicate")	
Arran Quay B	Grace Park	Duplicate	Duplicate
Arran Quay C	Grange A	Duplicate	Duplicate
Arran Quay D	Grange B	Duplicate	Duplicate
Arran Quay E	Grange C	Duplicate	Duplicate
Artagh North	Grange D	Unique	Unique
Artagh South	Grange E	Unique	Unique
Artramon	Harmonstown A	Unique	Unique
Arvagh	Harmonstown B	Unique	Unique
Ashfield	Hollywood	Unique	Unique
Ashtown A	Holmpatrick	Duplicate	Duplicate
Ashtown B	Howth	Duplicate	Duplicate

Lastly, data regarding General Election 2011 percentage turnout was obtained from ElectionsIreland.org and was added to the final data-set. One elimination was necessitated regarding the Phoenix Park ED as this was split into two parts in separate ELAs. The data relating to the Phoenix Park was also characterised with unusual data given that it is not a residential area of the city.

The final dataset consisted of 161 tuples with 113 columnar variables. The first four variables represented electoral data while the remaining variables represented educational data obtained from the 2011 census. The educational data consisted of the follow headings:

Population aged 15 years and over by age education ceased
Population aged 15 years and over whose education has not ceased
Population aged 15 years and over by field of study
Population aged 15 years and over by sex and highest level of education completed

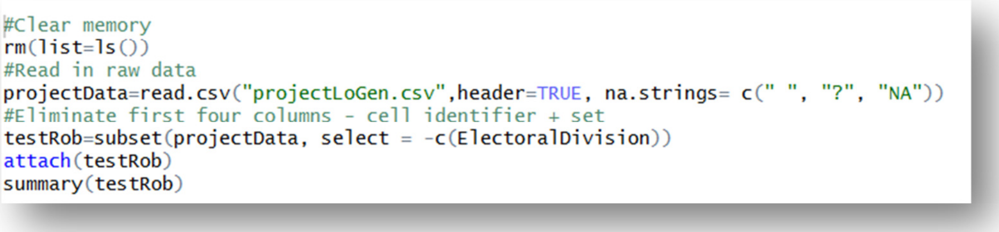
3.3R

3.4 Introduction

RStudio was used extensively throughout all facets of the assignment. R is an open source statistical integrated development environment for statistical computing. All graphs and graphics were output through the various in-built functions and packages available.

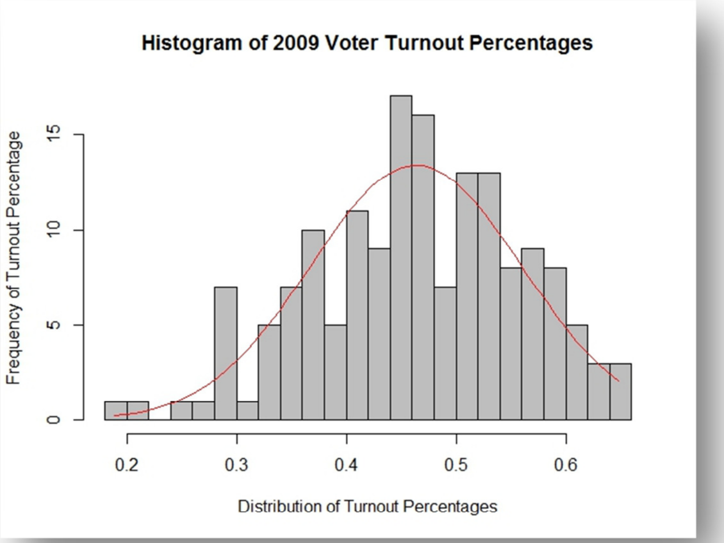
3.5 Importing Data and Preliminary investigation

Preliminary investigation of the compiled electoral data was first and foremost carried out by examining the distributions of potential candidate variables. This was achieved by importing the entire compiled dataset of EDs into R.

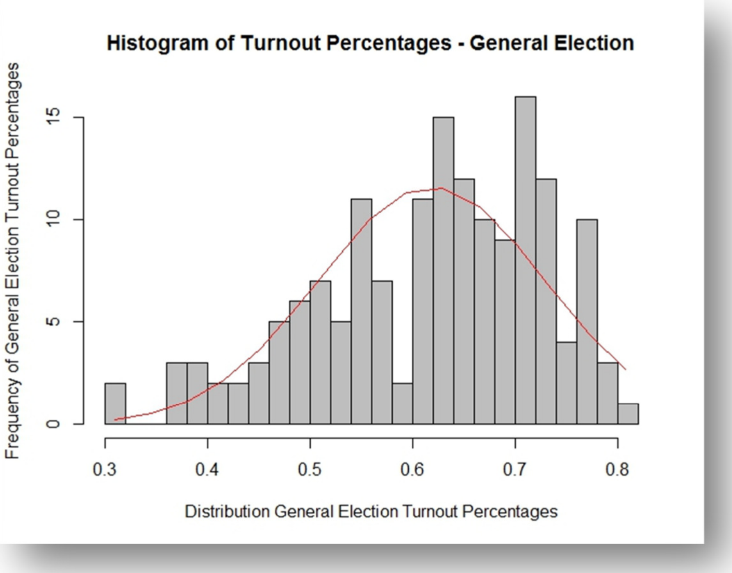


```
#Clear memory
rm(list=ls())
#Read in raw data
projectData=read.csv("projectLoGen.csv",header=TRUE, na.strings= c(" ", "?", "NA"))
#Eliminate first four columns - cell identifier + set
testRob=subset(projectData, select = -c(ElectoralDivision))
attach(testRob)
summary(testRob)
```

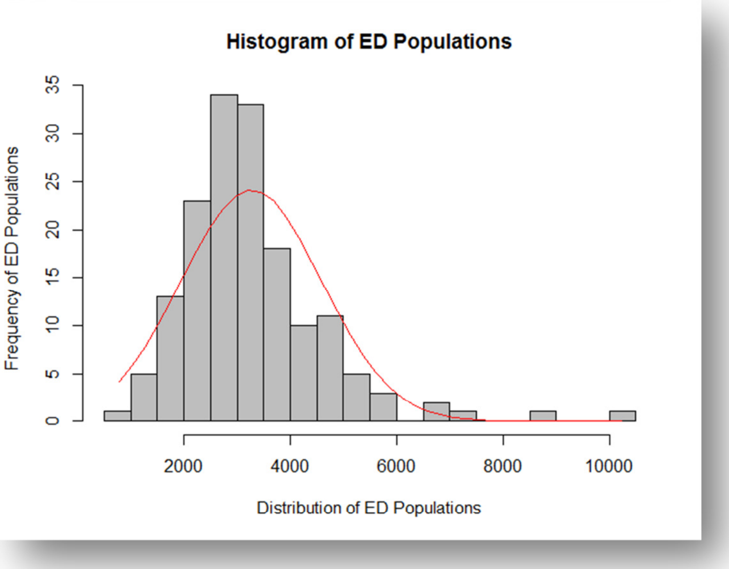
Upon importing the data, the first variable examined was the 2009 Local and European election *percentage turnout variable*. This revealed a relatively normal distribution with a slight left skew and similar mean and median of 46 percent. The minimum turnout was evidenced in North City ED at just fewer than 19 percent which is part of the North Inner City LEA. The highest percentage turnout was evidenced in Beaumont E electoral district, located within Clontarff LEA where turnout was recorded at just below 65 percent.



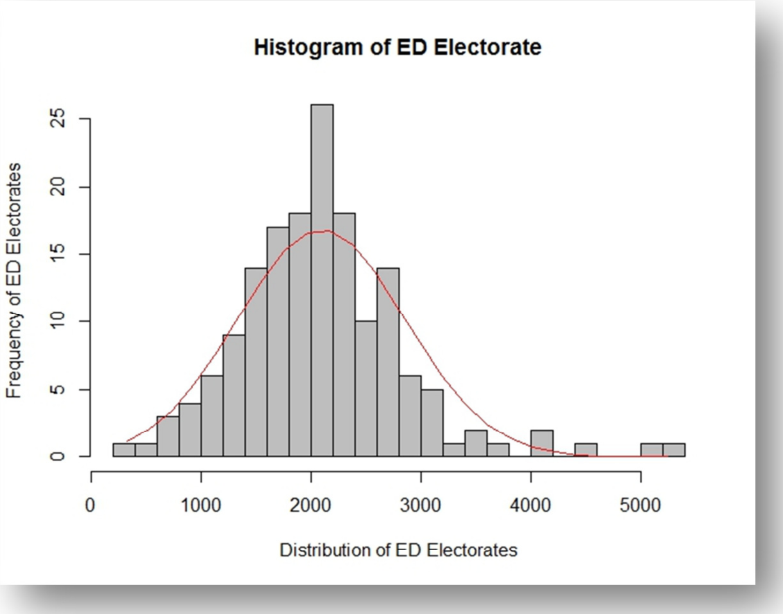
The second variable examined was the 2011 general election *percentage turnout variable*. The lowest observation recorded was again in the in the North City, this electoral district had a turnout of 31 percent. The largest percentage turnout was Raheny-St. Assam with a percentage turnout of just over 80 percent. Similar observations were recorded for both median and mean percentage turnouts respectively at 63 and 61 percent.



Further analysis of population and electorate statistics revealed important information in order to put in context turnout percentages. The minimum population evidenced across all EDs was observed to be Terenure D which had a recorded population of 779 and the maximum was Ashtown A with a population of 10,203. The median and mean populations across all Dublin City EDs were 3084 and 3268, respectively. The distributions of populations revealed a right skew and long tail.



The distribution of electorates across EDs bears similar proportions to the population distribution, as is to be expected. The Electoral District of Mansion House B reported the lowest electorate figure at 326, while Ashtown A reported the largest at 5,238. Both the median and mean were once again close at 2,039 and 2,093 respectively. That Mansion House B reports such a low electorate vis-à-vis population may be informed by the fact that there is most likely a high density of government buildings, offices and hotels in the area, thus the census captures a distorted picture in terms of actual residency levels.



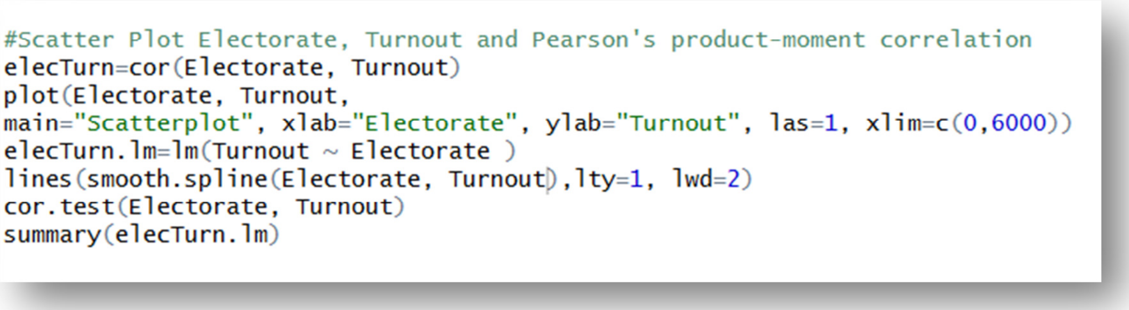
4. Experiment

4.1 Introduction

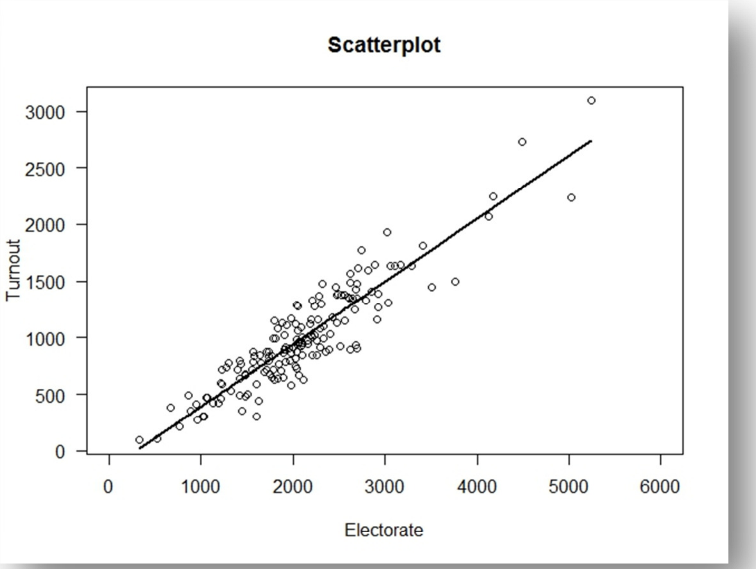
In this chapter experimentation was conducted to examine the extent of correlation between the amalgamated dataset. Also examined were the relationships and extent of correlation between population, electorates and election turnouts. Calculations and prediction experiments are also performed on the basis of R functions that approximate to *ordinary least square regression*, i.e., via obtaining a sum of the least values of the squared residuals.

Correlations – 2009 Local Election Data

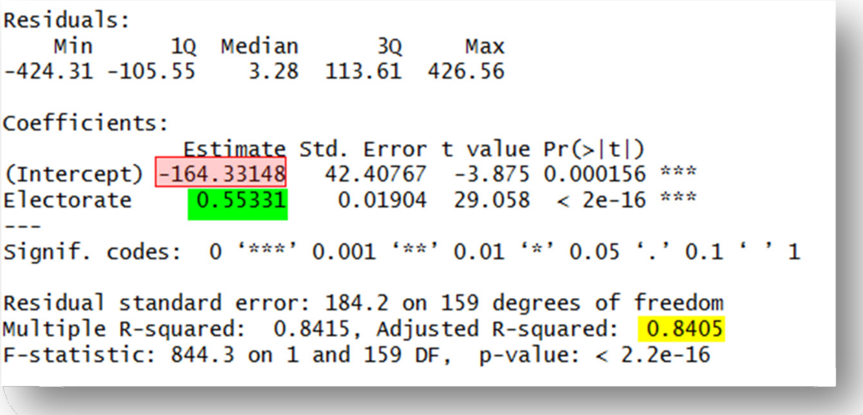
Correlation is the statistical relationships concerning dependence, where the dependence is some statistical relationship between the two sets of data in question. Correlation implies the strength of the association between the variables. In order to confirm the data's suitability and aggregation accuracy, the first correlation plot tested was performed for the two central 2009 election data variables, namely *Electorate* and *Turnout*. The following R code was used to generate a simple linear regression model to predict estimated turnout on the basis of numbers on the electoral registers:



```
#Scatter Plot Electorate, Turnout and Pearson's product-moment correlation
elecTurn=cor(Electorate, Turnout)
plot(Electorate, Turnout,
main="Scatterplot", xlab="Electorate", ylab="Turnout", las=1, xlim=c(0,6000))
elecTurn.lm=lm(Turnout ~ Electorate )
lines(smooth.spline(Electorate, Turnout),lty=1, lwd=2)
cor.test(Electorate, Turnout)
summary(elecTurn.lm)
```



The resulting output gave the following statistics:



```
Residuals:
    Min      1Q  Median      3Q     Max
-424.31 -105.55    3.28  113.61  426.56

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -164.33148   42.40767  -3.875 0.000156 ***
Electorate     0.55331    0.01904  29.058  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 184.2 on 159 degrees of freedom
Multiple R-squared:  0.8415, Adjusted R-squared:  0.8405
F-statistic: 844.3 on 1 and 159 DF,  p-value: < 2.2e-16
```

The adjusted *R–squared* value highlighted above is reasonably high at 0.8405, indicating the linear relationship between electorate and turnout. The *R–squared* coefficient of determination is a statistical measure of how well the regression line approximates the actual data points. An *R–squared* of 1 indicates that the regression line perfectly fits the data. A simple calculation was used to estimate a hypothetical scenario whereby if the persons registered on the electorate were increased by 10 percent this would result in a commensurate increase in turnout. The mean of Dublin City electorates is 2,093 and 2302.3 when added with a 10% increase. Through using the formula: **y = mx + c** (where **m** is the slope, **c** is the intercept and **x** is the number or level of turnout) yielded the following calculation and results:

$$\mathbf{y = mx + c}$$

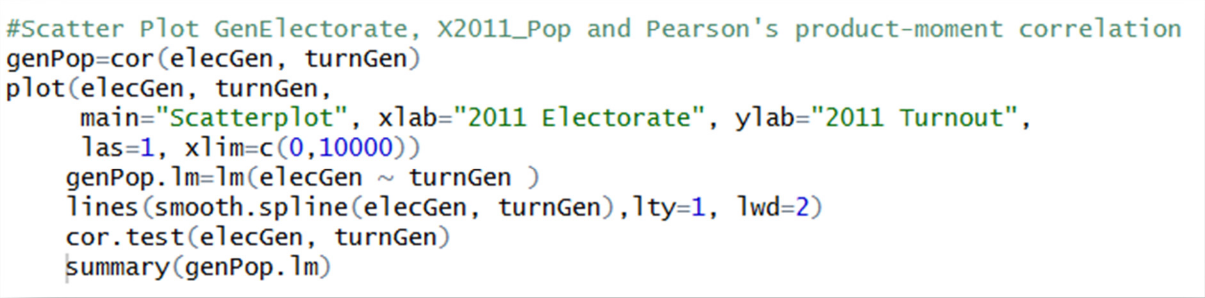
$$\mathbf{= 0.55331 \times 2302.3 + (-164.33148)}$$

$$\mathbf{= 1109.5}$$

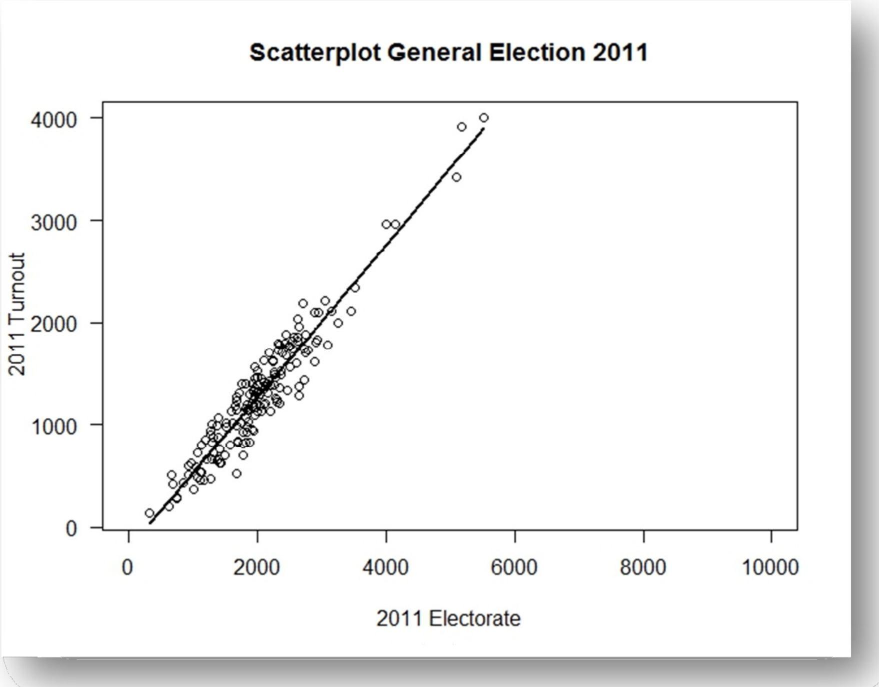
Bearing in mind the mean turnout across the sample is 993 or in percentage terms 47 percent. The 10 percent increase in with regard to the electoral register should expect and average turnout of 1109.5. In percentage terms this would be an increased percentage turnout from 47 percent to 48.2 percent.

Correlations 2011 General Election Data

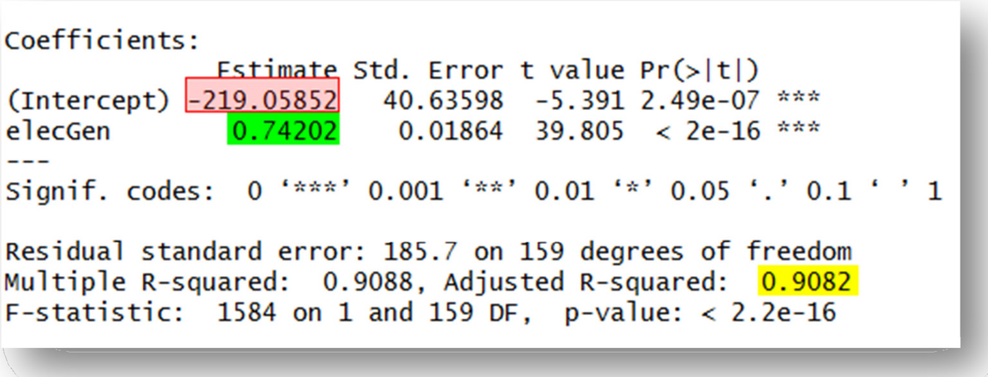
Given that the General Election and the Census 2011 occurred in the same year there was a similar expectation based upon the above results that that a similarly high level of correlation would be present in terms of the General Election electorate figures and the turnout for that election.



```
#Scatter Plot GenElectorate, X2011_Pop and Pearson's product-moment correlation
genPop=cor(elecGen, turnGen)
plot(elecGen, turnGen,
     main="Scatterplot", xlab="2011 Electorate", ylab="2011 Turnout",
     las=1, xlim=c(0,10000))
    genPop.lm=lm(elecGen ~ turnGen )
    lines(smooth.spline(elecGen, turnGen),lty=1, lwd=2)
    cor.test(elecGen, turnGen)
    summary(genPop.lm)
```



As anticipated the results we highly correlated:



```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -219.05852   40.63598  -5.391 2.49e-07 ***
elecGen        0.74202    0.01864  39.805  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 185.7 on 159 degrees of freedom
Multiple R-squared:  0.9088, Adjusted R-squared:  0.9082 
F-statistic:  1584 on 1 and 159 DF,  p-value: < 2.2e-16
```

The *R-squared* value can be seen to be greater than the preceding value for the 2009 Local Elections at 0.9082, thus indicating that slightly greater than 90 percent of the total variation is explained by the model. The same calculation was used to estimate the hypothetical scenario where persons registered on the electorate where increased by 10 percent, whether this would result in a commensurate increase in turnout. The mean of Dublin City electorates is 2,034 and 2237.4 when added with a 10% increase. Using the equation of a line formula: **y = mx + c** yielded the following result:

y = mx + c

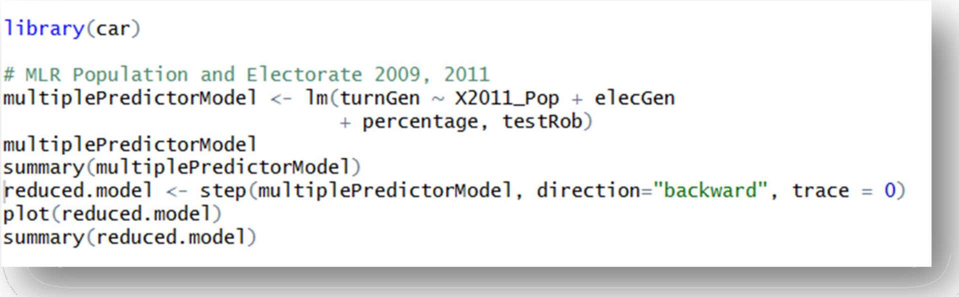
= 0.74202 x 2237.4 + (-219.05852)

= 1441.13 (2dp)

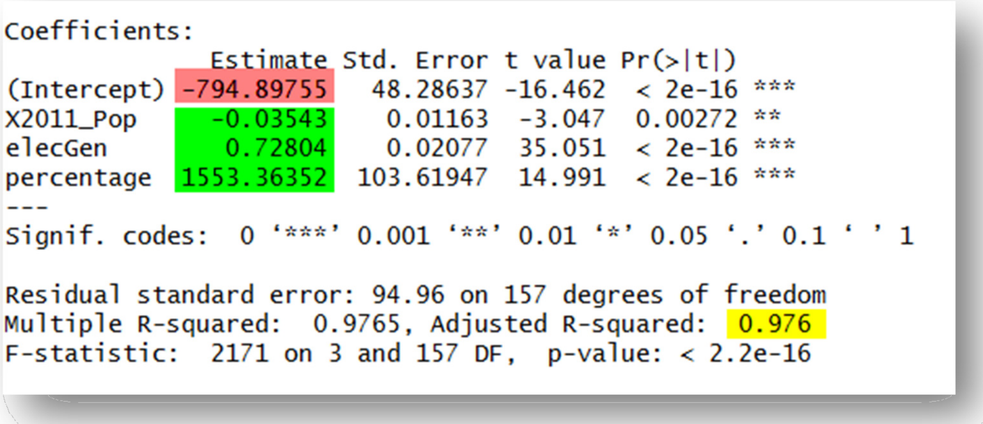
The mean turnout for 2011 in Dublin City was 1,290, or in percentage terms 63.4 percent. Inclusive of a 10 percent increase to the electorate we should expect and average turnout of 1,441. In percentage terms this would be an increased percentage turnout to 64.4 percent.

Multi-Linear Regression Model – Turnout Prediction

Through the implementation of a multi-linear regression model via taking more variables into account, such as the local election percentage turnout results along with the 2011 population and electoral register figures, these gave a better fit for the model. The code used and output is shown below:



```
library(car)

# MLR Population and Electorate 2009, 2011
multiplePredictorModel <- lm(turnGen ~ X2011_Pop + elecGen
                             + percentage, testRob)
multiplePredictorModel
summary(multiplePredictorModel)
reduced.model <- step(multiplePredictorModel, direction="backward", trace = 0)
plot(reduced.model)
summary(reduced.model)
```

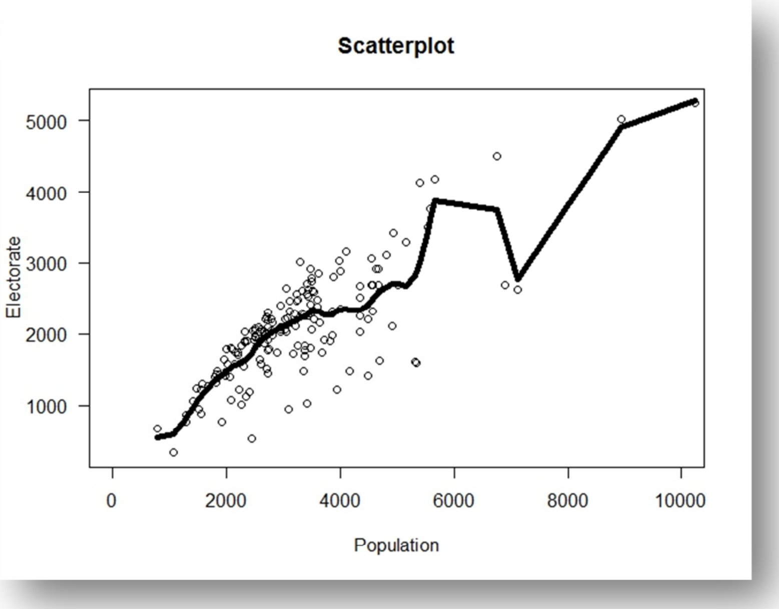


```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -794.89755   48.28637 -16.462  < 2e-16 ***
X2011_Pop     -0.03543    0.01163  -3.047  0.00272 **
elecGen        0.72804    0.02077  35.051  < 2e-16 ***
percentage  1553.36352  103.61947  14.991  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 94.96 on 157 degrees of freedom
Multiple R-squared:  0.9765, Adjusted R-squared:  0.976
F-statistic:  2171 on 3 and 157 DF,  p-value: < 2.2e-16
```

As can be seen from the output above, the *Adjusted R-squared* value is 0.976, which is very high. Through multiplying the coefficients subtracting the intercept, and then adding this result to the mean electorate (plus 10 percent), the average turnout that could be expected across Dublin City EDs goes from a percentage turnout of 63.4 percent to 66.2 percent.

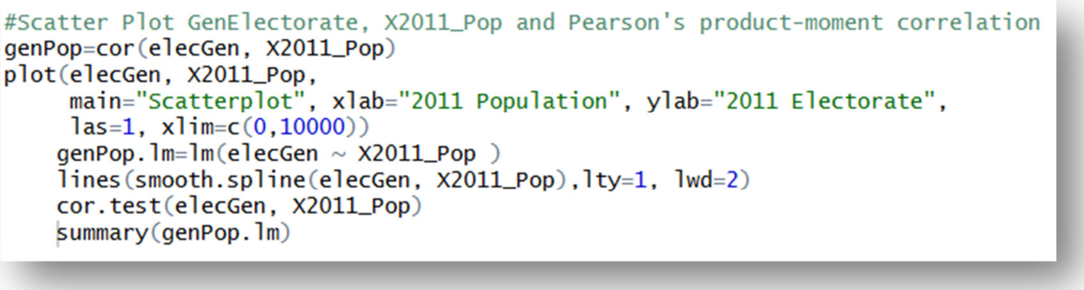
Correlations between Populations and Electorates

In order to better understand the relationship between ED electorates and ED populations, further calculations were used to determine the correlation between variables. For the 2009 electoral register and 2011 census population figures *Pearson product-moment correlation coefficient* was calculated to measure the dependence between these variables and a smoothing spline was used to display the linearity over the data to visualise the results in the graph below:



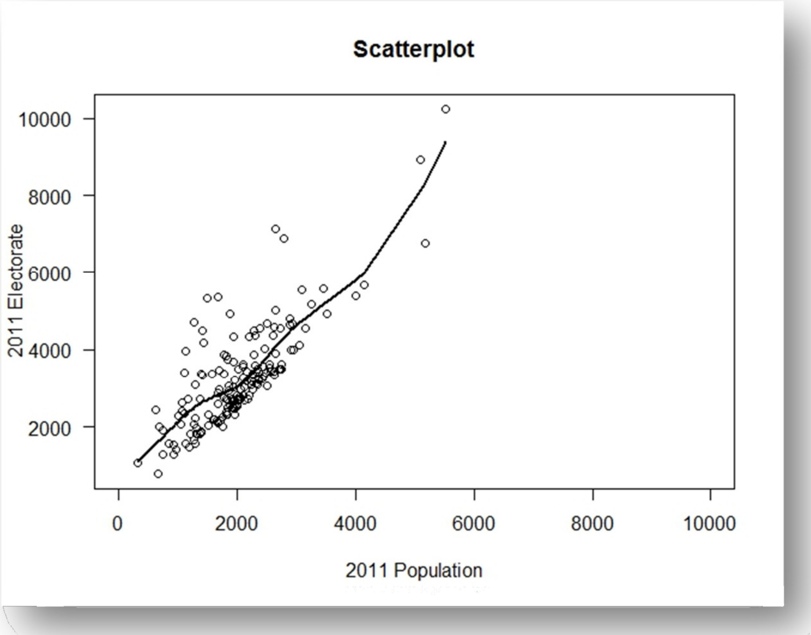
The results measured a 0.78 correlation with a 95% confidence interval of 0.716 - 0.837, implying a reasonably strong linear relationship between these variables.

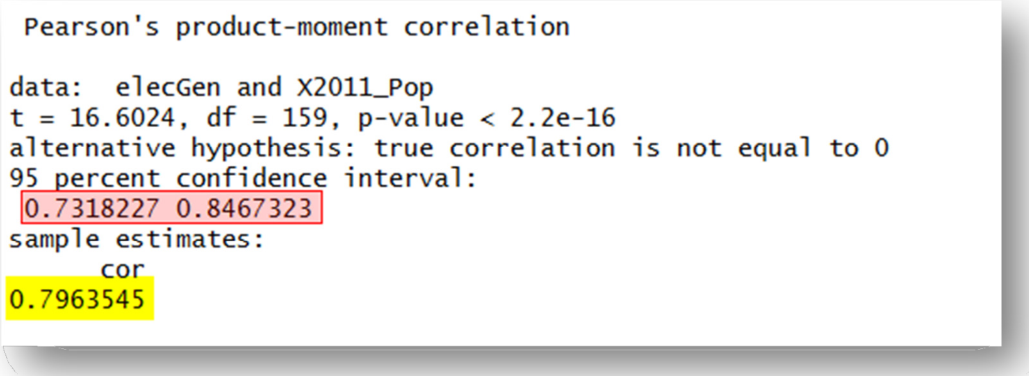
The same analysis was applied to the 2011 electoral register data and the population figures from the census population records using the following code:



```
#Scatter Plot GenElectorate, X2011_Pop and Pearson's product-moment correlation
genPop=cor(elecGen, X2011_Pop)
plot(elecGen, X2011_Pop,
     main="Scatterplot", xlab="2011 Population", ylab="2011 Electorate",
     las=1, xlim=c(0,10000))
    genPop.lm=lm(elecGen ~ X2011_Pop )
    lines(smooth.spline(elecGen, X2011_Pop),lty=1, lwd=2)
    cor.test(elecGen, X2011_Pop)
    summary(genPop.lm)
```

This produced the following graph again with a smoothing spline:





```
 Pearson's product-moment correlation

data:  elecGen and X2011_Pop
t = 16.6024, df = 159, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7318227 0.8467323
sample estimates:
      cor
0.7963545
```

A correlation of 0.731(3dp) with a 95 percent confidence interval of 0.731 (3dp) - 0.846(3dp) was observed. This was a slightly higher correlation between the data than the previous analysis, perhaps owing to the general election and the census being held in the same year.

5. Results and Evaluation - Introduction

Having established the level consistency with regard to the amalgamated datasets, the next steps taken were to begin to build multiple linear regression models taking into account education levels in order to predict voter turnout. Various combinations of variables were examined in order to determine the best predator variables.

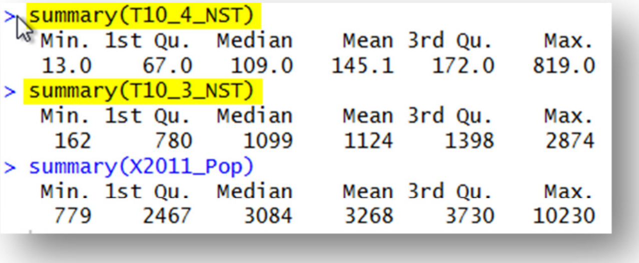
5.1 Results and Evaluation

On the basis that a strong correlation was determined between population and the electorates with regard to predicting turnout, the next step was to examine whether there were observable correlations with education statistics across EDs in order to determine whether they had any correlation to the dependent variable; voter turnout.

In order to build the multi-linear regression model, all of the relevant variables pertaining to totals i.e. levels of education attainment by discipline for both males and females, were first isolated from the dataset. These were:

DESCRIPTION OF FIELD	Code
Not Stated (Total) (Post Grad?)	T10_4_NST
Other (Total)	T10_2_OTHT
Still At School (Total)	T10_2_SAST
Agriculture and Veterinary (Total)	T10_3_AGRT
Art (Total)	T10_3_ARTT
Education and teacher training (Total)	T10_3_EDUT
Engineering, Manufacturing and Construction (Total)	T10_3_ENGT
Health and Welfare (Total)	T10_3_HEAT
Humanities (Total)	T10_3_HUMT
Not Stated (Total) (All?)	T10_3_NST
Other subjects (Total)	T10_3_OTHT
Science, Mathematics and Computing (Total)	T10_3_SCIT
Services (Total)	T10_3_SERT
Social Sciences, Business and Law (Total)	T10_3_SOCT
Advanced Certificate/Completed Apprenticeship (Total)	T10_4_ACCAT
Doctorate(Ph.D) or higher (Total)	T10_4_DT
Higher Certificate (Total)	T10_4_HCT
Honours Bachelor Degree, Professional Qualification or both (Total)	T10_4_HDPQT
Lower Secondary (Total)	T10_4_LST
No Formal Education (Total)	T10_4_NFT
Ordinary Bachelor Degree or National Diploma (Total)	T10_4_ODNDT
Postgraduate Diploma or Degree (Total)	T10_4_PDT
Primary Education (Total)	T10_4_PT
Technical or Vocational qualification (Total)	T10_4_TVT
Upper Secondary (Total)	T10_4_UST

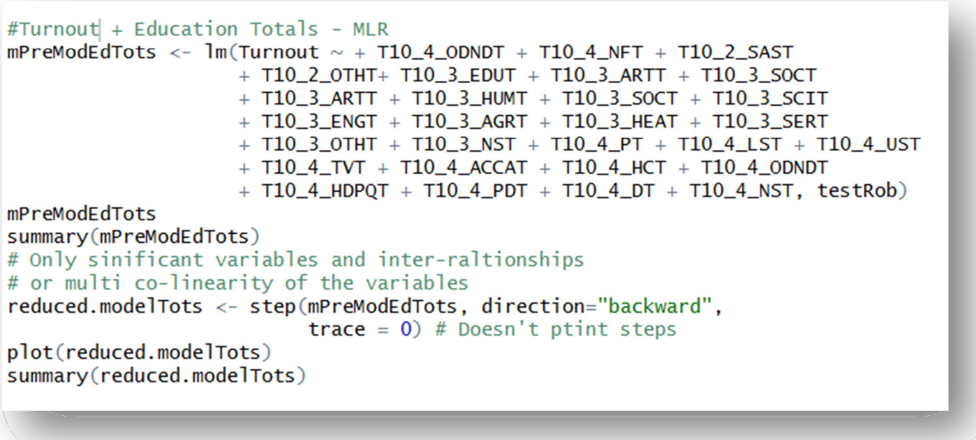
As can be seen from the table, *Not Stated* appears twice as different variables. When examining the documentation regarding the census no clarification was given. By using R summary statistics however, a clearer distinction was given regarding the difference between these variables:



```
> summary(T10_4_NST)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   13.0    67.0   109.0   145.1   172.0   819.0 
> summary(T10_3_NST)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
    162     780    1099    1124    1398    2874 
> summary(X2011_Pop)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
    779    2467    3084    3268    3730   10230 
```

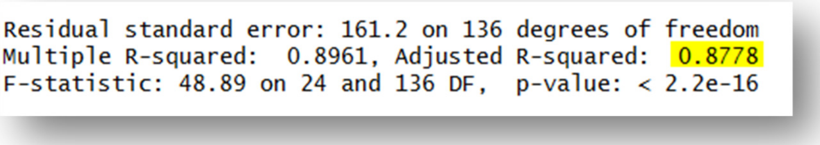
Given that the values for the Min, 1st Qu., Median, 3rd Qu. and Max are all significantly different, the assumption can be safely drawn that the data is not duplicated. Owing to the grouping of the variables in the dataset, it can be speculated that one category of *Not Stated,* T10_3_NST refers to the overall total of respondents that did not make a statement regarding any level of education, while T10_4_NST mostly likely referred to 3rd level qualification not stated.

The first model attempted pertained to the 2009 election data along with the 2011 education data. The following code was used to generate a multi-linear model. The predicted or dependent value is again turnout while the independent predictor variables are the variables referring to the education statistics.



```
#Turnout + Education Totals - MLR
mPreModEdTots <- lm(Turnout ~ + T10_4_ODNDT + T10_4_NFT + T10_2_SAST
                    + T10_2_OTHT+ T10_3_EDUT + T10_3_ARTT + T10_3_SOCT
                    + T10_3_ARTT + T10_3_HUMT + T10_3_SOCT + T10_3_SCIT
                    + T10_3_ENGT + T10_3_AGRT + T10_3_HEAT + T10_3_SERT
                    + T10_3_OTHT + T10_3_NST + T10_4_PT + T10_4_LST + T10_4_UST
                    + T10_4_TVT + T10_4_ACCAT + T10_4_HCT + T10_4_ODNDT
                    + T10_4_HDPQT + T10_4_PDT + T10_4_DT + T10_4_NST, testRob)
mPreModEdTots
summary(mPreModEdTots)
# Only sinificant variables and inter-raltionships
# or multi co-linearity of the variables
reduced.modelTots <- step(mPreModEdTots, direction="backward",
                          trace = 0) # Doesn't ptint steps
plot(reduced.modelTots)
summary(reduced.modelTots)
```

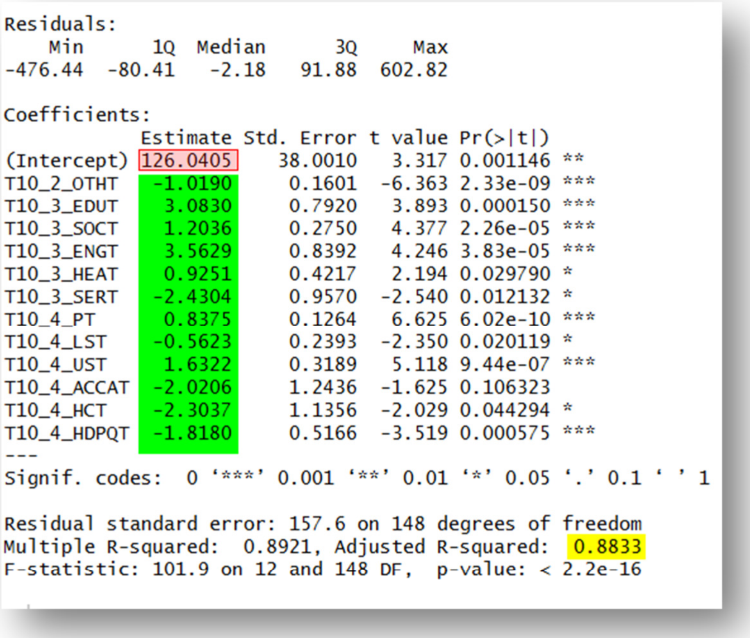
This produced the following output:



```
Residual standard error: 161.2 on 136 degrees of freedom
Multiple R-squared:  0.8961, Adjusted R-squared:  0.8778 
F-statistic: 48.89 on 24 and 136 DF,  p-value: < 2.2e-16
```

As can be seen the model produced a relatively high *R-Squared* value. Given that an *R-Squared* of 1 indicates a regression line perfectly fitting the data, we see that the model above an R-Squared value accounting for nearly 90 percent of the variance in the data. Given that *Adjusted R-Squared* increases when variables are added to improve the model more than would be expected by chance, we see here that the *Adjusted R-Squared* is less, thus indicating that the large number of variables is detracting from the model.

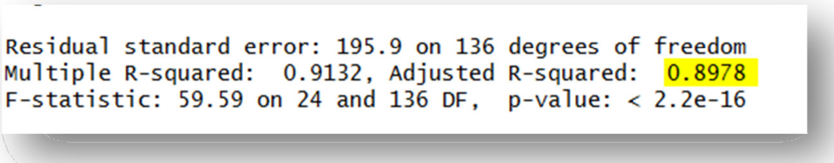
In order to select the most statistically significant variables, a backward stepwise variable selection was performed. This is the process whereby having started with all candidate variables we then test the deletion of each variable, (as to determine whether the deletion of that variable improves the model) and the repeating this process until no further improvement is possible. The backward stepwise elimination produced the following results:



```
Residuals:
    Min      1Q  Median      3Q     Max 
-476.44  -80.41   -2.18   91.88  602.82 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 126.0405    38.0010   3.317 0.001146 ** 
T10_2_OTHT   -1.0190     0.1601  -6.363 2.33e-09 ***
T10_3_EDUT    3.0830     0.7920   3.893 0.000150 ***
T10_3_SOCT    1.2036     0.2750   4.377 2.26e-05 ***
T10_3_ENGT    3.5629     0.8392   4.246 3.83e-05 ***
T10_3_HEAT    0.9251     0.4217   2.194 0.029790 *  
T10_3_SERT   -2.4304     0.9570  -2.540 0.012132 *  
T10_4_PT      0.8375     0.1264   6.625 6.02e-10 ***
T10_4_LST    -0.5623     0.2393  -2.350 0.020119 *  
T10_4_UST     1.6322     0.3189   5.118 9.44e-07 ***
T10_4_ACCAT  -2.0206     1.2436  -1.625 0.106323    
T10_4_HCT    -2.3037     1.1356  -2.029 0.044294 *  
T10_4_HDPQT  -1.8180     0.5166  -3.519 0.000575 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 157.6 on 148 degrees of freedom
Multiple R-squared:  0.8921,	Adjusted R-squared:  0.8833 
F-statistic: 101.9 on 12 and 148 DF,  p-value: < 2.2e-16
```

As can be seen in the above output, the model has been reduced from 25 independent predictor variables to 12. Asterisks are put beside the most statistically significant p-values on the right. These are all below the measure of statistical significance of 0.05, except for the variable T10_4_ACCAT (Advanced Certificate/Completed Apprenticeship). The *Adjusted R-Squared* value has gone from .8778 in the previous model to a slightly more improved value of 0.8833.

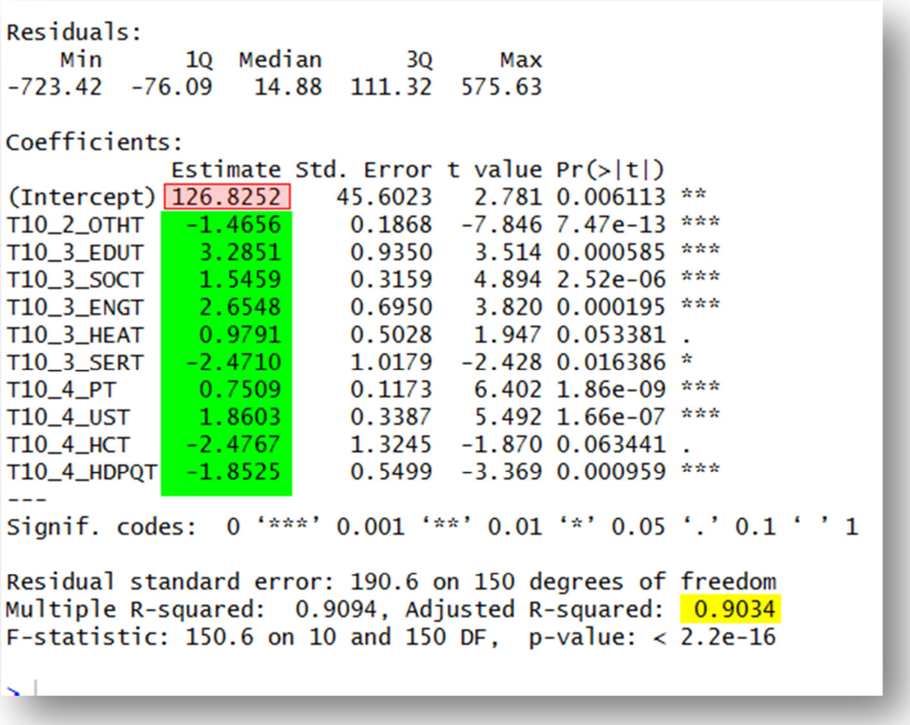
In respect to the 2011 general election turnout data, the same procedure was again performed:



```
Residual standard error: 195.9 on 136 degrees of freedom
Multiple R-squared:  0.9132,	Adjusted R-squared:  0.8978 
F-statistic: 59.59 on 24 and 136 DF,  p-value: < 2.2e-16
```

The model included the same 25 predictor education variables as before. The *R-Squ*ared and *Adjusted R-Squared* values were slightly higher than the model for the 2009 election data.

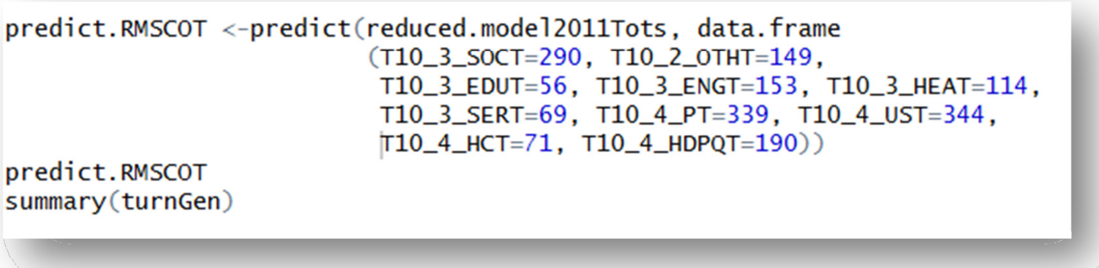
Employing the backward stepwise model again performed better however in terms of the *Adjuster R-Squared* value:



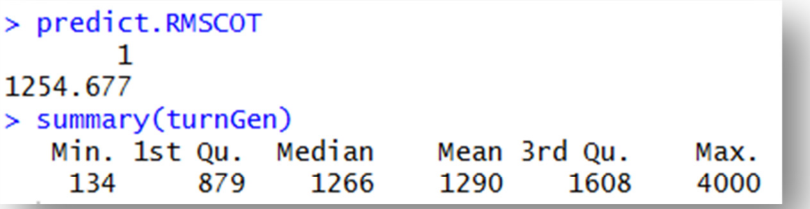
```
Residuals:
    Min      1Q  Median      3Q     Max 
-723.42  -76.09   14.88  111.32  575.63 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 126.8252    45.6023   2.781 0.006113 ** 
T10_2_OTHT   -1.4656     0.1868  -7.846 7.47e-13 ***
T10_3_EDUT    3.2851     0.9350   3.514 0.000585 ***
T10_3_SOCT    1.5459     0.3159   4.894 2.52e-06 ***
T10_3_ENGT    2.6548     0.6950   3.820 0.000195 ***
T10_3_HEAT    0.9791     0.5028   1.947 0.053381 .  
T10_3_SERT   -2.4710     1.0179  -2.428 0.016386 *  
T10_4_PT      0.7509     0.1173   6.402 1.86e-09 ***
T10_4_UST     1.8603     0.3387   5.492 1.66e-07 ***
T10_4_HCT    -2.4767     1.3245  -1.870 0.063441 .  
T10_4_HDPQT  -1.8525     0.5499  -3.369 0.000959 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 190.6 on 150 degrees of freedom
Multiple R-squared:  0.9094,	Adjusted R-squared:  0.9034 
F-statistic: 150.6 on 10 and 150 DF,  p-value: < 2.2e-16

>
```

This multi-linear model with stepwise regression selected 10 dependent variables from the original 25 included in the original model. All bar two were statistically significant; T10_3_HEAT (Health and Welfare) and H10_4_HTC (Higher Certificate), though these variables were not far from statistical significance at 0.53(3dp) and 0.063(3dp) respectfully. The *R-Squared* and *Adjusted R-Squared* values are similar and account for approximately 90-91% of the variability in the data.

Using the selected variables above, a linear predictor was coded in R that allowed for the input of certain values for these variables. By taking the median value for all 10 of the variables across all EDs in Dublin and then inputting them into the code, the model predicted a close approximation to the median general election turnout.

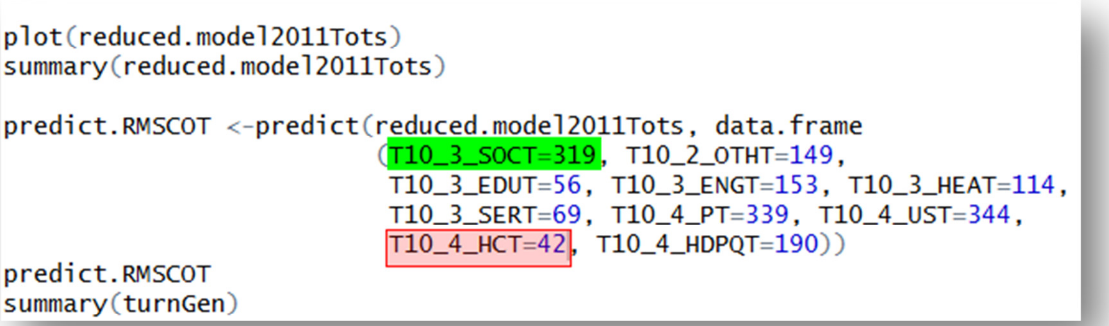


```
predict.RMSCOT <-predict(reduced.model2011Tots, data.frame
                         (T10_3_SOCT=290, T10_2_OTHT=149,
                          T10_3_EDUT=56, T10_3_ENGT=153, T10_3_HEAT=114,
                          T10_3_SERT=69, T10_4_PT=339, T10_4_UST=344,
                          T10_4_HCT=71, T10_4_HDPQT=190))
predict.RMSCOT
summary(turnGen)
```

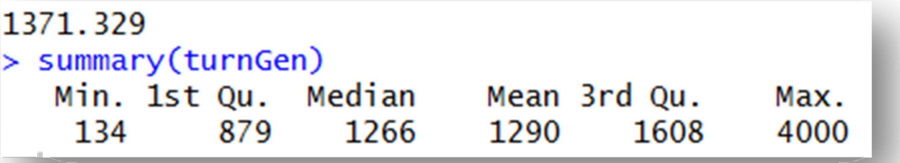


```
> predict.RMSCOT
       1 
1254.677 
> summary(turnGen)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
    134     879    1266    1290    1608    4000 
```

As can be seen from the output, the model specified with median values for the independent predictor variables predicted a turnout of 1254.677, while the actual median value was 1266. Through tweaking the values used for the predictor variables it is possible to hypothesise scenarios whereby an increase in certain levels of education at the expense of others predict alternate scenarios. Thus should we add 10 percent more *Social Sciences, Business and Law* graduates to the median, at the expense of a commensurate reduction in the value of less statistically significant variables such as *Higher Certificate,* we should observe variable results. In the scenario outlined about the following code was used and output observed (changes in the code are highlighted).



```
plot(reduced.model2011Tots)
summary(reduced.model2011Tots)

predict.RMSCOT <-predict(reduced.model2011Tots, data.frame
                          (T10_3_SOCT=319, T10_2_OTHT=149,
                           T10_3_EDUT=56, T10_3_ENGT=153, T10_3_HEAT=114,
                           T10_3_SERT=69, T10_4_PT=339, T10_4_UST=344,
                           T10_4_HCT=42, T10_4_HDPQT=190))
predict.RMSCOT
summary(turnGen)
```



```
1371.329 
> summary(turnGen)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
    134     879    1266    1290    1608    4000 
```

As can be seen from the output, given a 10 percent increase in the specified variable highlighted green and a decrease in the other specified variable highlighted red, we can observe a predicted median of 1371.329 vis-à-vis the actual median value of 1266.

6. Conclusions and Future Work

6.1 Introduction

In this chapter conclusions and insights with regard to observations that have been made by analysing the educational attainment data from the CSO and electoral data from the Dublin Returning Officer will be detailed. Also considered are possible areas for which future research and analysis might be directed.

6.2 Conclusions

Having investigated the election data thoroughly, strong correlations exist between levels of turnout and the numbers of persons of the electoral register within the respective electoral districts. Strong correlations also existed over time whereby the 2009 local elections showed strong correlation with general election results. By merging the datasets and taking populations into account, multi-linear models were able to explain 97 percent of the variation in the data when predicting turnout. The turnout for general election of 2011 proved to have a higher correlation to populations and the electoral register than the 2009 local elections had to those respective variables, perhaps owing to the coincidence of the 2011 census occurring in the same year as the general election.

Various multi-linear regression models using R's approximation to *ordinary least squares* proved effective when attempting to predict voter turnout on the basis of levels and types of educational attainment. Through processing and analysing the models using backward stepwise methods these models were refined to select statistically significant variables. On the basis of this selection method, a predictive model was constructed enabling the adjustment of values within variables, this allows for the exploration of counterfactual scenarios whereby the make-up of education profiles in relation to electoral districts and Dublin City can be examined.

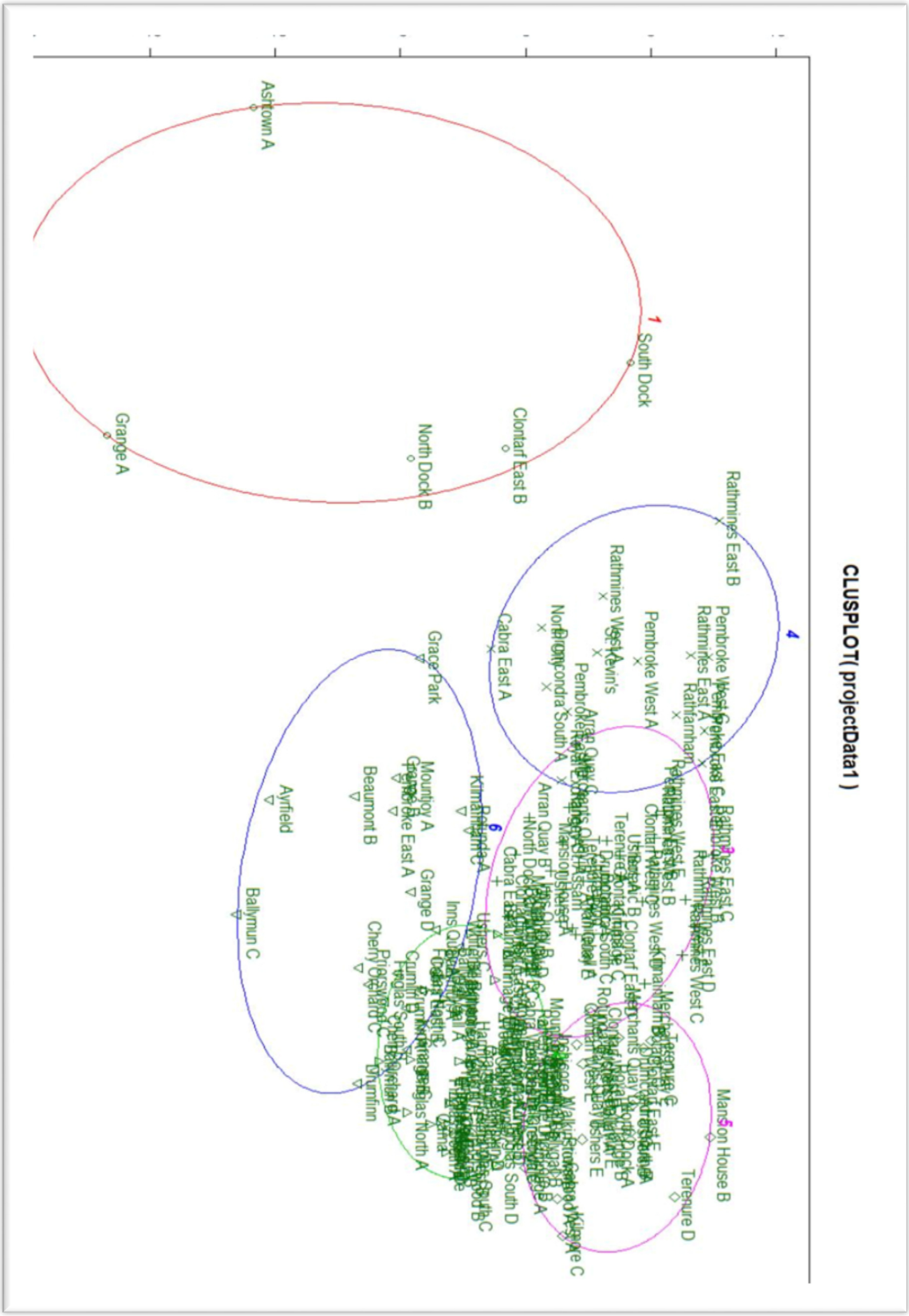
It should be noted that backward stepwise regression has advantages and disadvantages. On the positive side it allows for an automatic process for selecting statistically significant variables however it has also drawbacks as authors have pointed to instances whereby models may be over-simplified (Roecker 1991). Lastly, though strong correlations have been observed in respect of several variables, the truism should be restated here that correlation does not equal causation. Thus should we seek to move or attract individuals with certain educational qualifications into areas in order to enhance electoral turnout, the certainty of this endeavour cannot be assured as previous authors (Burden 2009) have note that increasing levels of education in other jurisdictions have not enjoyed commensurate increases in voter turnout as expected.

6.3 Future Work

The work undertaken heretofore was possible due to the exceptional and publically available election results published by the *Dublin City Returning Officer*. Unfortunately, similar granular results are not available for the rest of the country. Recent legislation has proposed the establishment of a national electoral commission and it is hoped that in the future that this detail information will be publically available. Indeed out of all information published by State agencies, it can easily be argued that information regarding elections should rank highest in terms of accessibility in order to underpin the transparency of the electoral process. Should electoral data become available I would like to apply similar analysis on a regional and national basis in order to confirm or disprove the findings made in this paper.

Other such work worthy of further investigation could include different predictive modelling techniques such as neural networks, *k*-nearest neighbours' algorithms and *k*-means clustering analysis. In the last instance a *k*-means clustering analysis was begun. (See Appendix for more information).

Appendix


CLUSPLOT(projectData1)
1
4
5
6
Ashtown A
South Dock
Grange A
Clontarf East B
North Dock B
Rathmines East B
Pembroke West C
Rathmines East A
Rathmines West A
St. Kevin's
Pembroke West A
Drumcondra South A
Cabra East A
Grace Park
Ayrfield
Ballymun C
Beaumont B
Mountjoy A
Pembroke East A
Grange D
Cherry Orchard C
Drumfinn
Finglas North A
Mansion House B
Terenure D
Rathmines West C
Arran Quay C
Cabra East C
Kilmainham C

With regard to the *k*-means clustering this is done via a complete cluster method using Euclidean distance (a way of measuring the distance between cases). In the diagram above each of the case or EDs are analysed on the basis of their dis-similarity. Seen in the diagram are 6 circles which are groups that I defined. This is based on the percentage sum of squares by cluster: In this example the percentage is 68.8%. On the basis of this approach, future work could canter around creating a categorical variable for the EDs; 1-6, etc. Further data analysis approaches such as decision trees could then be utilized in order to gain further insights from the data.

Bibliography

Barrington, T. (1991), 'Local Government in Ireland', in R. Batley and G. Stoker (eds.), Local Government in Europe – Trends and Developments, Houndmills, Basingstoke: Macmillan

Callanan, M. and Keogan J.F. (eds.) (2004), Local Government in Ireland: Inside Out, Dublin: Institute of Public Administration.

Coakley and M. Gallagher (eds.), Politics in the Republic of Ireland (4th edition), New York: Routledge.

Delli Carpini, M.X., Keeter, S., (1996), What Americans Know about Politics and Why It Matters. Yale University Press, New Haven, CT. in in Burden, B. (2009) The dynamic effects of education on voter turnout. Electoral Studies, 28, 540-549

Farrell, D.M. (2008), 'Electoral Systems: A Comparative Introduction', Palgrave: New York.

Highton, B. (2004) Voter registration and turnout in the United States. Perspectives on Politics 2, 506–515., in Burden, B. (2009), The dynamic effects of education on voter turnout. Electoral Studies, 28, 540-549

Marsh, M., Sinnott, R., Garry. J., and Kennedy. F. (2008), The Irish Voter: The nature of Electoral Competition in the Republic of Ireland, Manchester: Manchester University Press. P. 199

Miller, W.E. (1992), The puzzle transformed: explaining declining turnout. Political Behavior 14, 1–43 in Burden, B. (2009), The dynamic effects of education on voter turnout. Electoral Studies, 28, 540-549

Miller, W.E. and Shanks, J.M. (1996), The New American Voter. Harvard., in Burden, B. (2009) The dynamic effects of education on voter turnout. Electoral Studies, 28, 540-549 University Press, Cambridge, MA in Burden, B. (2009) The dynamic effects of education on voter turnout. Electoral Studies, 28, 540-549

Lyons, P. and Sinnott, R. (2003), 'Voter Turnout in the Republic of Ireland', Institute for the Study of Social Change Public Opinion & Political Behaviour. Accessed on the 15th June, 2014 *Citation request pending: http://www.ucd.ie/dempart/workingpapers/ireland.pdf

Lijphart, A. (1997), 'Unequal participation: democracy's unresolved dilemma', American Political Science Review.

Putnam, R.D. (2000), Bowling Alone: The Collapse and Revival of American Community. New York: Simon & Schuster

Roecker, Ellen B. (1991), Prediction error and its estimation for subset—selected models. Technometrics, 33, 459–468, in Burden, B. (2009) The dynamic effects of education on voter turnout. Electoral Studies, 28, 540-549

Rosenstone, S.J. and Hansen, J.M. (1993), Mobilization, Participation, and Democracy in America. Macmillan, New York, NY., in Burden, B. (2009) The dynamic effects of education on voter turnout. Electoral Studies, 28, 540-549

Timpone, R.J. (1998), Structure, behavior, and voter turnout in the United States. American Political Science Review 92, 145–158., in Burden, B. (2009) The dynamic effects of education on voter turnout. Electoral Studies, 28, 540-549

Van der Eij, C. and Franklin, M.N. (2009), Elections and Voters, Palgrave MacMillian: Hapshire.

Wolfinger, R.E. and Rosenstone, S.J. (1980), Who Votes? Yale University Press, New Haven, CT., in Burden, B. (2009) The dynamic effects of education on voter turnout. Electoral Studies, 28, 540-549

Zaller, J.R. (1992) ,The Nature and Origins of Mass Opinion. Cambridge University Press, New York, NY., in Burden, B. (2009) The dynamic effects of education on voter turnout. Electoral Studies, 28, 540-549