Final Dataset Analysis

Drai Paulen-Patterson 500986410

2022-05-25

Contents

1 Introduction

Voter turnout is a measurement of the eligible population that participates in a given election. In this dataset, voter turnout refers to the percentage of the eligible population that participated in each country's federal election. This is an important statistic to any democratic nation, as greater voter turnout results in a better representation of the will of the population. This report will attempt to find some key predictors for voter turnout, and prove their association through statistical analysis.

2   Hypothesis

I believe a government that mandates elections will create larger voter turnout through threat of consequences, and that a better educated population will be more likely to vote, as they may be more invested in political discourse. Furthermore, I believe that a country's election type will be correlated with the compulsory status of their elections, as stricter governments may have similar election systems.

These theories will be analysed by examining the statistical significance of the correlation between voter turnout, education and compulsory status of elections, as well as the significance of the correlation between election mandates and election types.

3   Data

This dataset will be referred to as the Vote dataset. It contains voter turnout data for 76 different countries, as well as entries for election year, compulsory status of the election, election type, democracy type, bicameral status of the election, district magnitude, effective number of political parties, a measure of competitiveness, years of continuous democracy, a measure of corruption, a measure of education level, and population. Some countries in the dataset have missing entries for some of these variables.

3.1   Wrangling and Tidying Data

For the purposes of this report, all the unneeded columns were removed, leaving only the country, compulsory status, election type, and education levels for analysis.

4   Univariate Analysis

4.1   Turnout

As can be seen in Table 1 and Table 2, most of the data is clustered from approximately 60% to 80%, with some exceptions far below 40%. The range of values is fairly large, as some countries have less than 30% election participation in comparison with other countries that have more than 95% voter turnout. The large standard deviation shows that the distribution is not tightly concentrated. As the mean is very close to the median, any outliers in the dataset are not skewing the data.

Table 1

count	mean	sd	median	min	max	range
76	70.1	15.5	69.5	21.6	98	76.5

Table 2: Turnout
Tendencies

	V1
Min.	21.6
1st Qu.	60.3
Median	69.5
Mean	70.1
3rd Qu.	81.4
Max.	98

Figure 1 shows that the distribution of voter turnout is mostly normal, with the median turnout being fairly centered in the distribution at approximately 70. It also confirms the small amount of countries with very low turnout. This graph, along with Table 1 also show that there is a single value (the minimum value of 21.6) that can be classified as an outlier, as it is more than 3 standard deviations from the mean.
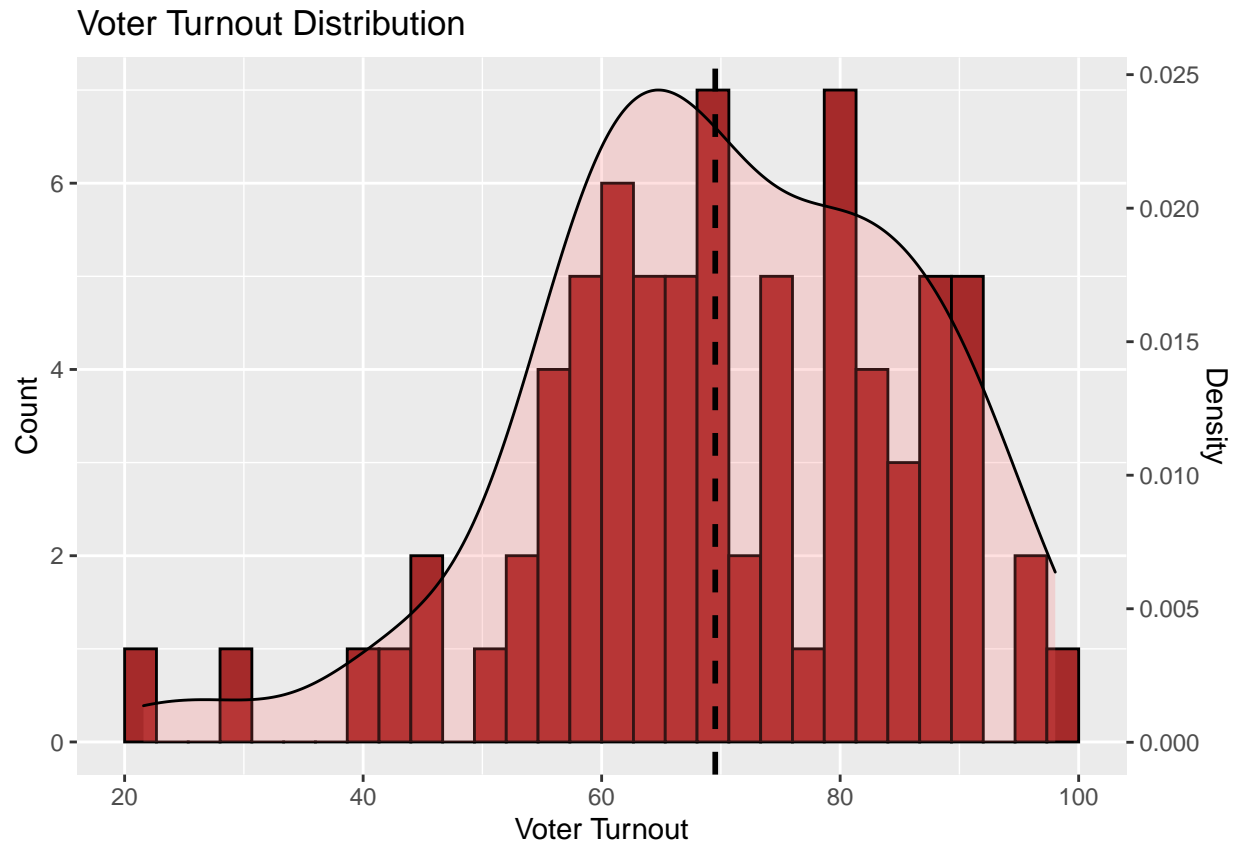

Figure 1: Turnout Distribution

4.2 Education Level

Table 3 and Table 4 show how the distribution of education levels is biased to the higher side of the scale with a median of 8.55, as well as the lower quartile being above the halfway point of the education measurement. The data is also widely dispersed, with a large standard deviation of 2.56. The range of values shows that there are no data points further than 3 stadard deviations from the mean, and therefore no outliers.

Table 3

count	mean	sd	median	range
49	7.97	2.56	8.55	11.2

Table 4: Education Tendencies

	V1
Min.	0.876
1st Qu.	6.14
Median	8.55
Mean	7.97
3rd Qu.	9.6
Max.	12
NA's	27

Figure 2 confirms that the distribution is biased towards the higher side, but also depicts how the distribution of education levels is very wide. This means that the levels of education of each country vary greatly.
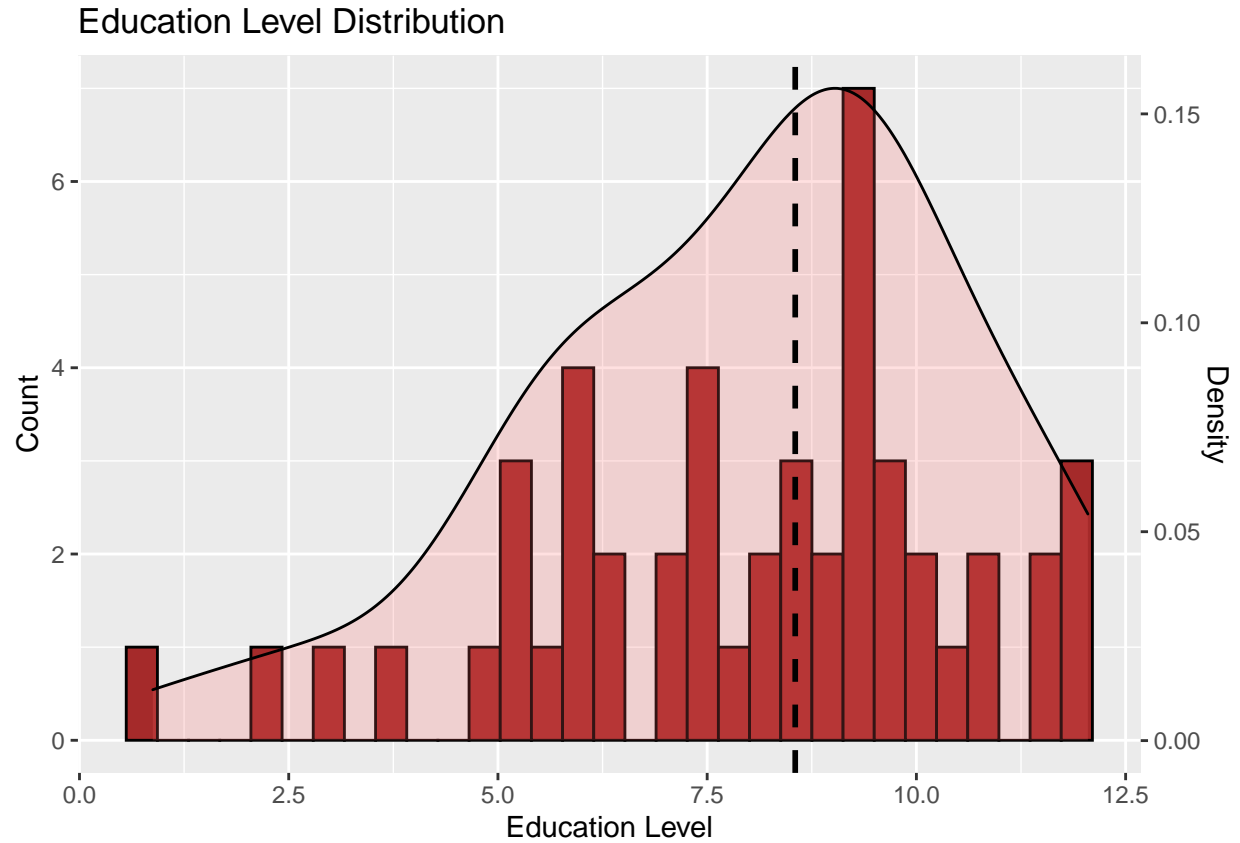
Figure 2: Education Level Distribution

4.3 Compulsory Status

Figure 3 depicts how most (60 of the 76) countries do not have mandatory elections. Furthermore, it shows how countries with mandatory elections are more likely to enforce them than allow these rules to go unenforced.
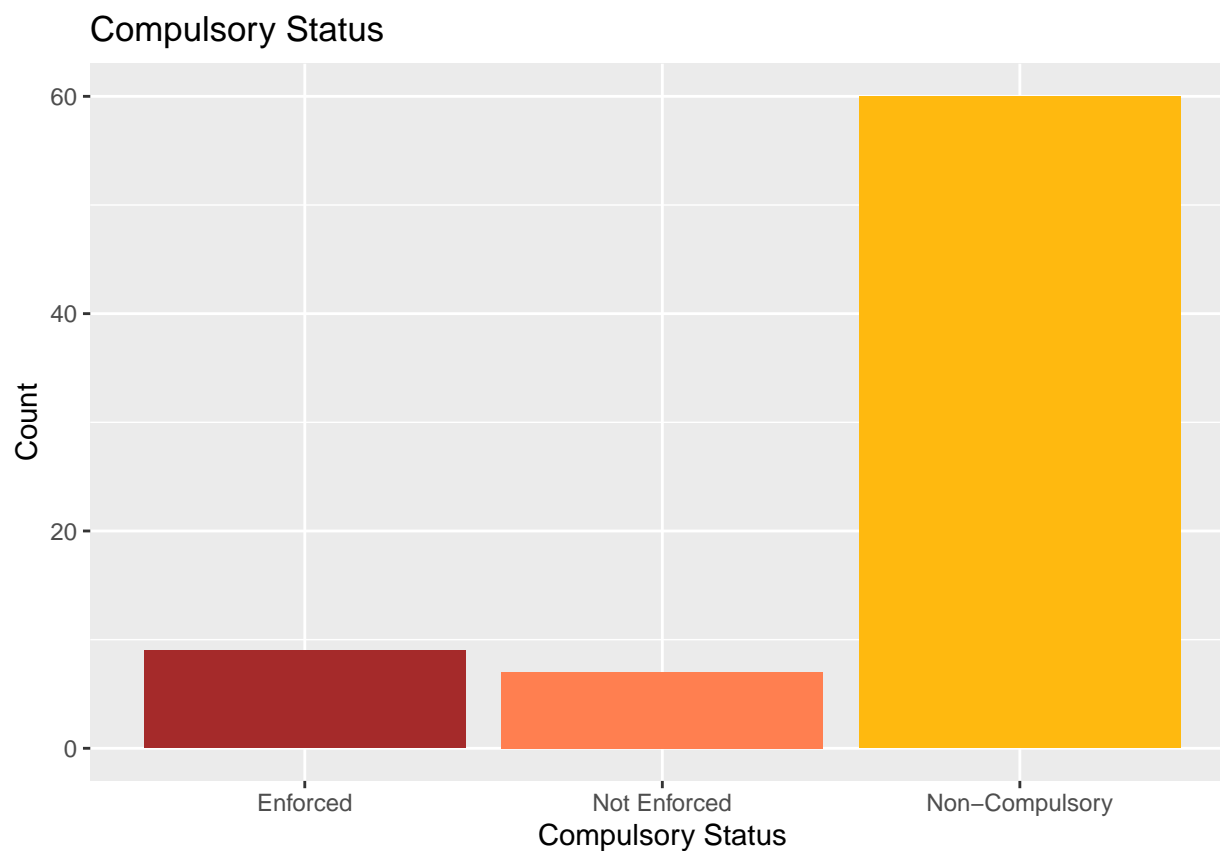
Figure 3: Number of Countries by Compulsory Status

Table 5 shows the exact values for the election compulsory statuses. Voluntary elections are more than 3 times as favoured over compulsory elections.

Table 5: Compulsory Status Counts

Compulsory	Count
Enforced	9
Not Enforced	7
Non-Compulsory	60

4.4 Election Type

Figure 4 depicts how most of the elections in the Vote dataset are of proportional type, with majoritarian types being second most likely, and mixed elections being least popular.
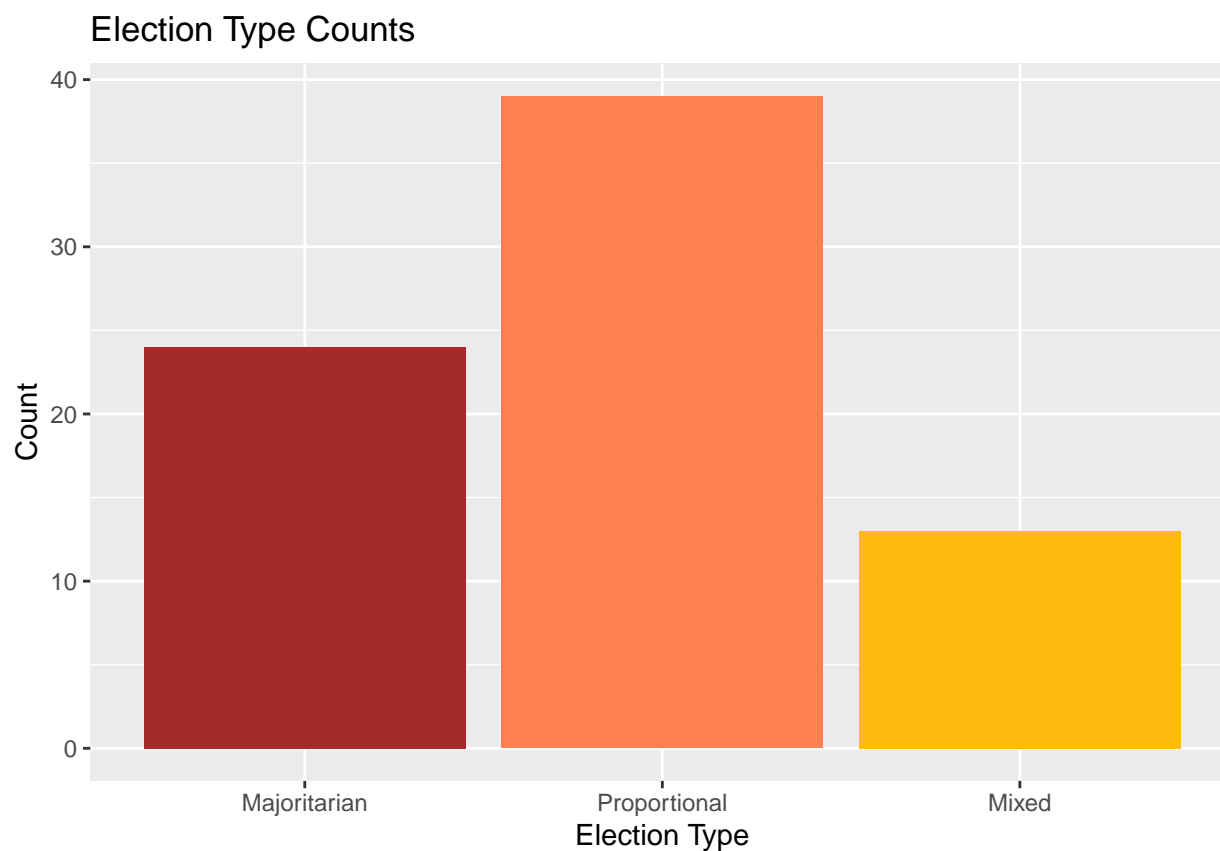
Figure 4: Number of Countries by Election Type

Table 6 shows the exact values of each type of election. Proportional elections are 3 times as popular as mixed elections.

Table 6: Election Type Counts

Election Type	Count
Majoritarian	24
Proportional	39
Mixed	13

5 Bivariate Analysis

5.1 Education

5.1.1 Correlation

Calculating the R correlation coefficient shows that there is a weak positive correlation between education level and turnout, as can be seen in Table 7.

Table 7

correlation
0.356

5.1.2 Linear Regression Graph

Plotting education level against voter turnout in Figure 5 shows that the points are widely dispersed, and the regression line does not pass through many points, which helps confirm that the correlation is fairly weak. Further, the regression line has a gradual slope, which suggests that increases in education level do not have drastic increases in voter turnout.
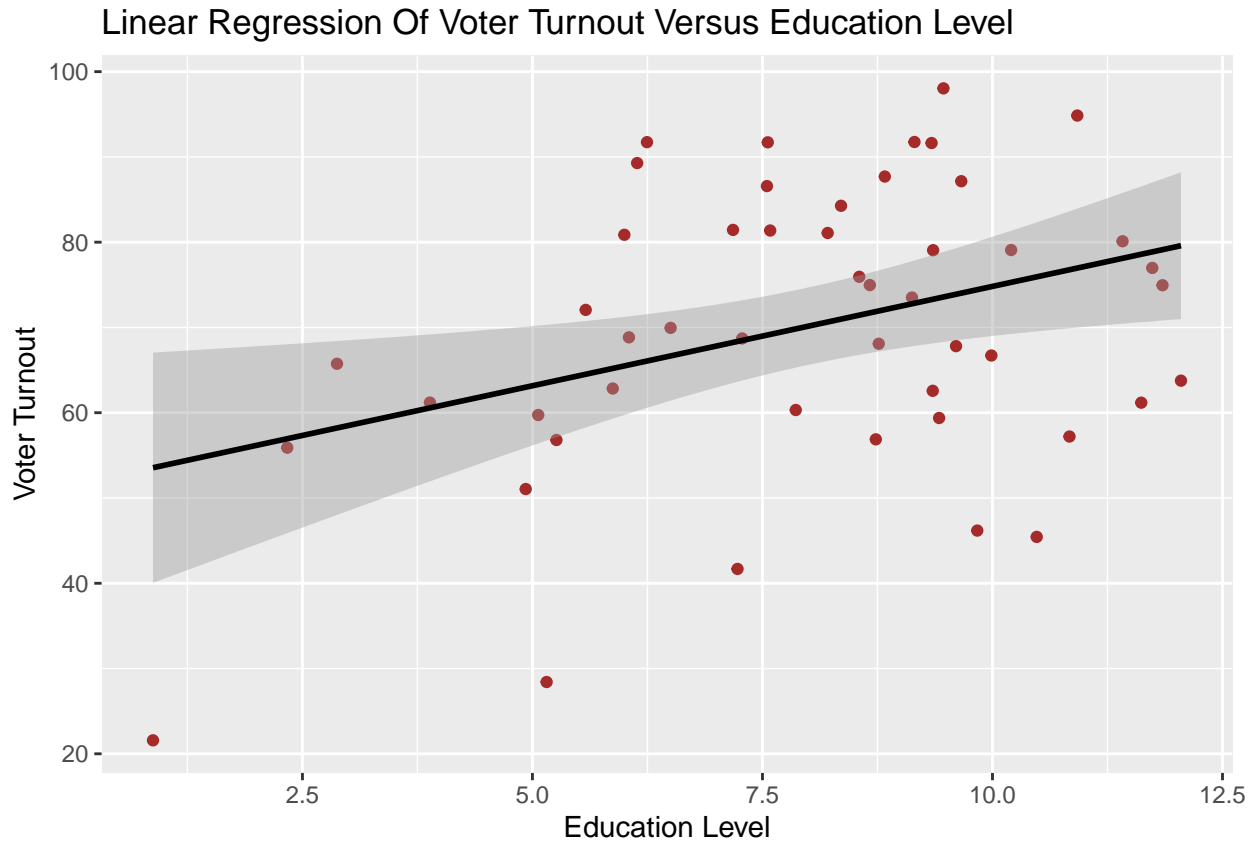

Figure 5: Linear Regression of Education

5.1.3 Regression Table

The regression table at Table 8 confirms that increasing education level by one is correlated with a small increase in voter turnout, with every point increase of education level augmenting voter turnout by 2.3 points on average. The regression table does also show that this correlation, although weak, is statistically significant with a p-value of 0.01. Because the confidence interval does not include 0, the null hypothesis can be successfully rejected.

Table 8

term	estimate	std_error	statistic	p_value	lower_ci	upper_ci
intercept	51.5	7.45	6.91	0	36.5	66.5
educ	2.33	0.891	2.62	0.012	0.538	4.12

5.2 Compulsory Status

5.2.1 Summarise data

A cursory look of the median turnout in comparison with compulsory status gives the impression that turnout seems to rise with mandatory elections, and that enforcement of mandatory elections creates the largest turnout.

Table 9

compulsory	median_turnout	count
Enforced	88.8	9
Not Enforced	70	7
Non-Compulsory	66.7	60

5.2.2 Graphs

Analysing the distributions of the voter turnout by compulsory status in Figure 6 and Figure 7 seem to confirm that stricter enforcement of election mandates does increase voter turnout.
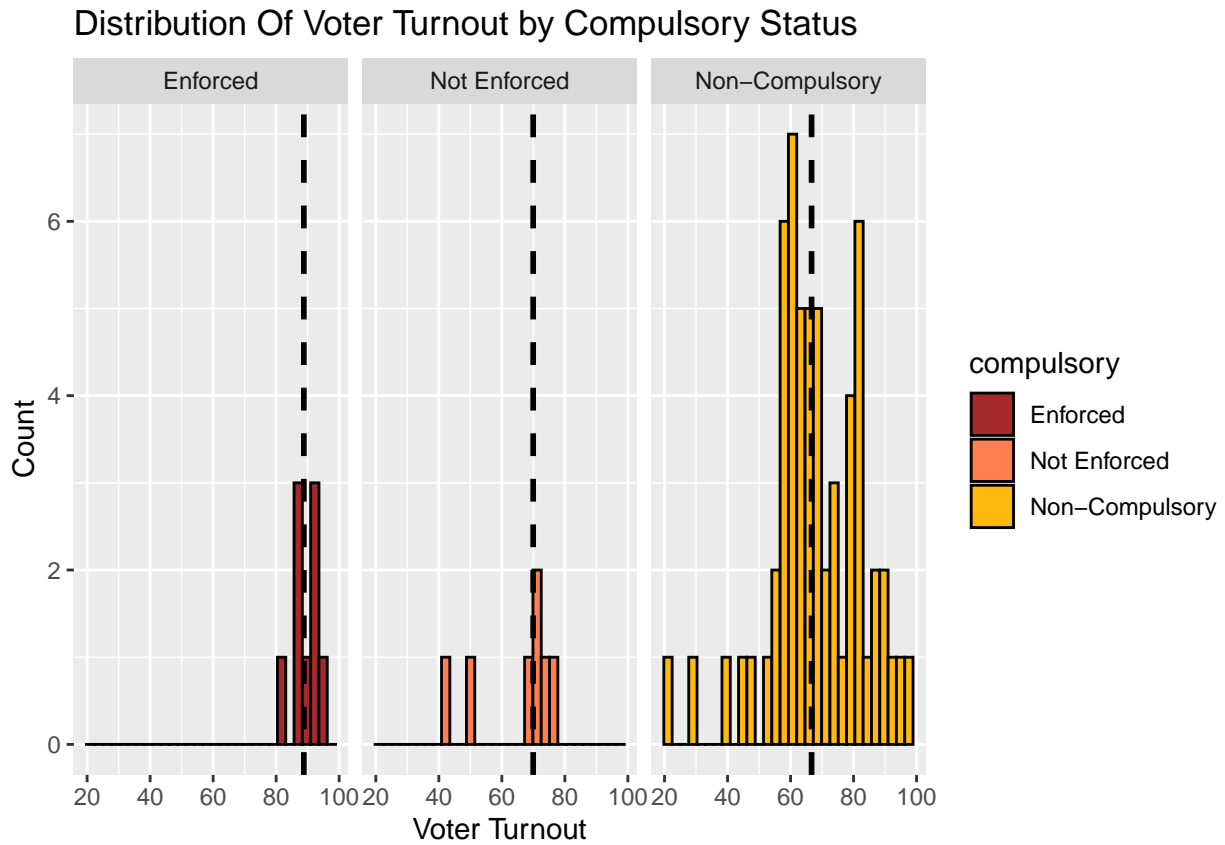
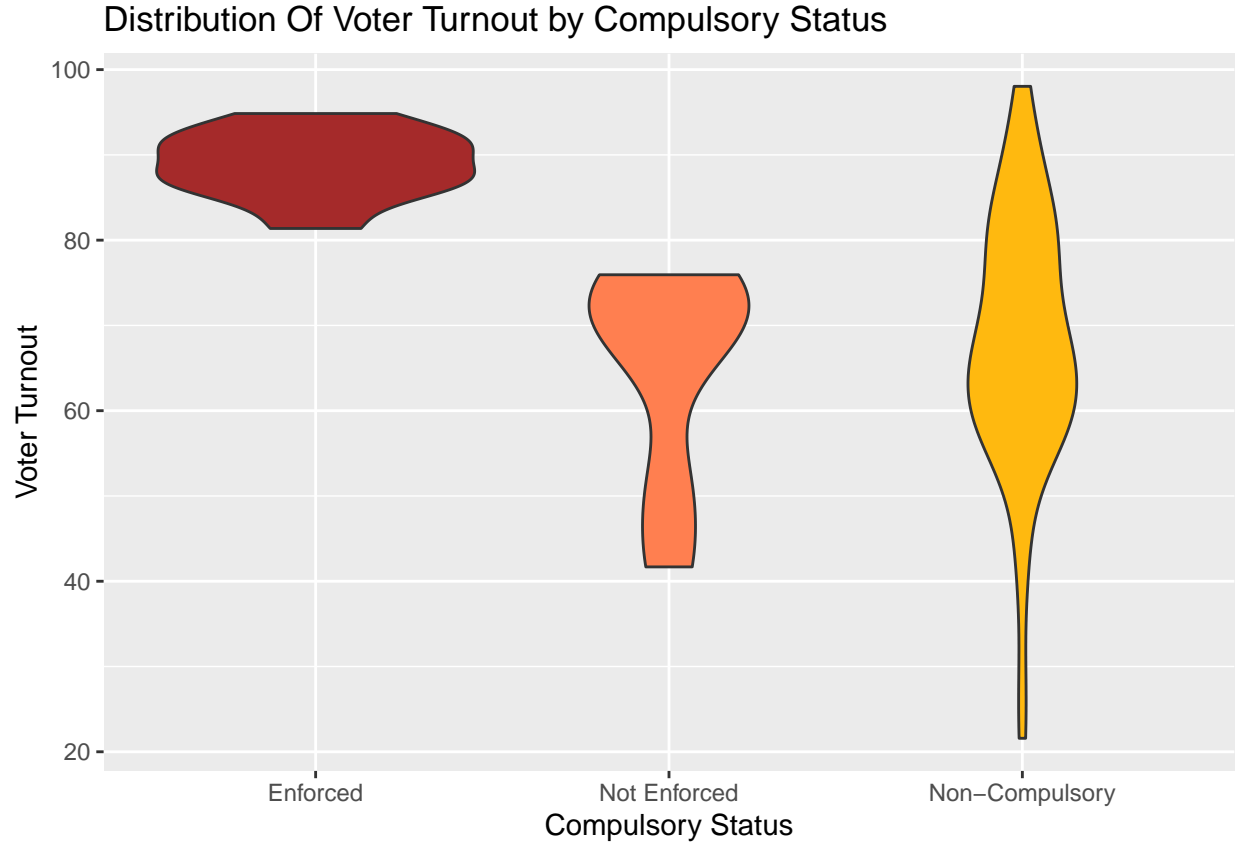Figure 6: Distribution of Turnout by Compulsory Status

Figure 7: Distibution of Turnout by Compulsory Status

5.2.3 T-Test

To conduct a t-test, the categorical variable must be a binary variable. To reduce compulsory status into a single binary value, the compulsory status variable was reduced to mandatory and non-mandatory elections. This t-test cannot speak to the effectiveness of enforcement of mandatory elections, as enforced and unenforced compulsory elections were grouped together.

This t-test in Table 10 confirms that the correlation between turnout and mandatory elections is significant (above the threshold of 95%), as the p-value is approximately 0.02. The t-test also shows that the confidence interval does not include 0, therefore the null hypothesis can be successfully rejected.

Table 10

statistic	t_df	p_value	alternative	estimate	lower_ci	upper_ci
2.47	23.5	0.0214	two.sided	10.5	1.7	19.3

5.2.4 Regression Table

Finally, the regression table at Table 11 confirms the statistical significance with the p-values below 0.05, and quantifies the strength of the relationship between compulsory status and voter turnout. Enforced elections were found to have an average 88.9% of voter turnout, in comparison with a change of -24 points

in unenforced compulsory elections and a change of -21 points in voluntary elections. This seems to indicate that mandatory elections that go unenforced are correlated with lower voter turnout than voluntary elections.

Table 11

term	estimate	std_error	statistic	p_value	lower_ci	upper_ci
intercept	88.9	4.69	19	0	79.5	98.2
compulsory: Not Enforced	-24	7.09	-3.38	0.001	-38.1	-9.83
compulsory: Non-Compulsory	-21	5.03	-4.17	0	-31	-10.9

5.3 Election Type

5.3.1 Bar Chart

The bar chart of election type in Figure 8 depicts how enforced compulsory elections are either of proportional or majoritarian type, and never of mixed type in our dataset. However the majoritarian and proportional election types seem to have a fairly even distribution in enforced, unenforced and non-compulsory elections.
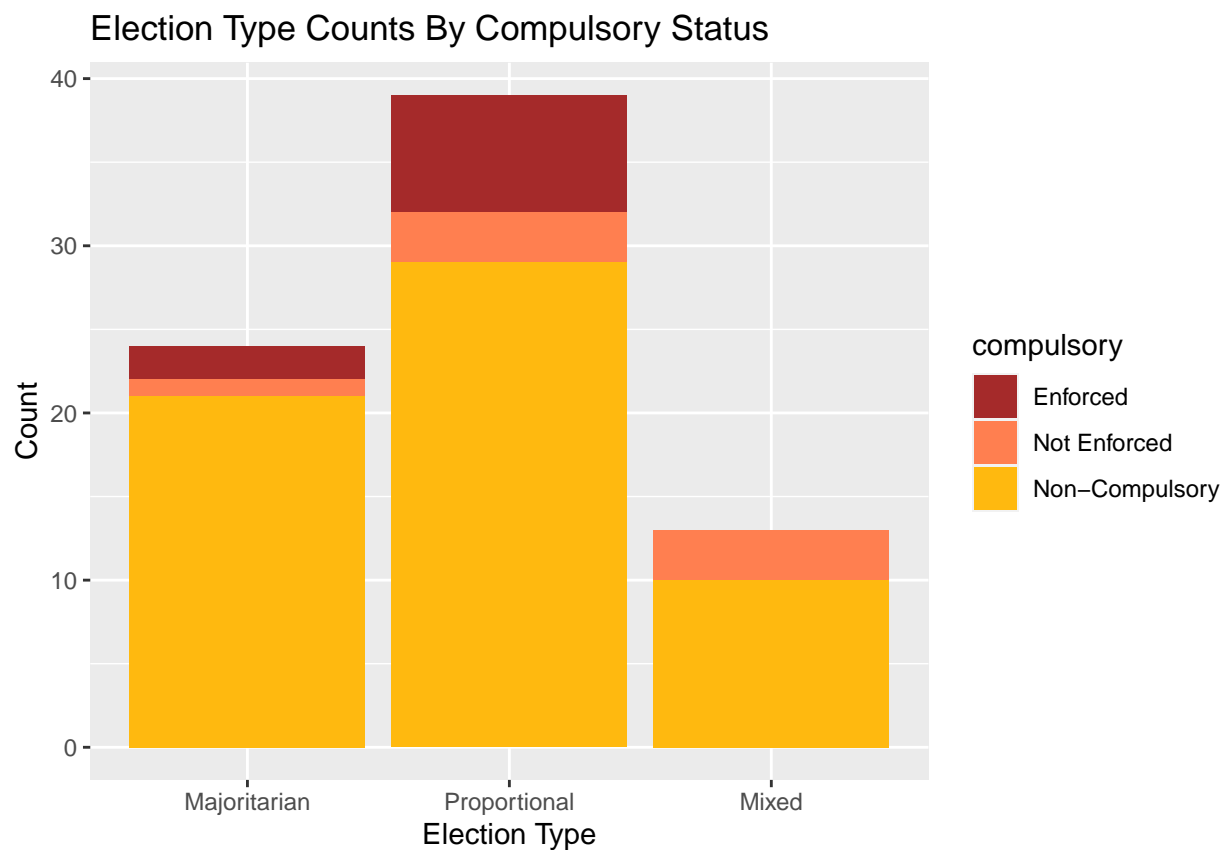

Figure 8: Election Type Counts By Compulsory Status

5.3.2 Chi Squared Test

The chi squared test in Table 12 between election type and compulsory status of elections shows that while there is a correlation between the two variables, it does not meet the threshold of statistical significance, as the p-value is 0.145. In this case, as the threshold of statistical significance is not met, and the null hypothesis cannot be successfully rejected.

Table 12

statistic	chisq_df	p_value
6.82	4	0.145

6 Interpretation

Upon completion of the bivariate analysis, there is a clear and significant correlation between voter education level and voter turnout in elections. Similarly, there is also a strong and significant correlation between an election's compulsory status and it's voter turnout.

Moreover, the hypothesis that election type is correlated with compulsory status of elections has been rejected, as the correlation was found to be insignificant.

To find the strongest predictor of voter turnout, we will use multivariate anlaysis on education level, election compulsory status and voter turnout.

6.1 Multivariate Analysis

6.1.1 Linear Regression Graph

As can be seen in Figure 9, all the countries with enforced compulsory elections are above the regression line for education. This may be an indicator that compulsory status is a strong predictor for voter turnout.
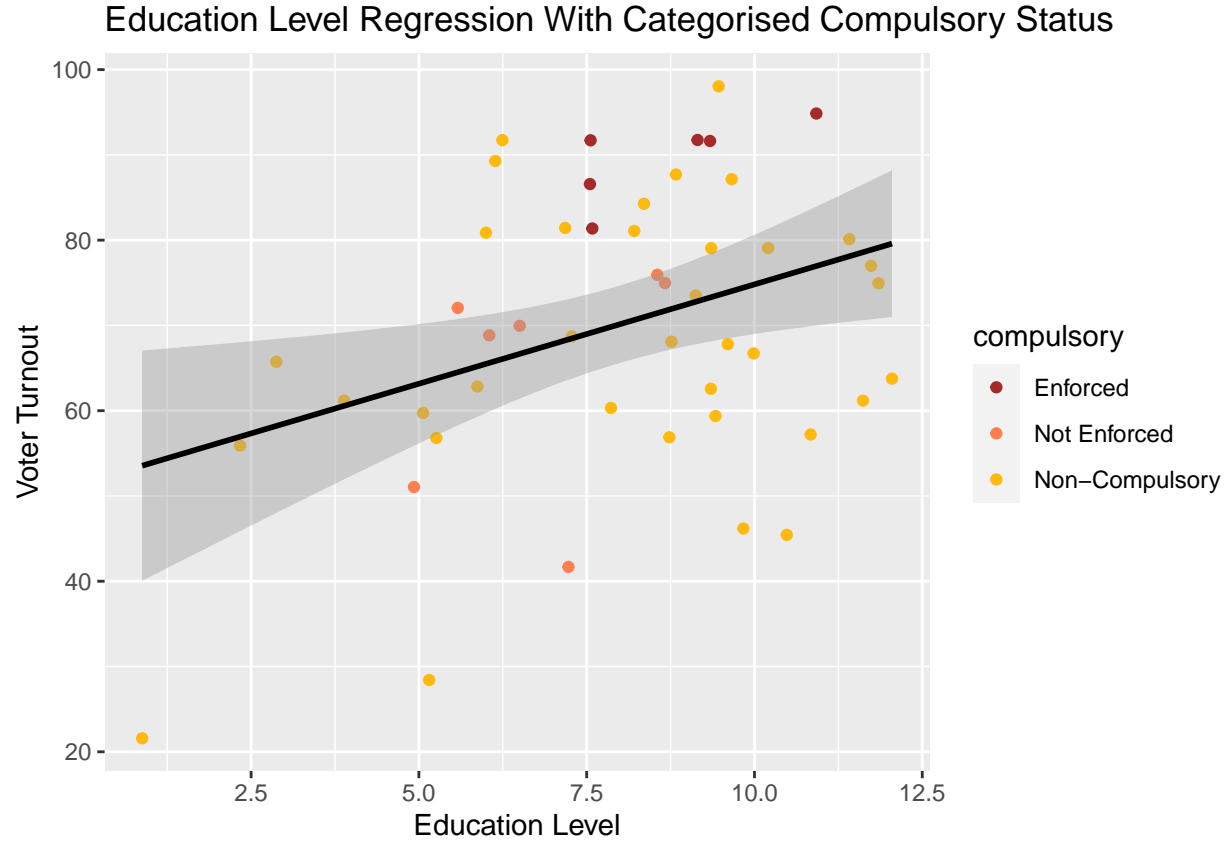
Figure 9: Education Level Regression With Categorised Compulsory Status

6.1.2 Multi-Regression Table

The Multi-Regression table in Table 13 shows that both education level and compulsory status are statistically significant predictors of voter turnout, as they all have p-values smaller than 0.05. Therefore, because the confidence intervals do not include 0, the null hypothesis for both variables can be successfully rejected.

Table 13

term	estimate	std_error	statistic	p_value	lower__ci	upper__ci
intercept	71.9	9.39	7.66	0	53	90.8
educ	2.04	0.838	2.44	0.019	0.355	3.73
compulsory: Not Enforced	-20.8	8.24	-2.53	0.015	-37.4	-4.25
compulsory: Non-Compulsory	-20.6	6.43	-3.2	0.003	-33.5	-7.64

6.1.3 Model Comparison

Table 14 compares the significance of the correlation of the independent variables with voter turnout simultaneously as well as individually. It can be seen that both education level and compulsory status remain significant predictors of voter turnout even when both are being used as independent variables. At the same

time, when both the independent variables are considered, all of their respective coefficients diminish in substantive significance as more of their predictive value is shared between themselves. Because education levels span from 0.876 to 12, an increase from the lowest recorded education level to the highest recorded education level will have a similar increase in the magnitude of voter turnout as a change from enforced compulsory elections to unenforced compulsory election or non-compulsory elections. When the education level is accounted for, the difference between coefficients of unenforced compulsory elections and voluntary elections greatly diminishes, indicating that unenforced compulsory elections have little association with increased voter turnout in comparison with voluntary elections.

Table 14: Voter Turnout Regression Results

	Model 1	Model 2	Model 3
(Intercept)	51.509***	88.880***	71.907***
	[36.517, 66.501]	[79.539, 98.221]	[52.993, 90.821]
Education	2.331*		2.043*
	[0.538, 4.123]		[0.355, 3.731]
Voting: Compulsory, Not Enforced		-23.953**	-20.845*
		[-38.075, -9.831]	[-37.445, -4.245]
Voting: Not Compulsory		-20.961***	-20.592**
		[-30.978, -10.944]	[-33.542, -7.642]
Num.Obs.	49	76	49
R2	0.127	0.202	0.292
RMSE	15.79	14.06	14.54

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

6.2 Results

Through statistical analysis, this report has proven that the measurement of education level has a similar impact on voter turnout percentage as enforced compulsory elections. In a similar fashion, this report has shown that both education level and the compulsory status of an election are both statistically significant factors in the prediction of voter turnout. This analysis has also shown that the election type variable is not a statistically significant predictor of the compulsory status of an election.

7 Conclusion

To conclude, the hypothesis was correct in stating that voter turnout is positively correlated with education level, as higher education rates are associated with higher voter participation. Similarly, the hypothesis was correct in stating that voter turnout can be predicted by the compulsory status of an election, although unenforced mandatory elections are associated with similar voter turnout rates as voluntary elections when controlling for education levels. Therefore, only enforced compulsory elections have a substantive correlation with increased voter turnout in comparison with unenforced compulsory elections. Finally, the hypothesis

that the compulsory status of an election is correlated with the type of election was not substantiated, as the correlation was not found to be statistically significant.

To increase voter turnout in Canadian elections, this report has shown that enforcing mandatory elections or increasing education levels are both significantly associated with increased election participation. To avoid impinging on the freedoms of Canadians, I would recommend that elections remain voluntary, and that the Canadian government should focus it's resources on improving the average education level of Canadians to raise levels of voter turnout.