Sample of Final Project

AUTHOR
Nathan Valle

```
Turnout <- read.csv("C:/Users/Nathan Valle/Downloads/Turnout.csv")
library(ggplot2)
library(stargazer)
```

```
Please cite as:

 Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.

 R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```
suppressWarnings(library(car))
```

```
Loading required package: carData
```

Below is a sample of a research project that I completed during the 2023 Fall Semester at SFU. The prompt that was given to us was that we were hired by the Canadian government to analyze different factors of voter turnout with specific attention given to Canada. This sample highlights my writing and communication abilities, as well as critical thinking and issue analysis.

```
Figure.10 <- with(Turnout[!is.na(Turnout$TURNOUT)&!is.na(Turnout$CORRUPTION),], lm(TURNOUT~CORR
Figure.11 <- with(Turnout[!is.na(Turnout$TURNOUT)&!is.na(Turnout$CORRUPTION),], lm(TURNOUT~CORR
Figure.12 <- with(Turnout[!is.na(Turnout$TURNOUT),], lm(TURNOUT~COMPULSORY))
Figure.13 <- with(Turnout[!is.na(Turnout$TURNOUT),], lm(TURNOUT~DEMOCRACY))
Figure.14 <- with(Turnout[!is.na(Turnout$TURNOUT),], lm(TURNOUT~ENPP))
```

```
stargazer(Figure.10, Figure.11, Figure.12, Figure.13, Figure.14, type="html",
          digits=2,
          covariate.labels=c("Corruption Level","Compulsory Voting", "Democracy", "Non-Democrat
          title = "Figure 10: Predicting Voter Turnout",
          dep.var.labels=c("Voter Turnout"))
```

Figure 10: Predicting Voter Turnout

	Dependent variable:				
	Voter Turnout				
	(1)	(2)	(3)	(4)	(5)
Corruption Level	-0.20**	-0.22***			
	(0.08)	(0.07)			

	(0.08)	(0.07)			
Compulsory Voting	-21.31***		-23.13***		
	(5.02)		(4.61)		
Democracy	-25.46***		-28.74***		
	(6.77)		(6.39)		
Non-Democratic Countries	-2.37			2.51	
	(2.93)			(2.68)	
Effective Number of Political Parties					0.33
					(0.73)
Intercept	97.02***	76.55***	86.16***	62.69***	64.23***
	(6.96)	(4.09)	(4.42)	(2.17)	(2.91)
Observations	166	166	179	179	108
R^2	0.16	0.06	0.14	0.005	0.002
Adjusted R^2	0.14	0.05	0.13	-0.001	-0.01
Residual Std. Error	15.87 (df = 161)	16.66 (df = 164)	15.95 (df = 176)	17.07 (df = 177)	14.71 (df = 106)
F Statistic	7.70*** (df = 4; 161)	10.12*** (df = 1; 164)	13.81*** (df = 2; 176)	0.88 (df = 1; 177)	0.21 (df = 1; 106)

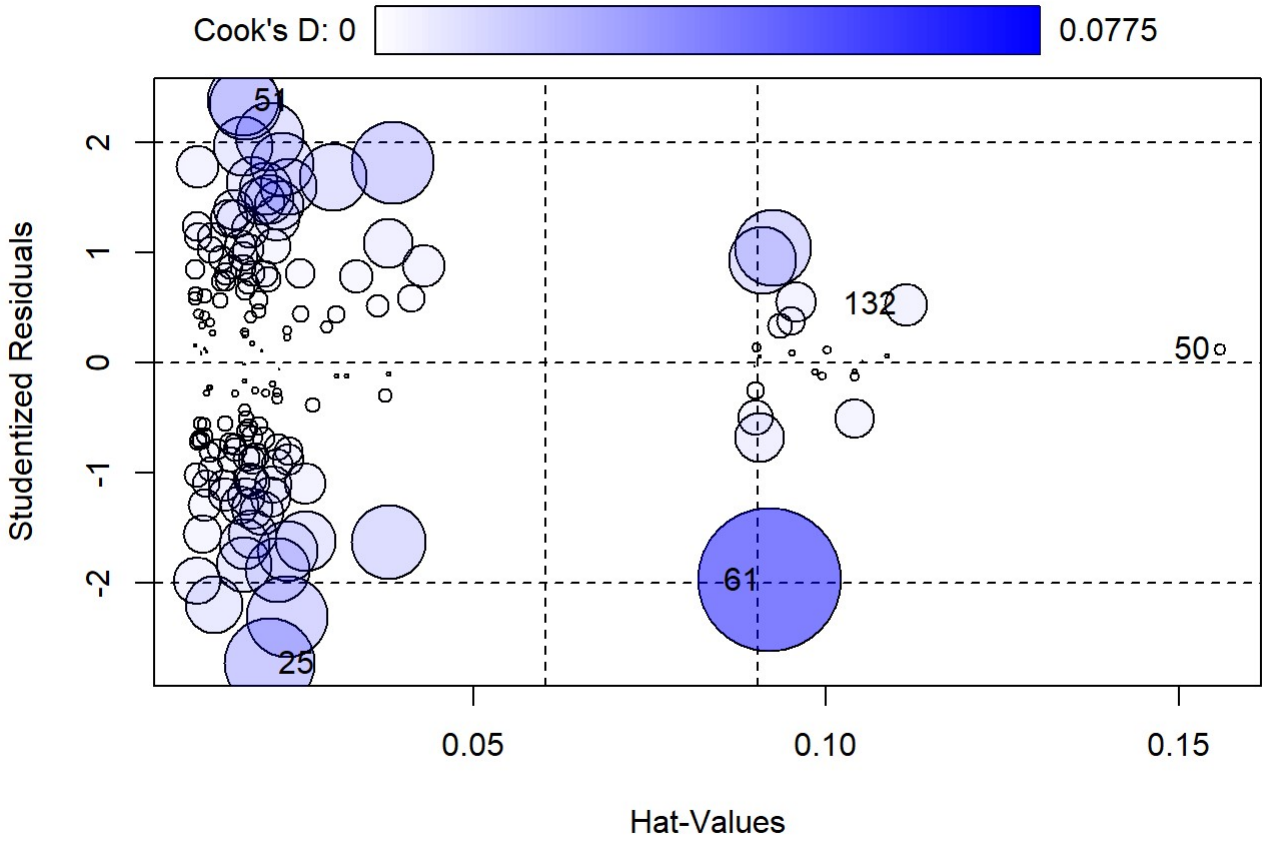
Note: *p<0.1;* ***p<0.05;*** p<0.01

Above shows that based on the regression output, the variables with positive effects on voter turnout are enforced compulsory voting, and non-democratic societies. Their effects are 21.313, and 2.3673 increase on voter turnout, respectively. The other variables, that of not enforced compulsory voting, as well as corruption, have negative effects on voter turnout. A 1 unit increase in corruption corresponds to a -0.199 change in voter turnout; countries that do not enforce voter turnout see a -4.145 change in voter turnout. Countries that are not democratic, however, see a positive increase in voter turnout, a 2.367 change. Using the hypotheses put forth earlier, we can reject the null hypothesis for corruption, and still conclude that the alternative hypothesis is correct, that corruption has a negative affect on voter turnout. We can also reject the null hypothesis for compulsory voting, and conclude that the alternative hypothesis was also correct in stating that compulsory voting has a positive effect on voter turnout. Comparing the 2 categories in the regression model, those that enforced voter turnout see a greater increase in voter turnout than those that do not.

Looking at the adjusted r-squared value (0.14), about 14% of the variance is explained by the model. What this tells us is that the 3 variables chosen are actually not good predictors of voter turnout when used together. I ran the tests again, but with using only 1 independent variable as a predictor of voter turnout. Looking solely at the adjusted r-squared value, the variables that explain the most variance by rank are compulsory voting (explaining 12.6% of the variance) and the corruption level (5.23% of the variance). The final variable, democracy, has a negative adjusted r-squared, meaning that it is difficult to fit a model. To test its fit as well, I used the earlier discarded variable of ENPP. It also returned a negative r-squared value. One thing to take note is that in all cases, Canada is in the majority of all the data sets,

meaning that these variables will not be good predictors for Canada's voter turnout as well. However, there could be many influence points that are skewing the data.

```
mod.1 <- with(Turnout[!is.na(Turnout$TURNOUT)&!is.na(Turnout$CORRUPTION),], lm(TURNOUT~CORRUPTI
my.res <- as.data.frame(rstudent(mod.1))
Figure.15 <- influencePlot(mod.1)
```




```
Figure.15
```

```
       StudRes        Hat        CookD
25  -2.7280828 0.02106267 0.0307938452
50   0.1205112 0.15582759 0.0005394666
51   2.3755102 0.01742413 0.0194527658
61  -1.9741314 0.09196807 0.0775483185
132  0.5272223 0.11135845 0.0069978735
```

Figure 15 is an influence plot of all the dependent variable values from the linear model of figure 10. The x-axis represents a point's leverage on the regression model, while the y-axis is the studentized residual (observation's distance from predicted value if removed from analysis). The farther along the x-axis a plot is, the more influence it has on the model; the higher or lower a plot is on the y-axis, is its studentized residual; the Cook's Distance at the very top of the plot shows the influence of an observation, the darker the bubble, the more influence it has. There are a lot of plots that have a

observation, the darker the bubble, the more influence it has. There are a lot of plots that have a studentized residual value of greater than 1, and a number of points that have a great amount of leverage on the regression equation. There are points that have a large amount of influence, while also being an outlier. A majority of the data is concentrated are over +1 and -1 studentized residual in the regression model. It is unfeasible to remove all these points from the regression equation because then there would be a small number of variables considering how much would be removed. Even though there are some plots whose residuals are close to 0, they have a lot of influence on the regression line.

What does this mean for voter turnout in the world as well as for Canada? The independent variables used for the analyses are very weak predictors of voter turnout. The analyses show that the variables have a weak relationship to voter turnout, as well as being poor predictors. Corruption levels are poor predictors of voter turnout, having a very weak relationship; the same can be said for democracies, and whether or a not a democratic society affects turnout. However, the predictor with the most effect (albeit weak) on voter turnout is compulsory voting. Looking back to the box plot, there is a stark difference between the countries that enforce it, and those that do not have it, and do not enforce it. Statistically speaking, countries that have enforce compulsory voting see a greater turnout in their elections. All the variables tested in this report are either not related to voter turnout at all, or there exists a weak relationship that if implemented or changed, would not help with Canada's level of voter turnout.

The findings also suggest that it is hard to pinpoint what exactly effects voter turnout. The variables that were tested have a weak relationship, as already stated above. Perhaps there are other variables not listed here that effect voter turnout, such as one's interest in politics. It is extremely difficult to find specific variables that effect voter turnout, in the world and in Canada.

My recommendation is that Canada should implement some type of compulsory voting if the government wants to see an increase in voter turnout. Lowering the already low corruption level of Canada will not provide any type of increase, nor will the effective number of parties, or removing Canada's democracy. If the goal is to see an increase in voter turnout, then enforced compulsory voting should be implemented. Looking again at the previous analyses, although there are less countries that enforce voter turnout, they do see a significant increase in voter turnout. If Canada were to implement this type of voting, it could also bolster voter turnout.