Assignment 5 Question 1

Jack Ogle in collaboration with Eva Haque, Matt Lohrs, and Jack Knickrehm

(a)

The assumptions underlying the RDD in this paper are that in order for the school to attain autonomy (GM) there must be a 50% vote from the community to have the school be GM. Clark also assumes that the school performance, test scores, is increasing in autonomy and school effort. Schools that are already GM at the start of a period are fully autonomous and therefore decide only how much effort to exert. Effort in turn improves school performance, test scores for example, but effort is costly. Schools that are not GM at the start period must decided how much effort to exert and whether or not to become Gm. For given effort, non GM schools performance is assumed lower than GM school performance; hence schools have an incentive to become GM. There are cost associated with GM status and the decision to become one is non trivial.

The conceptual framework assumed that schools were identical. In practice, schools differ along many dimensions, and certain types of school may be more likely to hold and win a GM vote (e.g., those with more entrepreneurial head teachers). Clark's empirical approach overcomes this selection problem by focusing on the jump in performance among schools at the 50 percent win threshold. Specifically, Clark considers variants of the fuzzy regression discontinuity model for school i voting on GM

(b)

These figures represent figure 8 from the paper. They are all visualizing:

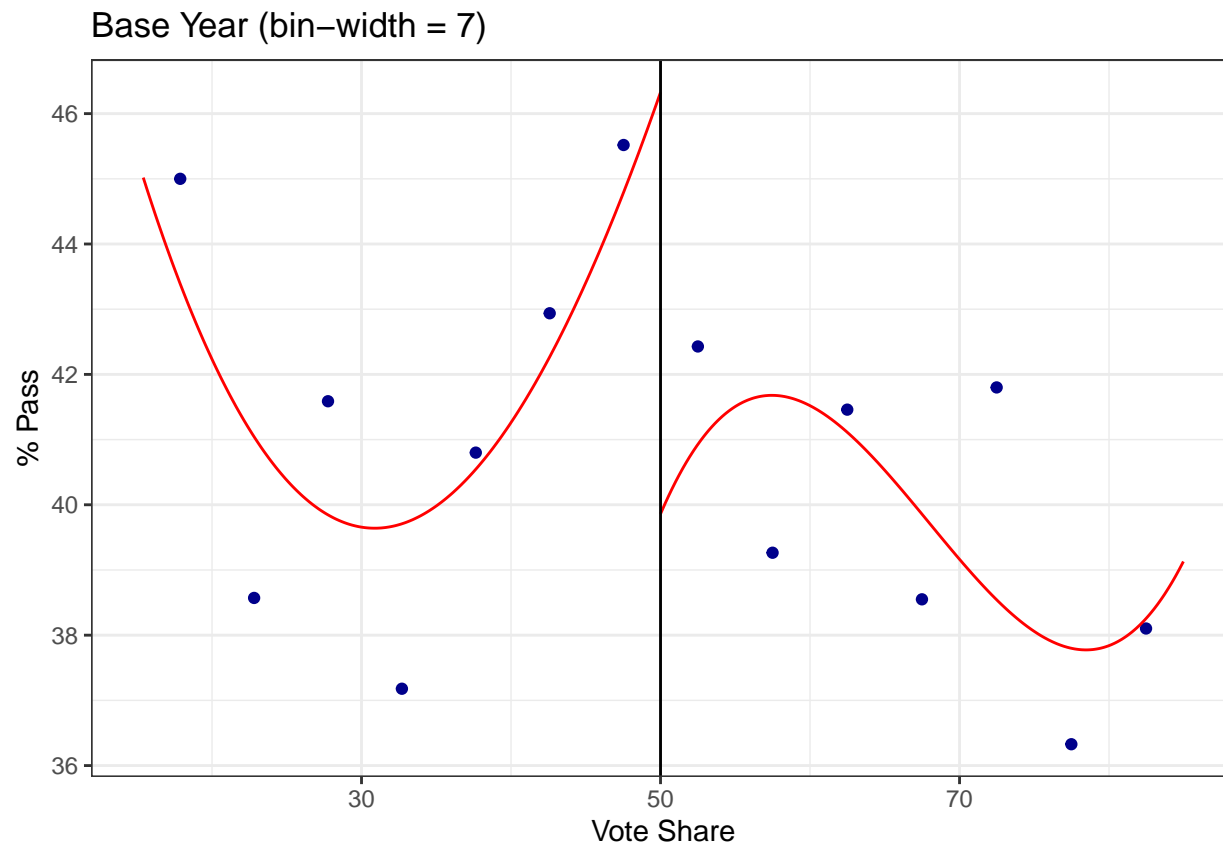The Impact of GM Status on Schools that Become Grant-Maintained

```r
# Restricting Data to (15,85)
restrict = subset(damon, vote <= 85 & vote >= 15)

# calculating the % change
restrict$percentage_change = restrict$passrate2 - restrict$passrate0

# Base Year
rdplot(y = restrict$passrate0, x = restrict$vote , c = 50, p = 3, nbins = 7, title = "Base Year (bin-width = 7)", x.l
```
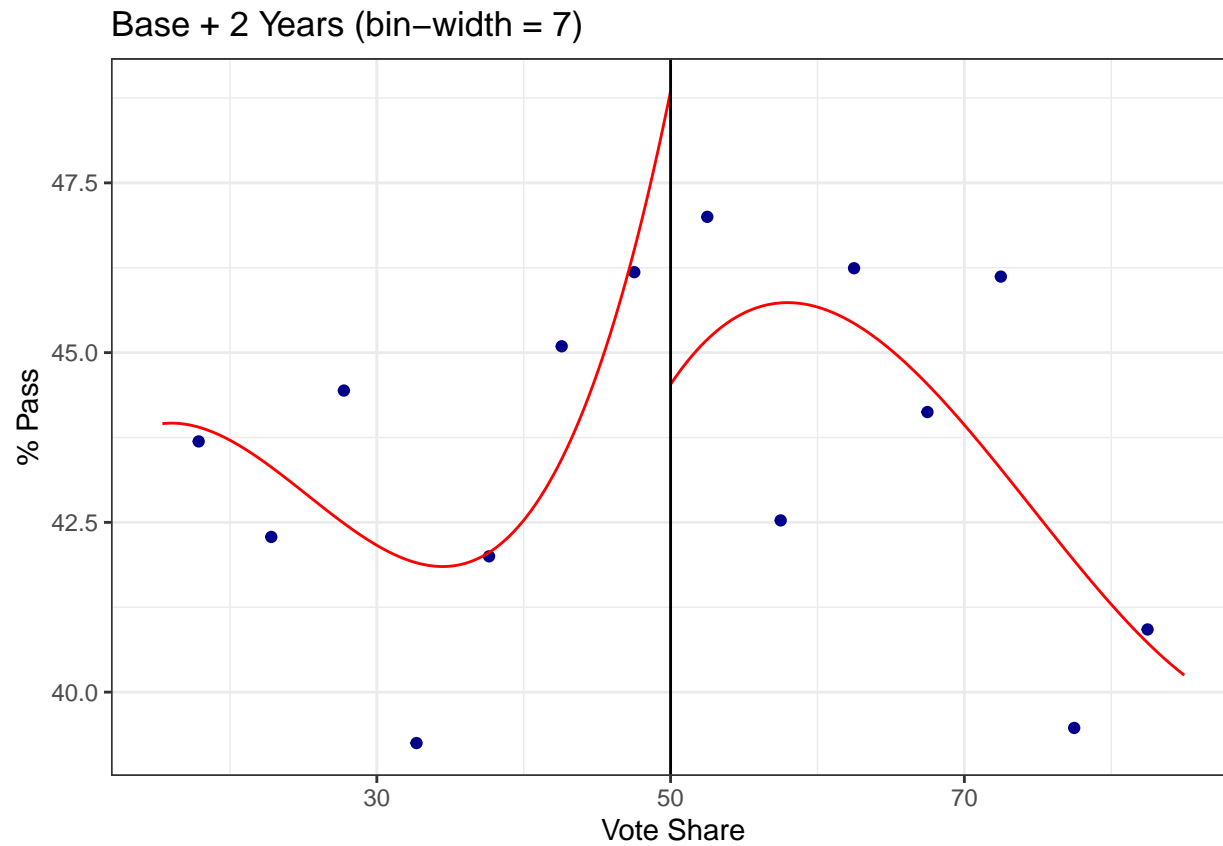
Base Year (bin−width = 7)

```
# Base +2 Year
rdplot(y = restrict$passrate2, x = restrict$vote, c = 50, p = 3, nbins = 7, title = "Base + 2 Years (bin-width = 7)",
```

Base + 2 Years (bin−width = 7)


```
# % change
rdplot(y = restrict$percentage_change, x = restrict$vote, p = 1, nbins = 7, c = 50, title = "Change (bin-width = 7)",
```

Change (bin−width = 7)


```r
# Base Year
rdplot(y = restrict$passrate0, x = restrict$vote , c = 50, p = 3, nbins = 2, title = "Base Year (bin-width = 2)", x.l
```
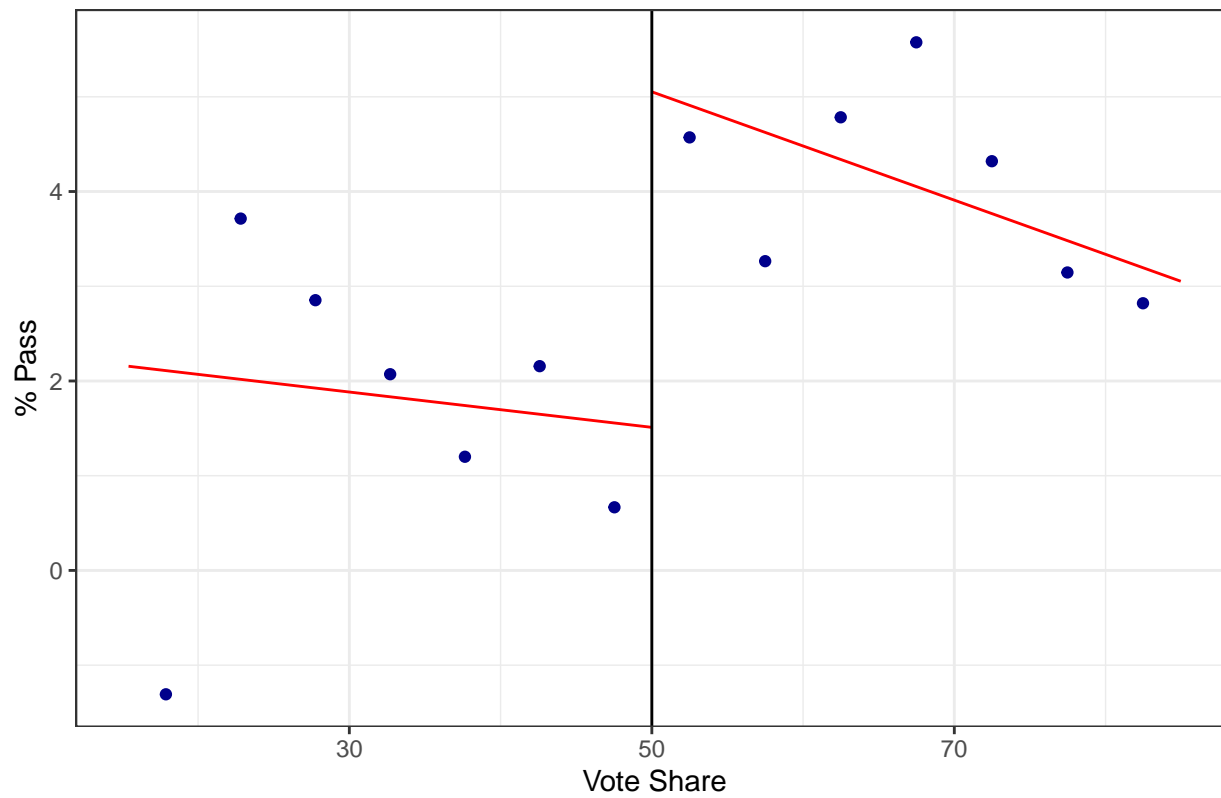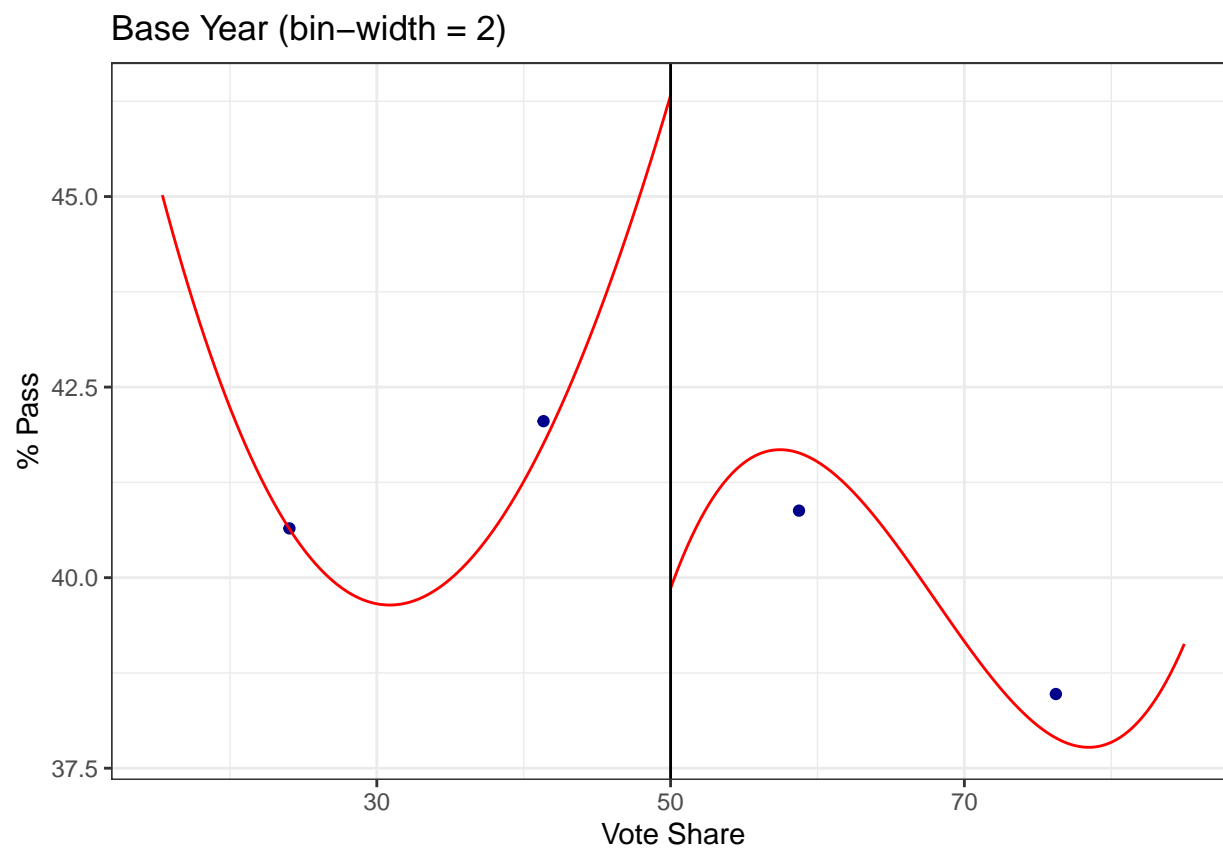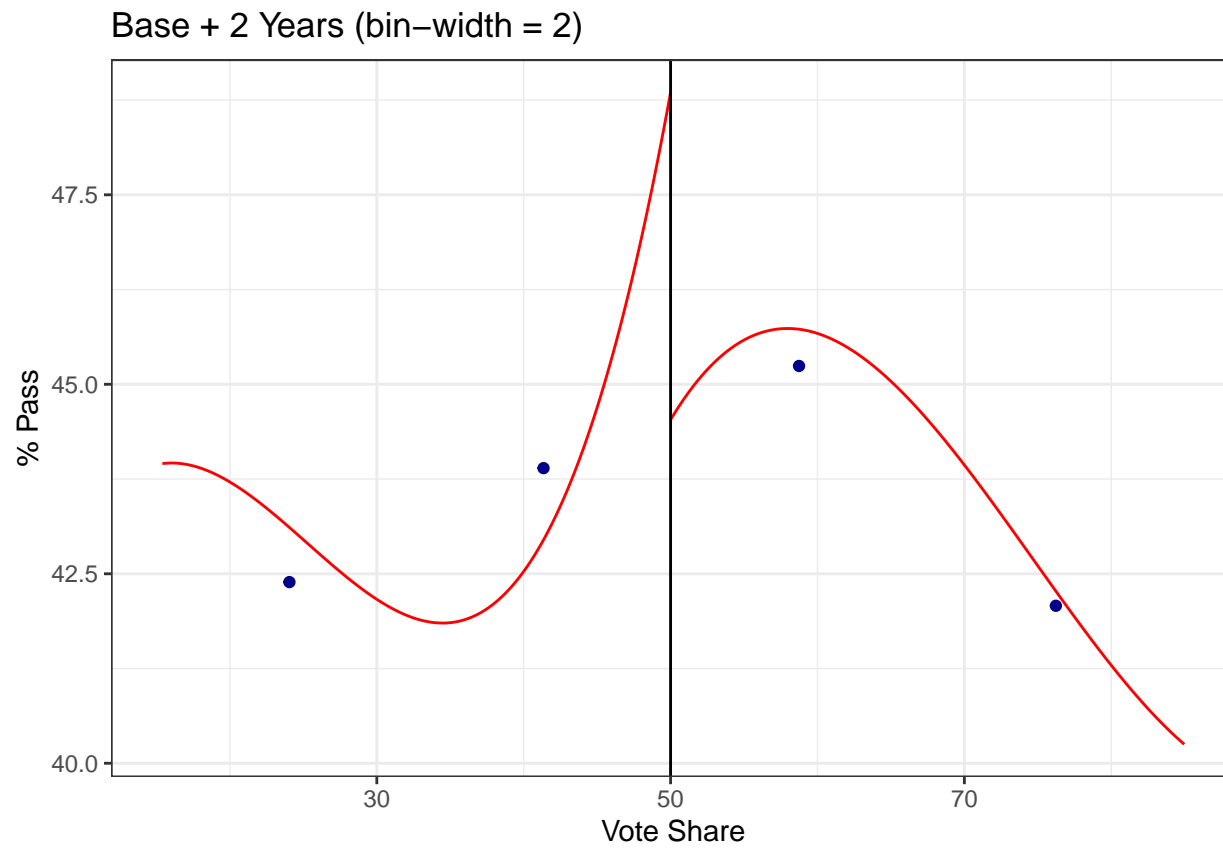
Base Year (bin−width = 2)
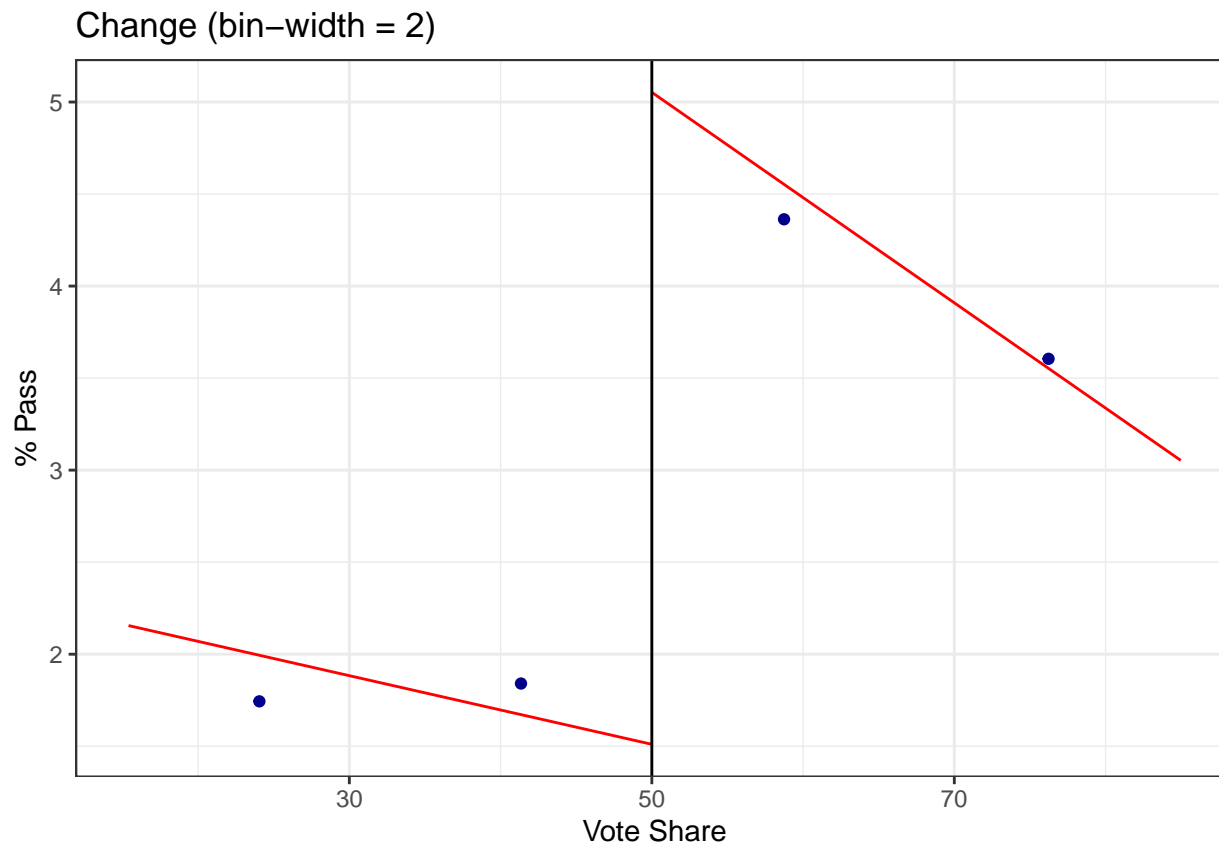
```
# Base +2 Year
rdplot(y = restrict$passrate2, x = restrict$vote, c = 50, p = 3, nbins = 2, title = "Base + 2 Years (bin-width = 2)",
```

Base + 2 Years (bin−width = 2)


```
# % change
rdplot(y = restrict$percentage_change, x = restrict$vote, p = 1, nbins = 2, c = 50, title = "Change (bin-width = 2)",
```

Change (bin–width = 2)

From the graphs above we can see that as you decrease the bin size from 7 which is what Clark uses in the paper for Figure 8. We can see that there is no difference in terms of the lines of best fit, however, we can better visualize the data with a higher bin width. We see more of the data points.

(c)

Clark restricts his sample to those schools with votes from 15% and 85% because he wants to reduce bias from outliers from schools who were very opposed or very for autonomy. Additionally, RD necessitates that in order to estimate the average effect of a treatment you look at an arbitrary cutoff and evaluate the treatment effects before and after the cutoff. Clark needs to

estimate around the cutoff, 50, which is inside the (15,85) interval. Other columns have functions of vote share because vote share is the forcing variable in this Fuzzy Regression Discontinuity model. Fuzzy RD exploits discontinuities in the probability of treatment conditional on a variate. The discontinuity, here is Vote share, and it becomes the IV for the treatment status. Which is why Table 3a includes functions of vote share both on their own and interacted with the win/lose variable.

(d)

```
model_d_1 <- lm(vote ~ win + vote + vote*lose + vote*win + (vote^2)*lose + (vote^2)*win + percentage_change, data = r
```

```
## Warning in model.matrix.default(mt, mf, contrasts): the response appeared on the
## right-hand side and was dropped
```

```
## Warning in model.matrix.default(mt, mf, contrasts): problem with term 2 in
## model.matrix: no columns are assigned
```

```
stargazer(model_d_1, column.labels=c("Non-Grammar Schools with Vote Shares in [15,85] interval"), title="Regression R
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
## % Date and time: Sun, Feb 20, 2022 - 18:19:54
## % Requires LaTeX packages: dcolumn
## \begin{table}[!htbp] \centering
##   \caption{Regression Results (d)}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lD{.}{.}{-3} }
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{1}{c}{\textit{Dependent variable:}} \\
## \cline{2-2}
## \\[-1.8ex] & \multicolumn{1}{c}{vote} \\
```

```
##  & \multicolumn{1}{c}{Non-Grammar Schools with Vote Shares in [15,85] interval} \\
## \hline \\[-1.8ex]
##  win & -0.000^{***} \\
##    & (0.000) \\
##    & \\
##  lose &  \\
##    &  \\
##    & \\
##  percentage\_change & 0.000 \\
##    & (0.000) \\
##    & \\
##  vote:lose & 1.000^{***} \\
##    & (0.000) \\
##    & \\
##  vote:win & 1.000^{***} \\
##    & (0.000) \\
##    & \\
##  Constant & -0.000^{***} \\
##    & (0.000) \\
##    & \\
## \hline \\[-1.8ex]
## Observations & \multicolumn{1}{c}{524} \\
## R$^{2}$ & \multicolumn{1}{c}{1.000} \\
## Adjusted R$^{2}$ & \multicolumn{1}{c}{1.000} \\
## Residual Std. Error & \multicolumn{1}{c}{0.000 (df = 519)} \\
## F Statistic & \multicolumn{1}{c}{34,064,124,303,065,597,294,021,672,173,568.000$^{***}$ (df = 4; 519)} \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{l}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
##  & \multicolumn{1}{l}{Standard errors in parentheses} \\
## \end{tabular}
## \end{table}
```

Table 1: Regression Results (d)

	Dependent variable:
	vote
	Non-Grammar Schools with Vote Shares in [15,85] interval
win	−0.000***
	(0.000)
lose	
percentage_change	0.000
	(0.000)
vote:lose	1.000***
	(0.000)
vote:win	1.000***
	(0.000)
Constant	−0.000***
	(0.000)
Observations	524
R^2	1.000
Adjusted R^2	1.000
Residual Std. Error	0.000 (df = 519)
F Statistic	34,064,124,303,065,597,294,021,672,173,568.000*** (df = 4; 519)

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors in parentheses

(e)

(f)

(g)

(h)

(i)

(j)