Milestone 6

Molly Chiang

4/3/2020

Grapic

Below is a beautiful graphic which uses the data from Marshall 2015 and is based on Figure 3 in the paper. I used inspiration/guidance from (Gary King 2000) to create it.

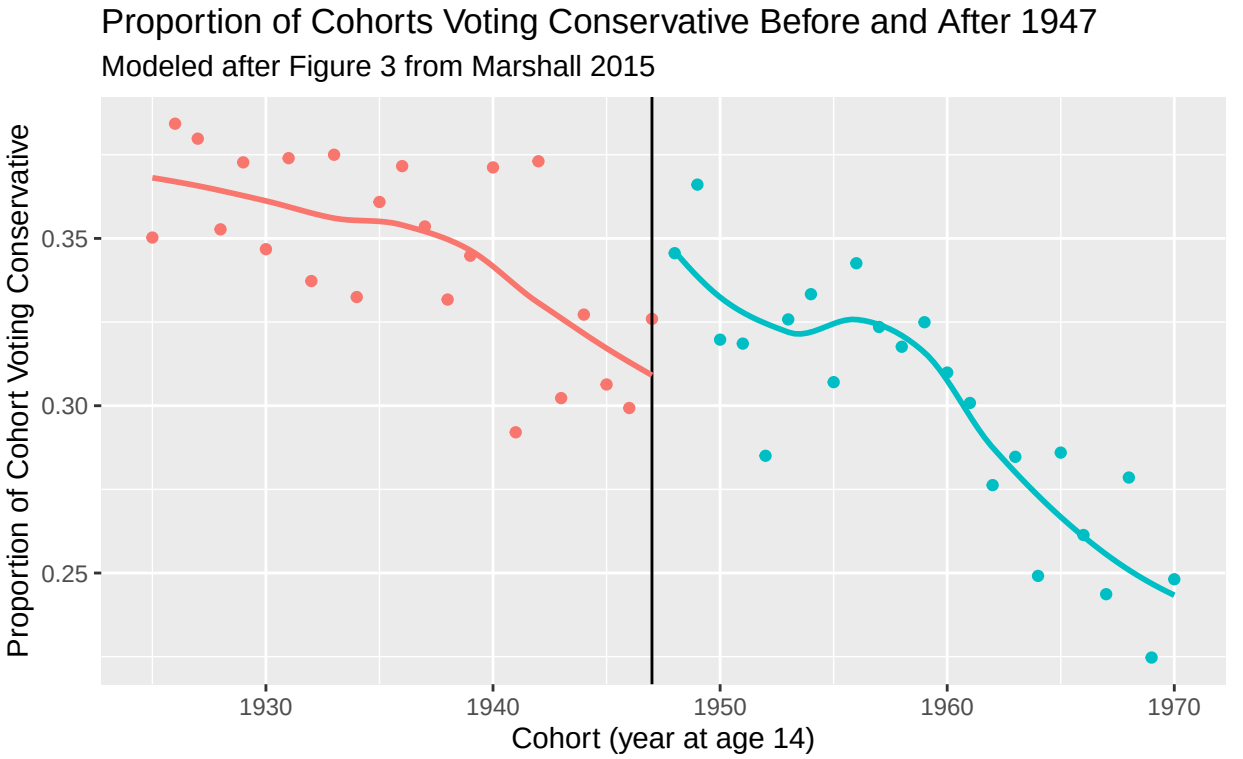


Figure 1: This figure illustrates the jump in proportion of a cohort, or class, of British individuals (classified by which they were 14), after the education reform in 1947 which increased the year at which you could legally school from 14 to 15.

Overview of Replication Paper

This paper by John Marshall uses data from the results of the 1947 high school leaving age reform in Great Britain, to analyze how additional years in high school affect political preferences (Marshall 2015). This paper looked specifically in how additional years in high school effected voting for the Conservative Party (Marshall 2015). In 1947, Great Britain changed the high school leaving age from 14 to 15, this induced almost half the student population to stay in school for at least 1 or 2 more years (Marshall 2015). Data from the 10 British elections between 1947 and 2010 was then used to compare voters young enough to have been effected by the reform to those who were too old to have been effected, using regression discontinuity (Marshall 2015). Regression discontinuity is usually used for determining if a program/treatment is effective, and essentially is is a pretest-posttest program-comparison group design strategy (Trochim 2020). Regression discontinuity is unique in that individuals are assigned to one of two groups, just based on if they are on either side of a pre-determined cut-off (Trochim 2020).The results of data analysis revealed staying in high

school for longer substantially increased likelihood to vote for the Conservative Party (staying one extra year increased probailitiy of voting Conservative by almost 12 percentage points) (Marshall 2015). This supports the previously studied fact that high school is extremely pertinent to political opinions later in life (Marshall 2015). In addition, this significant finding indicates the education reform of 1947 may have had an even greater affect on politics and election results nationwide over many years than ever expected (Marshall 2015).

All analysis for this paper is available in my github repo.[1].

Extension

500 words about your proposed extension. You do not have to have done the extension yet. (That comes next week.) But it is time to start thinking about what your contribution to human knowledge will be. You seek admission to the School of Athens. What do you have to offer us?

Replication Process

I was able to replicate nearly all of the figures in the paper in stata. The code I ran and the output figures are included in the appendix. In terms of the tables, I was able to run all of the code in stata almost directly from Marshall's replication code (some aspects I had to change because some of the functions he used have been updated since he wrote the paper). However, I had a lot of trouble going from the raw results of the models Marshall ran to the polished tables I saw in his paper, as only code for the models were included in his .do file. Thus, I have included screenshots of the raw code and the relevant aspects of the output that would be included in the table in the paper. I hope to learn more basics of stata in order to compile and nicely display these raw outputs from Marshall's code.

For brevity, I did only included the screenshots from tables 1 and 2 and not table 3 (as it would be a lot of seperate screenshots to include), but I have included the table 3 code, which was run in the same way. Table 4 was not able to be replicated as there was no code for table 4 included in Marshall's replication code.

Most coefficients exactly or almost exactly matched Marshall's published results. A few were a few decimal places off (potentially due to different iterations of the regression), and a even fewer were quite different, which still needs to be investigated.

This paper relies heavily on rdrobust models, and the reference material for learning about this function and its package is from (Sebastian Calonico 2020).

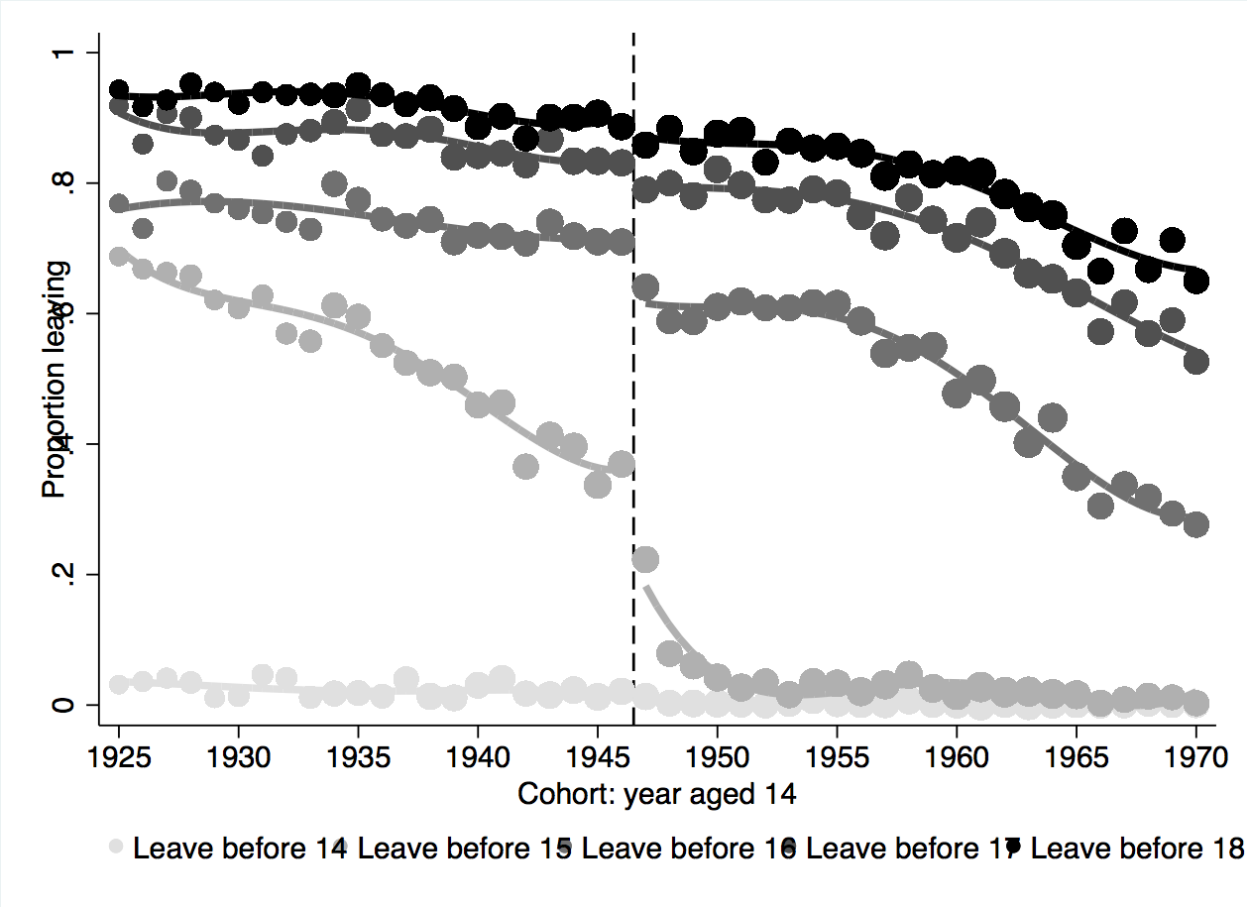


Figure 1: Marshall Figure 1

```
*** Figure 1: Trends in school leaving age
twoway (lpoly leave_l8 yearat14 if yearat14<1947 & yearat14>=1925, lcolor(gs14) clwidth(thick) degree(4)) ///
  (lpoly leave_l8 yearat14 if yearat14>=1947 & yearat14<=1970, lcolor(gs14) clwidth(thick) degree(4)) ///
  (scatter meanleave_l8 yearat14 if yearat14>=1925 & yearat14<=1970 [weight=weight_14], msize(small) mcolor(gs14)) ///
  (lpoly leave_l9 yearat14 if yearat14<1947 & yearat14>=1925, lcolor(gs11) clwidth(thick) degree(4)) ///
  (lpoly leave_l9 yearat14 if yearat14>=1947 & yearat14<=1970, lcolor(gs11) clwidth(thick) degree(4)) ///
  (scatter meanleave_l9 yearat14 if yearat14>=1925 & yearat14<=1970 [weight=weight_14], msize(small) mcolor(gs11)) ///
  (lpoly leave_l10 yearat14 if yearat14<1947 & yearat14>=1925, lcolor(gs7) clwidth(thick) degree(4)) ///
  (lpoly leave_l10 yearat14 if yearat14>=1947 & yearat14<=1970, lcolor(gs7) clwidth(thick) degree(4)) ///
  (scatter meanleave_l10 yearat14 if yearat14>=1925 & yearat14<=1970 [weight=weight_14], msize(small) mcolor(gs7)) ///
  (lpoly leave_l11 yearat14 if yearat14<1947 & yearat14>=1925, lcolor(gs5) clwidth(thick) degree(4)) ///
  (lpoly leave_l11 yearat14 if yearat14>=1947 & yearat14<=1970, lcolor(gs5) clwidth(thick) degree(4)) ///
  (scatter meanleave_l11 yearat14 if yearat14>=1925 & yearat14<=1970 [weight=weight_14], msize(small) mcolor(gs5)) ///
  (lpoly leave_l12 yearat14 if yearat14<1947 & yearat14>=1925, lcolor(black) clwidth(thick) degree(4)) ///
  (lpoly leave_l12 yearat14 if yearat14>=1947 & yearat14<=1970, lcolor(black) clwidth(thick) degree(4)) ///
  (scatter meanleave_l12 yearat14 if yearat14>=1925 & yearat14<=1970 [weight=weight_14], msize(small) mcolor(black)), ///
  graphregion(fcolor(white) lcolor(white)) ylab(,nogrid) ytitle(Proportion leaving) xtitle(Cohort: year aged 14) xline(1946.5,
lcolor(black) lpattern(dash)) xlab(1925(5)1970) ///
  legend(nobox region(fcolor(white) margin(zero) lcolor(white)) lab(3 "Leave before 14") lab(6 "Leave before 15") lab(9 "Leave
before 16") lab(12 "Leave before 17") lab(15 "Leave before 18") order(3 6 9 12 15) row(1))
```

Figure 2: Marshall Figure 1 code

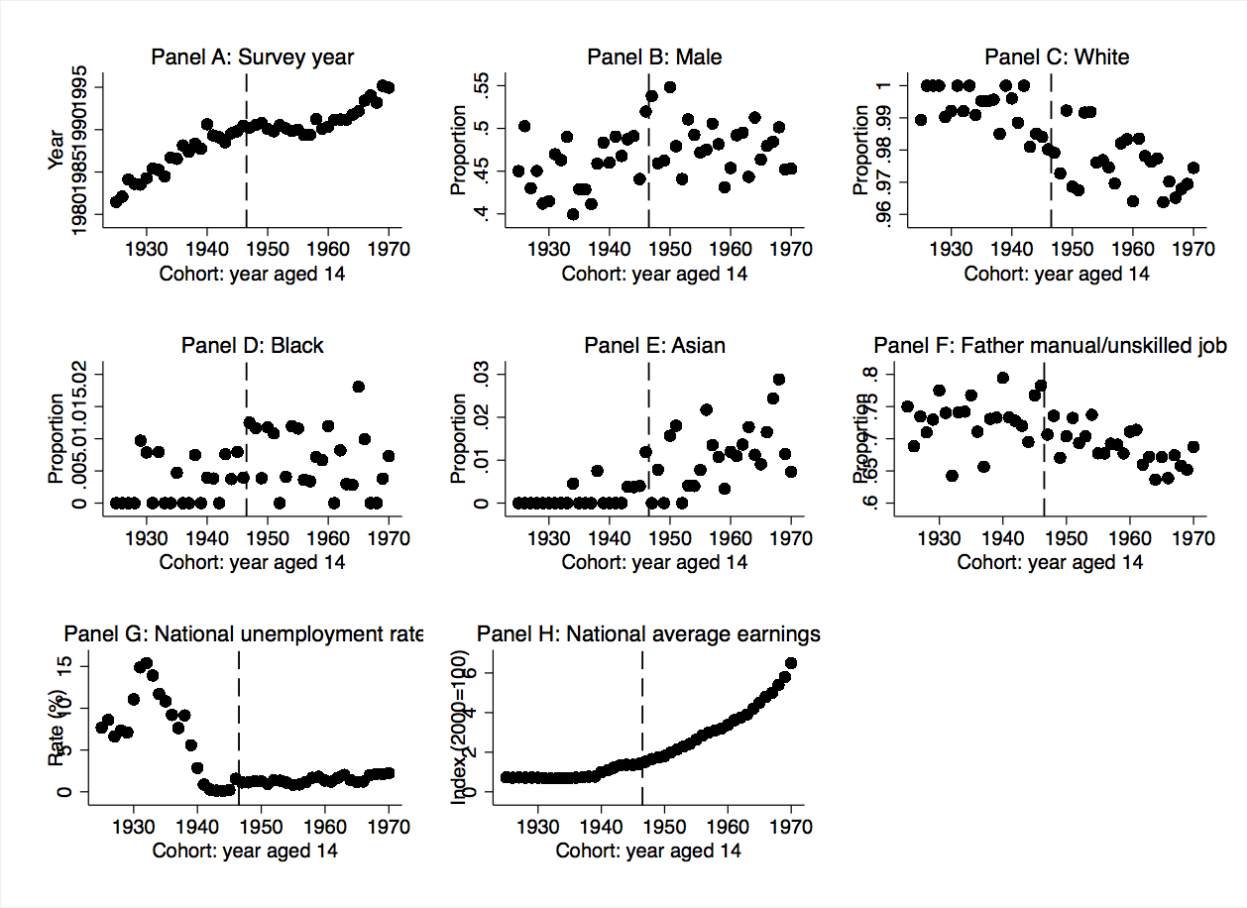


Figure 3: Marshall Figure 2

```
*** Figure 2: Continuity graphs
capture by yearat14, sort : egen meanyear = mean(year)
twoway (scatter meanyear yearat14 if yearat14>=1925 & yearat14<=1970, mcolor(black) msize(medsmall)), ///
  graphregion(fcolor(white) lcolor(white)) ylab(,nogrid) ytitle(Year) xtitle(Cohort: year aged 14) xline(1946.5, lcolor(black) lpattern(dash)) ///
  legend(off) title(Panel A: Survey year, color(black) size(medium)) xlab(1930[10]1970)
graph save Graph "g1.gph", replace

capture by yearat14, sort : egen meanmale = mean(male)
twoway (scatter meanmale yearat14 if yearat14>=1925 & yearat14<=1970, mcolor(black) msize(medsmall)), ///
  graphregion(fcolor(white) lcolor(white)) ylab(,nogrid) ytitle(Proportion) xtitle(Cohort: year aged 14) xline(1946.5, lcolor(black) lpattern(dash)) ///
  legend(off) title(Panel B: Male, color(black) size(medium)) xlab(1930[10]1970)
graph save Graph "g2.gph", replace

capture by yearat14, sort : egen meanwhite = mean(white)
twoway (scatter meanwhite yearat14 if yearat14>=1925 & yearat14<=1970, mcolor(black) msize(medsmall)), ///
  graphregion(fcolor(white) lcolor(white)) ylab(,nogrid) ytitle(Proportion) xtitle(Cohort: year aged 14) xline(1946.5, lcolor(black) lpattern(dash)) ///
  legend(off) title(Panel C: White, color(black) size(medium)) xlab(1930[10]1970)
graph save Graph "g3.gph", replace

capture by yearat14, sort : egen meanblack = mean(black)
twoway (scatter meanblack yearat14 if yearat14>=1925 & yearat14<=1970, mcolor(black) msize(medsmall)), ///
  graphregion(fcolor(white) lcolor(white)) ylab(,nogrid) ytitle(Proportion) xtitle(Cohort: year aged 14) xline(1946.5, lcolor(black) lpattern(dash)) ///
  legend(off) title(Panel D: Black, color(black) size(medium)) xlab(1930[10]1970)
graph save Graph "g4.gph", replace

capture by yearat14, sort : egen meanasian = mean(asian)
twoway (scatter meanasian yearat14 if yearat14>=1925 & yearat14<=1970, mcolor(black) msize(medsmall)), ///
  graphregion(fcolor(white) lcolor(white)) ylab(,nogrid) ytitle(Proportion) xtitle(Cohort: year aged 14) xline(1946.5, lcolor(black) lpattern(dash)) ///
  legend(off) title(Panel E: Asian, color(black) size(medium)) xlab(1930[10]1970)
graph save Graph "g5.gph", replace

capture by yearat14, sort : egen meanmanual = mean(fathermanual)
twoway (scatter meanmanual yearat14 if yearat14>=1925 & yearat14<=1970, mcolor(black) msize(medsmall)), ///
  graphregion(fcolor(white) lcolor(white)) ylab(,nogrid) ytitle(Proportion) xtitle(Cohort: year aged 14) xline(1946.5, lcolor(black) lpattern(dash)) ///
  legend(off) title(Panel F: Father manual/unskilled job, color(black) size(medium)) xlab(1930[10]1970)
graph save Graph "g6.gph", replace

capture by yearat14, sort : egen meanurate = mean(urate)
twoway (scatter meanurate yearat14 if yearat14>=1925 & yearat14<=1970, mcolor(black) msize(medsmall)), ///
  graphregion(fcolor(white) lcolor(white)) ylab(,nogrid) ytitle(Rate (%)) xtitle(Cohort: year aged 14) xline(1946.5, lcolor(black) lpattern(dash)) ///
  legend(off) title(Panel G: National unemployment rate, color(black) size(medium)) xlab(1930[10]1970)
graph save Graph "g7.gph", replace

twoway (scatter average_earnings yearat14 if yearat14>=1925 & yearat14<=1970, mcolor(black) msize(medsmall)), ///
  graphregion(fcolor(white) lcolor(white)) ylab(,nogrid) ytitle(Index (2000=100)) xtitle(Cohort: year aged 14) xline(1946.5, lcolor(black) lpattern(dash)) ///
  legend(off) title(Panel H: National average earnings, color(black) size(medium)) xlab(1930[10]1970)
graph save Graph "g8.gph", replace

gr combine "g1" "g2" "g3" "g4" "g5" "g6" "g7" "g8", rows(3) cols(3) subtitle(, color(black) fcolor(white) lcolor(white)) graphregion(fcolor(white) lcolor(white) ifcolor(white)
ilcolor(white))
```

Figure 4: Marshall Figure 2 code

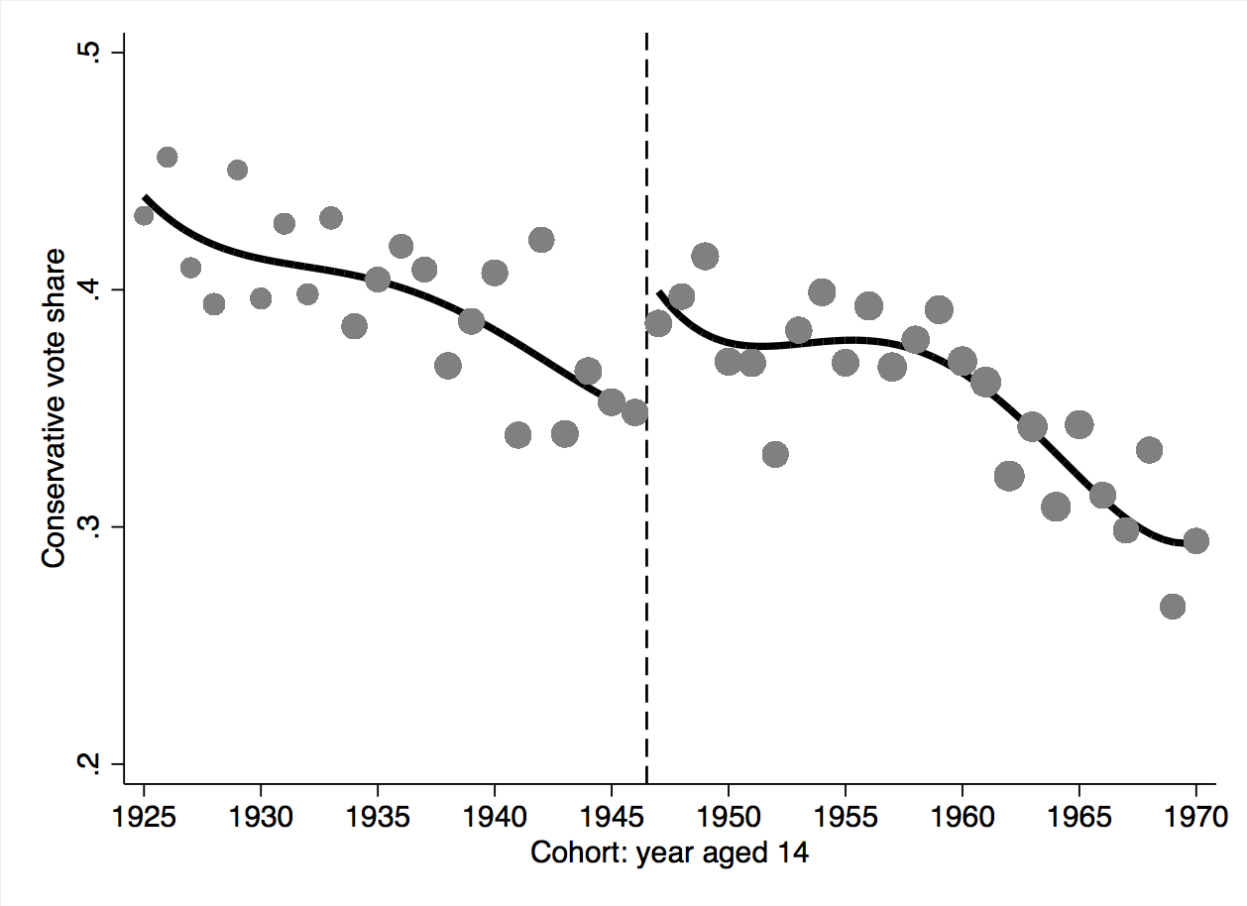


Figure 5: Marshall Figure 3

```
*** Figure 3: Reduced form
twoway (lpoly con yearat14 if yearat14>=1925 & yearat14<1947, lcolor(black) clwidth(thick) degree(4)) ///
  (lpoly con yearat14 if yearat14>=1947 & yearat14<=1970, lcolor(black) clwidth(thick) degree(4)) ///
  (scatter meancon14 yearat14 if yearat14>=1925 & yearat14<=1970 [weight=weight_14], msize(small) mcolor(gray)), ///
  graphregion(fcolor(white) lcolor(white)) ylab(,nogrid) ytitle(Conservative vote share) xtitle(Cohort: year aged 14) xline(1946.5,
lcolor(black) lpattern(dash)) ///
  yscale(range(.2 .5)) ylabel(.2[0.1]0.5) xlab(1925[5]1970) legend(off)
```

Figure 6: Marshall Figure 3 code

Table 1. Estimates of Schooling's Effect on Voting Conservative

	Years of Schooling LLR (1)	Attend University LLR (2)	Vote Conservative LLR (3)	Vote Conservative LLR IV (4)	Vote Conservative OLS (5)	Vote Conservative OLS (6)	Vote Labour LLR IV (7)	Vote Liberal LLR IV (8)
Post-1947 reform	.381***	.009	.044**					
	(.076)	(.013)	(.020)					
Years of schooling				.116**	.021***		−.071	−.021
				(.056)	(.002)		(.052)	(.043)
8th year of schooling						−.020		
						(.036)		
9th year of schooling						Baseline		
10th year of schooling						.126***		
						(.013)		
11th year of schooling						.213***		
						(.014)		
12th year of schooling						.289***		
						(.017)		
13th year of schooling						.306***		
						(.018)		
14th year of schooling						.281***		
						(.020)		
Outcome range	0–40	0 or 1	0 or 1	0 or 1	0 or 1	0 or 1	0 or 1	0 or 1
Outcome mean	10.6	.11	.38	.38	.36	.36	.36	.19
Outcome standard deviation	1.86	.32	.49	.49	.48	.48	.48	.39
First stage *F*-statistic	25.4	.5		25.4			25.4	25.4
Observations	11,068	11,068	11,068	11,068	16,757	16,757	11,068	11,068

Figure 7: Original Marshall Table 1

Appendix

Column 1

Coef.	Std. Err.
.4112	.09145

Variable	Obs	Mean	Std. Dev.	Min	Max
leave	11,068	10.58312	1.858667	0	40

Column 2

Coef.	Std. Err.
.03414	.01798

Variable	Obs	Mean	Std. Dev.	Min	Max
uni	11,068	.1145645	.31851	0	1

Column 3

Coef.	Std. Err.
.04548	.01829

Variable	Obs	Mean	Std. Dev.	Min	Max
leave	11,068	10.58312	1.858667	0	40

Column 4

Coef.	Std. Err.
.11581	.05539

Variable	Obs	Mean	Std. Dev.	Min	Max
leave	11,068	10.58312	1.858667	0	40

Column 5

Coef.	Robust Std. Err.
.021135	.0018537

Variable	Obs	Mean	Std. Dev.	Min	Max
con	16,757	.3587158	.4796379	0	1

Column 6

8	-.0200463	.0364474
10	.1257948	.0134744
11	.2126346	.0139522
12	.2891507	.0166584
13	.3061406	.0176145
14	.2810211	.0200969

Variable	Obs	Mean	Std. Dev.	Min	Max
con	16,757	.3587158	.4796379	0	1

Column 7

Coef.	Std. Err.
-.06892	.05082

Variable	Obs	Mean	Std. Dev.	Min	Max
lib	11,068	.1857608	.3889311	0	1

Column 8

Coef.	Std. Err.
-.00924	.04584

Variable	Obs	Mean	Std. Dev.	Min	Max
lib	11,068	.1857608	.3889311	0	1

[1]Link to Github Repo

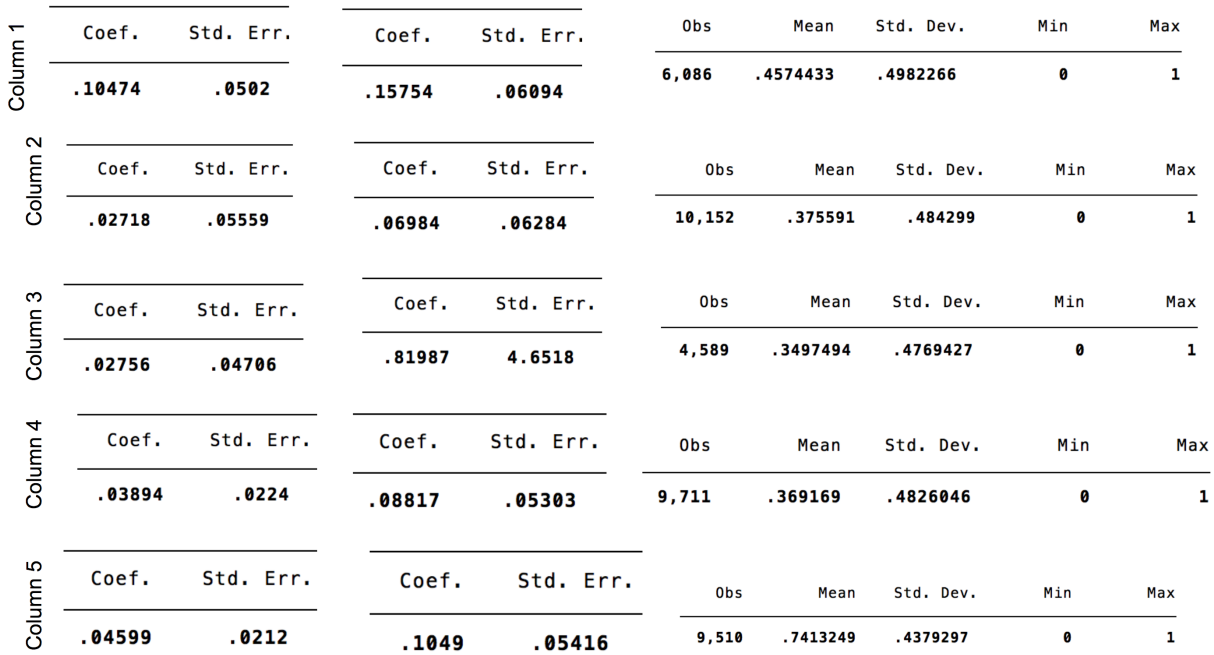
```
*** Table 1: Main estimates
rdrobust leave yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(mserd)
sum leave if yearat14>=1933 & yearat14<=1961
rdrobust uni yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(mserd)
sum uni if yearat14>=1933 & yearat14<=1961
rdrobust con yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(mserd)
rdrobust con yearat14, c(1947) fuzzy(leave) p(1) q(2) kernel(tri) bwselect(mserd)
sum leave if yearat14>=1933 & yearat14<=1961
areg con leave male white black asian sagesq-sagequart syearat14 syearat14sq syearat14cub syearat14quart, ro a(survey)
areg con ib9.leave male white black asian sagesq-sagequart syearat14 syearat14sq syearat14cub syearat14quart, ro a(survey)
summ con if e(sample)
rdrobust lab yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(mserd) fuzzy(leave)
sum lib if yearat14>=1933 & yearat14<=1961
rdrobust lib yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(mserd) fuzzy(leave)
sum lib if yearat14>=1933 & yearat14<=1961
```

Figure 8: Marshall Table 1 Code

Table 2. Schooling, Conservative Voting, and Income-Based Mechanisms

	Nonmanual Worker (below 60) (1)	Vote Conservative (below 60) (2)	Vote Conservative (60 or Above) (3)	Conservative Partisan (4)	Decided before Campaign (5)
A. Reduced form estimates:					
Post-1947 reform	.075**	.049**	.006	.039*	.041**
	(.030)	(.024)	(.029)	(.022)	(.019)
B. IV estimates:					
Years of schooling	.144***	.111**	.030	.092*	.103*
	(.055)	(.056)	(.153)	(.052)	(.053)
Outcome range	0 or 1	0 or 1	0 or 1	0 or 1	0 or 1
Outcome mean	.46	.38	.35	.37	.74
Outcome standard deviation	.50	.48	.48	.48	.44
Bandwidth	13.9	22.6	15.0	13.9	12.5
First stage *F*-statistic	28.1	28.3	2.7	26.4	23.4
Observations	6,086	10,152	4,589	9,711	9,510

Figure 9: Original Marshall Table 2



	Coef.	Std. Err.	Coef.	Std. Err.	Obs	Mean	Std. Dev.	Min	Max
Column 1	.10474	.0502	.15754	.06094	6,086	.4574433	.4982266	0	1
Column 2	.02718	.05559	.06984	.06284	10,152	.375591	.484299	0	1
Column 3	.02756	.04706	.81987	4.6518	4,589	.3497494	.4769427	0	1
Column 4	.03894	.0224	.08817	.05303	9,711	.369169	.4826046	0	1
Column 5	.04599	.0212	.1049	.05416	9,510	.7413249	.4379297	0	1

Figure 10: Replicated Table 2

```
*** Table 2: Raising social class, Heterogeneity by age (above 60), Become a Conservative partisan, and Decide before the electoral
campaign
rdrobust nonmanual yearat14 if age<60, c(1947) p(1) q(2) kernel(tri) bwselect(mserd)
rdrobust nonmanual yearat14 if age<60, fuzzy(leave) c(1947) p(1) q(2) kernel(tri) bwselect(mserd)
sum nonmanual if age<60 & yearat14>=1934 & yearat14<=1960

rdrobust con yearat14 if age<60, c(1947) p(1) q(2) kernel(tri) bwselect(mserd)
rdrobust con yearat14 if age<60, c(1947) p(1) q(2) kernel(tri) bwselect(mserd) fuzzy(leave)
sum con if age<60 & yearat14>=1923 & yearat14<=1969

rdrobust con yearat14 if age>=60, c(1947) p(1) q(2) kernel(tri) bwselect(mserd)
rdrobust con yearat14 if age>=60, c(1947) p(1) q(2) kernel(tri) bwselect(mserd) fuzzy(leave)
sum con if age>=60 & yearat14>=1932 & yearat14<=1962

rdrobust conpart yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(mserd)
rdrobust conpart yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(mserd) fuzzy(leave)
sum conpart if yearat14>=1934 & yearat14<=1960

rdrobust perm yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(mserd)
rdrobust perm yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(mserd) fuzzy(leave)
sum perm if yearat14>=1935 & yearat14<=1959
```

Figure 11: Marshall Table 2 Code

```
*** Table 3 Panel A: Economic policy preferences
rdrobust taxspendself yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(IK)
rdrobust welfaretoofar yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(IK)
rdrobust redist yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(IK)
rdrobust gender_not_too_much yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(IK)
sum gender_not_too_much if yearat14>=1931 & yearat14<=1963
rdrobust econ_values yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(IK)
sum econ_values if yearat14>=1930 & yearat14<=1964

rdrobust taxspendself yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(IK) fuzzy(leave)
rdrobust welfaretoofar yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(IK) fuzzy(leave)
rdrobust redist yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(IK) fuzzy(leave)
rdrobust gender_not_too_much yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(IK) fuzzy(leave)
rdrobust econ_values yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(IK) fuzzy(leave)



*** Table 3 Panel B: Non-economic policy preferences
rdrobust crime_rights_scale yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(IK)
rdrobust leave_europe yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(IK)
rdrobust end_priv_edu yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(IK)
rdrobust abortion_too_far yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(IK)
sum abortion_too_far if yearat14>=1933 & yearat14<=1961
rdrobust raceequ_too_far yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(IK)
sum raceequ_too_far if yearat14>=1927 & yearat14<=1967

rdrobust crime_rights_scale yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(IK) fuzzy(leave)
rdrobust leave_europe yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(IK) fuzzy(leave)
rdrobust end_priv_edu yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(IK) fuzzy(leave)
rdrobust abortion_too_far yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(IK) fuzzy(leave)
rdrobust raceequ_too_far yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(IK) fuzzy(leave)
```

Figure 12: Marshall Table 3 Code

References

Reference material to create this bibliography comes from (J Allaire 2016)

Gary King, Jason Whittenberg, Michael Tomz. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." Midwest Political Science Associatio. https://www-jstor-org.ezp-prod1.hul.harvard.edu/stable/pdf/2669316.pdf.

J Allaire, Jonathan McPherson, Yihui Xie. 2016. "Bibliographies and Citations." https://rmarkdown.rstudio.com/authoring_bibliographies_and_citations.html.

Marshall, John. 2015. "Education and Voting Conservative: Evidence from a Major Schooling Reform in Great Britain." Southern Political Science Association. https://www-journals-uchicago-edu.ezp-prod1.hul.harvard.edu/doi/pdfplus/10.1086/683848.

Sebastian Calonico, Max H. Farrell, Matias D. Cattaneo. 2020. "Package 'Rdrobust'." https://cran.r-project.org/web/packages/rdrobust/rdrobust.pdf.

Trochim, Prof William M.K. 2020. "The Regression-Discontinuity Design." https://socialresearchmethods.net/kb/regression-discontinuity-design/.