Lesson 6, Part 1:
Linear Mixed Effects Models

This Lesson's Goals

Learn about linear mixed effects models (LMEM)

Make figures for data for LMEMs

Run some preliminary LMEMs in R

Summarise results in an R Markdown document

This Lesson's Goals

Learn about linear mixed effects models (LMEM)

Make figures for data for LMEMs

Run some preliminary LMEMs in R

~~Summarise results in an R Markdown document~~

End of Lesson 5 Questions

But aren't percentages *really*
just summarized count data?

But we had to drop a bunch of Union states,
isn't that a problem?

But Alabama was missing one Democrat data point,
isn't it not balanced?

But what about the variance for 'year',
shouldn't we try and account for that too?

generalized linear mixed effects models

Math (Part 1)

$$y_i = a + b_1x_{1i} + b_2x_{2i} + b_3x_{1i}x_{2i} + e_i$$

How do I add factors for random variance?

(i.e. things we’re not directly testing)

$$y_i = a + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{1i} x_{2i} + e_i$$

$$y_i = a + \boxed{a_s} + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{1i} x_{2i} + e_i$$

random effect

s = state

$$y_i = a + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{1i} x_{2i} + e_i$$

$$y_i = a + \boxed{a_s} + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{1i} x_{2i} + e_i$$

random intercept

s = state

$$y_i = a + b_1x_{1i} + b_2x_{2i} + b_3x_{1i}x_{2i} + e_i$$

$$y_i = a + \boxed{a_s} + \boxed{a_y} + b_1x_{1i} + b_2x_{2i} + b_3x_{1i}x_{2i} + e_i$$

random intercept #1

random intercept #2

s = state

y = year

In this paper we tested the effect of time on weight. A total of 50 baby chicks were included in the study.

	ANOVA language	LMEM language
weight	dependent variable	dependent variable
time	independent variable	fixed effect
baby chick	error variable	random effect

mixed effects

$$\boxed{y_i} = \boxed{a} + \boxed{a_s} + \boxed{a_y} + \boxed{b_1}\boxed{x_{1i}} + \boxed{b_2}\boxed{x_{2i}} + \boxed{b_3} x_{1i} x_{2i} + \boxed{e_i}$$

y_i = specific y value

a = intercept

a_s = random intercept #1 for specific level

a_y = random intercept #2 for specific level

b_1 = slope of first variable

x_{1i} = specific x value for first variable

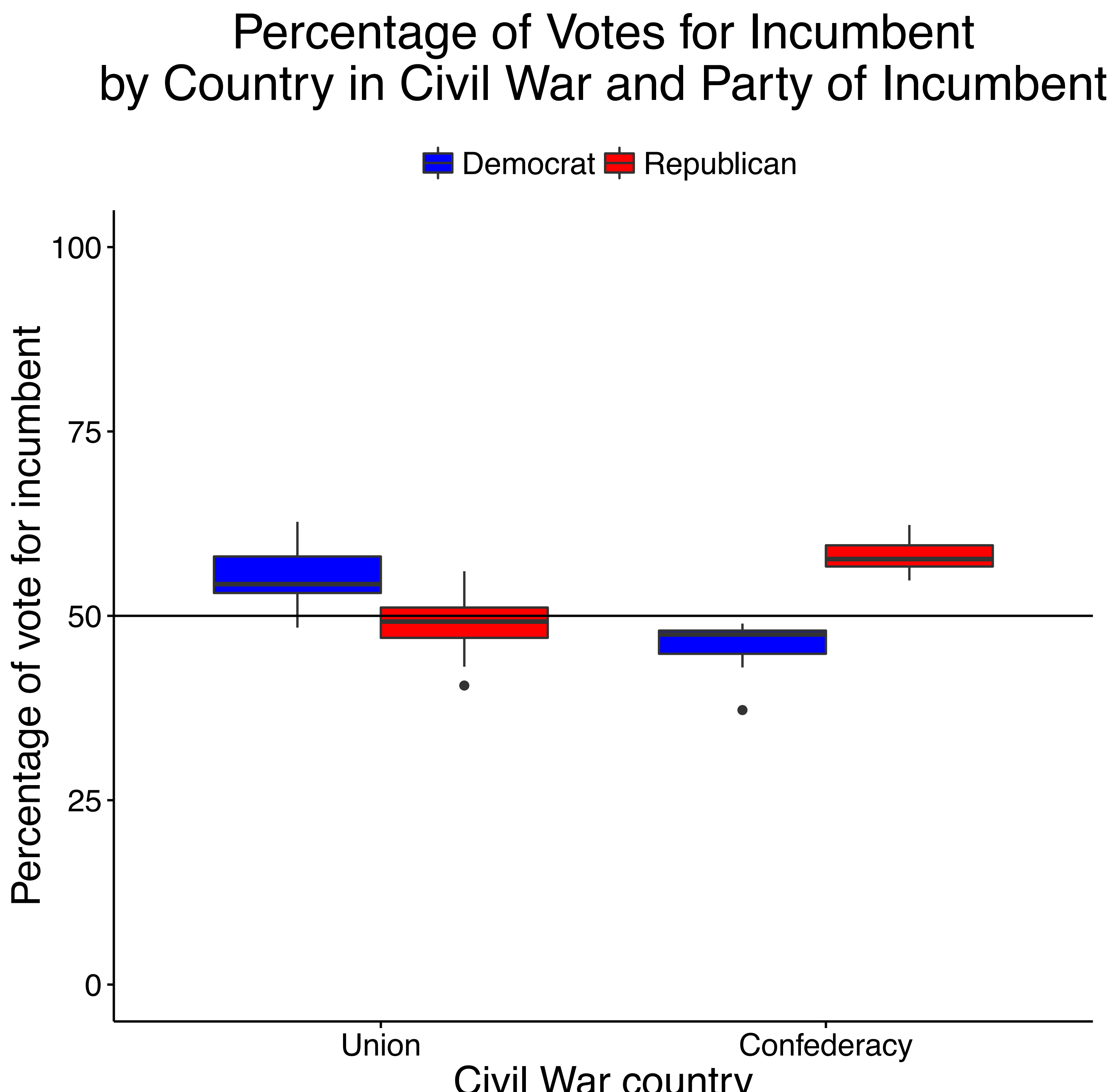
b_2 = slope of second variable

x_{2i} = specific x value for second variable

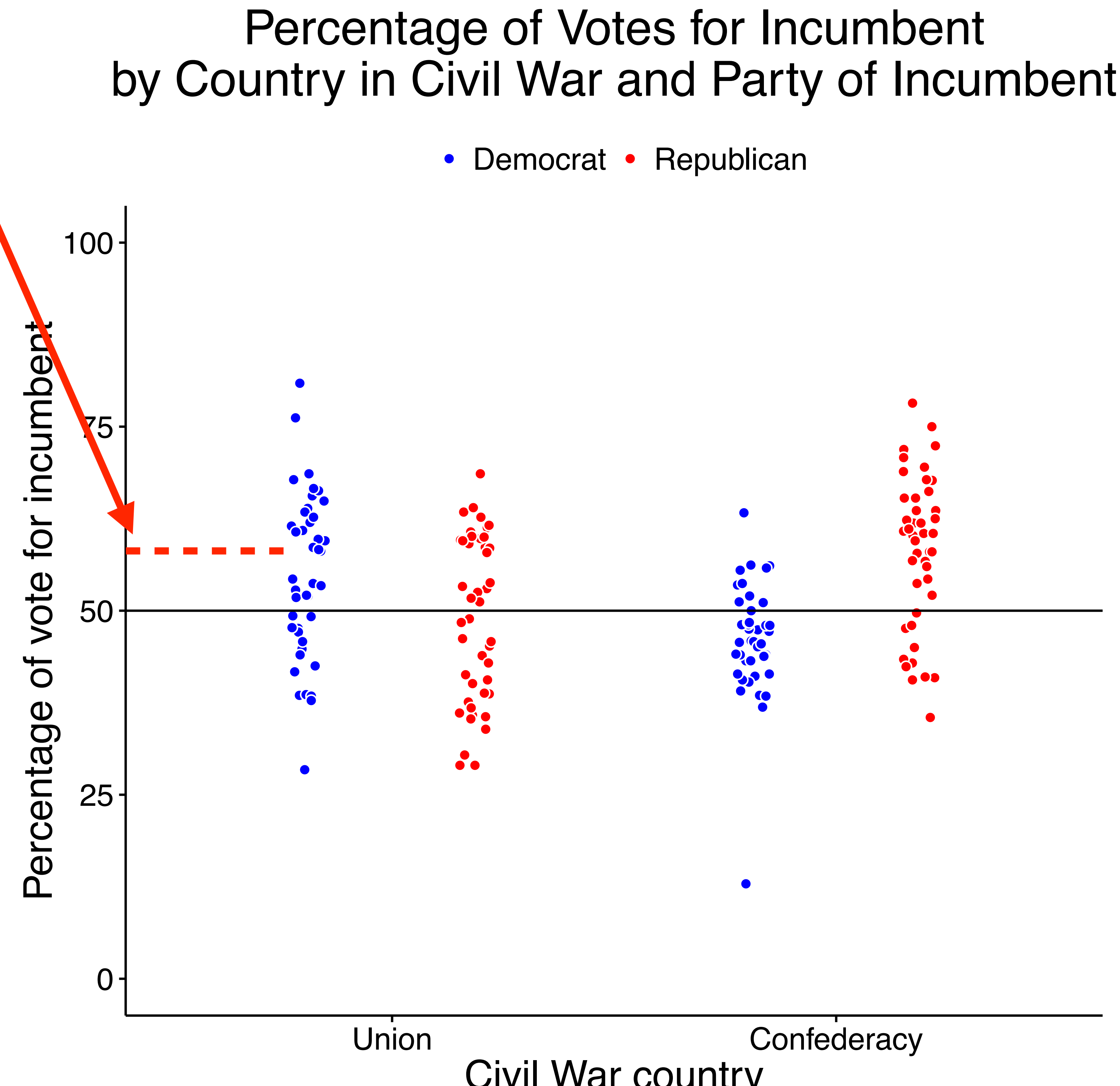
b_3 = slope of third variable (interaction)

e_i = random variance or the residual

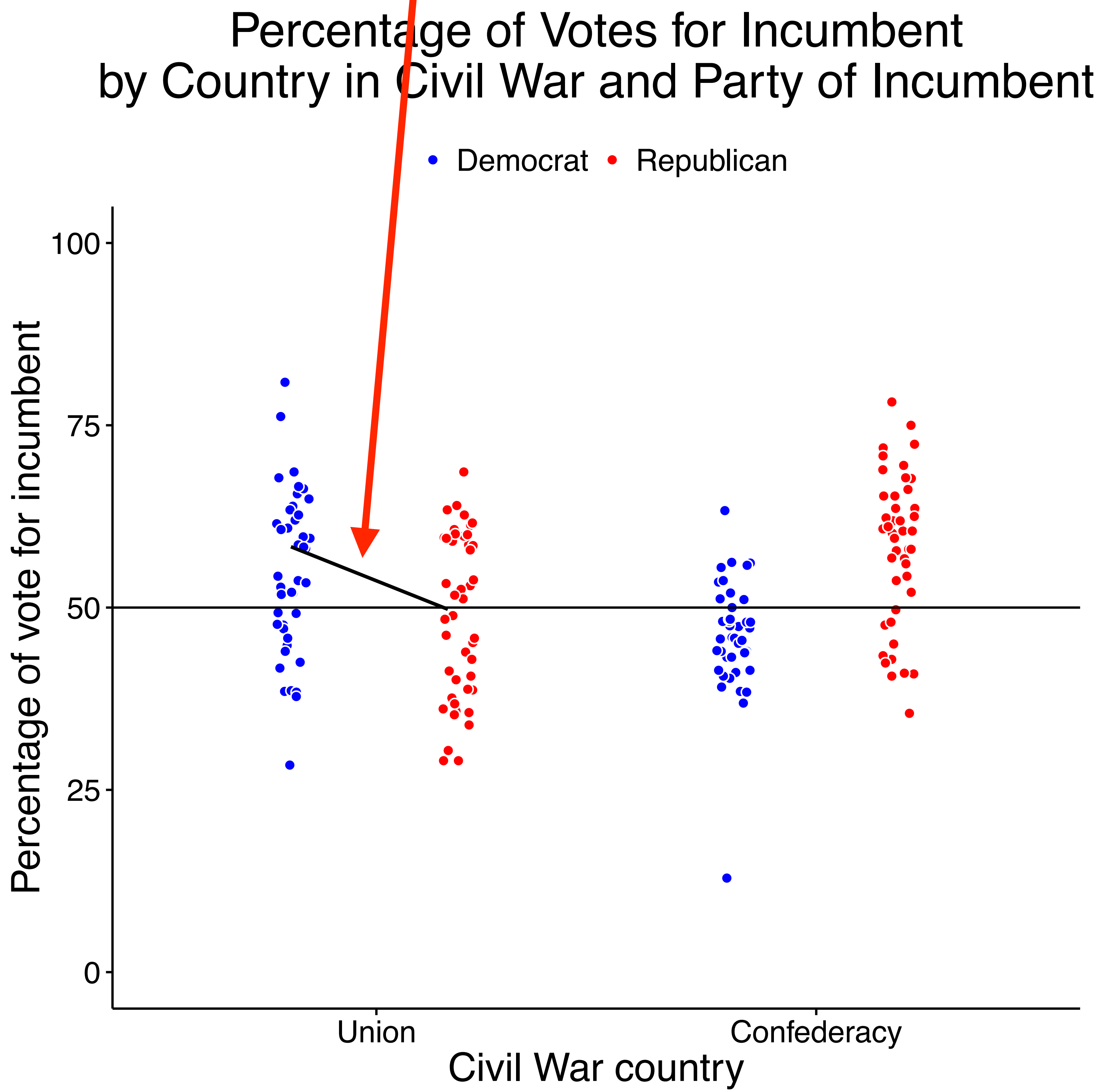
$$y_i = a + a_s + a_y + b_1x_{1i} + b_2x_{2i} + b_3x_{1i}x_{2i} + e_i$$



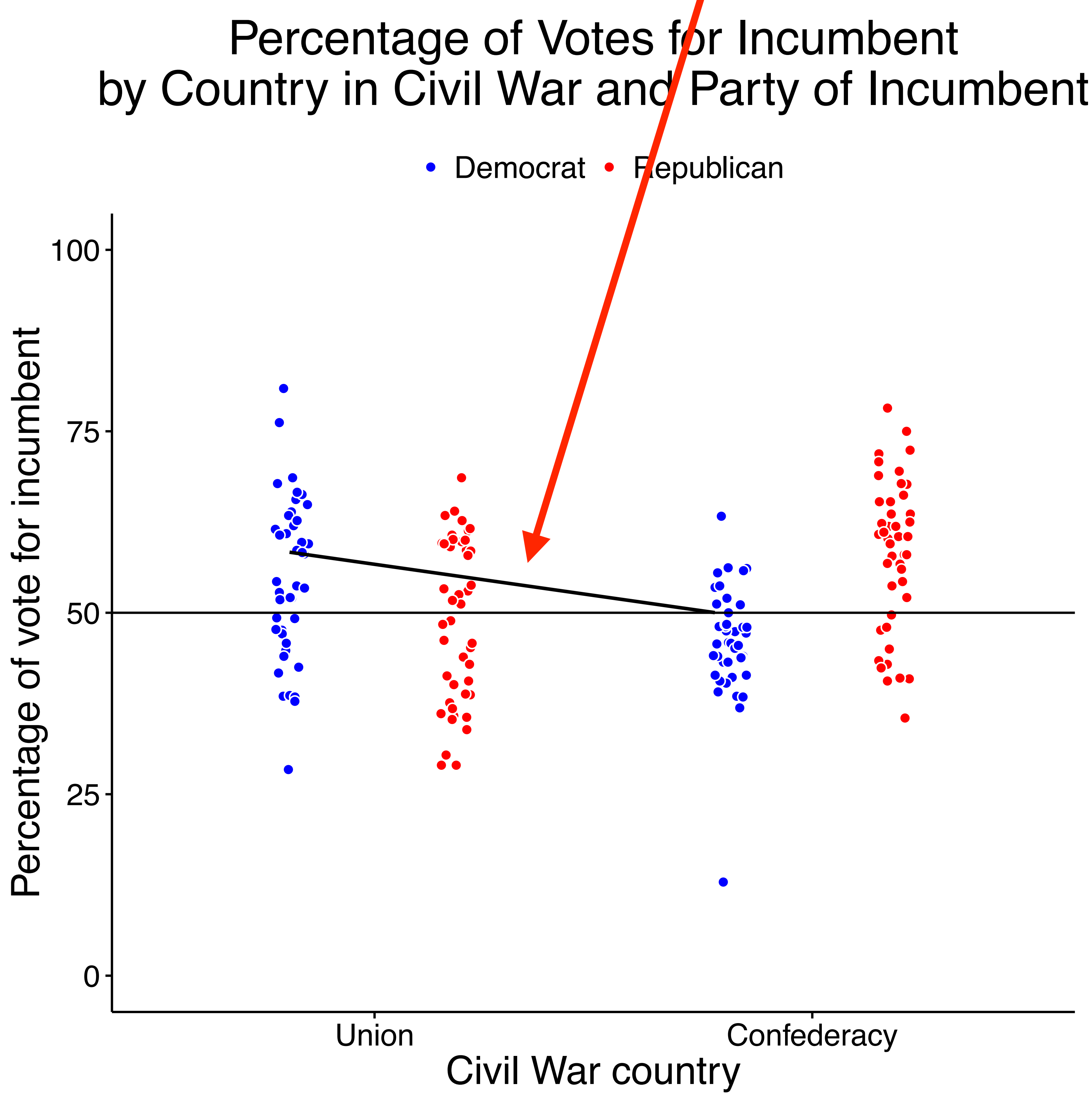
$$y_i = a + a_s + a_y + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{1i} x_{2i} + e_i$$



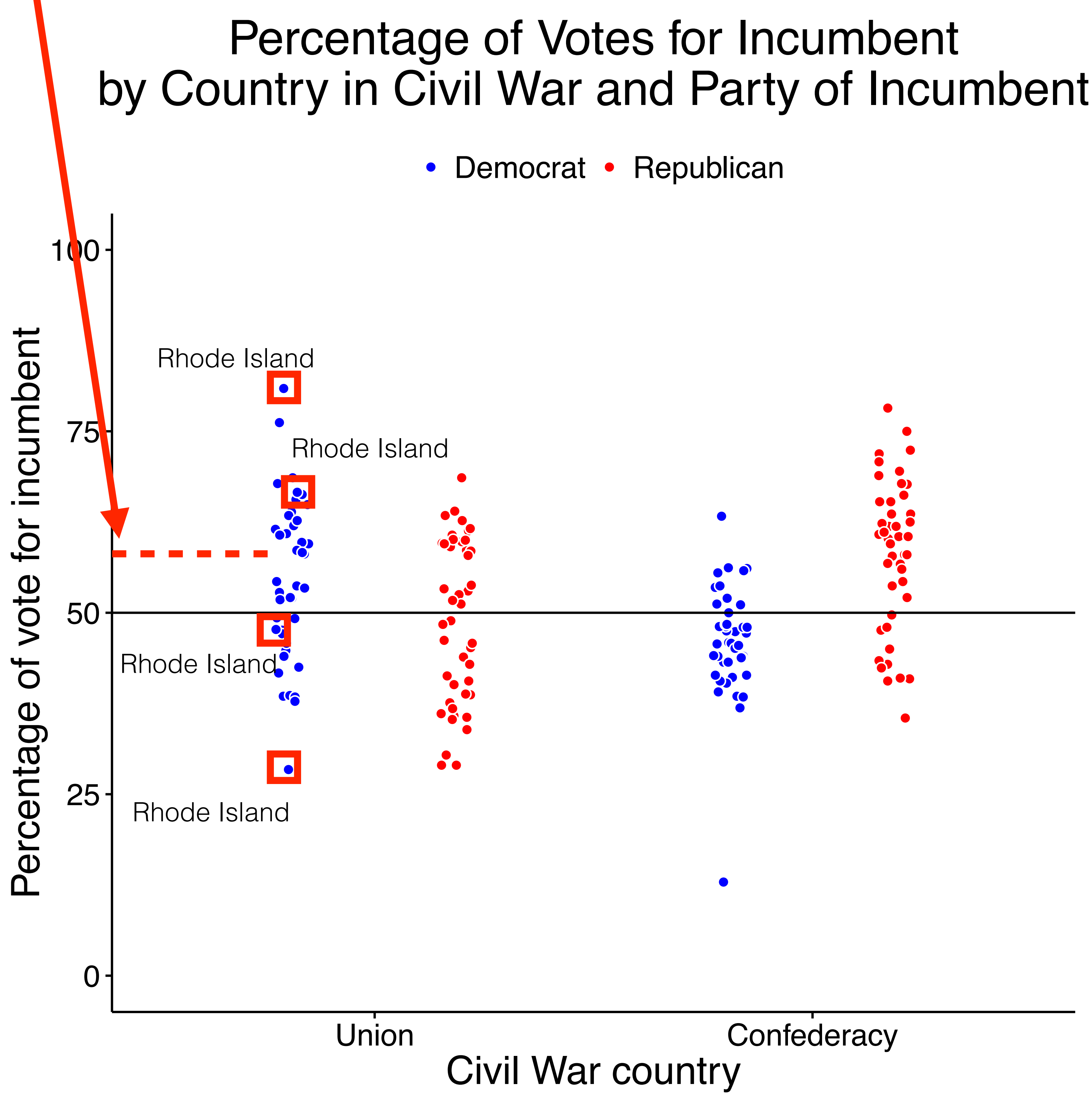
$$y_i = a + a_s + a_y + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{1i} x_{2i} + e_i$$



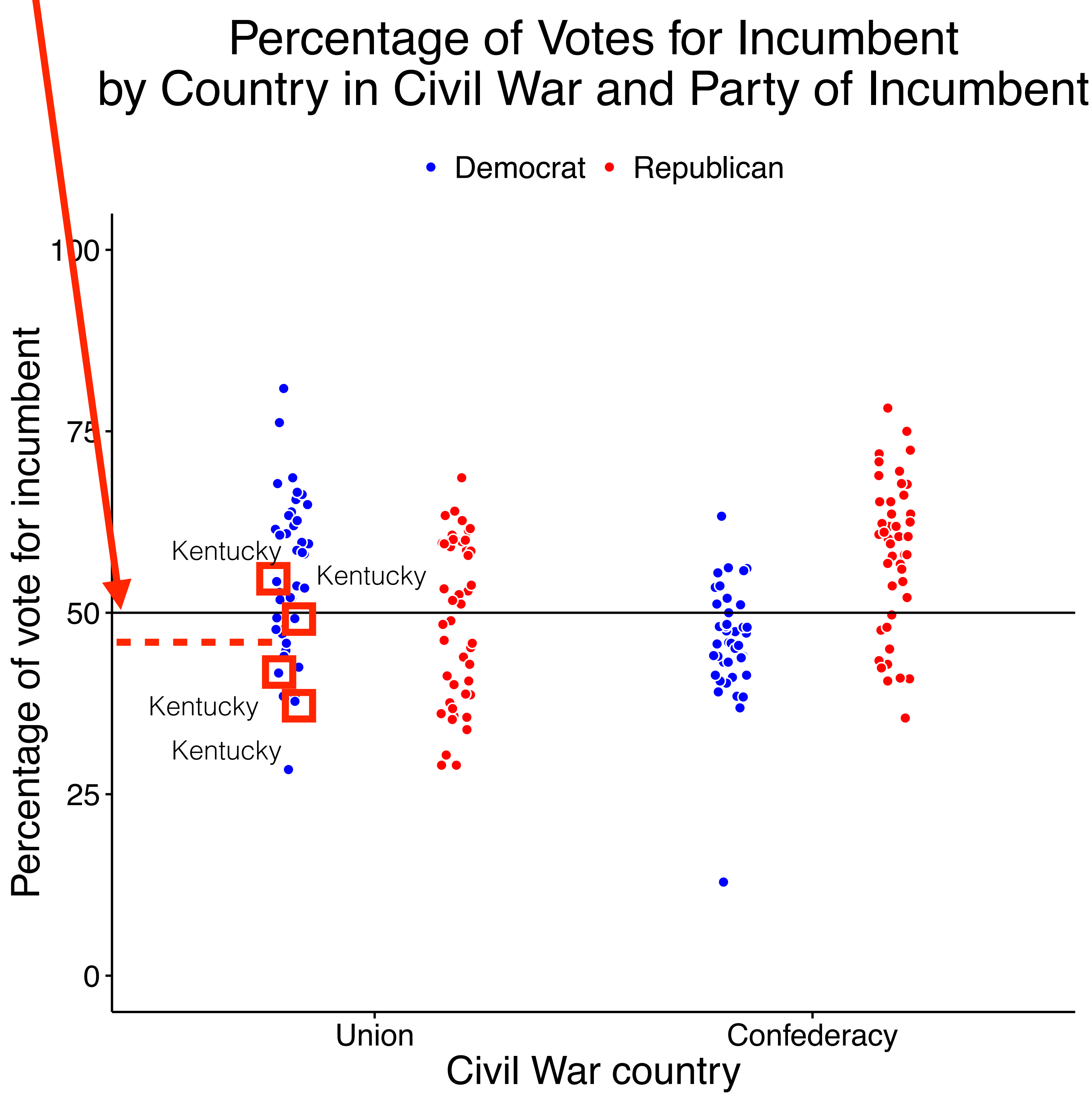
$$y_i = a + a_s + a_y + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{1i} x_{2i} + e_i$$



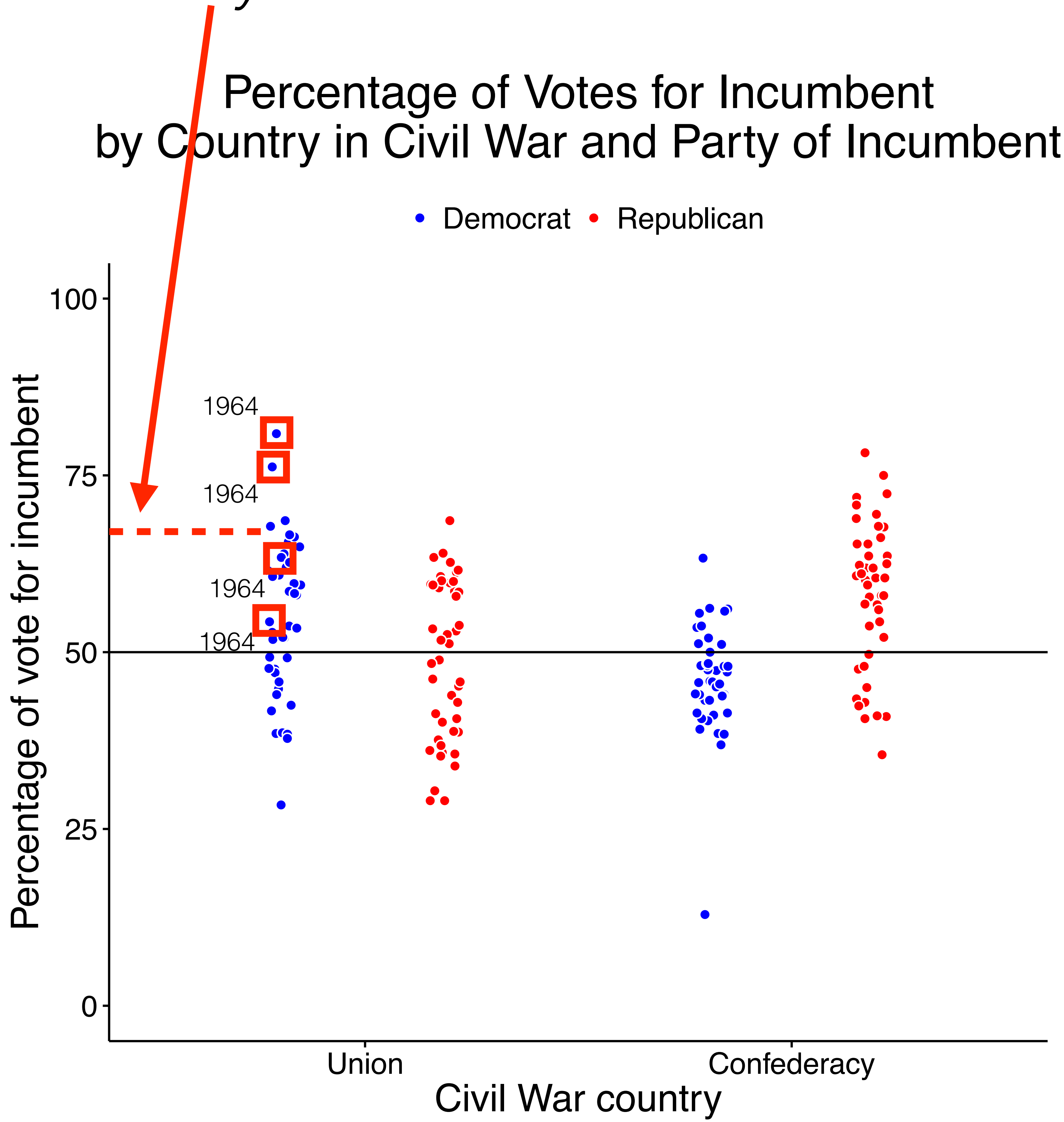
$$y_i = a + a_s + a_y + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{1i} x_{2i} + e_i$$



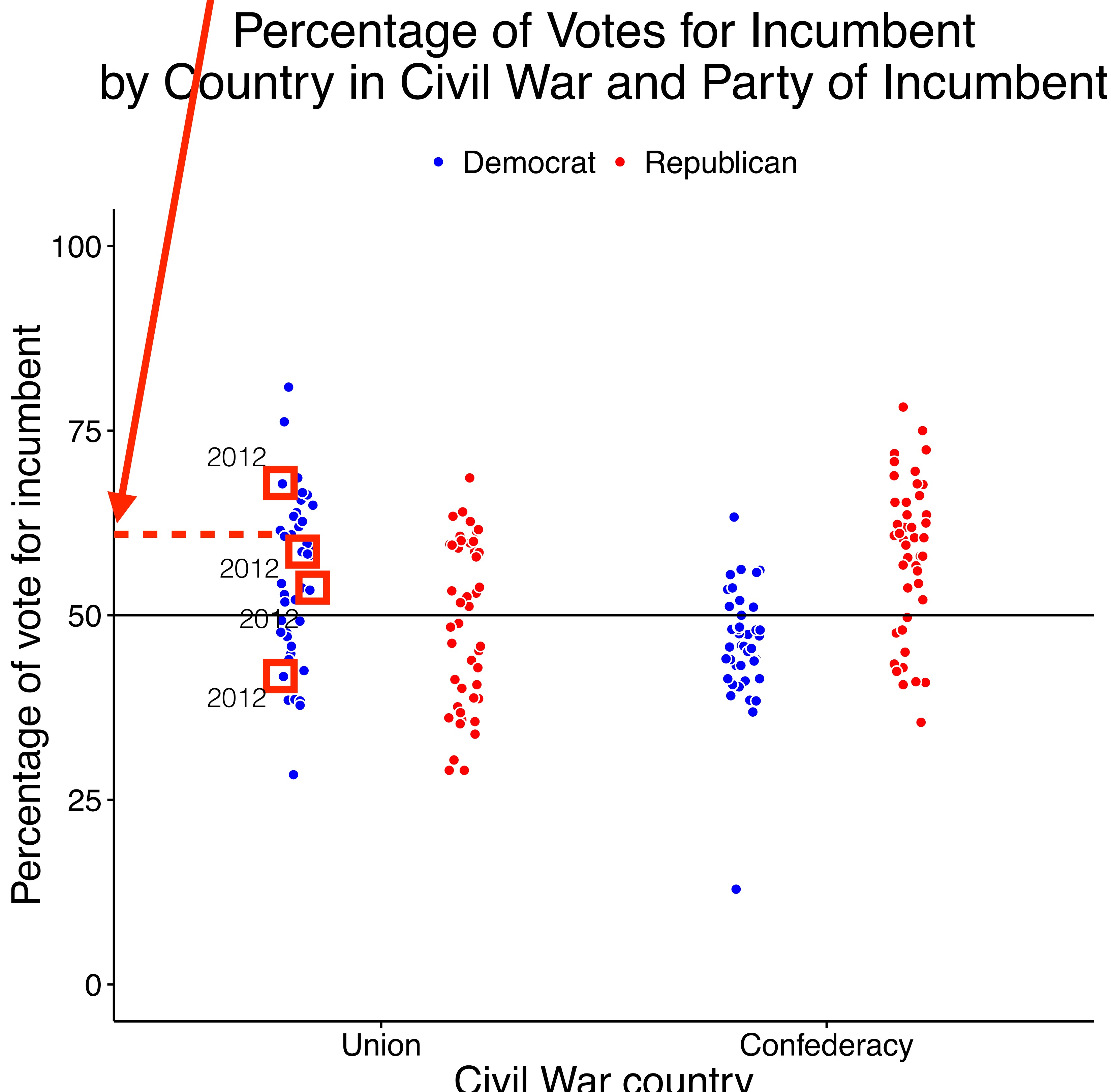
$$y_i = a + a_s + a_y + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{1i} x_{2i} + e_i$$



$$y_i = a + a_s + a_y + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{1i} x_{2i} + e_i$$



$$y_i = a + a_s + a_y + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{1i} x_{2i} + e_i$$



But, at the end of the last lesson we said this was bad, because it was a percentage of a count?

$$y_i = a + a_s + a_y + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{1i} x_{2i} + e_i$$

$$\boxed{\log[p/(1-p)]_i} = a + a_s + a_y + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{1i} x_{2i} + e_i$$

logistic regression —>

generalized linear mixed effects model

R Code (Part 1)

lme4

$$y_i = a + a_s + a_y + b_1x_{1i} + b_2x_{2i} + b_3x_{1i}x_{2i} + e_i$$

```
lmer(perc_votes_incumbent ~
incumbent_party * civil_war
          + (1|state)
          + (1|year))
```

```
Fixed effects:
                                         Estimate Std. Error t value
(Intercept)                                55.164      4.483  12.305
incumbent_partyRepublican                  -6.273      6.340  -0.989
civil_warConfederacy                       -8.990      1.444  -6.226
incumbent_partyRepublican:civil_warConfederacy   18.231      2.036   8.955
```

lme4

$$y_i = a + a_s + a_y + b_1x_{1i} + b_2x_{2i} + b_3x_{1i}x_{2i} + e_i$$

```
lmer(perc_votes_incumbent ~
incumbent_party * civil_war
        + (1|state)
        + (1|year))
```

	(Intercept)	incumbent_partyRepublican	civil_warConfederacy	incumbent_partyRepublican:civil_warConfederacy
Alabama	55.16364	-6.272727	-8.98988	18.23079
Arkansas	55.16364	-6.272727	-8.98988	18.23079
Connecticut	55.16364	-6.272727	-8.98988	18.23079
Delaware	55.16364	-6.272727	-8.98988	18.23079
Florida	55.16364	-6.272727	-8.98988	18.23079
Georgia	55.16364	-6.272727	-8.98988	18.23079

	(Intercept)	incumbent_partyRepublican	civil_warConfederacy	incumbent_partyRepublican:civil_warConfederacy
1964	62.33514	-6.272727	-8.98988	18.23079
1972	65.93855	-6.272727	-8.98988	18.23079
1980	48.65702	-6.272727	-8.98988	18.23079
1984	60.95054	-6.272727	-8.98988	18.23079
1992	41.19320	-6.272727	-8.98988	18.23079
1996	54.63202	-6.272727	-8.98988	18.23079

But, in the ANOVA we accounted for the fact that a variable could be within- or between-subject?

Math (Part 2)

$$y_i = a + a_s + a_y + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{1i} x_{2i} + e_i$$

$$y_i = a + a_s + a_y + (\boxed{b_{s1}} + b_1) x_{1i} + (\boxed{b_{y1}} + b_2) x_{2i} + b_3 x_{1i} x_{2i} + e_i$$

random slope

random slope

s = state

y = year

$$y_i = a + \boxed{a_s} + \boxed{a_y} + (\boxed{b_{s1}} + b_1)x_{1i} + (\boxed{b_{y1}} + b_2)x_{2i} + b_3 x_{1i} x_{2i} + e_i$$

y_i = specific y value

a = intercept

a_s = random intercept #1

a_y = random intercept #2

b_{s1} = slope of r.e. #1

b_1 = slope of variable #1

x_{1i} = x value for variable #1

b_{y1} = slope of r.e. #2

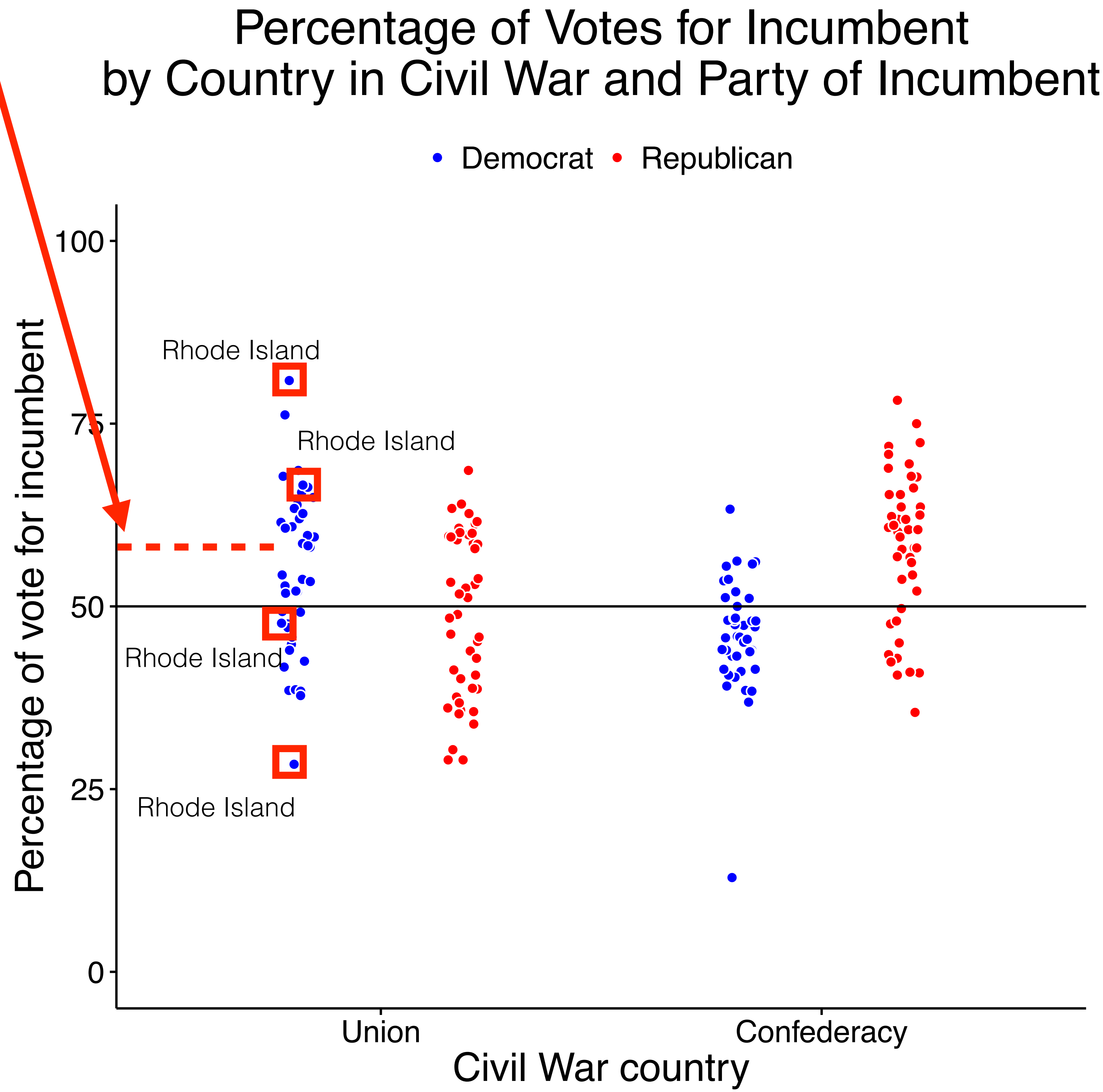
b_2 = slope of variable #2

x_{2i} = x value for variable #2

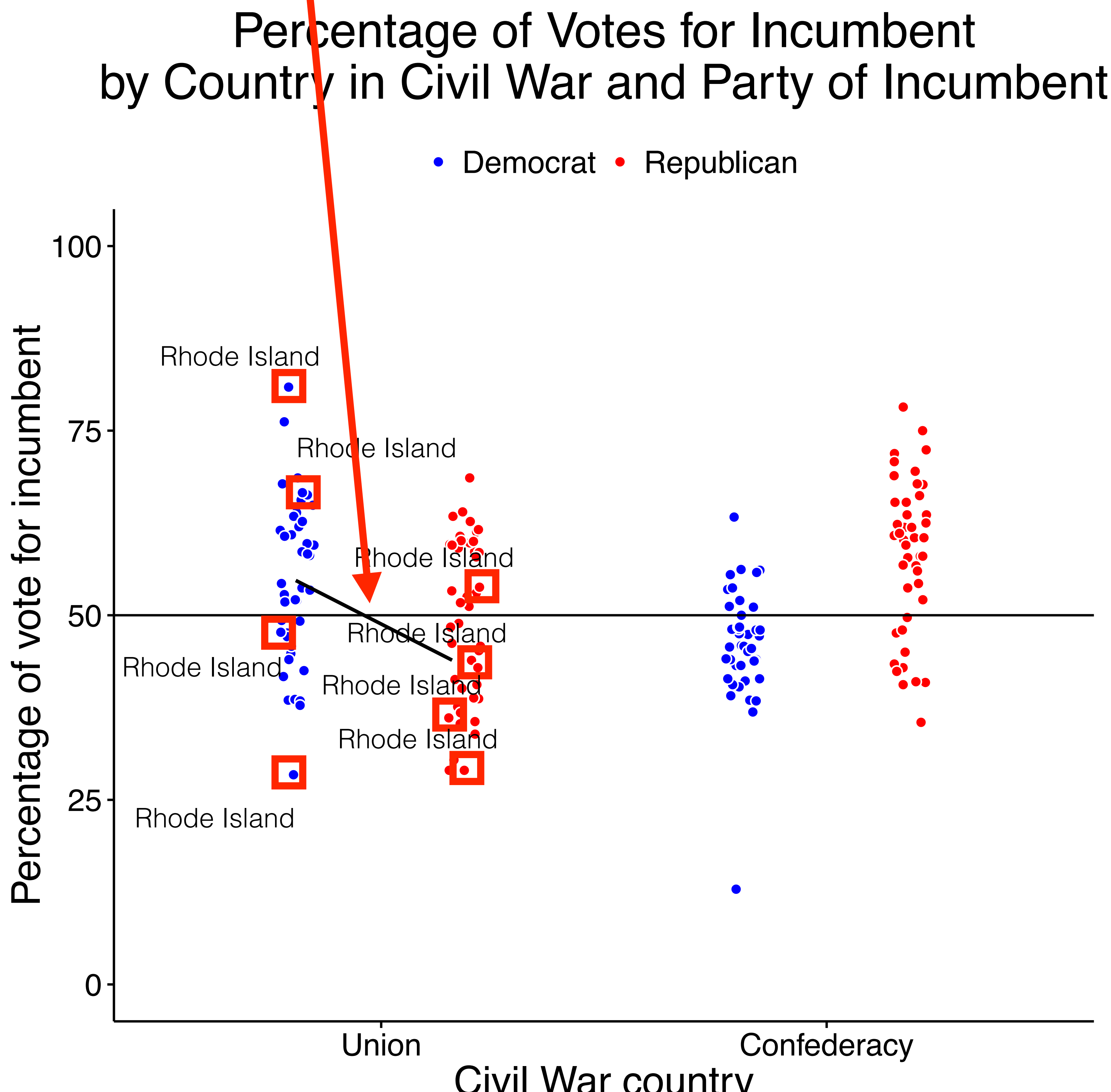
b_3 = slope of variable #3

e_i = random variance

$$y_i = a + a_s + a_y + (b_{s1}+b_1)x_{1i} + (b_{y1}+b_2)x_{2i} + b_3 x_{1i} x_{2i} + e_i$$



$$y_i = a + a_s + a_y + (b_{s1}+b_1)x_{1i} + (b_{y1}+b_2)x_{2i} + b_3x_{1i}x_{2i} + e_i$$



R Code (Part 2)

lme4

$$y_i = a + a_s + a_y + (b_{s1}+b_1)x_{1i} + (b_{y1}+b_2)x_{2i} + b_3x_{1i}x_{2i} + e_i$$

```
lmer(perc_votes_incumbent ~
incumbent_party * civil_war
+ (1+incumbent_party|state)
   + (1+civil_war|year))
```

```
Fixed effects:
                                           Estimate Std. Error t value
(Intercept)                                  55.164      5.591   9.866
incumbent_partyRepublican                    -6.273      8.012  -0.783
civil_warConfederacy                         -9.155      4.183  -2.189
incumbent_partyRepublican:civil_warConfederacy 18.396    6.189   2.972
```

lme4

$$y_i = a + a_s + a_y + (b_{s1}+b_1)x_{1i} + (b_{y1}+b_2)x_{2i} + b_3 x_{1i} x_{2i} + e_i$$

```
lmer(perc_votes_incumbent ~
incumbent_party * civil_war
+ (1+incumbent_party|state)
  + (1+civil_war|year))
```

	(Intercept)	incumbent_partyRepublican	civil_warConfederacy	incumbent_partyRepublican:civil_warConfederacy
Alabama	53.19379	-2.537193	-9.155457	18.39637
Arkansas	57.53616	-10.771874	-9.155457	18.39637
Connecticut	54.46253	-4.943172	-9.155457	18.39637
Delaware	54.20025	-4.445809	-9.155457	18.39637
Florida	55.77018	-7.422953	-9.155457	18.39637
Georgia	55.78214	-7.445637	-9.155457	18.39637

	(Intercept)	incumbent_partyRepublican	civil_warConfederacy	incumbent_partyRepublican:civil_warConfederacy
1964	67.30095	-6.272727	-19.271213	18.39637
1972	65.15338	-6.272727	-7.413460	18.39637
1980	42.24166	-6.272727	3.503531	18.39637
1984	63.33161	-6.272727	-13.732613	18.39637
1992	41.22931	-6.272727	-9.482726	18.39637
1996	53.79614	-6.272727	-7.361084	18.39637

LMEM with only random intercepts

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	55.164	4.483	12.305
incumbent_partyRepublican	-6.273	6.340	-0.989
civil_warConfederacy	-8.990	1.444	-6.226
incumbent_partyRepublican:civil_warConfederacy	18.231	2.036	8.955

LMEM with only random intercepts and slopes

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	55.164	5.591	9.866
incumbent_partyRepublican	-6.273	8.012	-0.783
civil_warConfederacy	-9.155	4.183	-2.189
incumbent_partyRepublican:civil_warConfederacy	18.396	6.189	2.972

Lab

Data set: Stroop Task

source: real students!

Say the color of the ink ***not*** the written word.

blue

Say the color of the ink ***not*** the written word.

blue

Say the color of the ink ***not*** the written word.

blue

word = ink color
congruent trial

blue

word ≠ ink color
incongruent trial

Data set: Stroop Task

Congruency: Are responses to incongruent trials less accurate and slower than to congruent trials?

Experiment half: Are responses more accurate and faster in the second half of the experiment than the first half of the experiment?

Congruency x Experiment half: Is there an interaction between these variables?

accuracy (logistic)

logit p_i = accuracy
x1 = congruency
x2 = experiment half
r1 = subject
r2 = item

reaction times (linear)

y_i = reaction times
x1 = congruency
x2 = experiment half
r1 = subject
r2 = item

source: real students!

dplyr

```
data_clean = data_results
```

dplyr

```
data_clean = data_results %>%
       rename(trial_number = SimpleRTBLock.TrialNr.)
```

change
variable
name

dplyr

```
data_clean = data_results %>%
       rename(trial_number = SimpleRTBLock.TrialNr.) %>%
       rename(congruency = Congruency) %>%
       rename(correct_response = StroopItem.CRESP.) %>%
       rename(given_response = StroopItem.RESP.) %>%
       rename(accuracy = StroopItem.ACC.) %>%
       rename(rt = StroopItem.RT.) %>%
```

change
variable
name

dplyr

```
data_clean = data_results %>%
      rename(trial_number = SimpleRTBLock.TrialNr.) %>%
      rename(congruency = Congruency) %>%
      rename(correct_response = StroopItem.CRESP.) %>%
      rename(given_response = StroopItem.RESP.) %>%
      rename(accuracy = StroopItem.ACC.) %>%
      rename(rt = StroopItem.RT.) %>%
      select(subject_id, block, item, trial_number,
                   congruency, correct_response,
                   given_response, accuracy, rt)
```

change
variable
name

choose subset
of variables

RColorBrewer

```
cols = brewer.pal(              )
```

call to
make palette

RColorBrewer

```
cols = brewer.pal(5          )
```

call to
make palette

number of
colors

RColorBrewer

```
cols = brewer.pal(5, “PuOr")
```

call to
make palette

number of
colors

palette
name

RColorBrewer

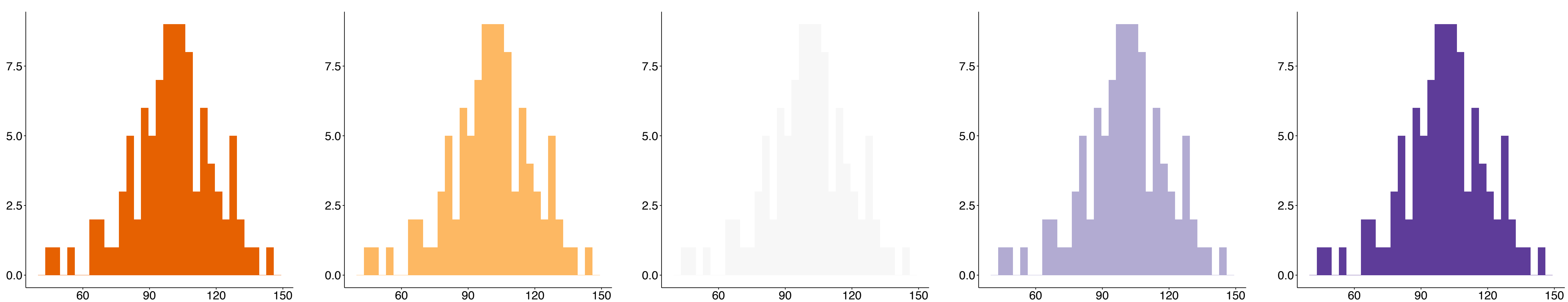
```
cols = brewer.pal(5, “PuOr")
```

call to make palette

number of colors

palette name

```
> cols
[1] "#E66101" "#FDB863" "#F7F7F7" "#B2ABD2" "#5E3C99"
```



RColorBrewer

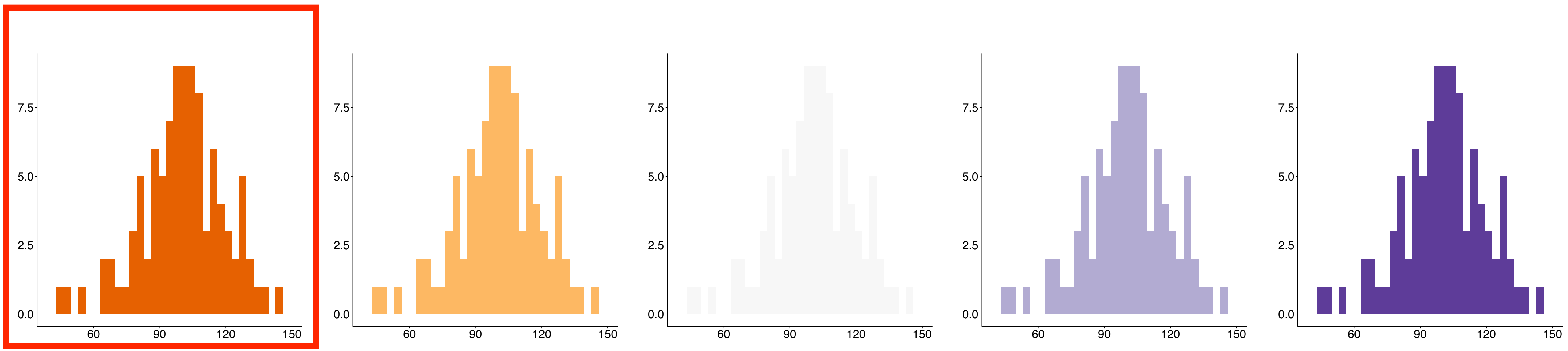
```
cols = brewer.pal(5, "PuOr")
col_con = cols[1]
```

call to make palette

number of colors

palette name

```
> cols
[1] "#E66101" "#FDB863" "#F7F7F7" "#B2ABD2" "#5E3C99"
```



RColorBrewer

```
cols = brewer.pal(5, “PuOr")
col_con = cols[1]
col_incon = cols[5]
```

call to make palette

number of colors

palette name

```
> cols
[1] "#E66101" "#FDB863" "#F7F7F7" "#B2ABD2" "#5E3C99"
```