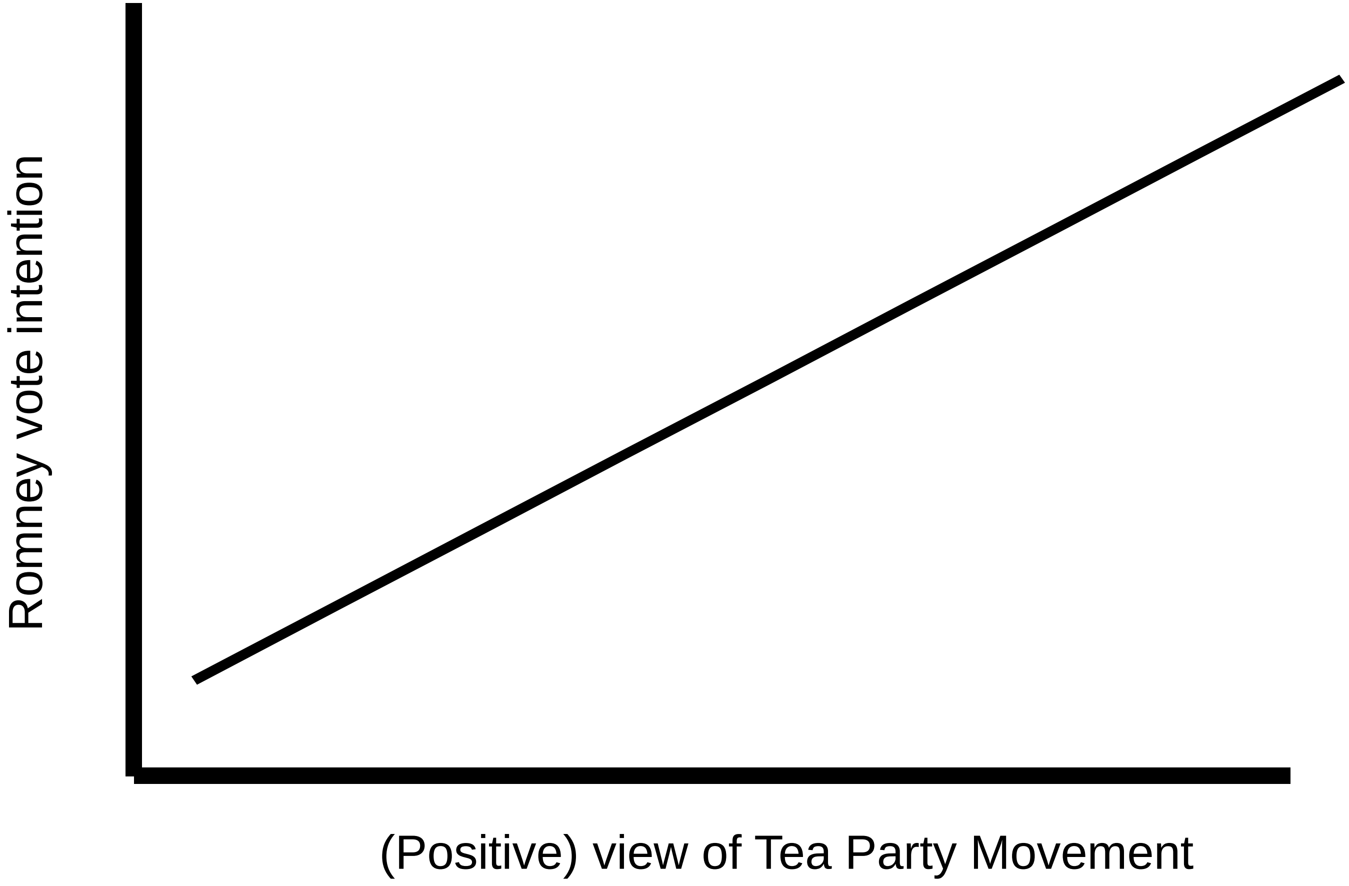
Addressing Alternative Explanations: Multiple Regression

17.871

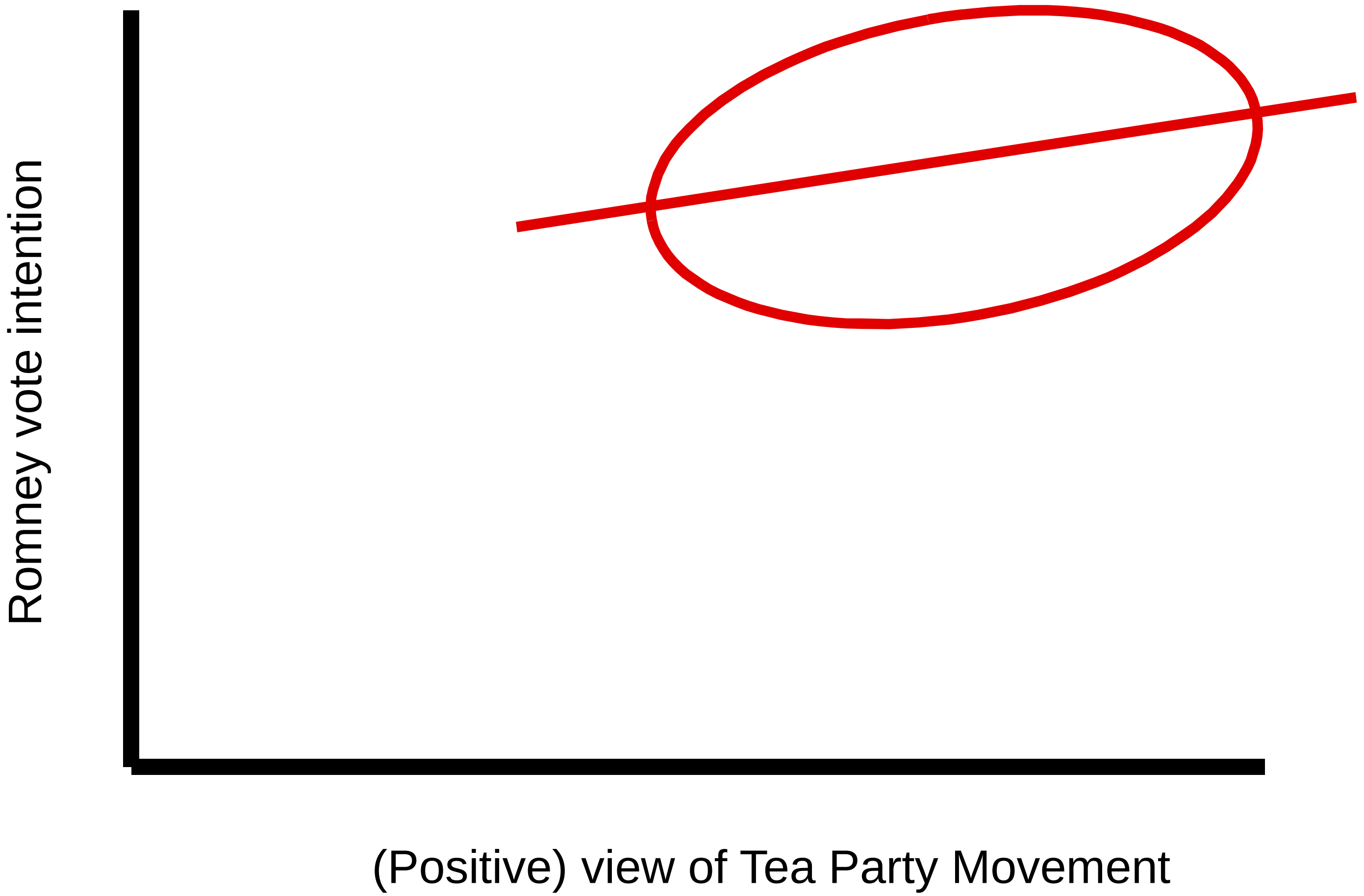
Spring 2015

Did the Tea Party Movement give a boost to Romney in 2012?

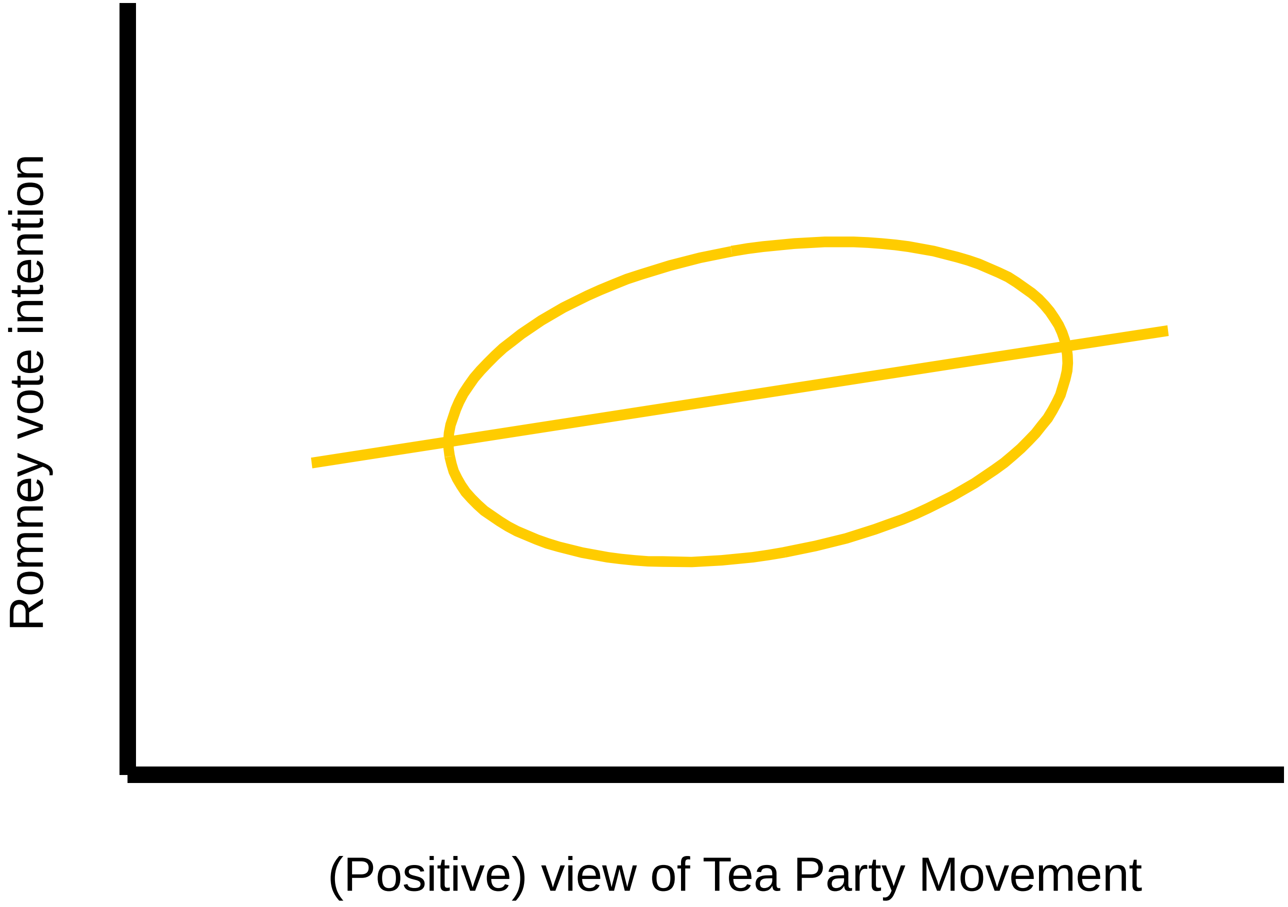
Bivariate regression of Romney vote intention on view about Tea Party



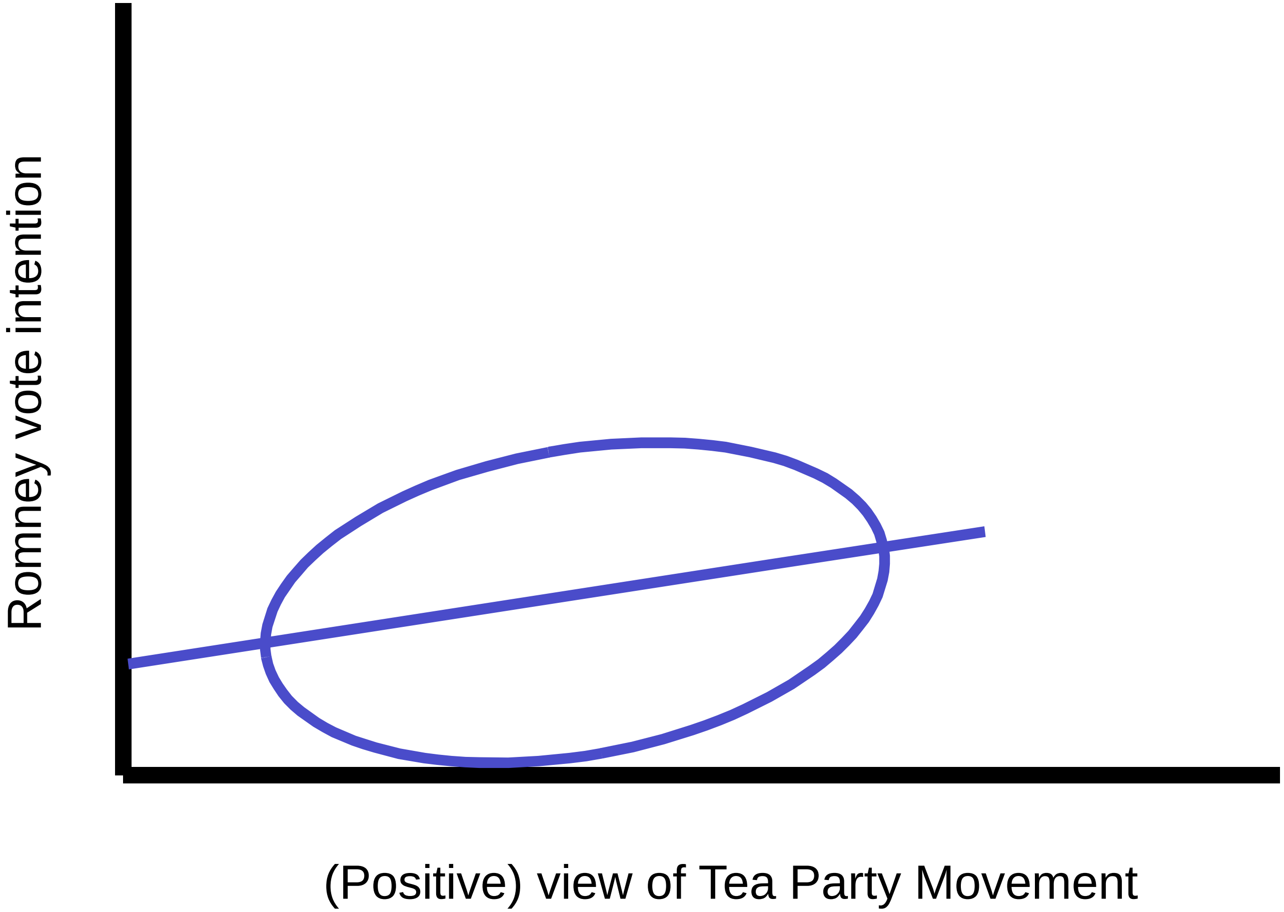
Republican picture



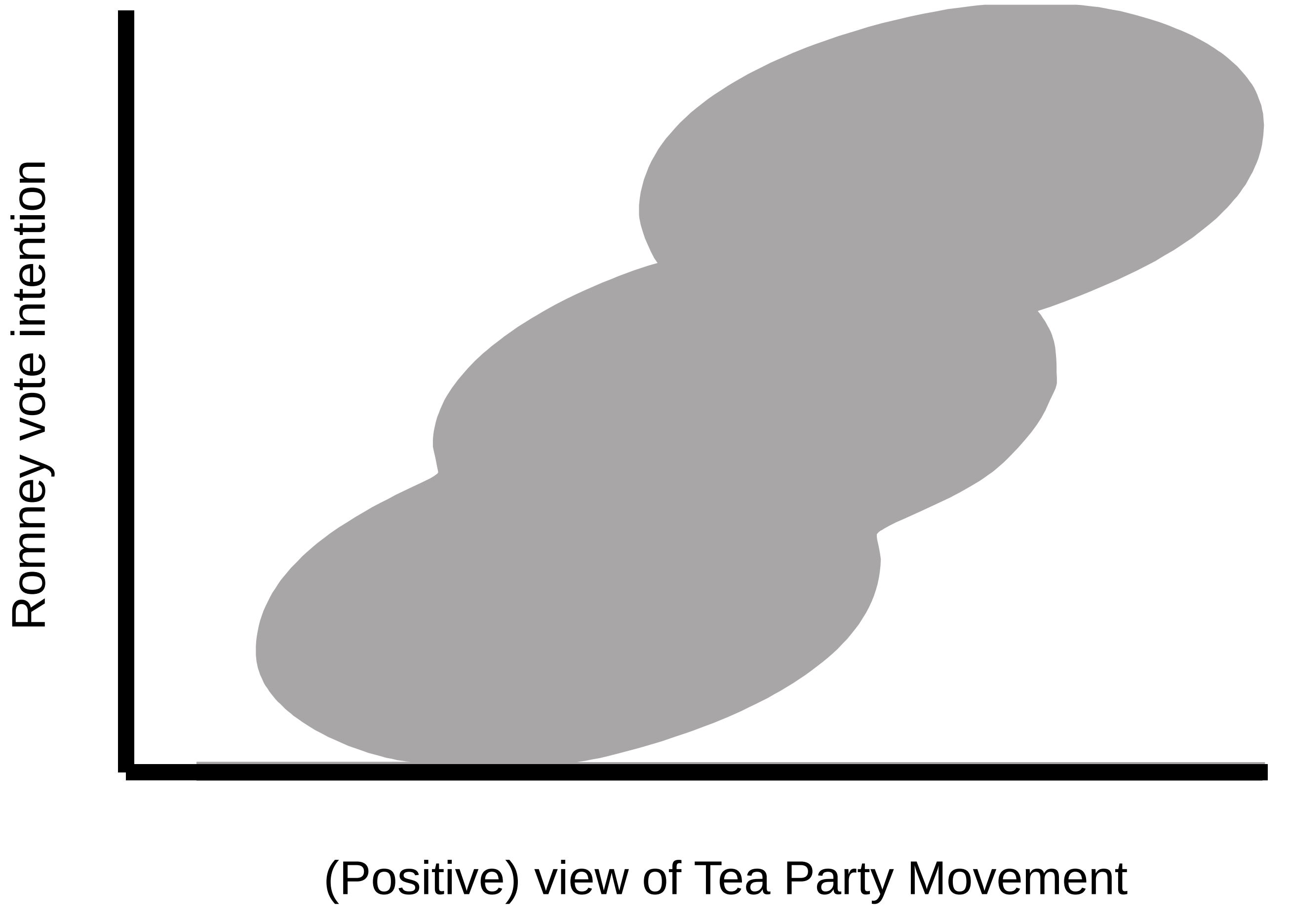
Independent picture



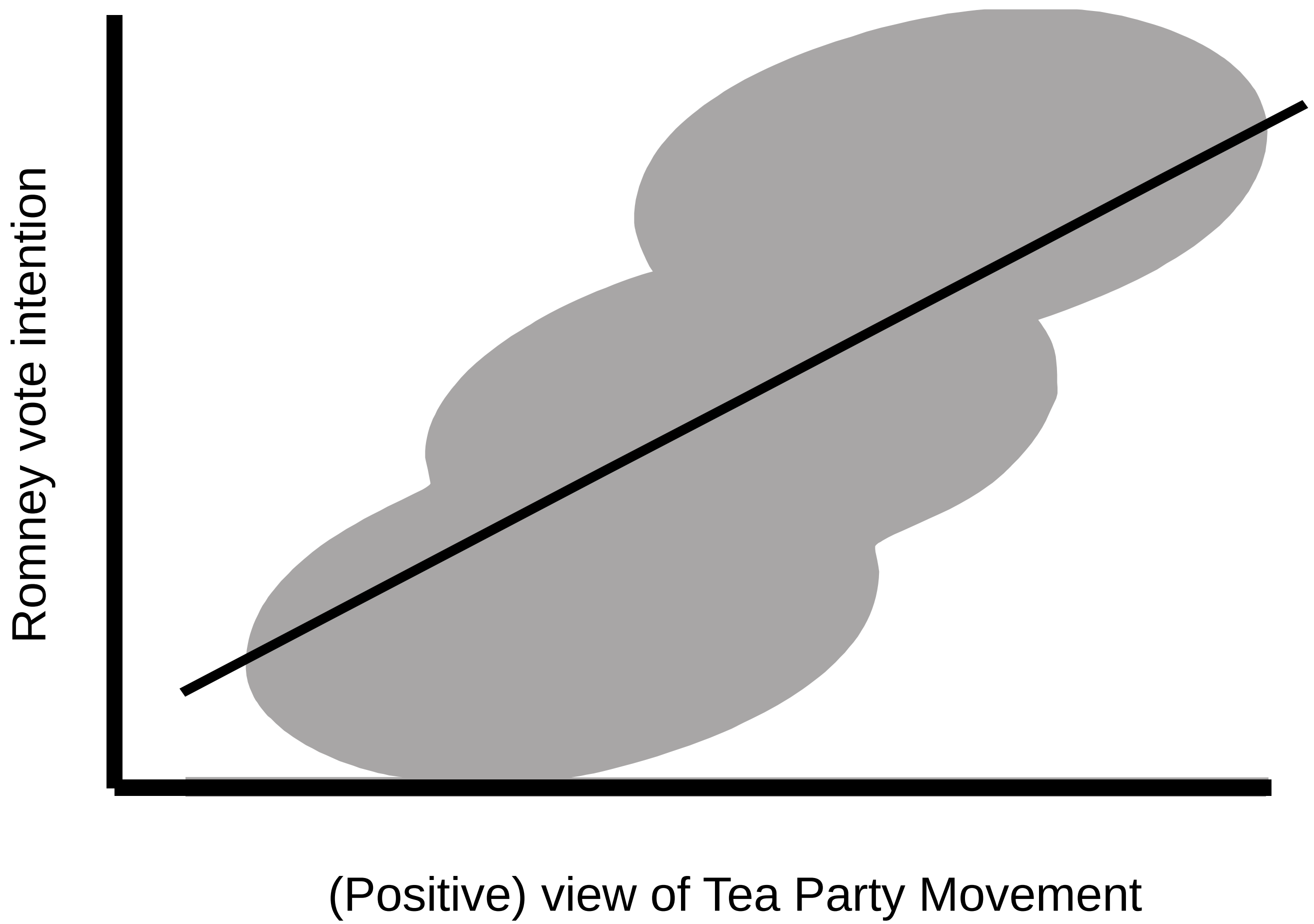
Democratic picture



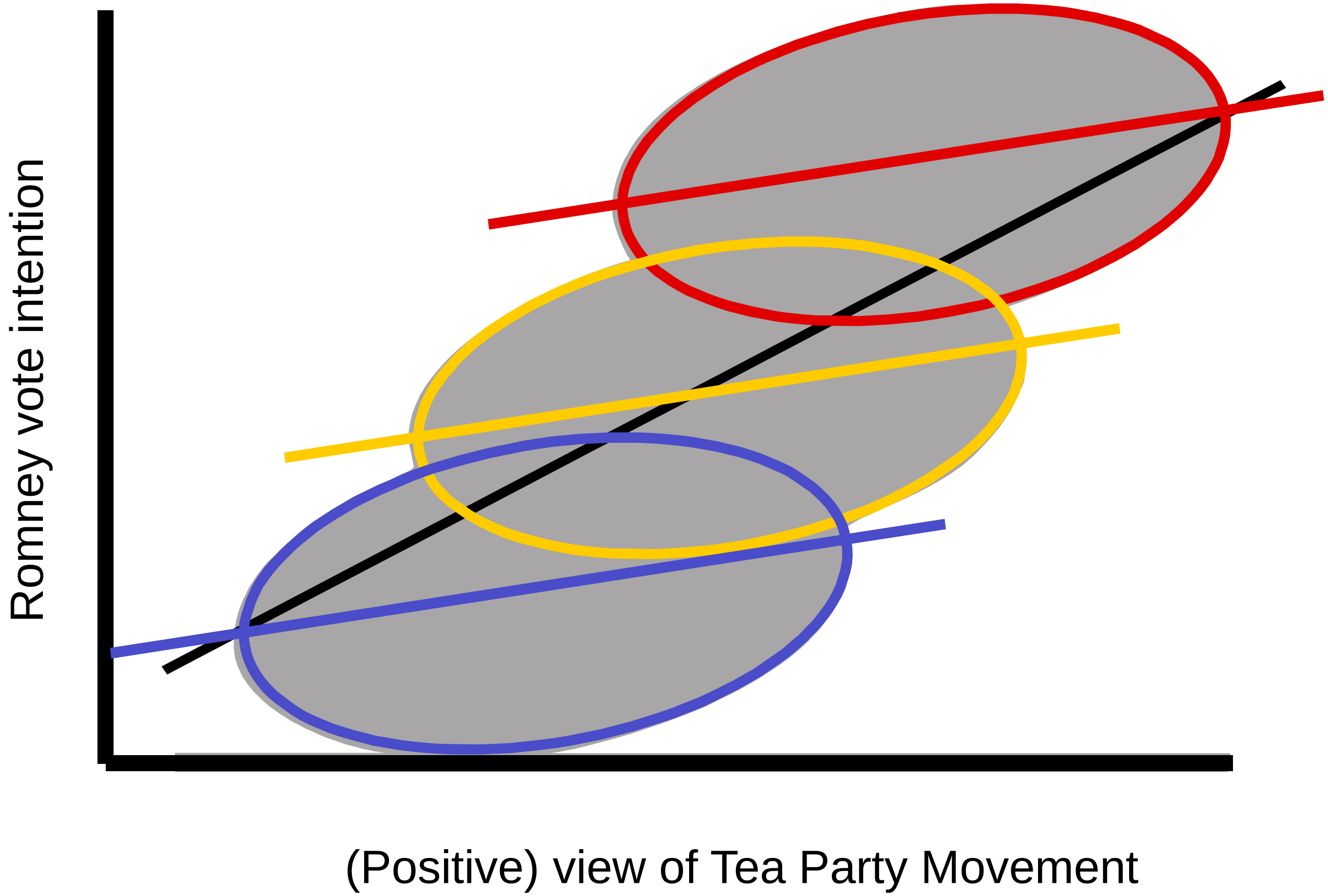
Combined data picture



Combined data picture with regression: bias!

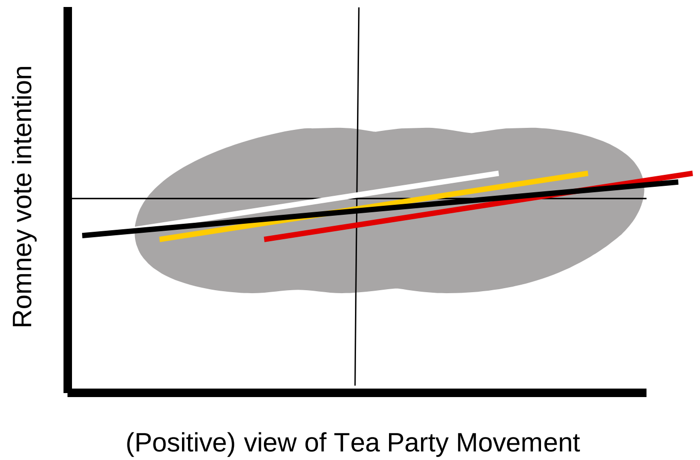


Combined data picture with “true” regression lines overlaid

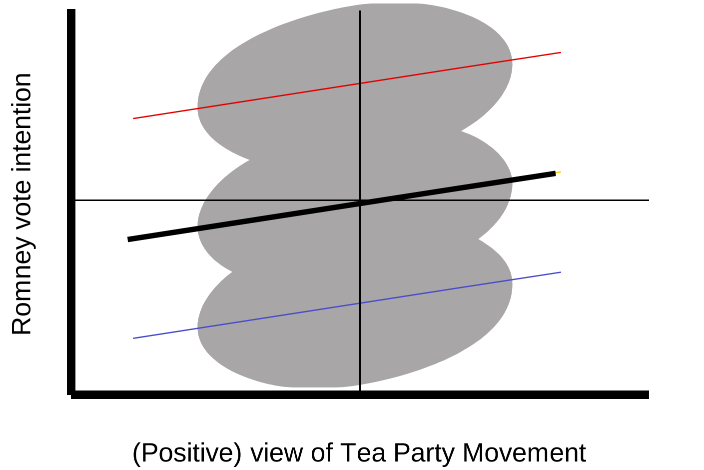


Tempting yet wrong normalizations

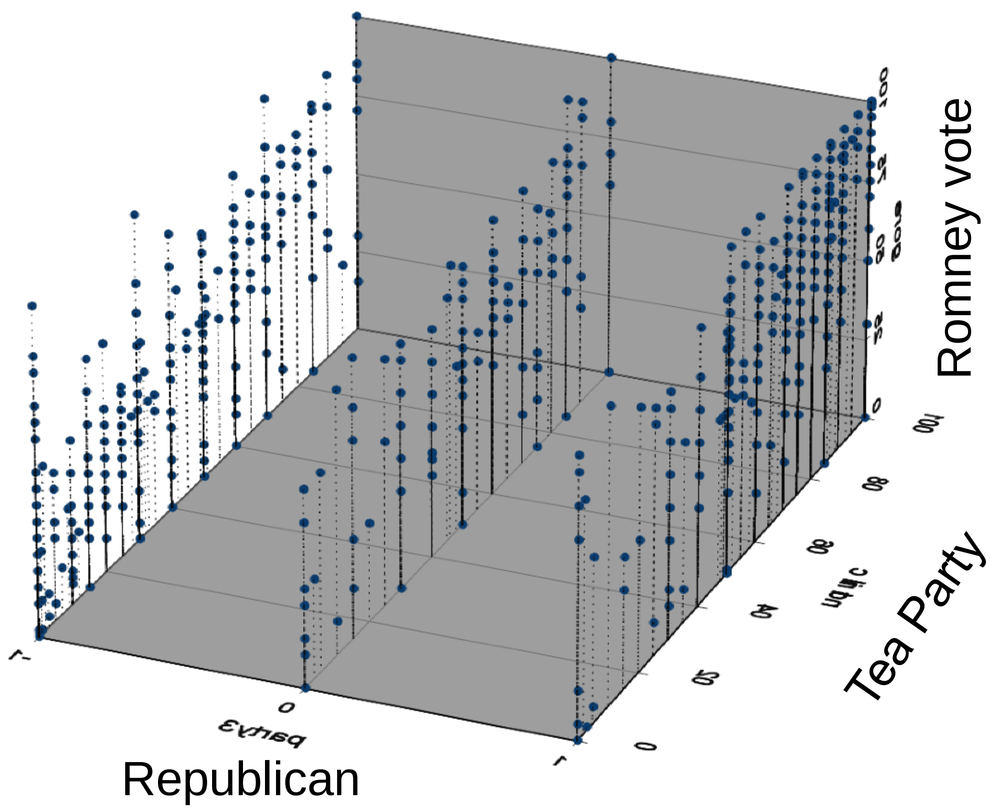
Subtract the Romney intention from the avg. Romney intention score



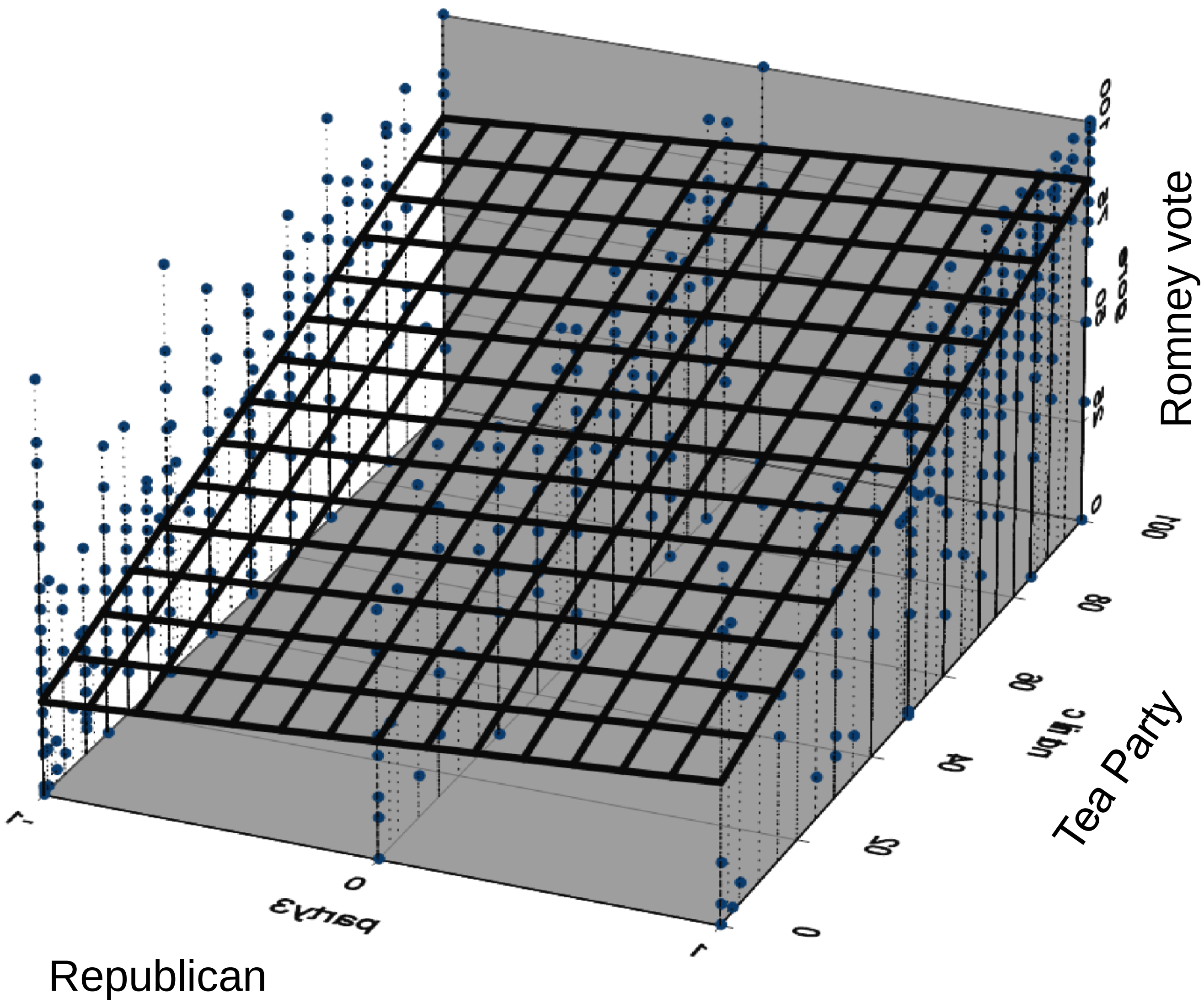
Subtract the Tea Party view from the avg. Tea Party view



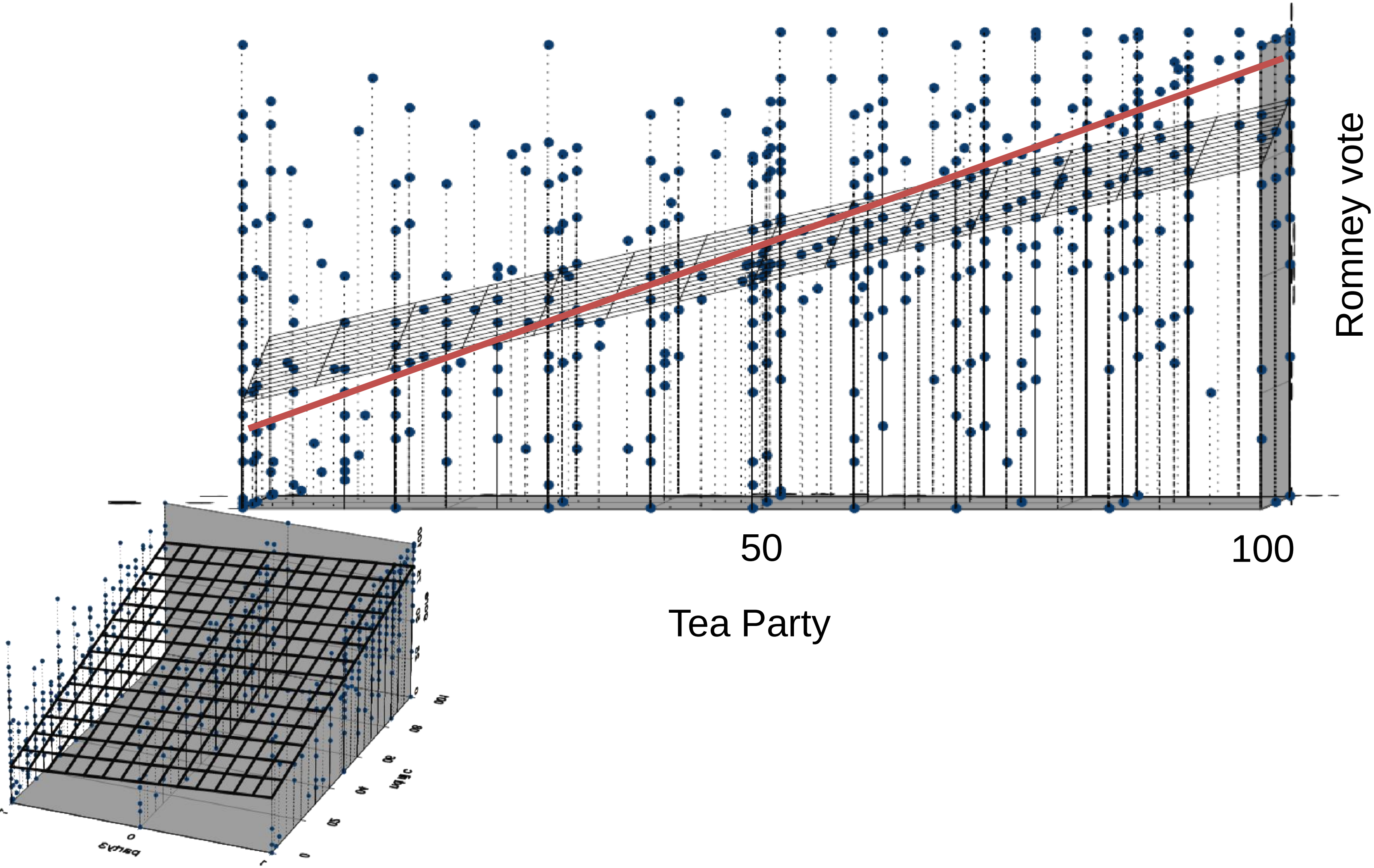
3D Relationship



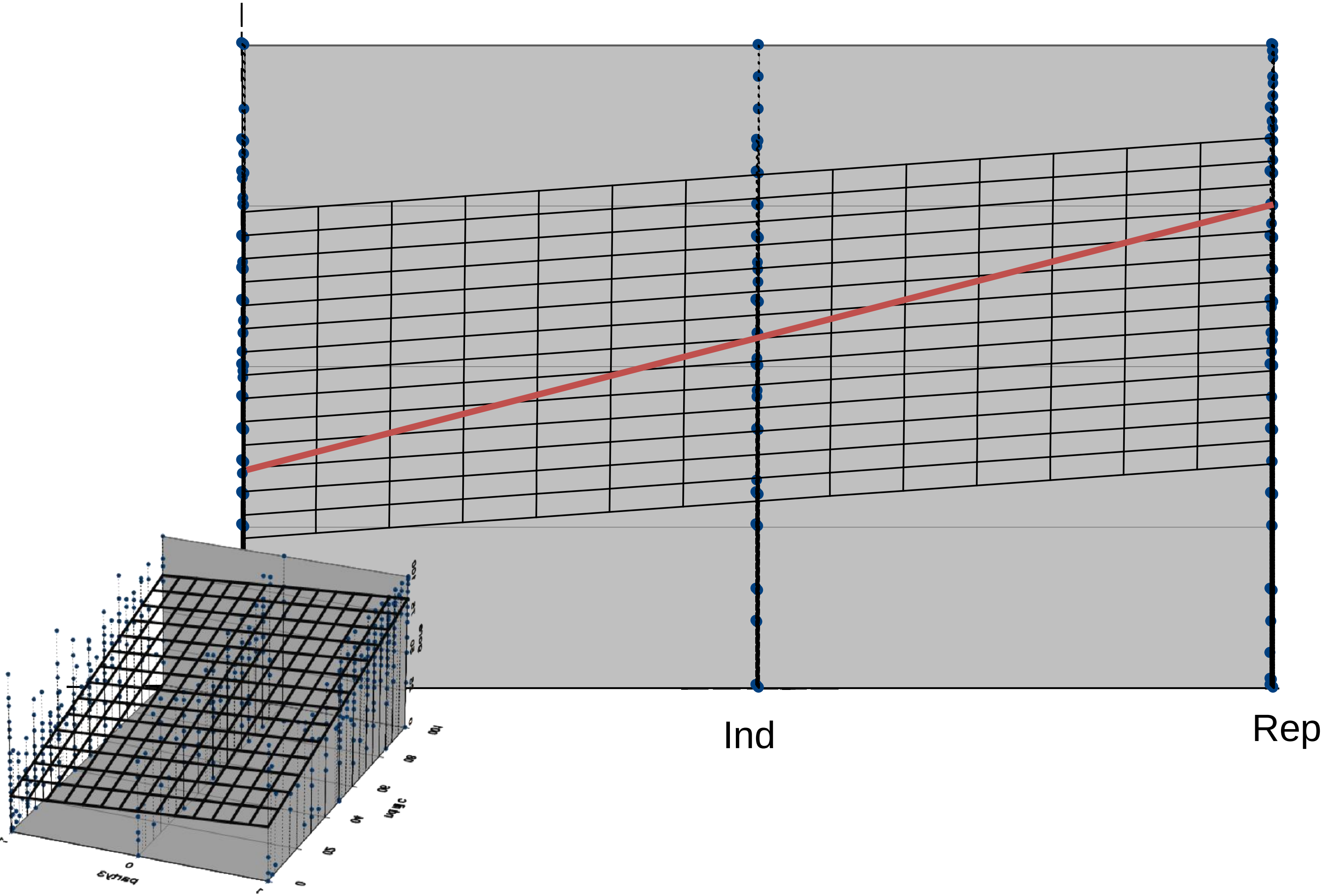
3D Linear Relationship



3D Relationship: Tea Party



3D Relationship: party



The Linear Relationship between Three Variables

Romney vote

Tea Party view

Party ID

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \varepsilon_i$$

The method of least squares (again)

Pick β_0, β_1, and β_2 to minimize

$$\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \text{ or}$$

$$\sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i - \beta_2 X_2)^2$$

The Slope Coefficients

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(\overline{Y} - Y_i)(\overline{X}_1 - X_{1,i})}{\sum_{i=1}^{n}(\overline{X}_1 - X_{1,i})^2} - \hat{\beta}_2 \frac{\sum_{i=1}^{n}(\overline{X}_1 - X_{1,i})(\overline{X}_2 - X_{2,i})}{\sum_{i=1}^{n}(\overline{X}_1 - X_{1,i})^2} \text{ and}$$

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n}(\overline{Y} - Y_i)(\overline{X}_2 - X_{1,i})}{\sum_{i=1}^{n}(\overline{X}_2 - X_{2,i})^2} - \hat{\beta}_1 \frac{\sum_{i=1}^{n}(\overline{X}_1 - X_{1,i})(\overline{X}_2 - X_{2,i})}{\sum_{i=1}^{n}(\overline{X}_2 - X_{2,i})^2}$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}_1 - \hat{\beta}_2 \overline{X}_2$$

X_1 is Tea Party view, X_2 is PID, and Y is Romney vote

The Matrix form

y_1		1	$x_{1,1}$	$x_{2,1}$	…	$x_{k,1}$
y_2		1	$x_{1,2}$	$x_{2,2}$	…	$x_{k,2}$
…		1	…	…	…	…
y_n		1	$x_{1,n}$	$x_{2,n}$	…	$x_{k,n}$

$$\beta = (X'X)^{-1} X'y$$

The Slope Coefficients (You've seen some of this before)

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(\bar{Y} - Y_i)(\bar{X}_1 - X_{1,i})}{\sum_{i=1}^{n}(\bar{X}_1 - X_{1,i})^2} - \hat{\beta}_2 \frac{\sum_{i=1}^{n}(\bar{X}_1 - X_{1,i})(\bar{X}_2 - X_{2,i})}{\sum_{i=1}^{n}(\bar{X}_1 - X_{1,i})^2} \text{ and}$$

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n}(\bar{Y} - Y_i)(\bar{X}_2 - X_{1,i})}{\sum_{i=1}^{n}(\bar{X}_2 - X_{2,i})^2} - \hat{\beta}_1 \frac{\sum_{i=1}^{n}(\bar{X}_1 - X_{1,i})(\bar{X}_2 - X_{2,i})}{\sum_{i=1}^{n}(\bar{X}_2 - X_{2,i})^2}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2$$

X_1 is Tea Party view, X_2 is PID, and Y is Romney vote

The Slope Coefficients More Simply

$$\hat{\beta}_1 = \frac{\operatorname{cov}(X_1, Y)}{\operatorname{var}(X_1)} - \hat{\beta}_2 \frac{\operatorname{cov}(X_1, X_2)}{\operatorname{var}(X_1)} \text{ and}$$

$$\hat{\beta}_2 = \frac{\operatorname{cov}(X_2, Y)}{\operatorname{var}(X_2)} - \hat{\beta}_1 \frac{\operatorname{cov}(X_1, X_2)}{\operatorname{var}(X_2)}$$

X_1 is Tea Party view, X_2 is PID, and Y is Romney vote

The Slope Coefficients More Simply (You've seen some of this before)

$$\hat{\beta}_1 = \frac{\operatorname{cov}(X_1, Y)}{\operatorname{var}(X_1)} - \hat{\beta}_2 \frac{\operatorname{cov}(X_1, X_2)}{\operatorname{var}(X_1)} \text{ and}$$

$$\hat{\beta}_2 = \frac{\operatorname{cov}(X_2, Y)}{\operatorname{var}(X_2)} - \hat{\beta}_1 \frac{\operatorname{cov}(X_1, X_2)}{\operatorname{var}(X_2)}$$

X_1 is Tea Party view, X_2 is PID, and Y is Romney vote

Multivariate slope coefficients

Tea Party effect (on Romney) in bivariate (*B*) regression

Are Romney and Party ID related?

Bivariate estimate: $\hat{\beta}_1^B = \dfrac{\text{cov}(X_1, Y)}{\text{var}(X_1)}$ vs.

Multivariate estimate: $\hat{\beta}_1^M = \dfrac{\text{cov}(X_1, Y)}{\text{var}(X_1)} - \hat{\beta}_2^M \dfrac{\text{cov}(X_1, X_2)}{\text{var}(X_1)}$

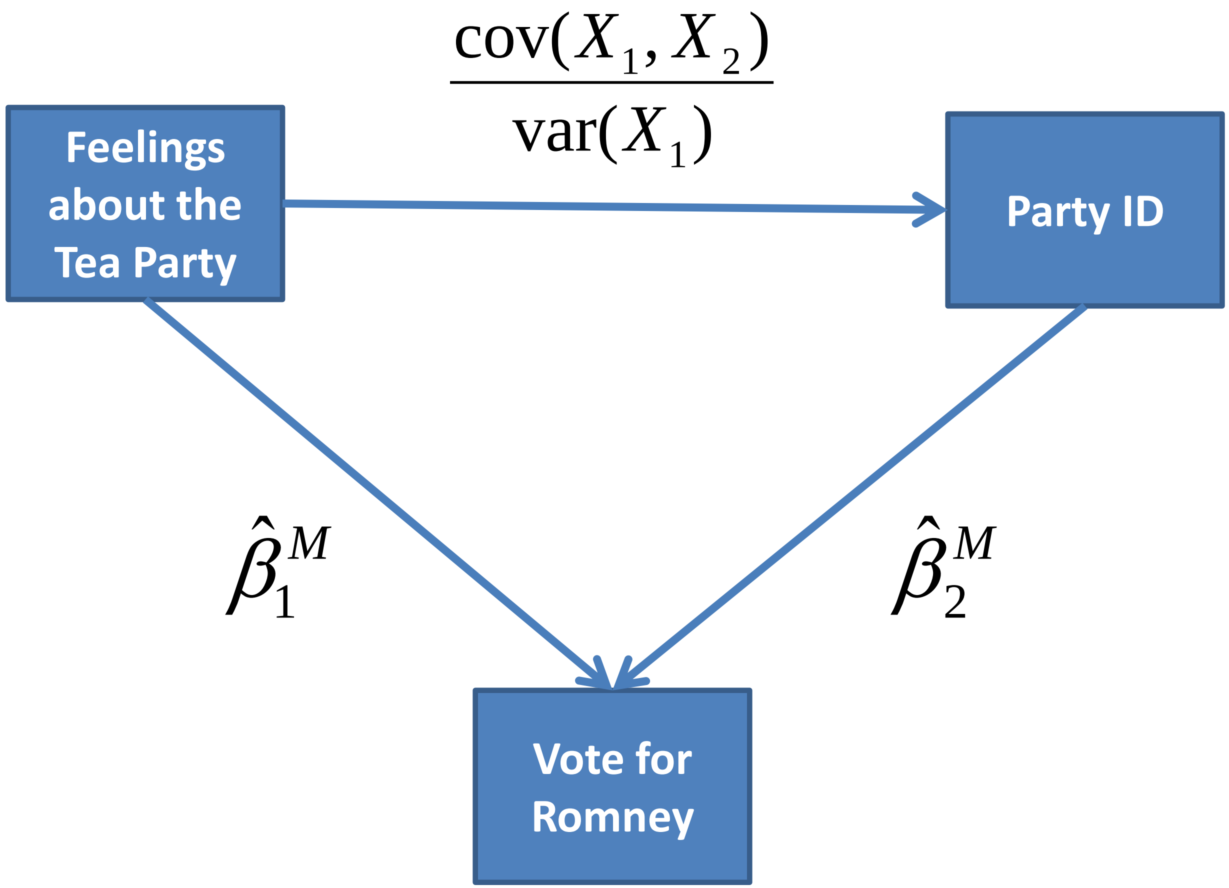
Tea Party effect (on Romney) in multivariate (*M*) regression

Are Tea Party and Party ID related?

When does $\hat{\beta}_1^B = \hat{\beta}_1^M$? Obviously, when $\hat{\beta}_2^M \dfrac{\text{cov}(X_1, X_2)}{\text{var}(X_1)} = 0$

X_1 is Tea Party view, X_2 is PID, and Y is Romney vote

The Graphical View

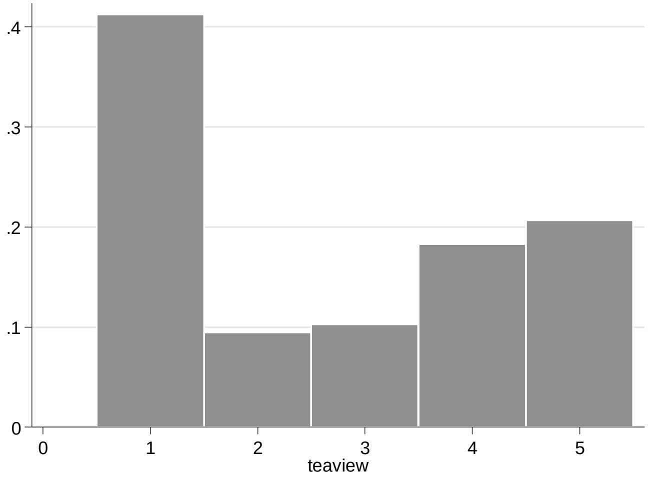


When does $\hat{\beta}_1^B = \hat{\beta}_1^M$?

Look at the data

```
. summm romneyvote teaview rep [aw=V103] if romneyvote~=.&teaview~=.&rep~=.
```

Variable	Obs	Weight	Mean	Std. Dev.	Min	Max
romneyvote	30959	30573.2277	.4705544	.4991403	0	1
teaview	30959	30573.2277	2.599286	1.553045	1	5
rep	30959	30573.2277	-.0146197	.8245732	-1	1




```
. tab rep if romneyvote~=.&teaview~=.&rep~=.
```

rep	Freq.	Percent	Cum.
-1	11,755	37.97	37.97
0	8,972	28.98	66.95
1	10,232	33.05	100.00
Total	30,959	100.00	

```
hist teaview if romneyvote~=.&teaview~=.&rep~=.,discrete
scheme(Tufte) fraction
```

The Output

```
. reg romney teaview rep [aw=V103]
(sum of wgt is   3.0573e+04)

      Source |       SS       df       MS              Number of obs =   30959
-------------+------------------------------           F(  2, 30956) =39398.65
       Model |  5537.47585     2  2768.73793           Prob > F      =  0.0000
    Residual |  2175.43137 30956  .070274951           R-squared     =  0.7179
-------------+------------------------------           Adj R-squared =  0.7179
       Total |  7712.90722 30958  .249141005           Root MSE      =  .26509

------------------------------------------------------------------------------
  romneyvote |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     teaview |   .1677419    .001252   133.98   0.000     .1652879    .1701958
         rep |   .2510075   .0023581   106.45   0.000     .2463856    .2556294
       _cons |   .0382149    .003606    10.60   0.000      .031147    .0452829
------------------------------------------------------------------------------
```

<u>Interpretation of `teaview` effect</u>: *Holding constant party identification*, a one-point increase in the Tea Party approval scale is associated with a .17 increase in the probability of voting for Romney.

Separate regressions

	(1)	(2)	(3)
Intercept	-0.18	0.48	0.04
Tea Party	0.25	--	0.17
Rep. party	--	0.44	0.25

$$\hat{\beta}_1 = \frac{\text{cov}(X_1, Y)}{\text{var}(X_1)} - \hat{\beta}_2 \frac{\text{cov}(X_1, X_2)}{\text{var}(X_1)} \text{ and}$$

$$\hat{\beta}_2 = \frac{\text{cov}(X_2, Y)}{\text{var}(X_2)} - \hat{\beta}_1 \frac{\text{cov}(X_1, X_2)}{\text{var}(X_2)}$$

Why did the Tea Party Coefficient change from 0.25 to 0.17?

```
. corr romney tea rep [aw=V103],cov
(sum of wgt is   3.0573e+04)
(obs=30959)

             | romney~e  teaview      rep
-------------+---------------------------
 romneyvote  |  .249141
    teaview  |  .607774  2.41195
        rep  |  .306451  .809492  .679921
```

The Calculations

$$\hat{\beta}_1^B = \frac{\text{cov}(romney, teaview)}{\text{var}(teaview)} = \frac{0.60774}{2.41195} = 0.2520$$

$$\hat{\beta}_1^M = \frac{\text{cov}(romney, teaview)}{\text{var}(teaview)} - \hat{\beta}_2^M \frac{\text{cov}(teaview, rep)}{\text{var}(teaview)}$$

$$= \frac{0.60774}{2.41195} - 0.2510 \frac{0.809492}{2.41195}$$

$$= 0.2520 - 0.0842$$

$$= 0.1678 \sim 0.1677$$

```
. corr romney tea rep [aw=V103],cov
(sum of wgt is   3.0573e+04)
(obs=30959)

             | romney~e  teaview      rep
-------------+---------------------------
 romneyvote  |  .249141
    teaview  |  .607774  2.41195
        rep  |  .306451  .809492  .679921
```

Another way of thinking about this

Rewrite

$$\hat{\beta}_1^M = \frac{\text{cov}(romney, teaview)}{\text{var}(teaview)} - \hat{\beta}_2^M \frac{\text{cov}(teaview, rep)}{\text{var}(teaview)}$$

as

$$\frac{\text{cov}(romney, teaview)}{\text{var}(teaview)} = \hat{\beta}_1^M + \hat{\beta}_2^M \frac{\text{cov}(teaview, rep)}{\text{var}(teaview)}$$

Total effect = Direct effect + indirect effect

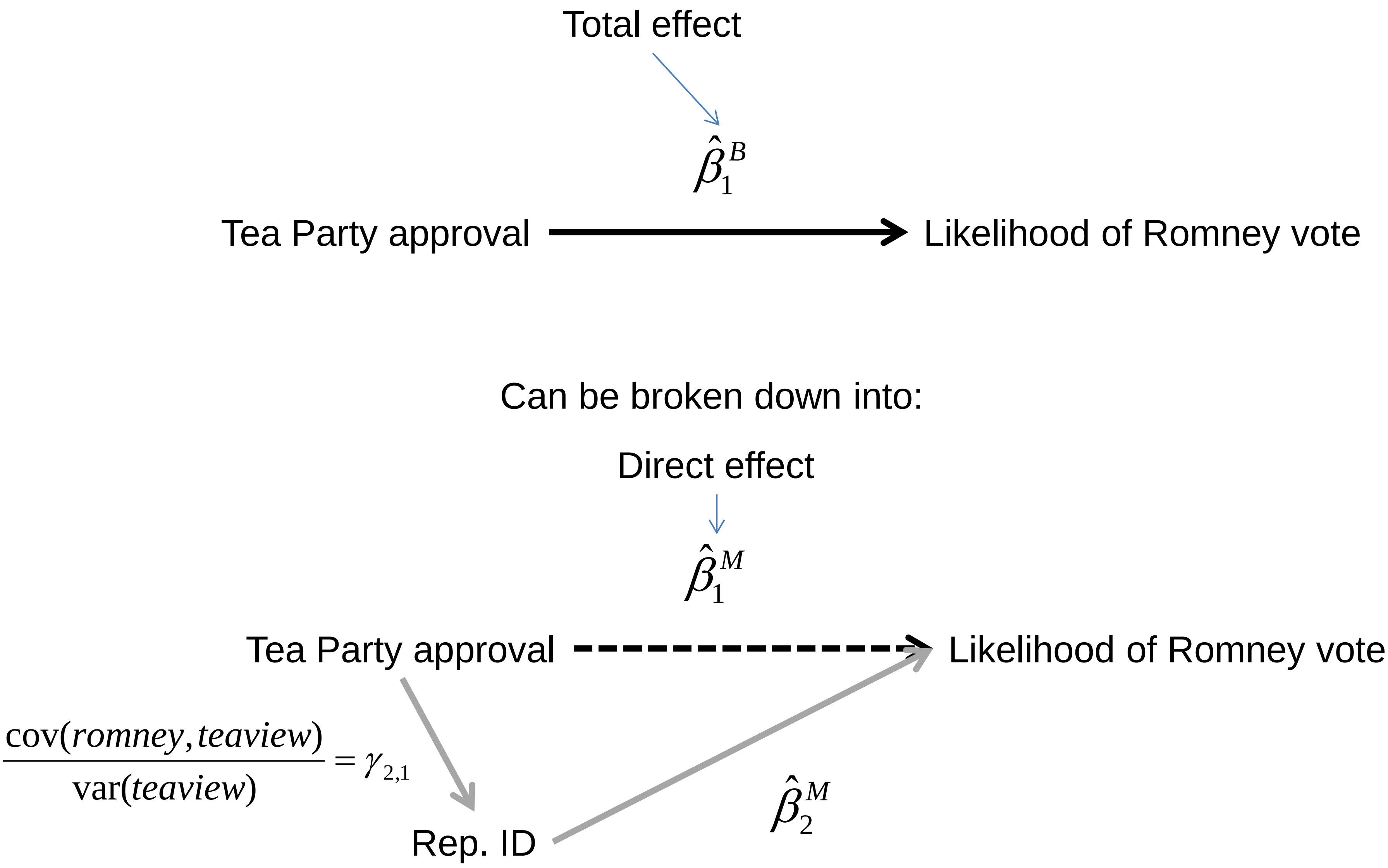
The Total Effect of the Tea Party view on the Romney vote (.25) can be Broken down into a direct effect of .17, plus an indirect effect (though party) of .08

Graphical way of thinking about this

Total effect

$$\text{Tea Party approval} \xrightarrow{\hat{\beta}_1^{B}} \text{Likelihood of Romney vote}$$

Graphical way of thinking about this



Graphical way of thinking about this

Total effect

$\hat{\beta}_1^B$

Tea Party approval $\longrightarrow$ Likelihood of Romney vote

Can be broken down into:

Direct effect

$\hat{\beta}_1^M$

Tea Party approval - - - - - → Likelihood of Romney vote

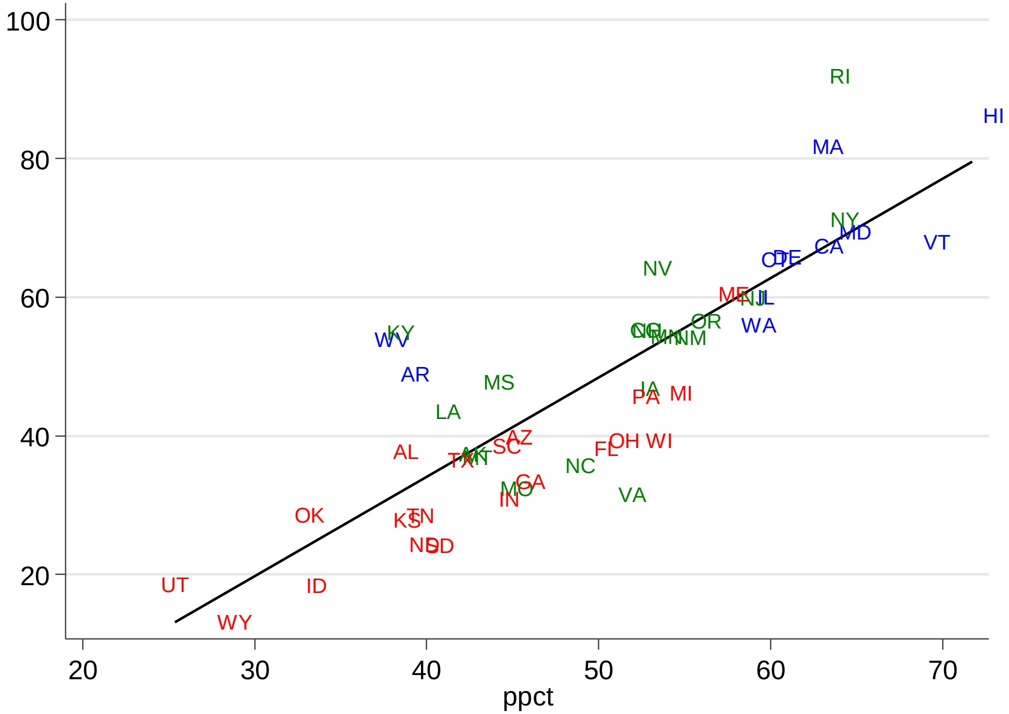
$$\frac{\text{cov}(romney, teaview)}{\text{var}(teaview)} = \gamma_{2,1}$$

Rep. ID

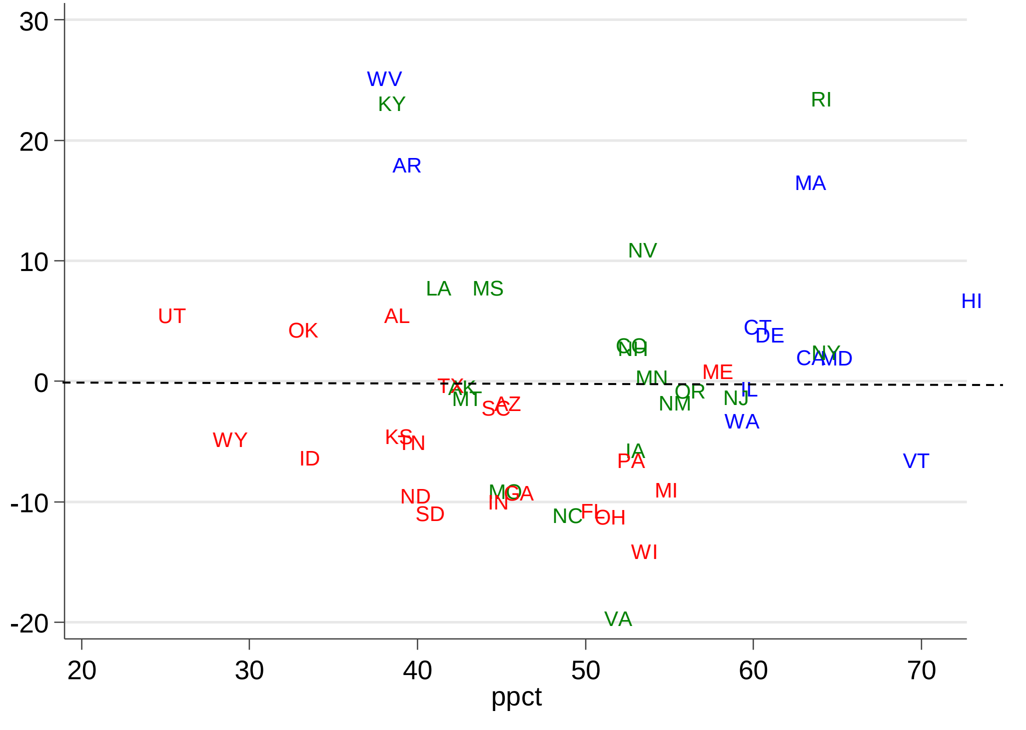
$\hat{\beta}_2^M$

Indirect effect

Return to the state legislative example



Red = redistricting controlled by Reps.
Blue = redistricting controlled by Dems.
Green = redistricting controlled by neither



30
20
10
0
-10
-20
20
30
40
50
60
70
ppct
WV
KY
RI
AR
MA
NV
LA
MS
HI
UT
OK
AL
CT
DE
CO
NH
NY
CA
MD
ME
TX
AK
MT
SC
AZ
MN
OR
IL
NJ
NM
WA
WY
KS
TN
ID
IA
PA
VT
MO
GA
MI
ND
IN
SD
NC
FL
OH
WI
VA

```
. list state ry after10 in 1/10

     +-------------------------------------+
     |          state          ry  after10 |
     |-------------------------------------|
  1. | West Virginia   25.17959          1 |
  2. |  Rhode Island   23.48404          0 |
  3. |      Kentucky   23.12402          0 |
  4. |      Arkansas   18.00081          1 |
  5. | Massachusetts   16.55731          1 |
     |-------------------------------------|
  6. |        Nevada   10.97821          0 |
  7. |   Mississippi   7.828268          0 |
  8. |     Louisiana   7.805305          0 |
  9. |        Hawaii    6.73896          1 |
 10. |          Utah   5.545974         -1 |
     +-------------------------------------+
```

```
     +---------------------------------------+
     |           state           ry  after10 |
     |---------------------------------------|
 41. |         Georgia    -9.231257       -1 |
 42. |    North Dakota    -9.460065       -1 |
 43. |         Indiana    -9.967912       -1 |
 44. |         Florida    -10.72338       -1 |
 45. |    South Dakota    -10.92215       -1 |
     |---------------------------------------|
 46. |  North Carolina    -11.10709        0 |
 47. |            Ohio    -11.19955       -1 |
 48. |       Wisconsin    -14.06958       -1 |
 49. |        Virginia    -19.61035        0 |
     +---------------------------------------+
```

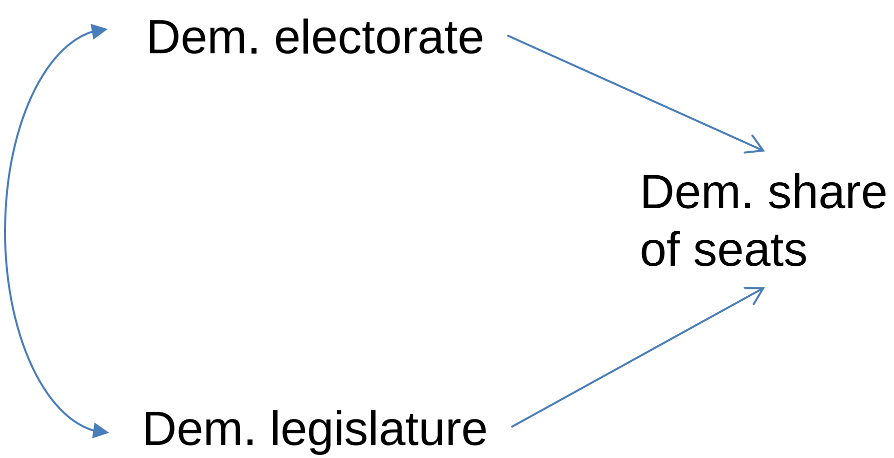
After10 = 1 if Dems control, -1 if Reps control, 0 if neither controls

	(1)	(2)	(3)
Obama vote	1.43 (0.14)	---	1.09 (0.14)
Dem. state	---	16.33 (2.33)	8.25 (1.82)
Intercept	-23.25 (6.81)	50.51 (1.85)	-4.93 (7.01)
N	49	49	49
S.E.R.	9.79	12.62	8.23
R^2	.71	.51	.80

Accounting for the total effect

	Total effect	Direct effect	Indirect effect
Obama vote	1.43	1.09 (76%)	0.34 (24%)
Party control of districting	16.33	8.25 (51%)	8.08 (49%)

	(1)	(2)	(3)
Obama vote	1.43 (0.14)	---	1.09 (0.14)
Dem. state	---	16.33 (2.33)	8.25 (1.82)
Intercept	-23.25 (6.81)	50.51 (1.85)	-4.93 (7.01)
N	49	49	49
S.E.R.	9.79	12.62	8.23
R^2	.71	.51	.80



Drinking and Greek Life Example

- Why is there a correlation between living in a fraternity/sorority house and drinking?
  - Greek organizations often emphasize social gatherings that have alcohol. The effect is being in the Greek organization itself, not the house.
  - There's something about the House environment itself.

Dependent variable: Times Drinking in Past 30 Days

C8. When did you last have a drink (that is more than just a few sips)?

- ◯ I have never had a drink ⟶ Skip to C22 (page 10)
- ◯ Not in the past year ⟶ Skip to C22 (page 10)
- ◯ More than 30 days ago, but in the past year ⟶ Skip to C17 (page 8)
- ◯ More than a week ago, but in the past 30 days ⟶ Go to C9
- ◯ Within the last week ⟶ Go to C9

C9. On how many occasions have you had a drink of alcohol in the <u>past 30 days</u>? (Choose one answer.)

1 ◯ Did not drink in the last 30 days
2 ◯ 1 to 2 occasions
3 ◯ 3 to 5 occasions
4 ◯ 6 to 9 occasions
5 ◯ 10 to 19 occasions
6 ◯ 20 to 39 occasions
7 ◯ 40 or more occasions

```
. infix age 10-11 residence 16 greek 24 screen 102
timespast30 103 howmuchpast30 104 gpa 278-279 studying 281
timeshs 325 howmuchhs 326 socializing 283 stwgt_99 475-493
weight99 494-512 using da3818.dat,clear
(14138 observations read)

. recode  timespast30 (1=0) (2=1.5) (3=4) (4=7.5) (5=14.5)
(6=29.5) (7=45)
(timespast30: 6571 changes made)

. replace timespast30=0 if screen<=3
(4631 real changes made)
```

```
. tab timespast30
```

timespast30	Freq.	Percent	Cum.
0	4,652	33.37	33.37
1.5	2,737	19.64	53.01
4	2,653	19.03	72.04
7.5	1,854	13.30	85.34
14.5	1,648	11.82	97.17
29.5	350	2.51	99.68
45	45	0.32	100.00
Total	13,939	100.00	

Key explanatory variables

- Live in fraternity/sorority house
  – Indicator variable (dummy variable)
  – Coded 1 if live in, 0 otherwise
- Member of fraternity/sorority
  – Indicator variable (dummy variable)
  – Coded 1 if member, 0 otherwise

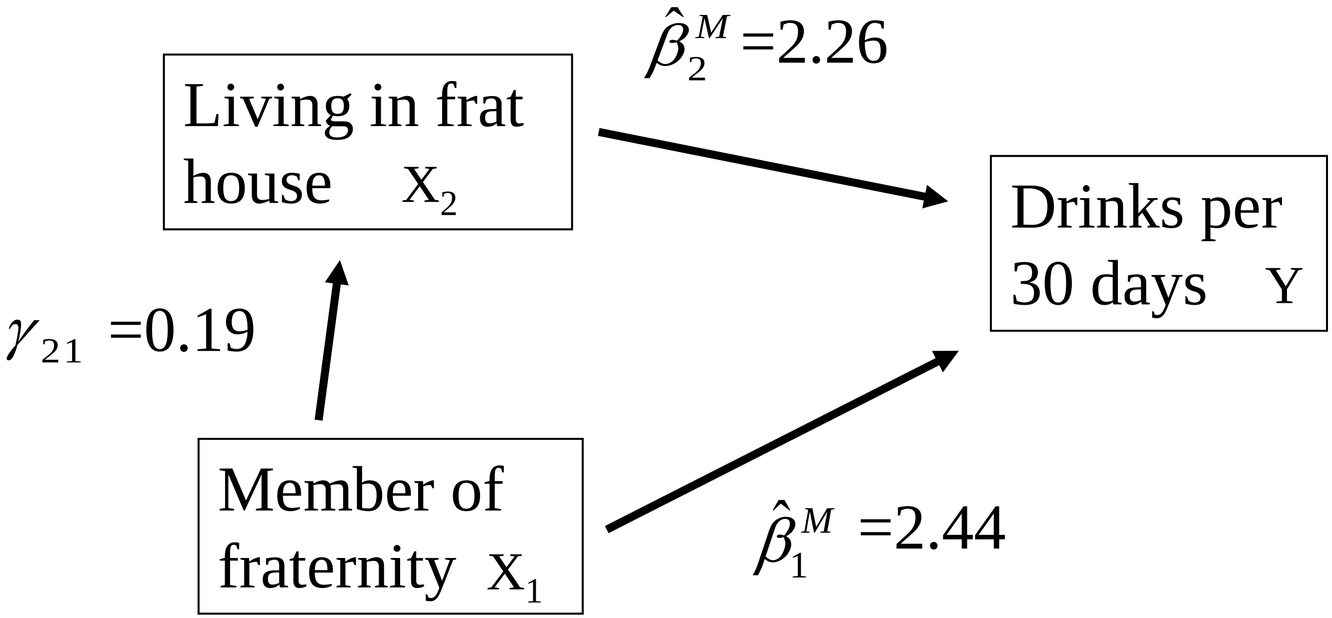
Three Regressions

Dependent variable: number of times drinking in past 30 days			
Live in frat/sor house (indicator variable)	4.44 (0.35)	---	2.26 (0.38)
Member of frat/sor (indicator variable)	---	2.88 (0.16)	2.44 (0.18)
Intercept	4.54 (0.56)	4.27 (0.059)	4.27 (0.059)
S.E.R.	6.49	6.44	6.44
R^2	.011	.023	.025
N	13,876	13,876	13,876

What is the substantive interpretation of the coefficients?

Note: Standard errors in parentheses. Corr. Between living in frat/sor house and being a member of a Greek organization is .42

The Picture



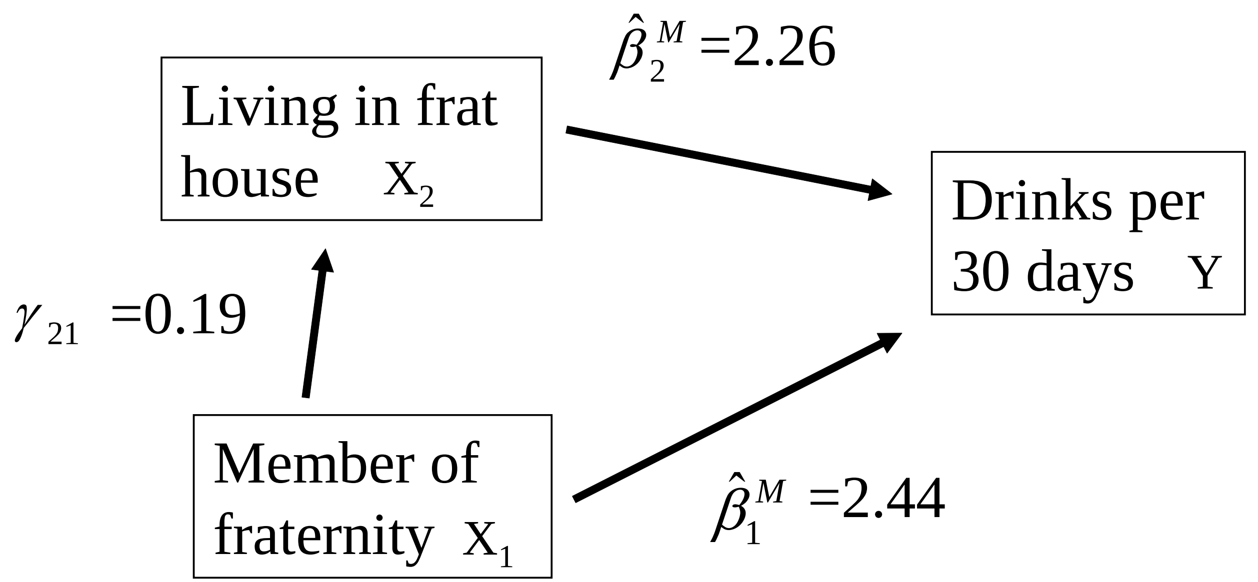
Remember that:

$$\hat{\beta}_1^B = 2.88$$

Accounting for the total effect

$$\hat{\beta}_1^B = \hat{\beta}_1^M + \hat{\beta}_2^M \gamma_{21}$$

Total effect = Direct effect + indirect effect



Accounting for the effects of frat house living and Greek membership on drinking

From bivariate regressions

From multiple regressions

From accounting identity: T=D+I

Effect	Total	Direct	Indirect
Member of Greek org.	2.88	2.44 (85%)	0.44 (15%)
Live in frat/ sor. house	4.44	2.26 (51%)	2.18 (49%)

Implications of for model-building

- Q: When do you decide whether to "control for" another variable?
  - A1: When excluding another variable(s) would lead to a biased estimate of the effect you are interested in
    - The omitted variable is correlated with the independent variable of interest **and**
    - The omitted variable is also related (statistically) to the dependent variable.*
  - A2: When theory or the question tell you to
  - A3: to deal with efficiency (covered after spring break)

*If you don't do this, you commit **omitted variables bias**

Standardized regression

- Used to try and judge which variables are "more important" in a multiple regression
- Other standardizations are possible (e.g., putting all variables into a 0,1 interval)
- Less informative than regressing on raw values or the 0,1 interval, but is useful to know about because it is so common.

The idea

- Transform **every** variable according to the following formula:

$$newvar = \frac{oldvar - \overline{oldvar}}{\sigma_{oldvar}}$$

- Do the regression on these “z-scores”
  - The intercept drops away
  - In bivariate regression, the standardized coefficient is equal to the correlation coefficient
  - The coefficients are sometimes called “BETA” coefficients (very confusingly)

Example: Influence over state legislature composition

- Variables:
  - **hpct:** =% of state House of Reps. that is Dem.
  - **ppct** = % of state vote that went to Obama
  - **after10** = 1 if government controlled by Dems after 2010, -1 if controlled by Reps., 0 is split control

```
. summ hpct ppct after10 if after10~=.

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        hpct |        49      47.508    17.86246   13.33333         92
        ppct |        49     49.3619    10.46804   25.37381   71.70385
     after10 |        49   -.1836735    .7819172         -1          1
```

Write out the transformed variables because the Office Equation Editor tried to destroy this presentation.

Comparison of regular regression and standardized regression

```
. reg hpct ppct after10,beta

      Source |       SS       df       MS              Number of obs =      49
-------------+------------------------------           F(  2,    46) =   89.98
       Model |  12197.5002     2  6098.7501           Prob > F      =  0.0000
    Residual |  3117.73096    46   67.77676           R-squared     =  0.7964
-------------+------------------------------           Adj R-squared =  0.7876
       Total |  15315.2312    48  319.067316           Root MSE      =  8.2327

------------------------------------------------------------------------------
        hpct |      Coef.   Std. Err.      t    P>|t|                     Beta
-------------+----------------------------------------------------------------
        ppct |   1.093083   .1361691     8.03   0.000                 .6405857
     after10 |   8.251803   1.822985     4.53   0.000                 .3612173
       _cons |  -4.933005    7.01148    -0.70   0.485                        .
------------------------------------------------------------------------------
```

Compare all this with normalizing variables to the 0-1 scale

$$X' = \frac{X - \min(X)}{\max(X) - \min(X)}$$

Comparison of 0-1 normalization and standardized regression

```
. reg hpct01 ppct01 after1001

      Source |       SS       df       MS              Number of obs =      49
-------------+------------------------------           F(  2,    46) =   89.98
       Model |  1.97099561     2  .985497803           Prob > F      =  0.0000
    Residual |  .503794544    46  .010952055           R-squared     =  0.7964
-------------+------------------------------           Adj R-squared =  0.7876
       Total |  2.47479015    48  .051558128           Root MSE      =  .10465

------------------------------------------------------------------------------
      hpct01 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ppct01 |   .6437588   .0801953     8.03   0.000     .4823341    .8051835
   after1001 |   .2097907   .0463469     4.53   0.000     .1164993    .3030822
       _cons |   .0154771   .0379184     0.41   0.685    -.0608486    .0918028
------------------------------------------------------------------------------
```

Comparison of 0-1 normalization and standardized regression, this time not transforming dep. var.

```
. reg hpct ppct01 after1001

      Source |       SS       df       MS              Number of obs =      49
-------------+------------------------------           F(  2,    46) =   89.98
       Model |  12197.5003     2  6098.75014           Prob > F      =  0.0000
    Residual |  3117.73089    46  67.7767585           R-squared     =  0.7964
-------------+------------------------------           Adj R-squared =  0.7876
       Total |  15315.2312    48  319.067316           Root MSE      =  8.2327

------------------------------------------------------------------------------
        hpct |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ppct01 |   50.64257    6.30872     8.03   0.000     37.94378    63.34137
   after1001 |   16.50361    3.64597     4.53   0.000      9.16465    23.84256
       _cons |   14.55087   2.982924     4.88   0.000     8.546553    20.55518
------------------------------------------------------------------------------
```

Three Ways to Normalize Vars

```
. summm hpct ppct after10 if after10~=.

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        hpct |        49      47.508    17.86246   13.33333         92
        ppct |        49     49.3619    10.46804   25.37381   71.70385
     after10 |        49   -.1836735    .7819172         -1          1
```

Write down results approach

```
. gen hpct01=(hpct-13.33333)/(92-13.333)
(1 missing value generated)

. gen ppct01=(ppct-25.37381)/(71.70385-25.37381)

. gen after1001=(after10+1)/(1+1)
(1 missing value generated)

. summ *01 if after10~=.
```

Brute force programming approach

```
. quietly summ hpct
. gen hpct_min=r(min)
. gen hpct_max=r(max)
. gen hpct01=(hpct-hpct_min)/(hpct_max-hpct_min)
(1 missing value generated)

. quietly summ ppct
. local ppct_min=r(min)
. local ppct_max=r(max)
. gen ppct01=(ppct-`ppct_min')/(`ppct_max'-`ppct_min')

. quietly summ after10
. local after10_min=r(min)
. local after10_max=r(max)
.
gen after1001=(after10-`after10_min')/(`after10_max'-`after10_min')
(1 missing value generated)
```

Three Ways to Normalize Vars

```
. summm hpct ppct after10 if after10~=.

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        hpct |        49      47.508    17.86246   13.33333         92
        ppct |        49     49.3619    10.46804   25.37381   71.70385
     after10 |        49   -.1836735    .7819172         -1          1
```

Brute force programming approach

```
. quietly summ hpct
. gen hpct_min=r(min)
. gen hpct_max=r(max)
. gen hpct01=(hpct-hpct_min)/(hpct_max-hpct_min)
(1 missing value generated)

. quietly summ ppct
. local ppct_min=r(min)
. local ppct_max=r(max)
. gen ppct01=(ppct-`ppct_min')/(`ppct_max'-`ppct_min')

. quietly summ after10
. local after10_min=r(min)
. local after10_max=r(max)
.
gen after1001=(after10-`after10_min')/(`after10_max'-`after10_min')
(1 missing value generated)
```

Elegant programming approach

```
foreach v of varlist hpct ppct after10 {
           quietly summ `v'
           gen `v'_min=r(min)
           gen `v'_max=r(max)
           gen `v'01=(`v'-`v'_min)/(`v'_max-`v'_min)
           drop `v'_min `v'_max
}
```