Research Questions

1. Is there an effect of news exposure on political knowledge?

2. Is there an effect of news exposure on political knowledge after controlling for demographic and political covariates?

3. Does education level moderate the effect of news exposure on political knowledge after controlling for demographic and political covariates?

Data and Analytic Strategy

DATA

The data in *pew.csv* come from a telephone survey conducted by The Pew Research Center for The People & The Press in February 2007. The data represent a probability sample of 100 adults in the U.S.

Look at the codebook every time you encounter data!

```
# A tibble: 100 × 9
      id knowledge  news   age education  male ideology party       engagement
   <dbl>     <dbl> <dbl> <dbl>     <dbl> <dbl>    <dbl> <chr>            <dbl>
 1     1        10    60  31.1        16     0     65.1 Democrat          19.8
 2     2        18    20  44.4        12     1     41   Independent       71.4
 3     3        88    59  53.7        16     1      3.4 Democrat          97.2
 4     4        75    89  62.5        13     1     50.6 Independent       61.5
 5     5        93    43  48.3        16     0     35.3 Democrat          53.4
 6     6        13    61  24.3         9     0      0.1 Independent       41.1
 7     7        37    45  59.2        12     0     13.7 Independent       96.3
 8     8        66    59  93.9        16     1     27.1 Democrat          86.8
 9     9        49    68  37          16     0     67.4 Independent       58.2
10    10        50    44  37.1        14     1     54.8 Democrat          45.2
```

4

CATEGORIZE ATTRIBUTES

<u>Outcome</u>
Political knowledge
(knowledge)

<u>Focal Predictor(s)</u>
News exposure
(news)

<u>Covariate(s)</u>
Demographic
- Age (age)
- Education level (education)
- Sex (male)

Political
- Political engagement (engagement)
- Political ideology (ideology)
- Political party (party)

GENERAL ANALYTIC STRATEGY

Explore Attributes

Examine attributes (plots, numerical summaries)

1. Outcome
2. Focal predictor
3. Covariates

Explore Relationships

Examine relationship between focal predictor/covariates and outcome
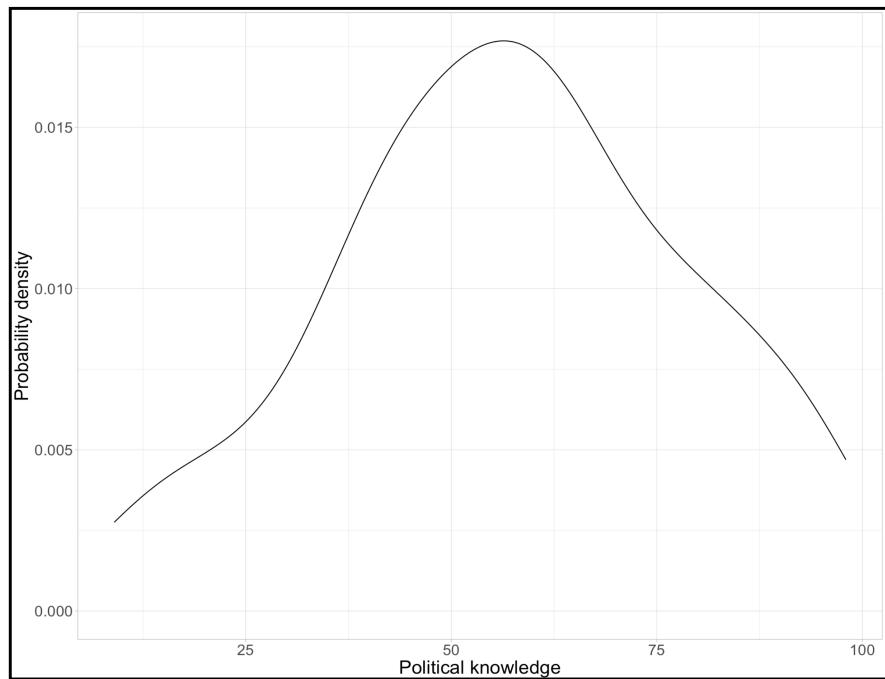
1. Scatterplot
2. Correlations

Modeling

Fit models

1. Model-level output (`glance()`)
2. Coefficient-level output (`tidy()`)
3. Examine residuals
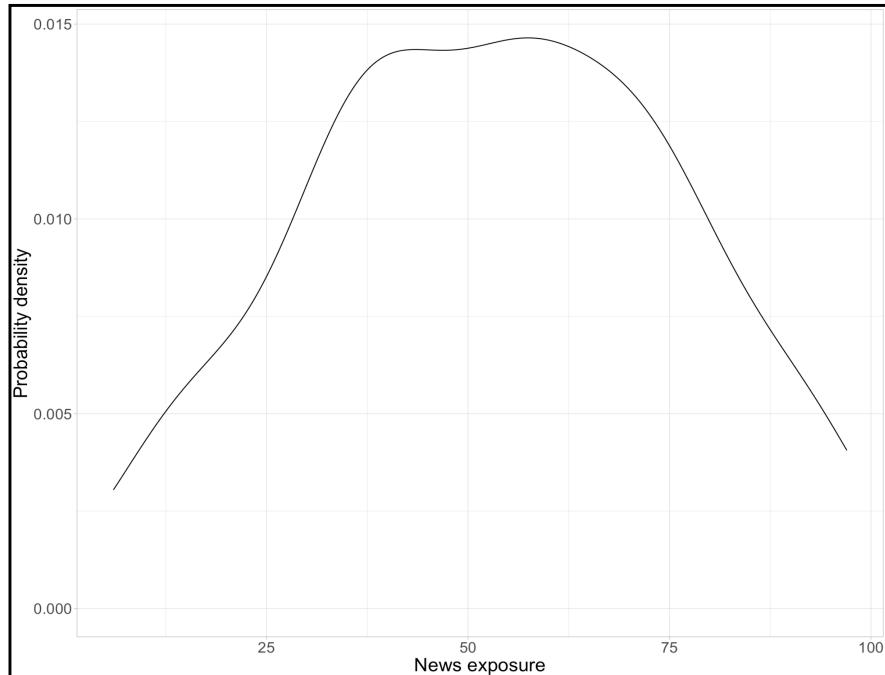
Exploration and Relationships

OUTCOME

M	SD
57.1	21.5

FOCAL PREDICTOR
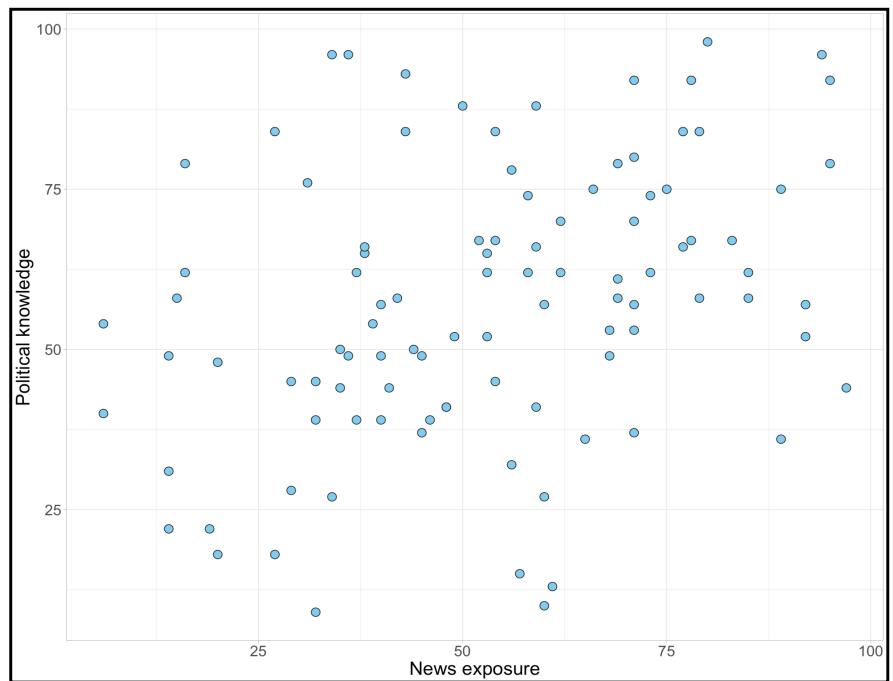
M	SD
53.3	22.6

COVARIATES
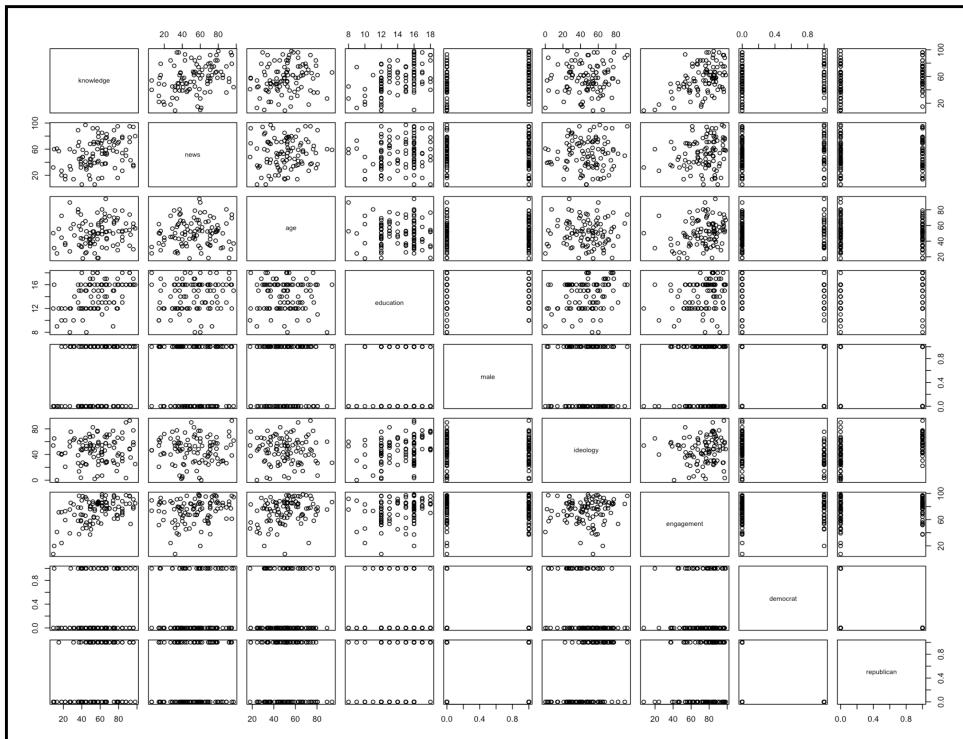
	skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
1	age	0	1	51.0	16.5	18	37.2	51	62.5	93.9	▁▅▇▇▃▁
2	education	0	1	14.2	2.44	8	12	15	16	18	▂▅▅▇▃
3	male	0	1	0.52	0.502	0	0	1	1	1	▇▁▁▁▇
4	ideology	0	1	46.9	20.1	0.1	30.9	48.4	59.7	92.8	▂▅▇▅▂
5	engagement	0	1	73.3	18.1	7.5	62.9	77	85.8	98.3	▁▁▂▅▇
6	democrat	0	1	0.29	0.456	0	0	0	1	1	▇▁▁▁▃
7	republican	0	1	0.43	0.498	0	0	0	1	1	▇▁▁▁▅

RELATIONSHIP BETWEEN OUTCOME AND FOCAL PREDICTOR

$r = 0.345$

RELATIONSHIP BETWEEN ALL NUMERIC ATTRIBUTES

This is called a scatterplot matrix. It shows the scatterplot for every pair of attributes.

CORRELATIONS BETWEEN ALL NUMERIC ATTRIBUTES

term	knowledge	news	age	education	male	ideology	engagement	democrat	republican
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1 knowledge	NA	0.345	0.171	0.471	0.264	0.0940	0.481	0.0844	0.245
2 news	0.345	NA	0.0841	0.150	-0.0151	0.0342	0.200	0.0735	0.0936
3 age	0.171	0.0841	NA	-0.0260	0.0307	-0.0231	0.269	0.0465	-0.0515
4 education	0.471	0.150	-0.0260	NA	0.0380	0.245	0.247	0.0473	0.212
5 male	0.264	-0.0151	0.0307	0.0380	NA	-0.0463	0.0285	0.173	-0.0954
6 ideology	0.0940	0.0342	-0.0231	0.245	-0.0463	NA	0.0754	-0.350	0.483
7 engagement	0.481	0.200	0.269	0.247	0.0285	0.0754	NA	0.0324	0.226
8 democrat	0.0844	0.0735	0.0465	0.0473	0.173	-0.350	0.0324	NA	-0.555
9 republican	0.245	0.0936	-0.0515	0.212	-0.0954	0.483	0.226	-0.555	NA

RQ1: Is there an effect of news exposure on political knowledge?

The positive correlation indicates a <u>sample</u> effect of news exposure on political knowledge. Is this due to sampling variation? Or is there a population effect?

Statistical Evidence: Statistically relevant effect of news exposure in model that explains variation in political knowledge.

$$\text{Knowledge}_i = \beta_0 + \beta_1(\text{News Exposure}_i) + \epsilon_i \qquad \text{where} \qquad \epsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$$

MODEL-LEVEL OUTPUT
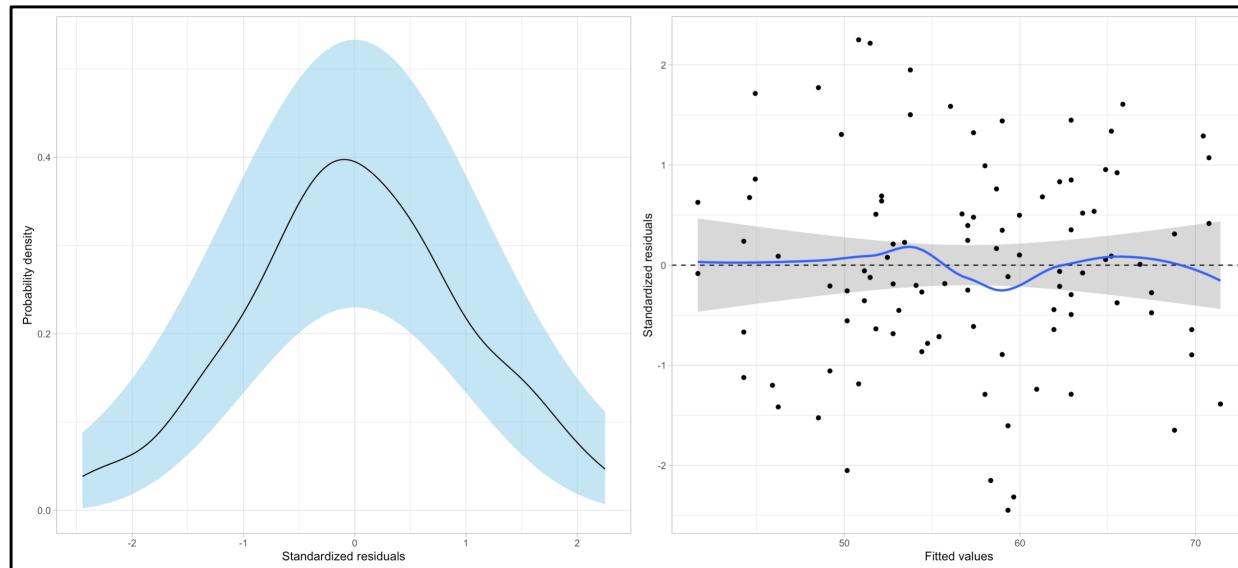
```
      r.squared adj.r.squared sigma statistic  p.value    df logLik
        <dbl>          <dbl> <dbl>     <dbl>     <dbl> <dbl>  <dbl>
    1    0.119          0.110  20.3      13.2 0.000442     1  -442.

        AIC   BIC deviance df.residual  nobs
      <dbl> <dbl>    <dbl>       <int> <int>
    1  890.  897.   40233.          98   100
```

Differences in the amount of news exposure explains 11.9% of the variation in political knowledge, which is more than we would expect because of chance; F(1, 98) = 13.20, *p* < .001.

COEFFICIENT-LEVEL OUTPUT

```
  term           estimate std.error statistic  p.value
  <chr>             <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)       39.7      5.20       7.63 1.56e-11
2 news             0.327    0.0899       3.64 4.42e- 4
```

RESIDUAL PLOTS FOR MODEL 1

There seems to be some violation of the linearity assumption.

Also the normality assumption may be violated.

Don't forget about the independence assumption!!

RQ2: Is there an effect of news exposure on political knowledge after controlling for demographic and political covariates?

Statistical Evidence: Statistically relevant effect of news exposure in main-effects model that explains variation in political knowledge that also includes all the demographic and political covariates.

$$\text{Knowledge}_i = \beta_0 + \beta_1(\text{News Exposure}_i) + \beta_2(\text{Age}_i) + \beta_3(\text{Education Level}_i) +$$
$$\beta_4(\text{Male}_i) + \beta_5(\text{Political Engagement}_i) + \beta_6(\text{Political Ideology}_i) +$$
$$\beta_7(\text{Democrat}_i) + \beta_8(\text{Republican}_i) + \epsilon_i$$

$$\text{where} \quad \epsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$$

MODEL-LEVEL OUTPUT

```
    r.squared adj.r.squared sigma statistic  p.value    df logLik
        <dbl>         <dbl> <dbl>    <dbl>    <dbl> <dbl>  <dbl>
        0.495         0.451  15.9     11.2 7.28e-11     8  -414.

    AIC   BIC deviance df.residual  nobs
  <dbl> <dbl>    <dbl>       <int> <int>
   848.  874.   23057.          91   100
```
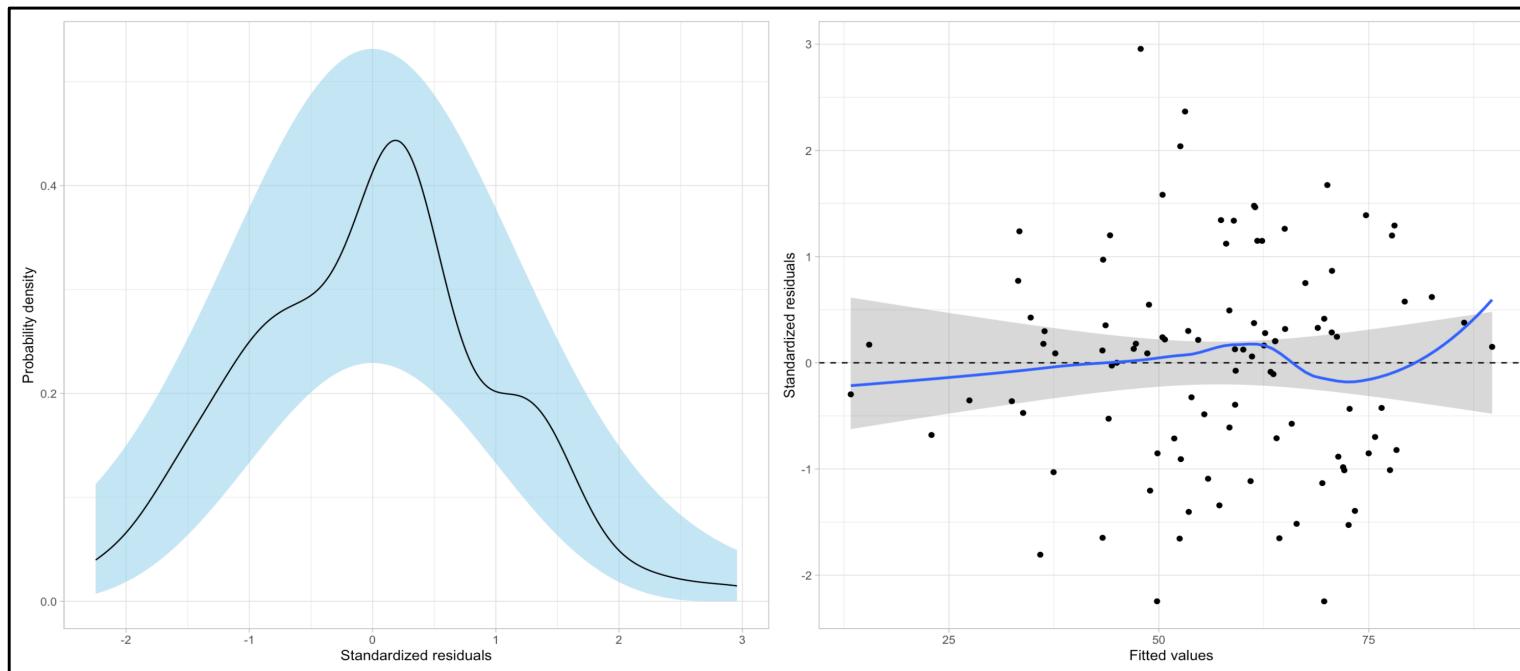
Differences in all of the predictors (i.e., the model) explains 49.5% of the variation in political knowledge, which is more than we would expect because of chance; $F_{(8, 91)} = 11.20$, $p < .001$.

COEFFICIENT-LEVEL OUTPUT

term	estimate	std.error	statistic	p.value
<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1 (Intercept)	-32.2	11.6	-2.78	0.00663
2 news	0.202	0.0731	2.76	0.00707
3 age	0.107	0.102	1.05	0.298
4 education	2.94	0.717	4.10	0.0000904
5 male	10.5	3.24	3.24	0.00165
6 engagement	0.337	0.0992	3.40	0.000997
7 ideology	-0.0743	0.0934	-0.795	0.428
8 democrat	3.89	4.47	0.871	0.386
9 republican	8.52	4.43	1.92	0.0575

We generally will keep covariates in the model (and report them) regardless of statistical significance!

RESIDUAL PLOTS FOR MODEL 2

RQ3: Does education level moderate the effect of news exposure on political knowledge after controlling for demographic and political covariates?

Statistical Evidence: Statistically relevant interaction effect of news exposure and education in model that explains variation in political knowledge that also includes all the demographic and political covariates.

$$\text{Knowledge}_i = \beta_0 + \beta_1(\text{News Exposure}_i) + \beta_2(\text{Age}_i) + \beta_3(\text{Education Level}_i)+$$
$$\beta_4(\text{Male}_i) + \beta_5(\text{Political Engagement}_i) + \beta_6(\text{Political Ideology}_i)+$$
$$\beta_7(\text{Democrat}_i) + \beta_8(\text{Republican}_i)+$$
$$\beta_9(\text{News Exposure}_i)(\text{Education Level}_i) + \epsilon_i$$

$$\text{where} \quad \epsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$$

MODEL-LEVEL OUTPUT

```
r.squared adj.r.squared sigma statistic  p.value    df logLik
   <dbl>          <dbl> <dbl>     <dbl>     <dbl> <dbl>  <dbl>
   0.521          0.473  15.6      10.9 2.92e-11     9  -411.

    AIC   BIC deviance df.residual  nobs
  <dbl> <dbl>    <dbl>       <int> <int>
   845.  873.   21892.          90   100
```

Differences in all of the predictors (i.e., the model) explains 52.1% of the variation in political knowledge, which is more than we would expect because of chance; F(9, 90) = 10.90, *p* < .001.

COEFFICIENT-LEVEL OUTPUT

term	estimate	std.error	statistic	p.value
<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1 (Intercept)	-79.9	24.6	-3.25	0.00161
2 age	0.107	0.0998	1.08	0.284
3 education	6.28	1.68	3.73	0.000329
4 male	11.3	3.20	3.55	0.000619
5 engagement	0.323	0.0974	3.31	0.00133
6 ideology	-0.0619	0.0917	-0.675	0.501
7 democrat	3.71	4.38	0.847	0.399
8 republican	8.29	4.34	1.91	0.0591
9 news	1.14	0.436	2.62	0.0103
10 education:news	-0.0651	0.0298	-2.19	0.0312

To better interpret the interaction effect, plot it.

The effect of news exposure on political knowledge depends on education level, after controlling for the set of demographic and political covariates. This interaction effect is larger than we expect because of chance (p = .031).

PLOTTING THE INTERACTION EFFECT

- Place the effect of news exposure on the *x*-axis (it is the focal predictor)
- Place political knowledge (outcome) on *y*-axis
- Choose multiple levels (2–3) of education level to show the interaction
- Choose interpretable values for the other covariates (1 value per covariate)

- Education level = {12, 16}
- Age = 50
- Sex = 0 (female)
- Political engagement = 50 (moderately engaged)
- Political ideology = 0 (Liberal)
- Democrat = 1
- Republican = 0

EDUCATION LEVEL = 12 (HIGH SCHOOL)

$\widehat{\text{Knowledge}}_i = -79.9 + 1.14(\text{News Exposure}_i) + 0.107(\text{Age}_i) + 6.28(\text{Education Level}_i)+$
$11.3(\text{Male}_i) + 0.323(\text{Political Engagement}_i) - 0.062(\text{Political Ideology}_i)+$
$3.71(\text{Democrat}_i) + 8.29(\text{Republican}_i)-$
$0.065(\text{News Exposure}_i)(\text{Education Level}_i)$

$\widehat{\text{Knowledge}}_i = -79.9 + 1.14(\text{News Exposure}_i) + 0.107(50) + 6.28(12) + 11.3(0)+$
$0.323(50) - 0.062(0) + 3.71(1) + 8.29(0) - 0.065(\text{News Exposure}_i)(12)$

$\widehat{\text{Knowledge}}_i = 6.135 + 0.3588(\text{News Exposure}_i)$
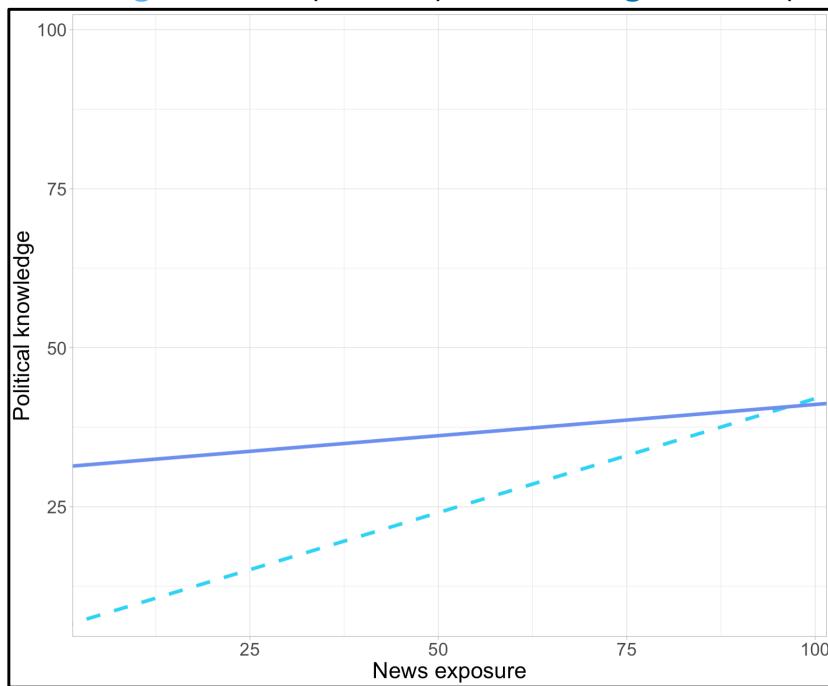
EDUCATION LEVEL = 16 (UNDERGRADUATE)

$$\hat{\text{Knowledge}}_i = -79.9 + 1.14(\text{News Exposure}_i) + 0.107(\text{Age}_i) + 6.28(\text{Education Level}_i) +$$
$$11.3(\text{Male}_i) + 0.323(\text{Political Engagement}_i) - 0.062(\text{Political Ideology}_i) +$$
$$3.71(\text{Democrat}_i) + 8.29(\text{Republican}_i) -$$
$$0.065(\text{News Exposure}_i)(\text{Education Level}_i)$$
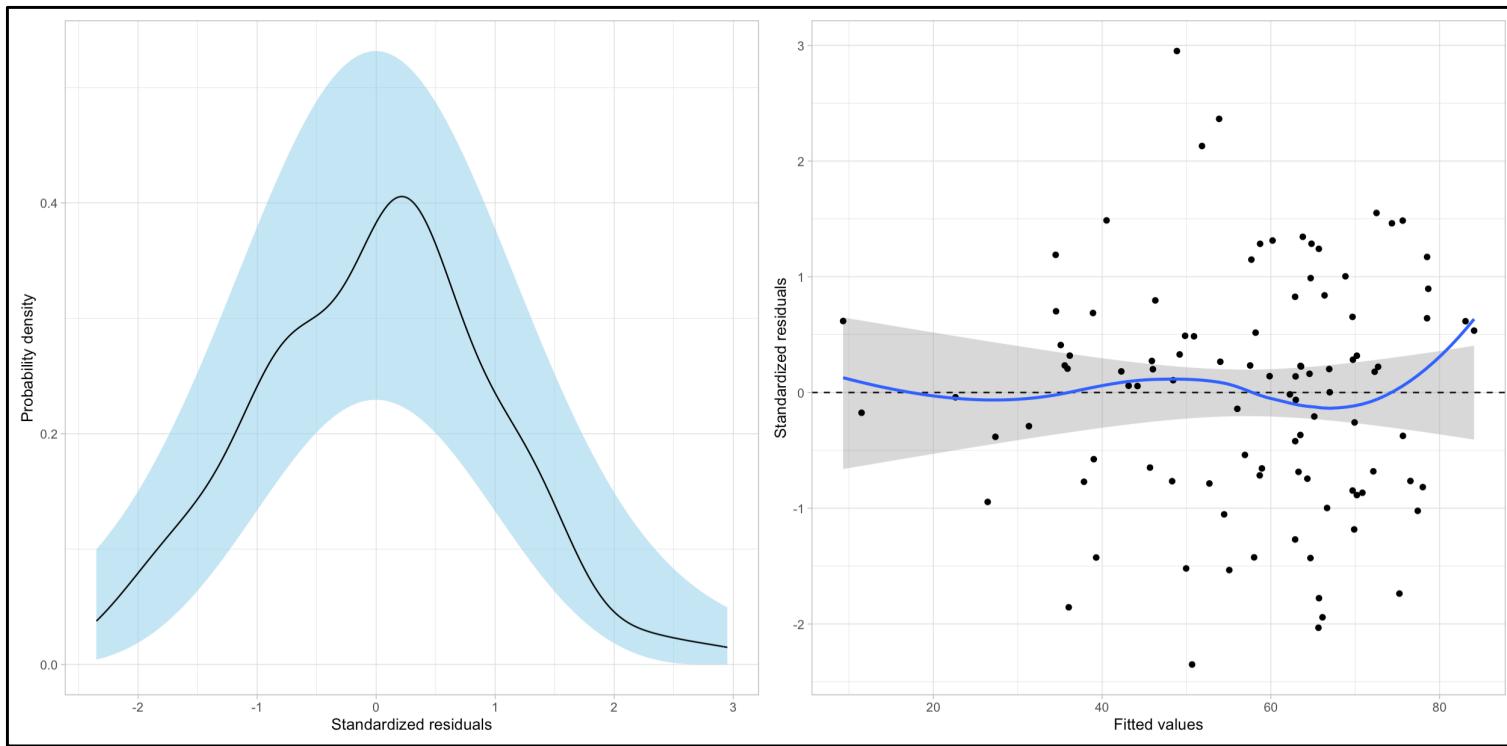
$$\hat{\text{Knowledge}}_i = -79.9 + 1.14(\text{News Exposure}_i) + 0.107(50) + 6.28(16) + 11.3(0) +$$
$$0.323(50) - 0.062(0) + 3.71(1) + 8.29(0) - 0.065(\text{News Exposure}_i)(16)$$

$$\hat{\text{Knowledge}}_i = 31.255 + 0.0984(\text{News Exposure}_i)$$

Predicted average political knowledge as a function of news exposure for U.S. extremely liberal, female, adult Democrats who are 50 years of age, are moderately engaged in politics. This is displayed for those with a a **high school** (dashed) and **undergraduate** (solid) level of education.

RESIDUAL PLOTS FOR MODEL 3

TABLES FOR PUBLICATION

Table 1. Summary statistics for the analytic variables.

Variable	M	SD	Min.	Max.
Political knowledge	57.11	21.48	9.0	98.0
News Exposure	53.27	22.64	6.0	97.0
Age	50.95	16.52	18.0	93.9
Education level	14.20	2.44	8.0	18.0
Male*	0.52	0.50	0.0	1.0
Political engagement	73.28	18.14	7.5	98.3
Political ideology	46.91	20.10	0.1	92.8
Democrat**	0.29	0.46	0.0	1.0
Independent**	0.28	0.45	0.0	1.0
Republican**	0.43	0.50	0.0	1.0

Note.
Sex (non-males, males) was dummy-coded.
**Each political party (Democrat, Independent, Republican) was dummy-coded.*

Table 2. Pairwise correlations between the analytic variables.

Variable	1	2	3	4	5	6	7	8	9
1. Political knowledge	1.00								
2. News exposure	.35	1.00							
3. Age	.17	.08	1.00						
4. Education level	.47	.15	−.03	1.00					
5. Male*	.26	−.02	.03	.04	1.00				
6. Political engagement	.48	.20	.27	.25	.03	1.00			
7. Political ideology	.09	.03	−.02	.25	−.05	.08	1.00		
8. Democrat**	.08	.07	.05	.05	.17	.03	−.35	1.00	
9. Republican**	.25	.09	−.05	.21	−.10	.23	.48	−.56	1.00

Table 3. Unstandardized regression coefficients (and SEs) for a set of models predicting variation in political knowledge.

Predictor	Model 1		Model 2		Model 3	
	B	SE	B	SE	B	SE
News Exposure	0.33	0.09	0.20	0.07	1.14	0.44
Age			0.11	0.10	0.11	0.10
Education level			2.94	0.72	6.28	1.68
Male[*]			10.5	3.24	11.3	3.20
Political engagement			0.34	0.10	0.32	0.10
Political ideology			–0.07	0.09	–0.06	0.09
Democrat[**]			3.89	4.47	3.71	4.38
Republican[**]			8.52	4.43	8.29	4.34
News Exposure x Education level					–0.07	0.03
Intercept	39.70	5.20	–32.2	11.6	–79.9	24.6
RMSE	20.3		15.9		15.6	
R^2	0.119		0.495		0.521	

Note.
[*]Sex was dummy-coded with non-males the reference group.
[**]Political party (Democrat, Independent, Republican) was dummy-coded with Independent as the reference group.

ANSWER THE RESEARCH QUESTIONS

There does seem to be a small, positive effect of news exposure on political knowledge.

After controlling for a set of demographic and political covariates, this effect persists, but diminishes slightly.

Education level, however, does seem to moderate this effect. The effect of news exposure on political knowledge is smaller for U.S. adults with higher levels of education, on average.

Consider your computer files and your organization of those files…

➡ Are your files organized into folders/directories? Or are they all in your Downloads folder?

➡ How did you organize all the data files, notes, etc. from EPsy 8251?

➡ If I asked you to find a specific file, could you locate it without using "Search"?

➡ Can you tell what is in a particular file by just looking at its name?

➡ Do your file names contain spaces? What about characters that aren't letters, numbers, dashes, or underscores?

➡ Are your file names consistent (all lower case letters, or all title case)? Or are they all different?