What’s the deal with regression?

API 202: TF Session 1

Nolan M. Kavanagh

Goals for today

1. **Get to know one another a little better.**
2. **Review regression notation, including the PRF vs. SRF.**
3. **Learn how to graph bivariate relationships.**
4. **Learn how to run regressions.**
5. **Review how to interpret regressions.**

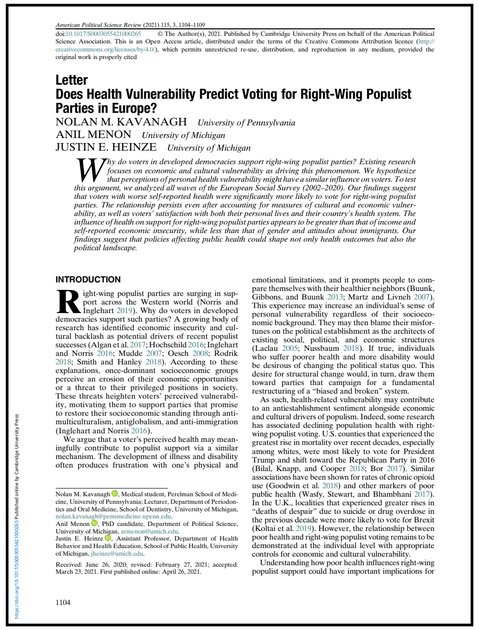
We'll treat this session like a workshop with interactive examples.

MD/PhD student in health policy.

“I don’t need backups. I’m going to Harvard.”

Hi, I’m Nolan.

GO BLUE!



American Political Science Review (2021) 115, 3, 1104–1109

doi:10.1017/S0003055421000265 © The Author(s), 2021. Published by Cambridge University Press on behalf of the American Political Science Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited

Letter

Does Health Vulnerability Predict Voting for Right-Wing Populist Parties in Europe?

NOLAN M. KAVANAGH *University of Pennsylvania*
ANIL MENON *University of Michigan*
JUSTIN E. HEINZE *University of Michigan*

Why do voters in developed democracies support right-wing populist parties? Existing research focuses on economic and cultural vulnerability as driving this phenomenon. We hypothesize that perceptions of personal health vulnerability might have a similar influence on voters. To test this argument, we analyzed all waves of the European Social Survey (2002–2020). Our findings suggest that voters with worse self-reported health were significantly more likely to vote for right-wing populist parties. The relationship persists even after accounting for measures of cultural and economic vulnerability, as well as voters’ satisfaction with both their personal lives and their country’s health system. The influence of health on support for right-wing populist parties appears to be greater than that of income and self-reported economic insecurity, while less than that of gender and attitudes about immigrants. Our findings suggest that policies affecting public health could shape not only health outcomes but also the political landscape.

INTRODUCTION

Right-wing populist parties are surging in support across the Western world (Norris and Inglehart 2019). Why do voters in developed democracies support such parties? A growing body of research has identified economic insecurity and cultural backlash as potential drivers of recent populist successes (Algan et al. 2017; Hochschild 2016; Inglehart and Norris 2016; Mudde 2007; Oesch 2008; Rodrik 2018; Smith and Hanley 2018). According to these explanations, once-dominant socioeconomic groups perceive an erosion of their economic opportunities or a threat to their privileged positions in society. These threats heighten voters’ perceived vulnerability, motivating them to support parties that promise to restore their socioeconomic standing through anti-multiculturalism, antiglobalism, and anti-immigration (Inglehart and Norris 2016).

We argue that a voter’s perceived health may meaningfully contribute to populist support via a similar mechanism. The development of illness and disability often produces frustration with one’s physical and emotional limitations, and it prompts people to compare themselves with their healthier neighbors (Buunk, Gibbons, and Buunk 2013; Martz and Livneh 2007). This experience may increase an individual’s sense of personal vulnerability regardless of their socioeconomic background. They may then blame their misfortunes on the political establishment as the architects of existing social, political, and economic structures (Laclau 2005; Nussbaum 2018). If true, individuals who suffer poorer health and more disability would be desirous of changing the political status quo. This desire for structural change would, in turn, draw them toward parties that campaign for a fundamental restructuring of a “biased and broken” system.

As such, health-related vulnerability may contribute to an antiestablishment sentiment alongside economic and cultural drivers of populism. Indeed, some research has associated declining population health with right-wing populist voting. U.S. counties that experienced the greatest rise in mortality over recent decades, especially among whites, were most likely to vote for President Trump and shift toward the Republican Party in 2016 (Bilal, Knapp, and Cooper 2018; Bor 2017). Similar associations have been shown for rates of chronic opioid use (Goodwin et al. 2018) and other markers of poor public health (Wasfy, Stewart, and Bhambhani 2017). In the U.K., localities that experienced greater rises in “deaths of despair” due to suicide or drug overdose in the previous decade were more likely to vote for Brexit (Koltai et al. 2019). However, the relationship between poor health and right-wing populist voting remains to be demonstrated at the individual level with appropriate controls for economic and cultural vulnerability.

Understanding how poor health influences right-wing populist support could have important implications for

Nolan M. Kavanagh, Medical student, Perelman School of Medicine, University of Pennsylvania; Lecturer, Department of Periodontics and Oral Medicine, School of Dentistry, University of Michigan, nolan.kavanagh@pennmedicine.upenn.edu.
Anil Menon, PhD candidate, Department of Political Science, University of Michigan, armenon@umich.edu.
Justin E. Heinze, Assistant Professor, Department of Health Behavior and Health Education, School of Public Health, University of Michigan, jheinze@umich.edu.

Received: June 26, 2020; revised: February 27, 2021; accepted: March 23, 2021. First published online: April 26, 2021.

https://doi.org/10.1017/S0003055421000265 Published online by Cambridge University Press

My research is on the politics of health.

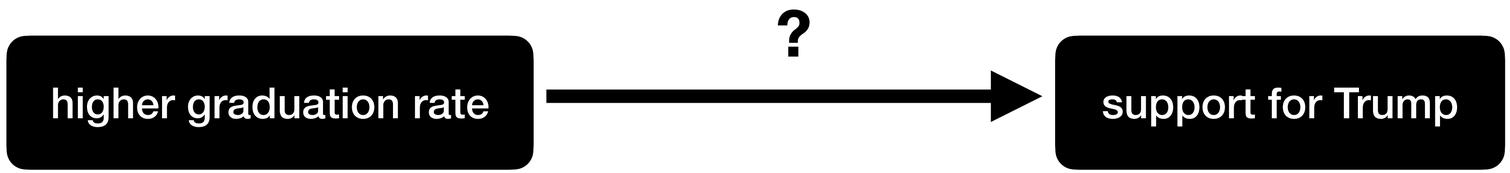
My go-to karaoke song is “Since U Been Gone.”

Overview of our sample data

`state`	**State of county**	*Administrative*
`county_fips`	**County FIPS identifier**	*Administrative*
`pc_under_18`	**Percent of county under age 18**	*American Community Survey (2016–2020)*
`pc_over_65`	**Percent of county over age 65**	*American Community Survey (2016–2020)*
`pc_male`	**Percent of county that is male**	*American Community Survey (2016–2020)*
`pc_black`	**Percent of county that is Black**	*American Community Survey (2016–2020)*
`pc_latin`	**Percent of county that is Hispanic/Latino**	*American Community Survey (2016–2020)*
`pc_hs_grad`	**Percent of county that graduated high school**	*American Community Survey (2016–2020)*
`unemploy_rate`	**County unemployment rate (%)**	*American Community Survey (2016–2020)*
`median_income`	**County median income ($)**	*American Community Survey (2016–2020)*
`pc_uninsured`	**Percent of county without health insurance**	*American Community Survey (2016–2020)*
`pc_trump`	**Percent of county votes for Trump in 2020**	*MIT Election Lab*

Tell me a story.

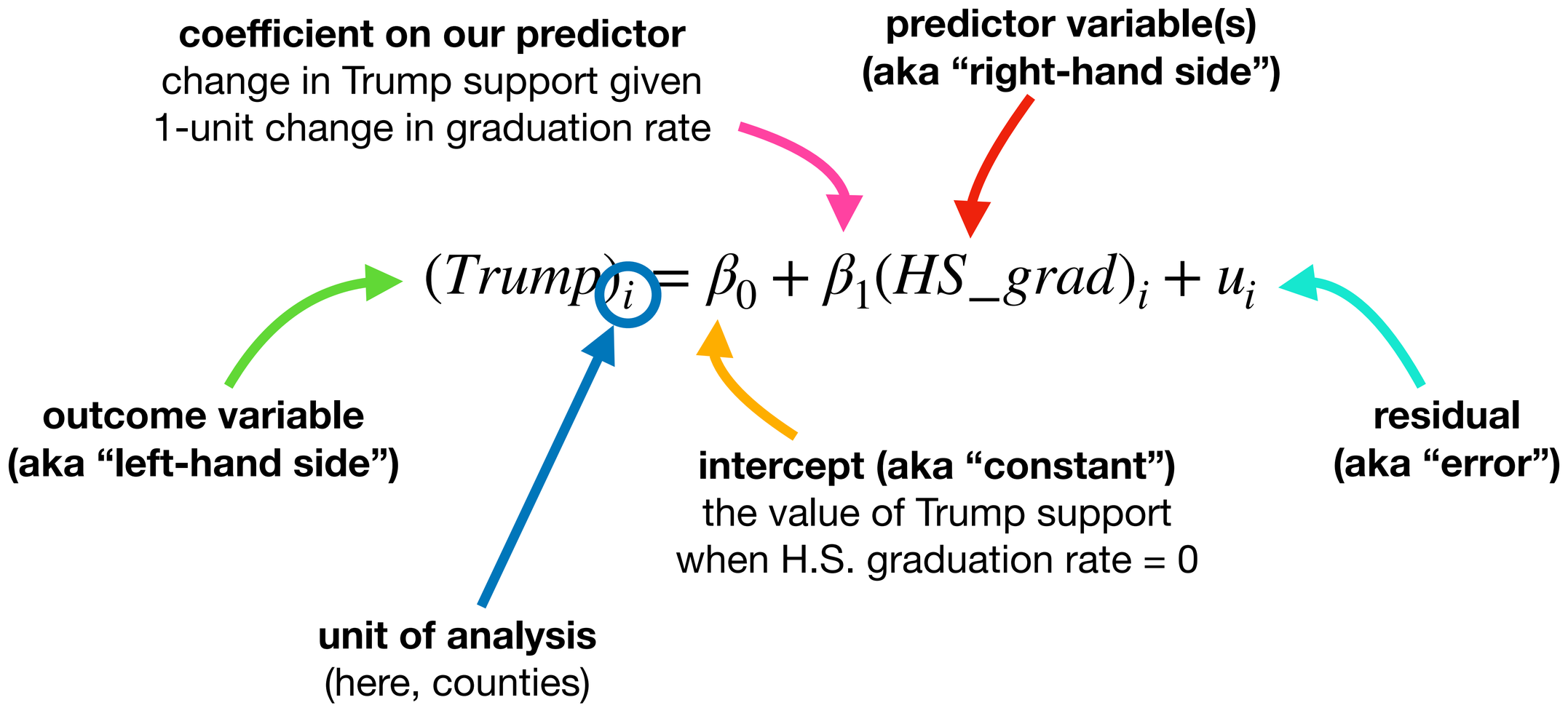
Let's say we're interested in the relationship between high school graduation and support for Trump.



What might be the mechanism?

More education = more income = prefer lower taxes?

Population regression function



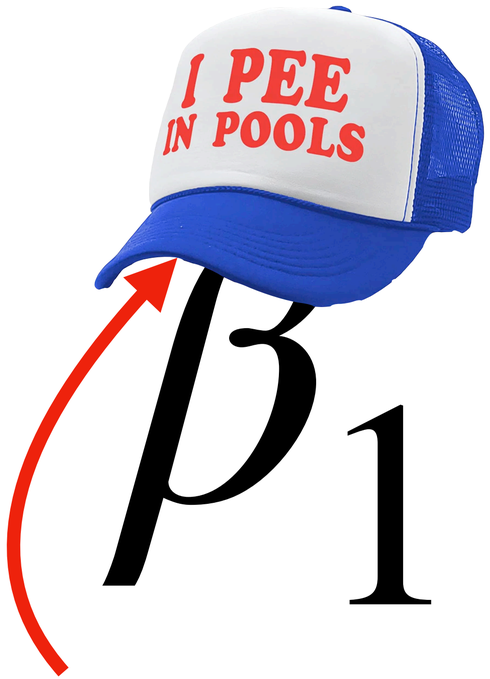
Population regression function

$$(Trump)_i = \beta_0 + \beta_1(HS\_grad)_i + u_i$$

We add "hats" to signify estimated values in our sample.

Sample regression function

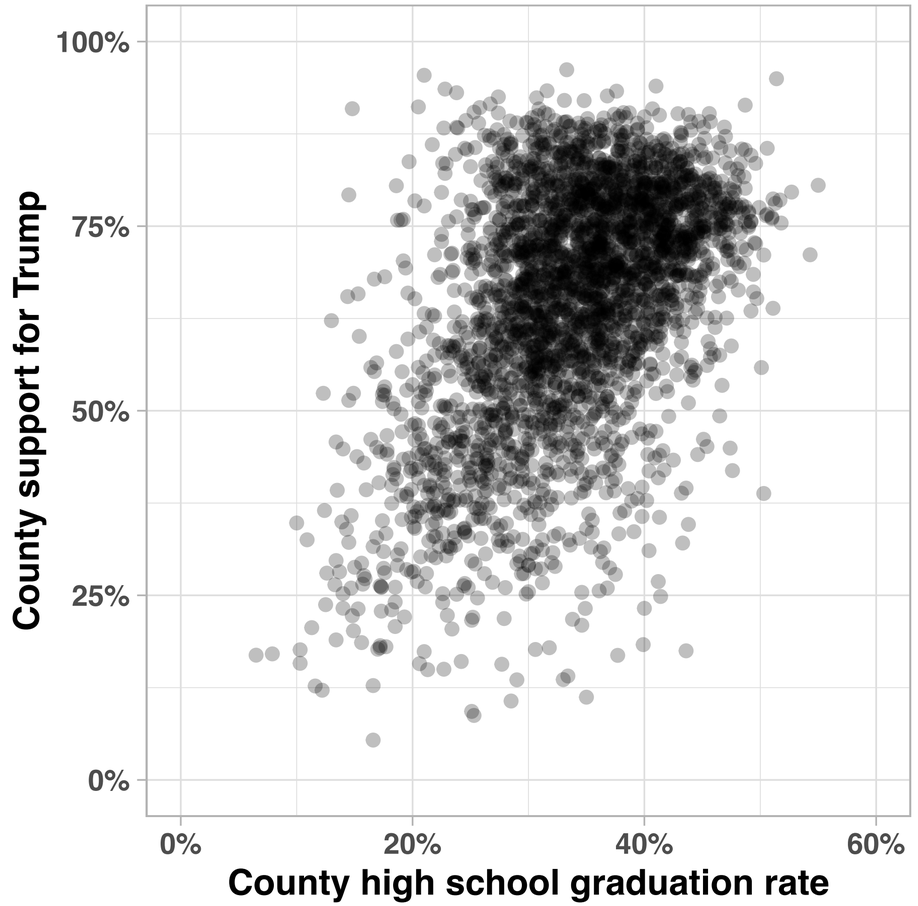
$$(Trump)_i = \hat{\beta}_0 + \hat{\beta}_1(HS\_grad)_i + \hat{u}_i$$



Only a <u>specific</u> sample would ever wear this hat.

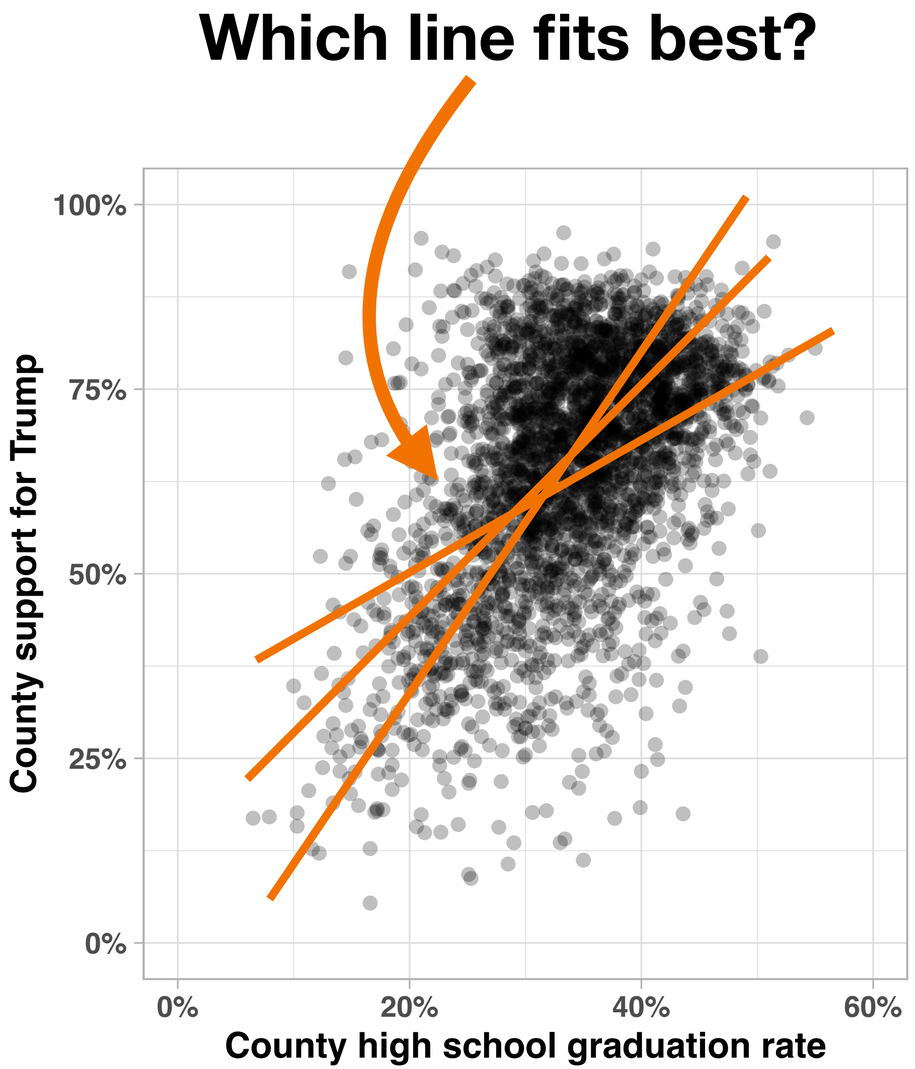
Let's graph our data.

```
# Graph high school graduation and Trump support
plot_1 <- ggplot(df, aes(x=pc_hs_grad, y=pc_trump)) +

  # Add scatterplot points
  geom_point(alpha=0.25) +

  # Labels of axes
  xlab("County high school graduation rate") +
  ylab("County support for Trump") +

  # Cosmetic changes
  theme_light() + theme(text = element_text(face="bold")) +
  scale_y_continuous(limits=c(0,100),
                     labels = function(x) paste0(x,"%")) +
  scale_x_continuous(limits=c(0,60),
                     labels = function(x) paste0(x, "%"))
```



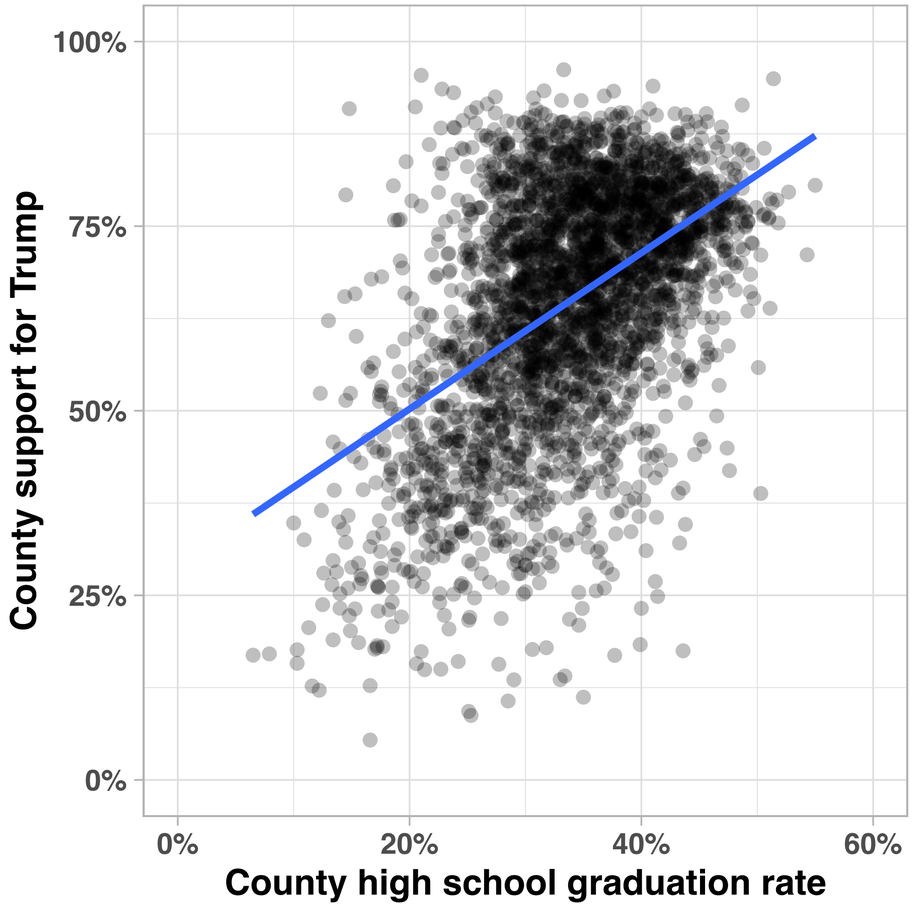
Let's graph our data.

```
# Graph high school graduation and Trump support
plot_1 <- ggplot(df, aes(x=pc_hs_grad, y=pc_trump)) +

  # Add scatterplot points
  geom_point(alpha=0.25) +

  # Labels of axes
  xlab("County high school graduation rate") +
  ylab("County support for Trump") +

  # Cosmetic changes
  theme_light() + theme(text = element_text(face="bold")) +
  scale_y_continuous(limits=c(0,100),
                     labels = function(x) paste0(x,"%")) +
  scale_x_continuous(limits=c(0,60),
                     labels = function(x) paste0(x, "%"))
```



Let's graph our data.

```
# Graph high school graduation and Trump support
plot_1 <- ggplot(df, aes(x=pc_hs_grad, y=pc_trump)) +

  # Add scatterplot points
  geom_point(alpha=0.25) +

  # Labels of axes
  xlab("County high school graduation rate") +
  ylab("County support for Trump") +

  # Cosmetic changes
  theme_light() + theme(text = element_text(face="bold")) +
  scale_y_continuous(limits=c(0,100),
                     labels = function(x) paste0(x,"%")) +
  scale_x_continuous(limits=c(0,60),
                     labels = function(x) paste0(x, "%")) +

  # Add line of best fit
  geom_smooth(method="lm", se=F, formula = y~x)
```



OK fine, let's run a regression.

Save it as object "reg_1"

```
# Estimate regression
# LHS = support for Trump
# RHS = unemployment rate
reg_1 <- lm(pc_trump ~ pc_hs_grad, data=df)
summary(reg_1)
```

Get regression output

$$(Trump)_i = \hat{\beta}_0 + \hat{\beta}_1(HS\_grad)_i + \hat{u}_i$$

OK fine, let's interpret a regression.

```
Call:
lm(formula = pc_trump ~ pc_hs_grad, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-57.708  -8.185   0.838   9.427  46.164

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 29.09825    1.19253   24.40   <2e-16 ***
pc_hs_grad   1.05721    0.03436   30.77   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14.13 on 3113 degrees of freedom
Multiple R-squared:  0.2332,            Adjusted R-squared:  0.233
F-statistic:   947 on 1 and 3113 DF,  p-value: < 2.2e-16
```

OK fine, let's interpret a regression.

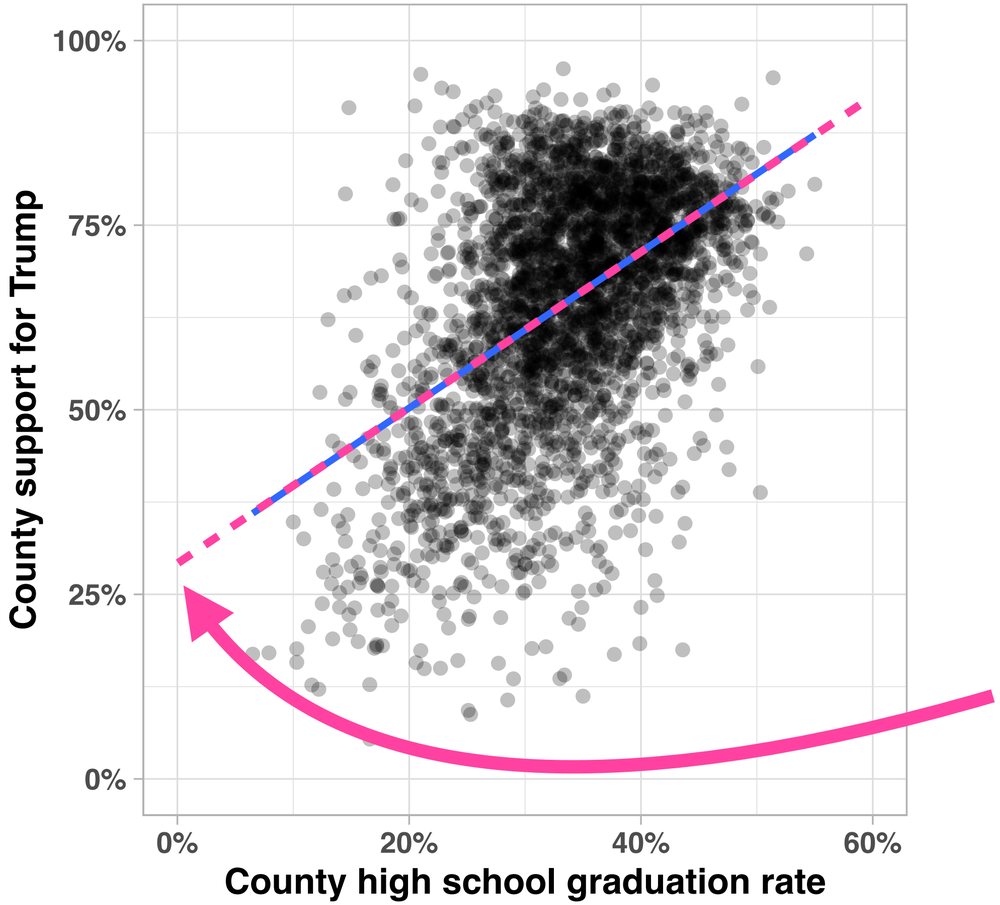
```
Call:
lm(formula = pc_trump ~ pc_hs_grad, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-57.708  -8.185   0.838   9.427  46.164

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 29.09825    1.19253   24.40   <2e-16 ***
pc_hs_grad   1.05721    0.03436   30.77   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.13 on 3113 degrees of freedom
Multiple R-squared:  0.2332,            Adjusted R-squared:  0.233
F-statistic:   947 on 1 and 3113 DF,  p-value: < 2.2e-16
```

We'll start with the intercept.

When 0% of a county has graduated high school, support for Trump is an estimated 29.1%.

(Is a graduation rate of 0% meaningful?)

OK fine, let's interpret a regression.



That looks about right!

We'll start with the intercept.

When 0% of a county has graduated high school, support for Trump is an estimated 29.1%.

(Is a graduation rate of 0% meaningful?)

OK fine, let's interpret a regression.

```
Call:
lm(formula = pc_trump ~ pc_hs_grad, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-57.708  -8.185   0.838   9.427  46.164

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 29.09825    1.19253   24.40   <2e-16 ***
pc_hs_grad   1.05721    0.03436   30.77   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.13 on 3113 degrees of freedom
Multiple R-squared:  0.2332,            Adjusted R-squared:  0.233
F-statistic:   947 on 1 and 3113 DF,  p-value: < 2.2e-16
```

We'll start with the intercept.

When 0% of a county has graduated high school, support for Trump is an estimated 29.1%.

(Is a graduation rate of 0% meaningful?)

The standard error is 1.2%. This gives us a 95% C.I. of 29.1 ± 1.96*1.2 = [26.8% to 31.4%].

The t-statistic is 24.4. The p-value is <0.05.

Thus, we can conclude that the intercept is significantly different from 0.

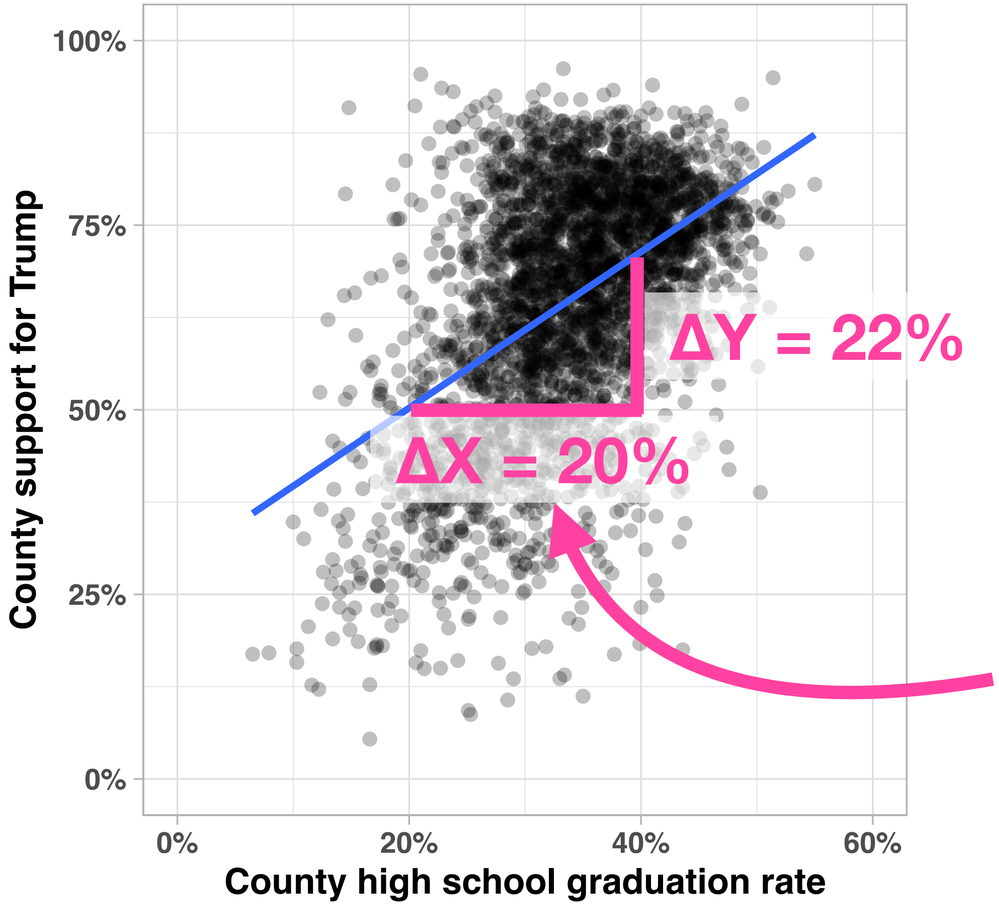
OK fine, let's interpret a regression.

```
Call:
lm(formula = pc_trump ~ pc_hs_grad, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-57.708  -8.185   0.838   9.427  46.164

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 29.09825    1.19253   24.40   <2e-16 ***
pc_hs_grad   1.05721    0.03436   30.77   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.13 on 3113 degrees of freedom
Multiple R-squared:  0.2332,            Adjusted R-squared:  0.233
F-statistic:   947 on 1 and 3113 DF,  p-value: < 2.2e-16
```

Now for the coefficient on `pc_hs_grad`.

For each 1 percentage point (pp) increase in a county's high school graduation rate, the estimated support for Trump increases by 1.1 pp.

OK fine, let's interpret a regression.



That looks about right!

Now for the coefficient on `pc_hs_grad`.

For each 1 percentage point (pp) increase in a county's high school graduation rate, the estimated support for Trump increases by 1.1 pp.

OK fine, let's interpret a regression.

```
Call:
lm(formula = pc_trump ~ pc_hs_grad, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-57.708  -8.185   0.838   9.427  46.164

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 29.09825    1.19253   24.40   <2e-16 ***
pc_hs_grad   1.05721    0.03436   30.77   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.13 on 3113 degrees of freedom
Multiple R-squared:  0.2332,            Adjusted R-squared:  0.233
F-statistic:   947 on 1 and 3113 DF,  p-value: < 2.2e-16
```

Now for the coefficient on `pc_hs_grad`.

For each 1 percentage point (pp) increase in a county's high school graduation rate, the estimated support for Trump increases by 1.1 pp.

The standard error is 0.03%. This gives us a 95% C.I. of 1.06 ± 1.96*0.03 = [0.99% to 1.12%].

The t-statistic is 30.8. The p-value is <0.05.

Thus, the coefficient is significantly different from 0. There's a positive association between graduation rates and Trump support.

That’s all good. But is it causal?

We’ll spend lots of time in API 202 asking this very question.

What problems with a causal interpretation come to mind?

What variables/influences are we missing?