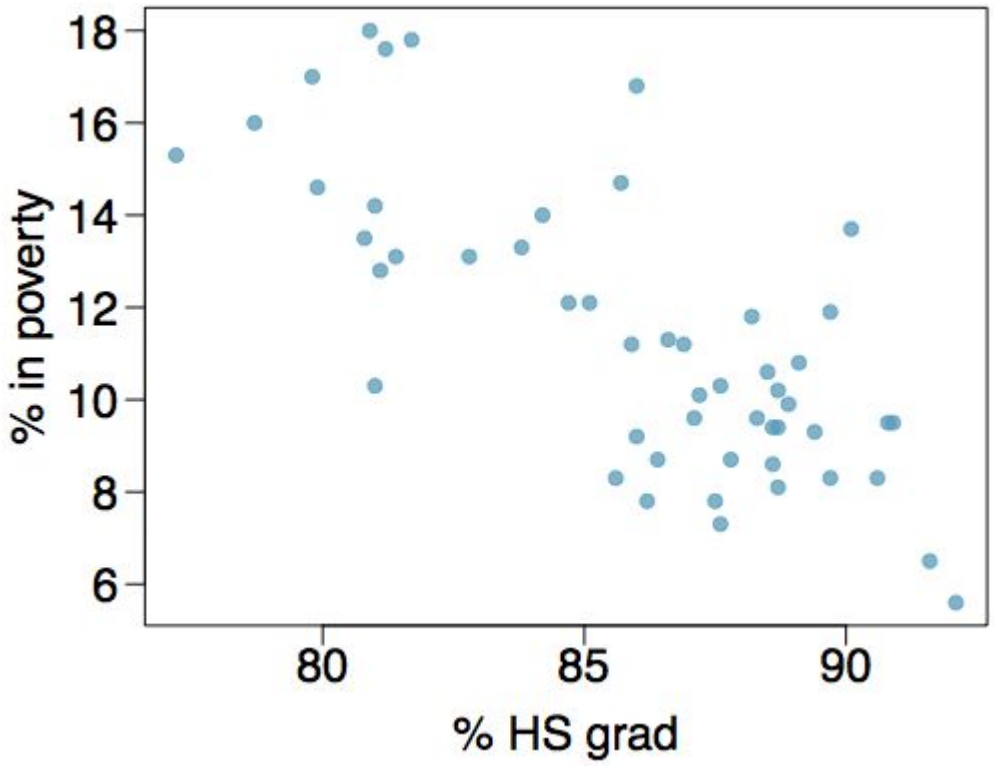
Correlation, Residuals and Line Fitting

Guessing the correlation

Which of the following is the best guess for the correlation between percent in poverty and percent HS grad?

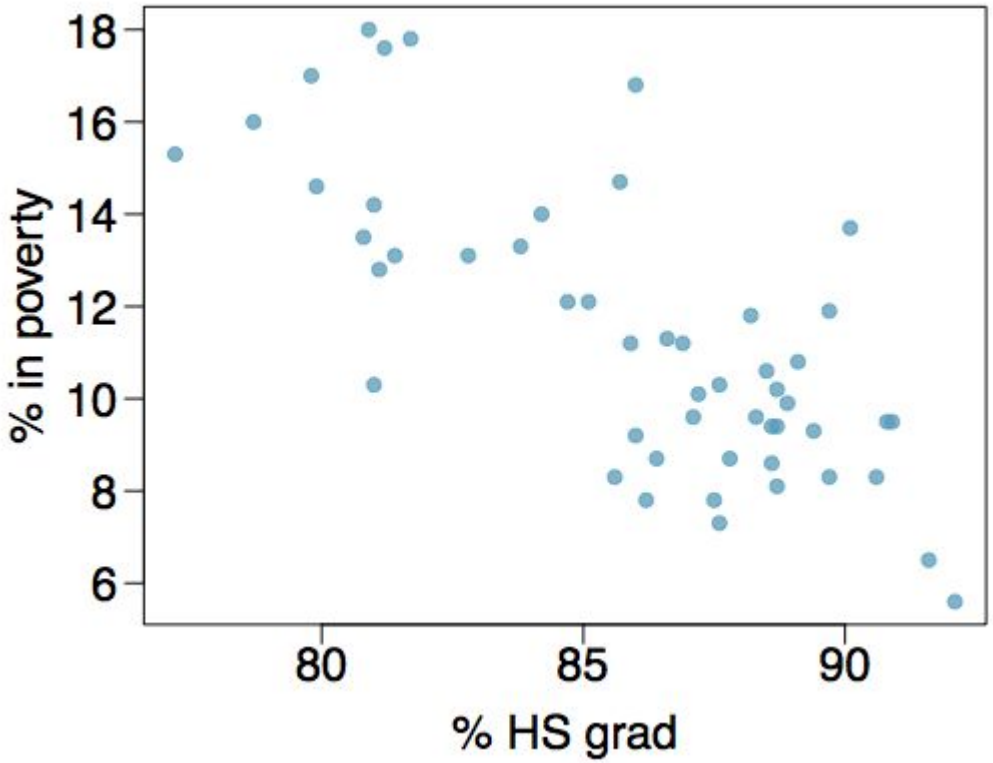
(a) 0.6
(b) -0.75
(c) -0.1
(d) 0.02
(e) -1.5



Guessing the correlation

Which of the following is the best guess for the correlation between percent in poverty and percent HS grad?

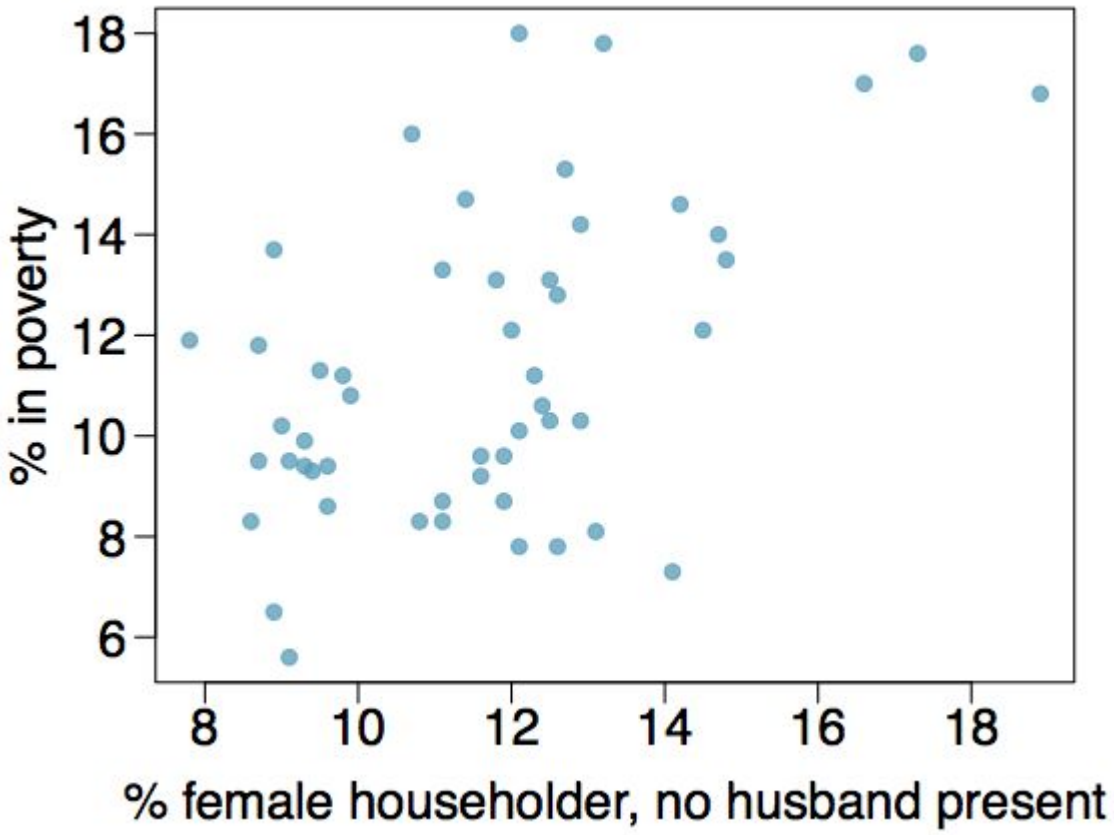
(a) 0.6
(b) -0.75
(c) -0.1
(d) 0.02
(e) -1.5



Guessing the correlation

Which of the following is the best guess for the correlation between percent in poverty and percent female householder?

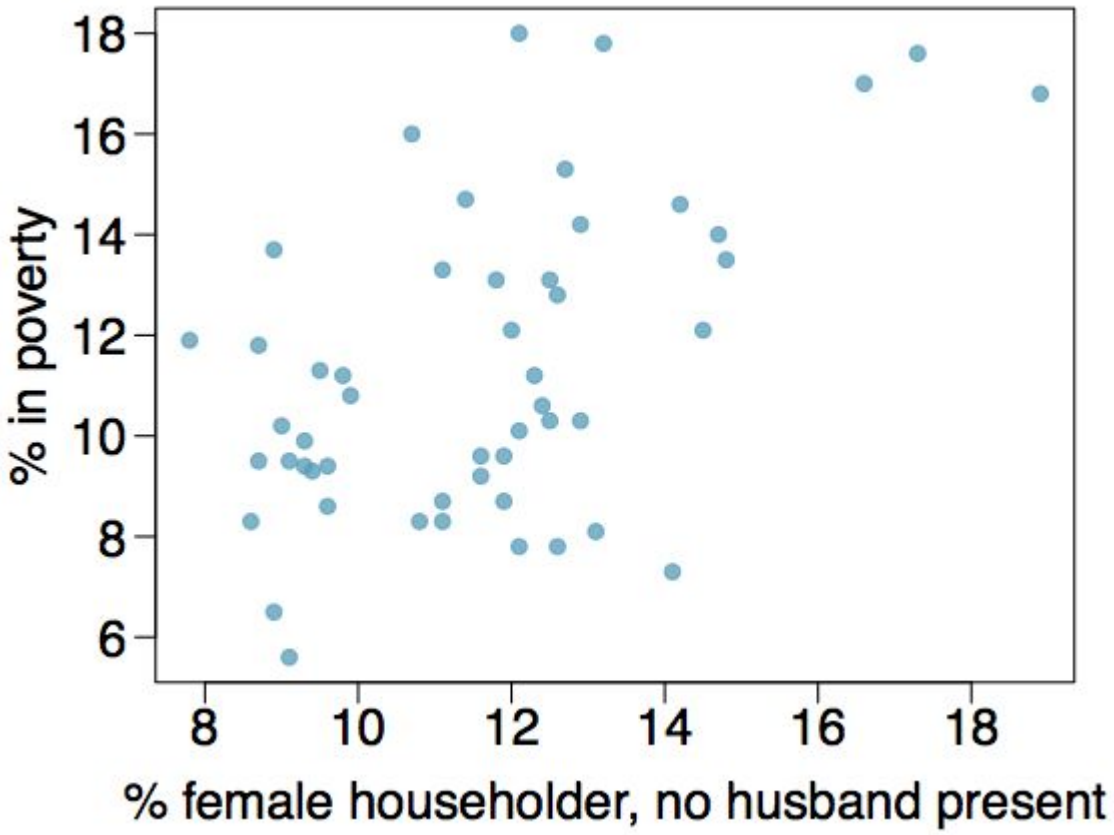
(a) 0.1
(b) -0.6
(c) -0.4
(d) 0.9
(e) 0.5



Guessing the correlation

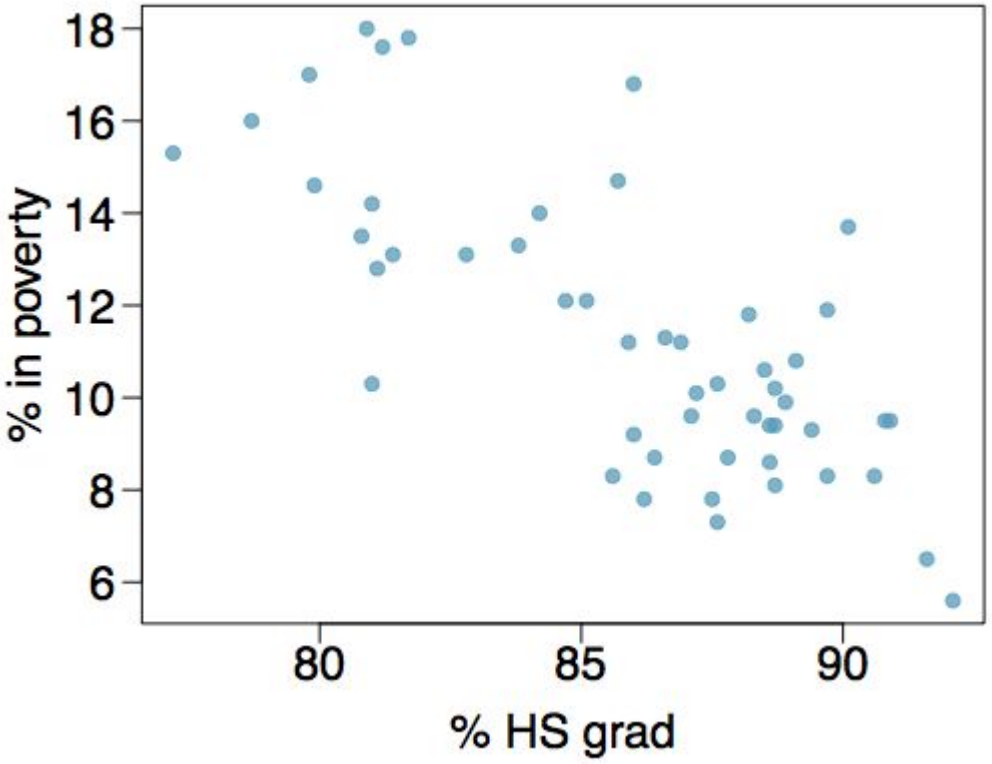
Which of the following is the best guess for the correlation between percent in poverty and percent female householder?

(a) 0.1
(b) -0.6
(c) -0.4
(d) 0.9
(e) 0.5



Poverty vs. HS graduate rate

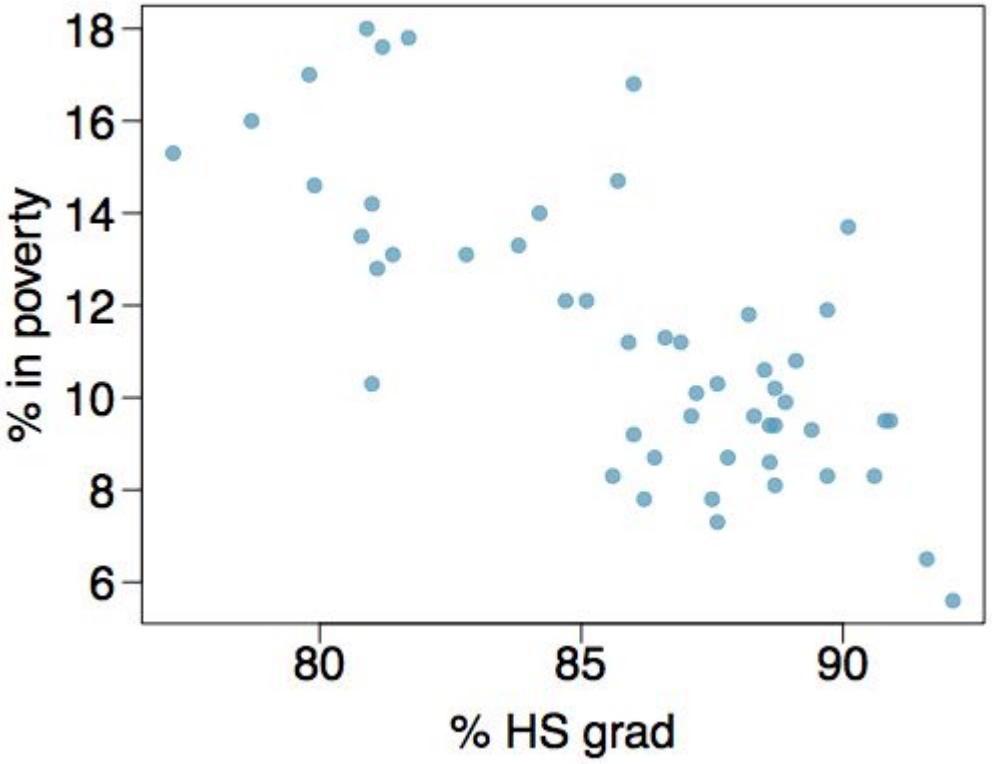
The *scatterplot* below shows the relationship between HS graduate rate in all 50 US states and DC and the percent of residents who live below the poverty line (income below $23,050 for a family of 4 in 2012).



Explanatory variable?

Poverty vs. HS graduate rate

The *scatterplot* below shows the relationship between HS graduate rate in all 50 US states and DC and the percent of residents who live below the poverty line (income below $23,050 for a family of 4 in 2012).

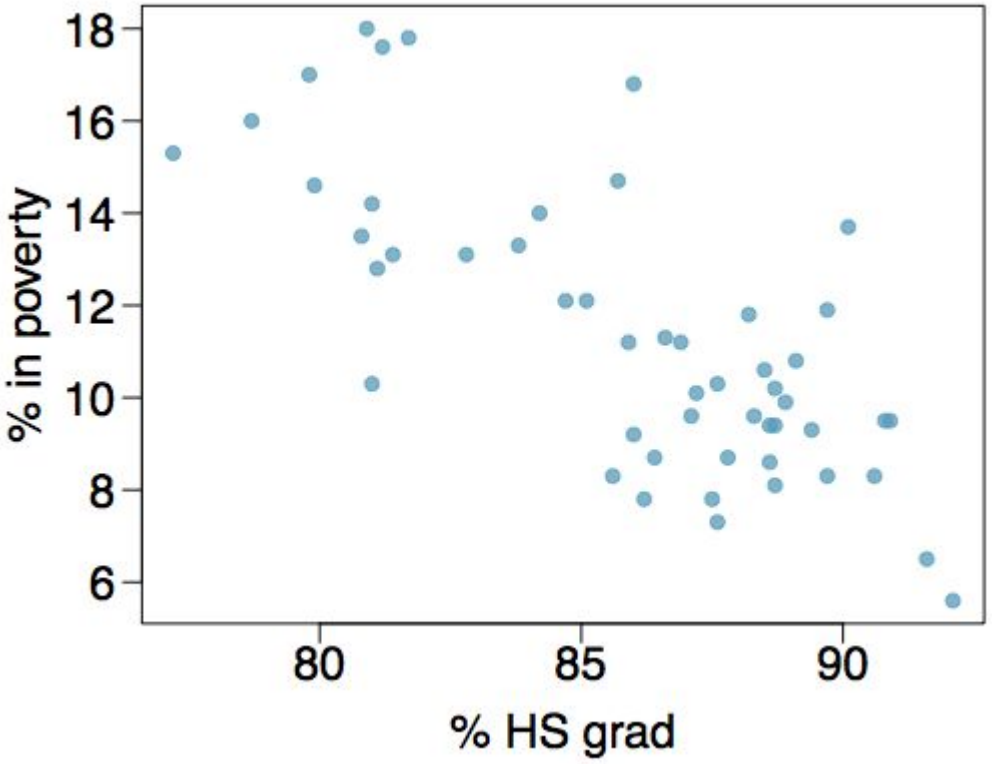


Explanatory variable?

% HS grad

Poverty vs. HS graduate rate

The *scatterplot* below shows the relationship between HS graduate rate in all 50 US states and DC and the percent of residents who live below the poverty line (income below $23,050 for a family of 4 in 2012).



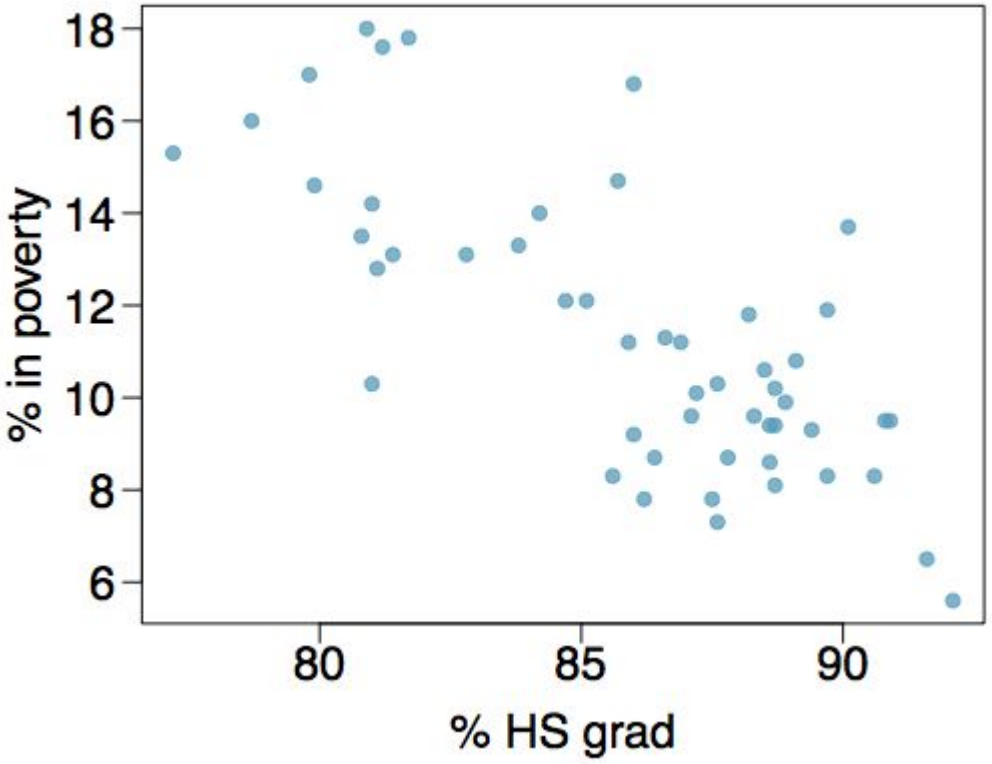
Explanatory variable?

% HS grad

Response variable?

Poverty vs. HS graduate rate

The *scatterplot* below shows the relationship between HS graduate rate in all 50 US states and DC and the percent of residents who live below the poverty line (income below $23,050 for a family of 4 in 2012).



Explanatory variable?

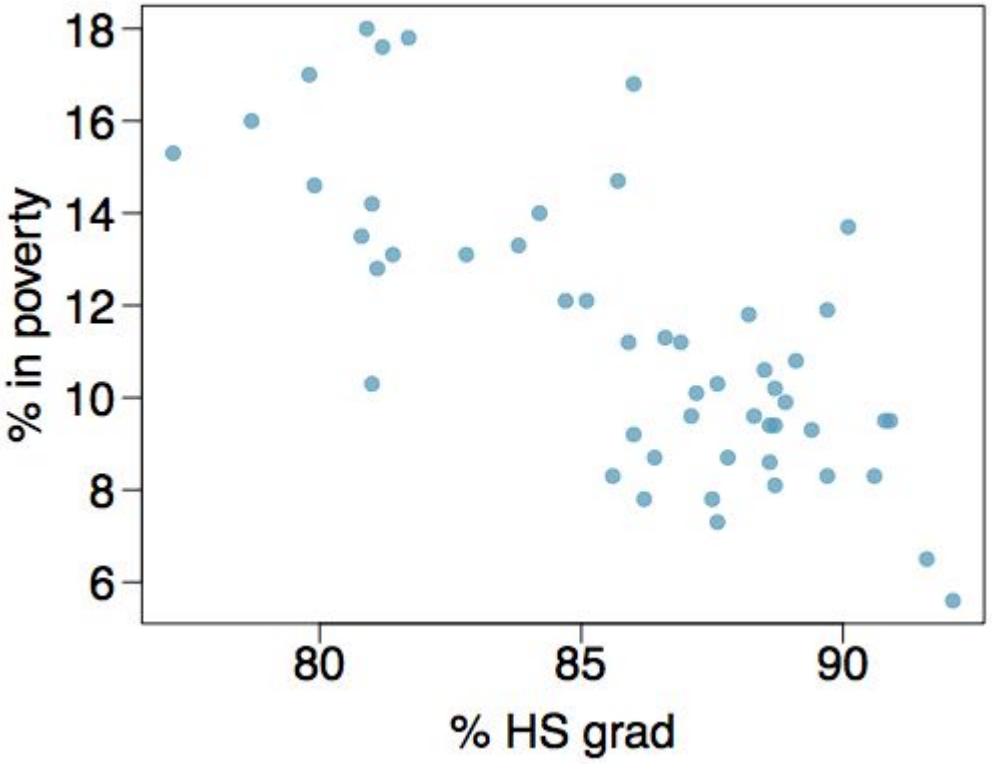
% HS grad

Response variable?

% in poverty

Poverty vs. HS graduate rate

The *scatterplot* below shows the relationship between HS graduate rate in all 50 US states and DC and the percent of residents who live below the poverty line (income below $23,050 for a family of 4 in 2012).



Explanatory variable?

% HS grad

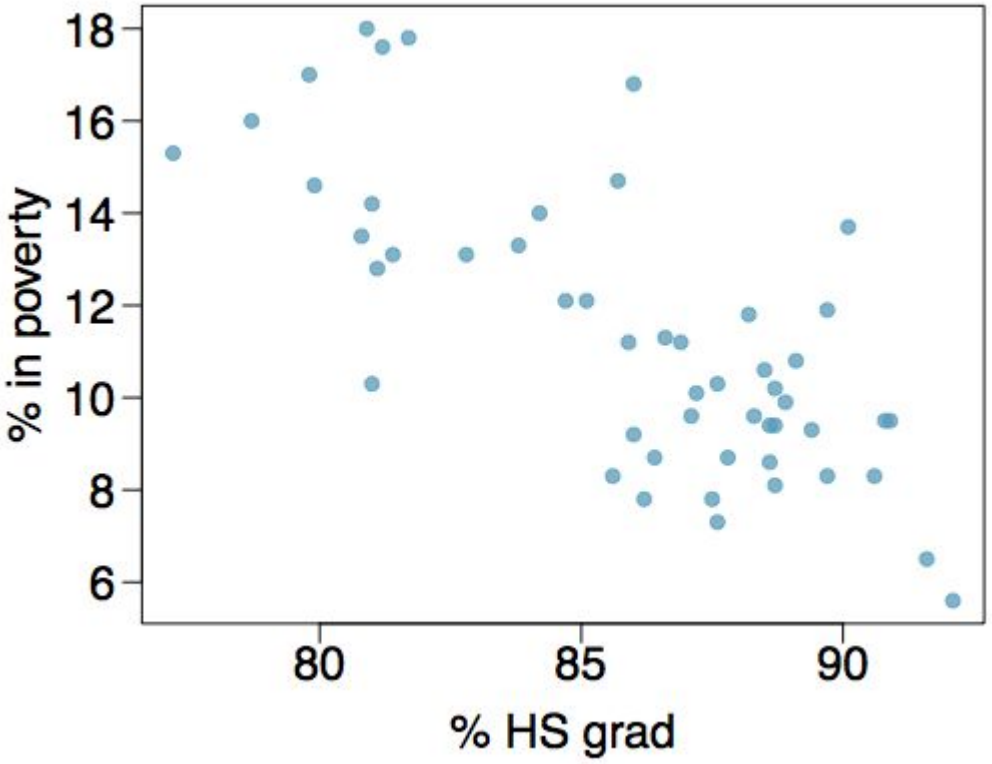
Response variable?

% in poverty

Relationship?

Poverty vs. HS graduate rate

The *scatterplot* below shows the relationship between HS graduate rate in all 50 US states and DC and the percent of residents who live below the poverty line (income below $23,050 for a family of 4 in 2012).



Explanatory variable?

% HS grad

Response variable?

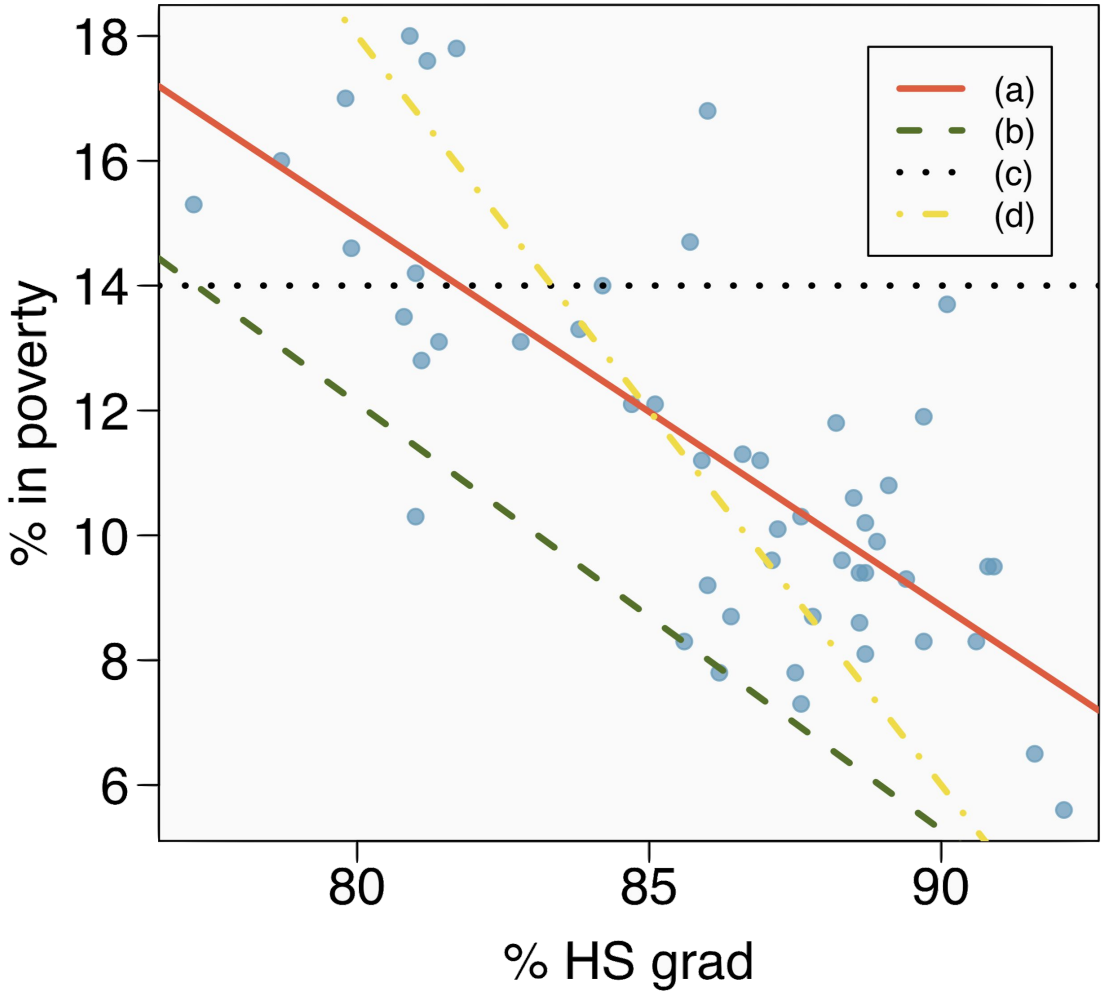
% in poverty

Relationship?

linear, negative, moderately strong

Eyeballing the line

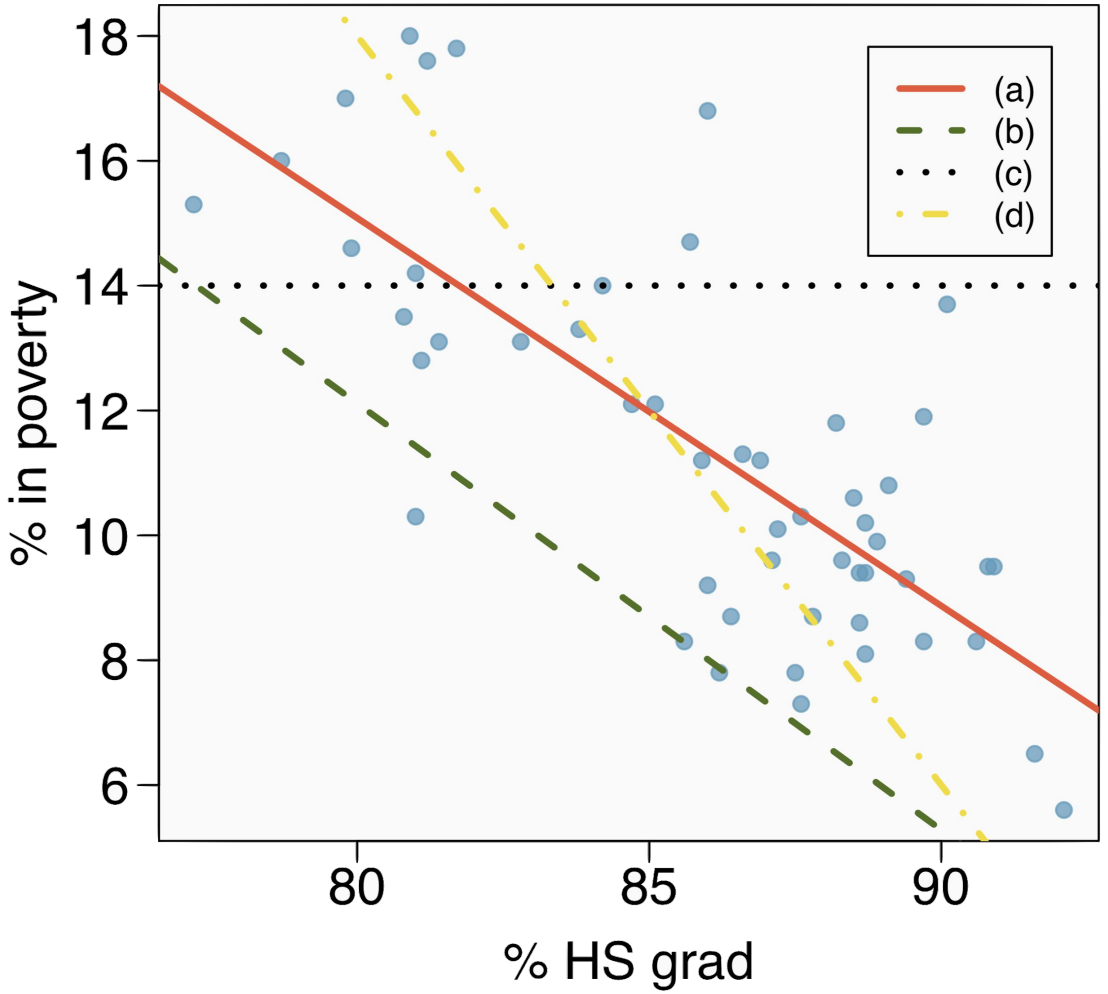
Which of the following appears to be the line that best fits the linear relationship between % in poverty and % HS grad? Choose one.



Eyeballing the line

Which of the following appears to be the line that best fits the linear relationship between % in poverty and % HS grad? Choose one.

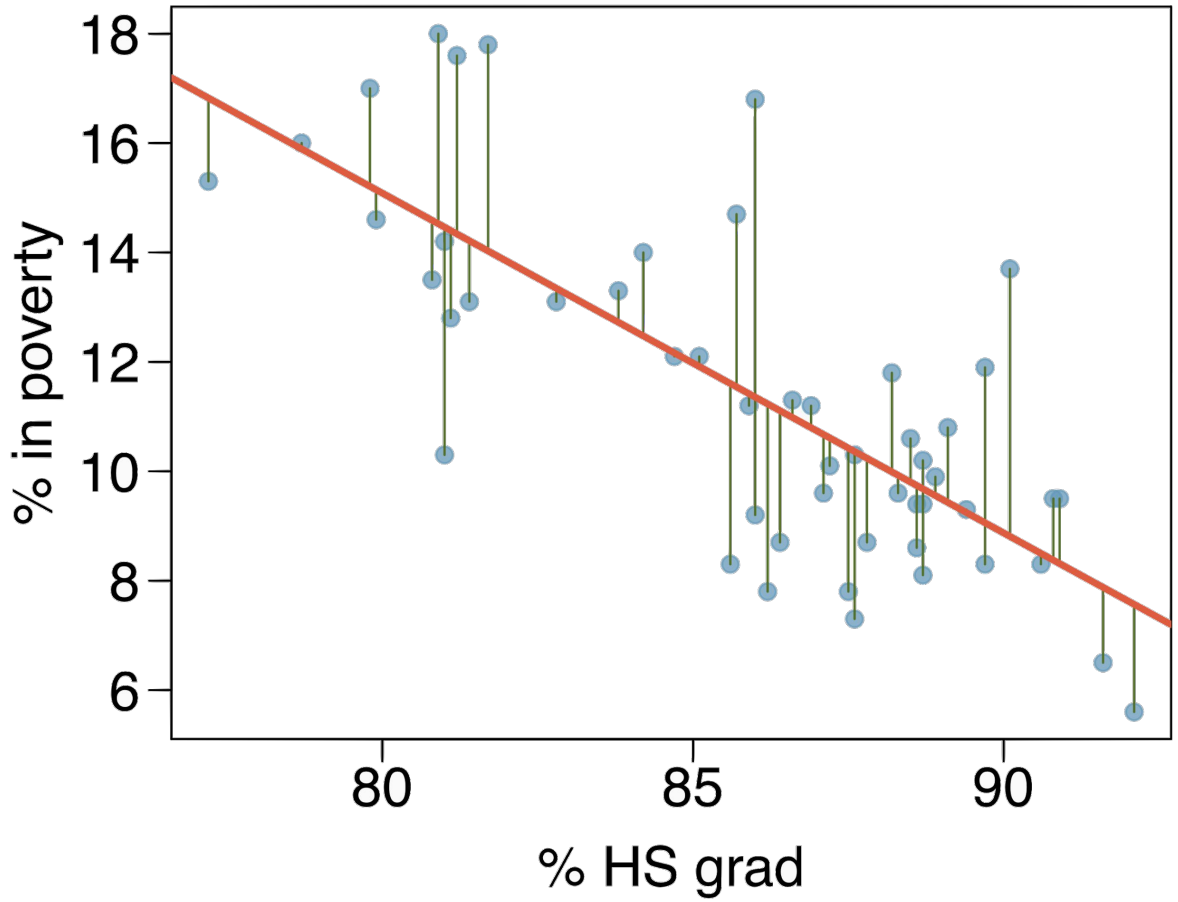
(a)



Residuals

Residuals are the leftovers from the model fit:

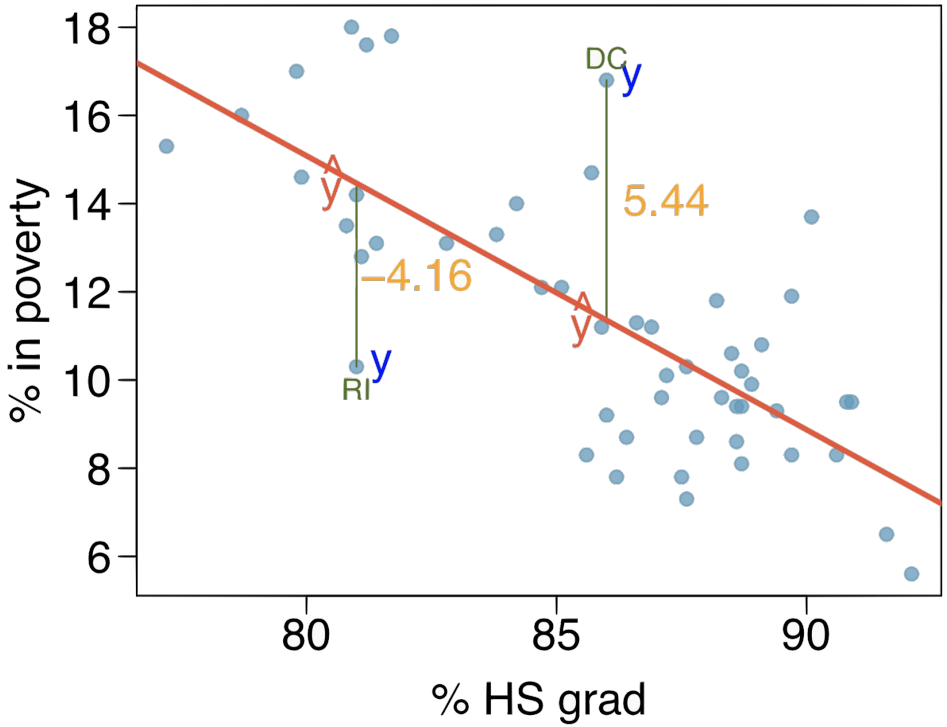
Data = Fit + Residual



Residuals (cont.)

Residual is the difference between the observed (y_i) and predicted $\hat{y}_i$.

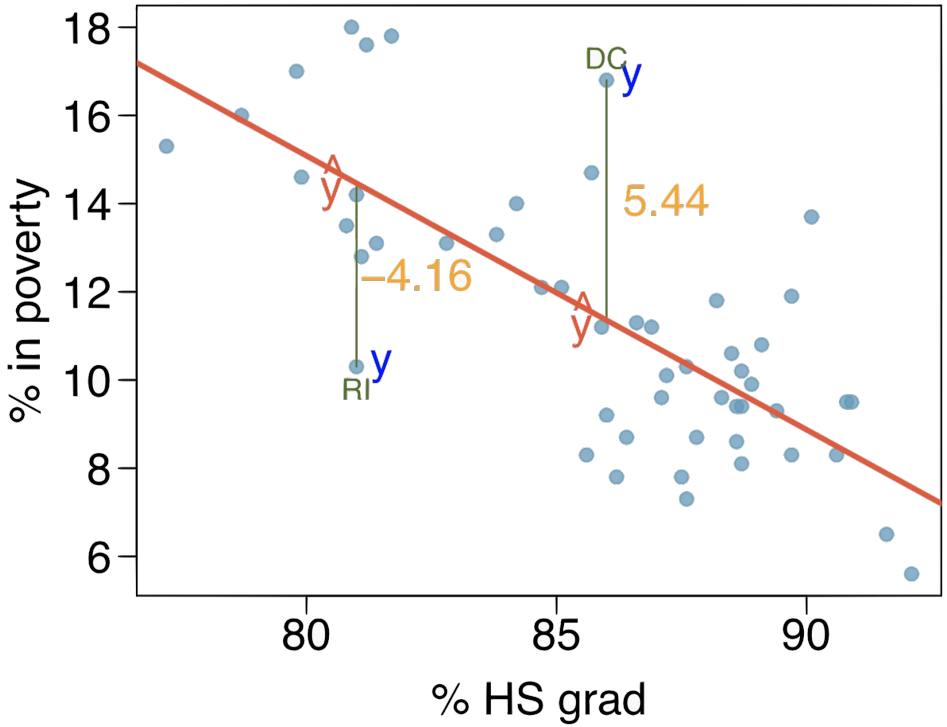
$$e_i = y_i - \hat{y}_i$$



Residuals (cont.)

Residual is the difference between the observed (y_i) and predicted $\hat{y}_i$.

$$e_i = y_i - \hat{y}_i$$

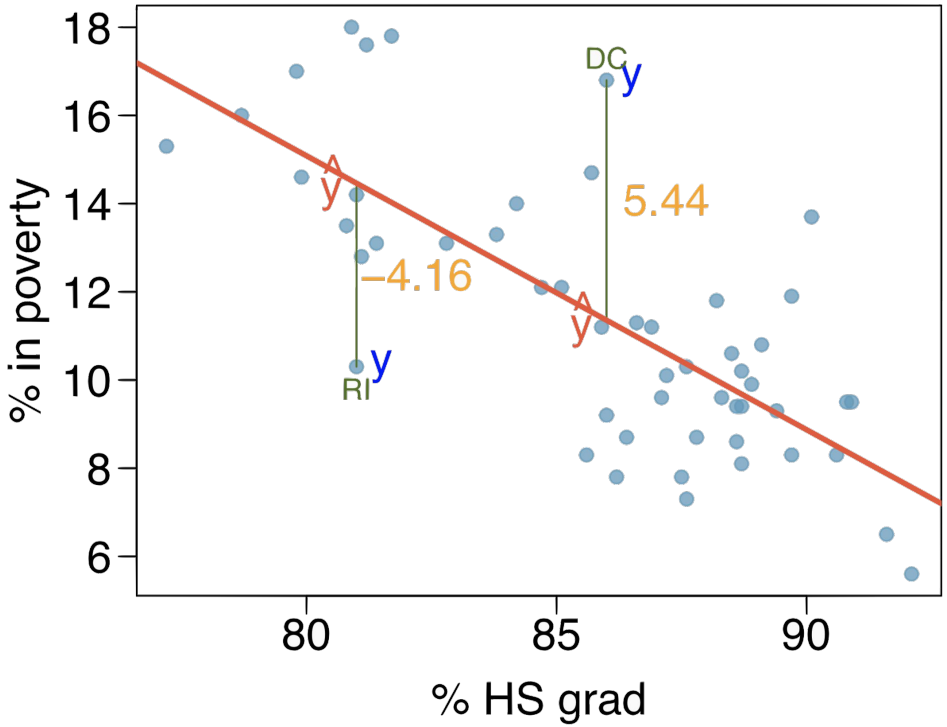


% living in poverty in DC is 5.44% more than predicted.

Residuals (cont.)

Residual is the difference between the observed (y_i) and predicted $\hat{y}_i$.

$$e_i = y_i - \hat{y}_i$$



% living in poverty in DC is 5.44% more than predicted.

% living in poverty in RI is 4.16% less than predicted.

Quantifying the relationship

- *Correlation* describes the strength of the *linear* association between two variables.

Poverty vs. HS graduate rate

The linear model for predicting poverty from high school graduation rate in the US is

$$\hat{poverty} = 64.78 - 0.62 * HS_{grad}$$

The "hat" is used to signify that this is an estimate.

Poverty vs. HS graduate rate

The linear model for predicting poverty from high school graduation rate in the US is

$$\hat{poverty} = 64.78 - 0.62 * HS_{grad}$$

The "hat" is used to signify that this is an estimate.

The high school graduate rate in Georgia is 85.1%. What poverty level does the model predict for this state?

64.78 − 0.62 x 85.1 = 12.018