

Activity: regression modeling

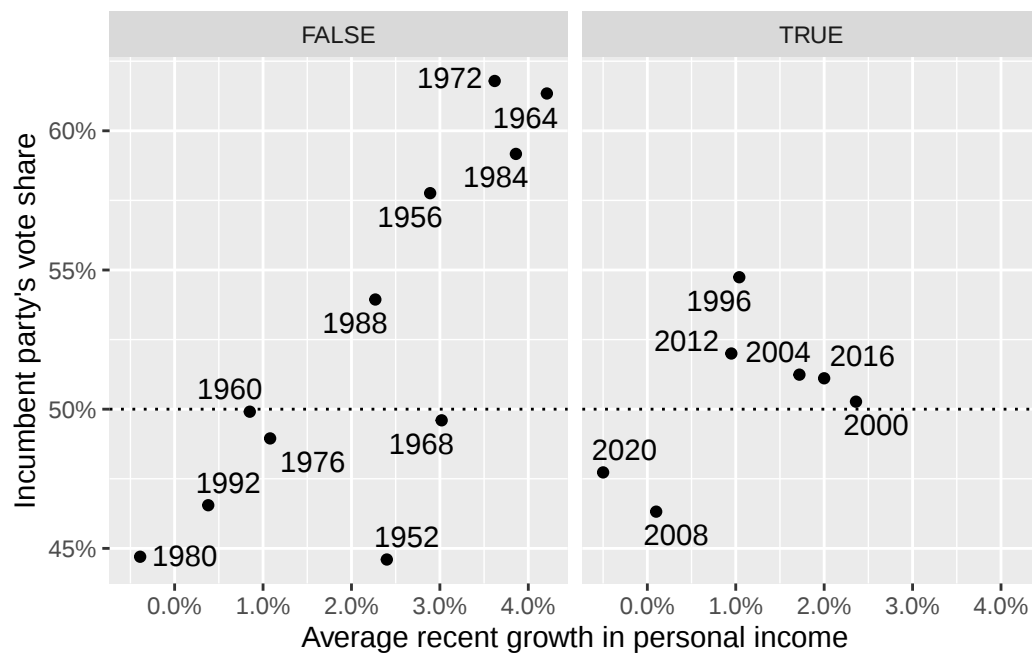
Consider the following data showing the relationship between the vote share of the president's incumbent party and the growth in average personal income over the preceding four-year period.

1. Draw (with a pencil) the **straight line** that you think “best” captures the data (on the plot below).



1. Estimate the **slope** and **intercept** of the line that you drew. Enter them in [this Google Sheet](#) (see link on Slack).
2. Write down the equation for the line (e.g, $\widehat{vote} = ? + ? \cdot incomeGrowth$).
3. Does the line intersect every data point exactly? If not, how far from the line is each point?
4. What criteria did you use to determine which line was “best”?

5. [Instructor] Use the `lm()` function to compute the least squares regression line.
6. What is the meaning of the intercept?
7. What is the meaning of the slope?
8. Political scientists have identified an increase in [political polarization](#) since 1990 (Gelman 2024). How would this affect the relationship between personal income growth and incumbent vote share?
9. Consider splitting the data into two parts for before and after 1990. What best-fit line(s) would you draw now?



10. To what extent does the data support the claim of increase political polarization since 1990?

References

Gelman, Andrew. 2024. *Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do*. Princeton University Press. https://www.google.com/books/edition/Red_State_Blue_State_Rich_State_Poor_Sta/jbingNQfdLgC?hl=en&gbpv=0.