PSC 202
SYRACUSE UNIVERSITY

INTRODUCTION TO POLITICAL ANALYSIS

BIVARIATE HYPOTHESIS TESTING PART 2

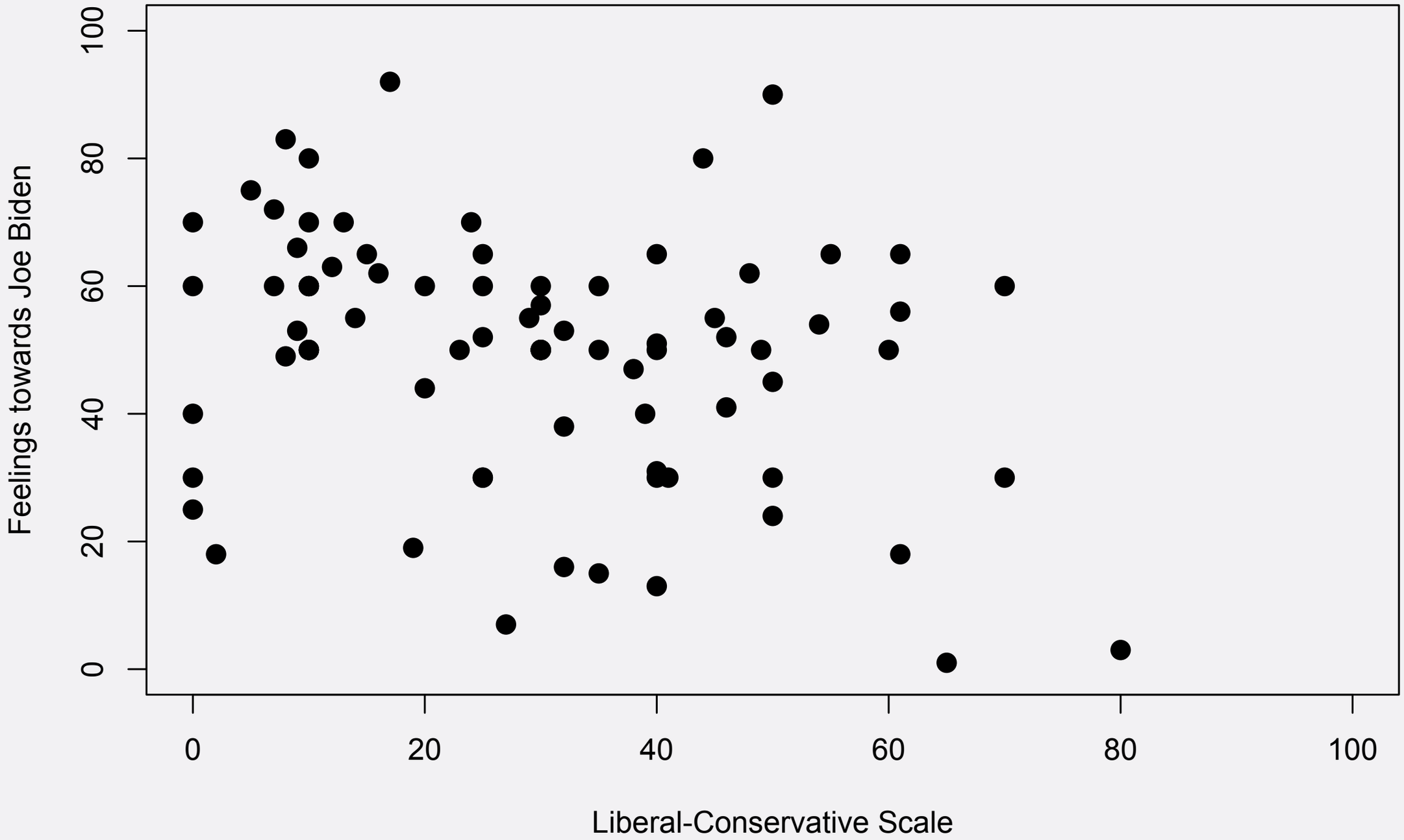
WHERE WE ARE

- Is there a credible causal mechanism that connects X to Y?
- Can we rule out the possibility that Y could cause X?
- Is there covariation between X and Y?
- Have we controlled for all confounding variables (Z) that might make the association between X and Y spurious?

BIVARIATE RELATIONSHIPS

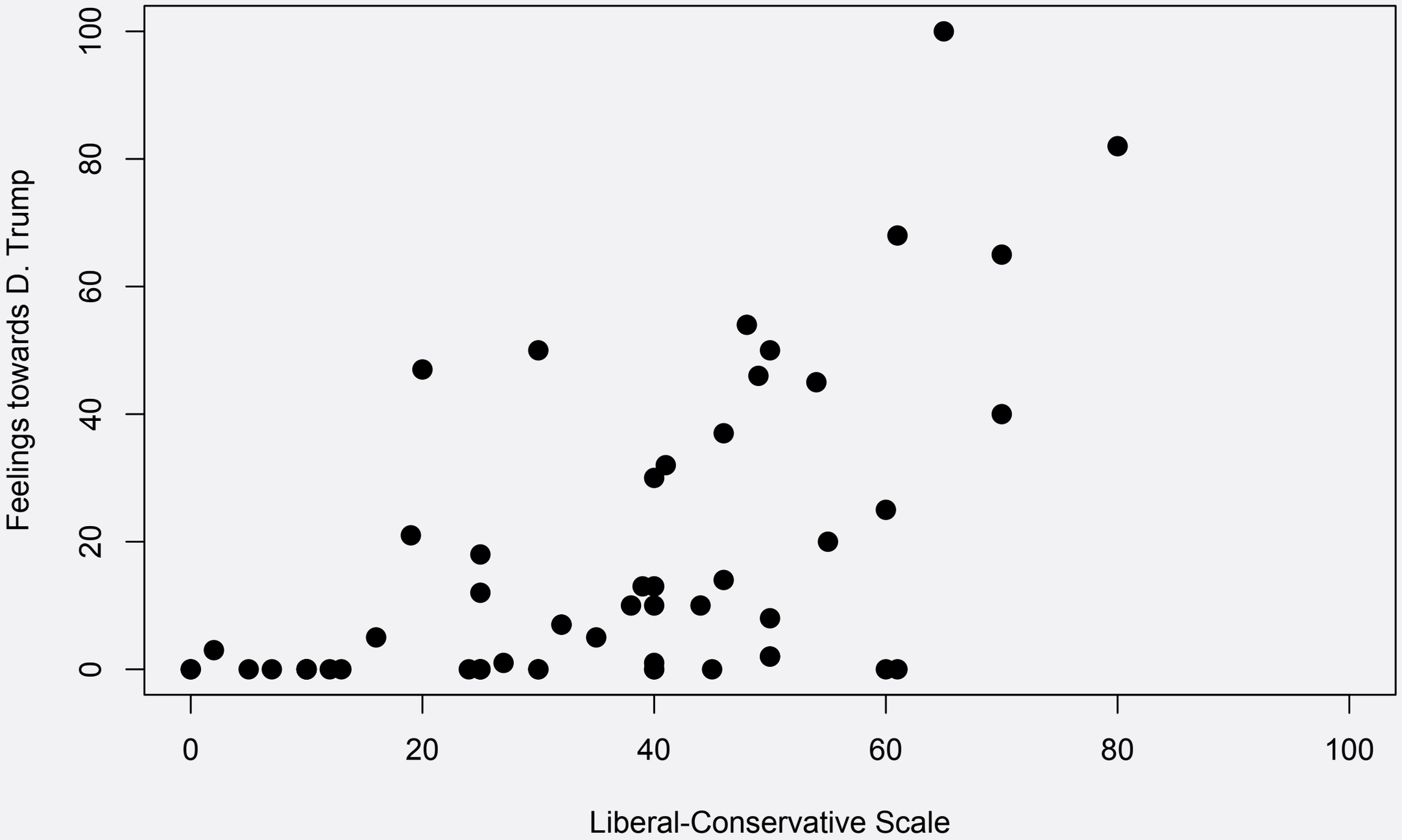
Dependent Variable \ Independent Variable	Nominal/Ordinal	Interval
Nominal/Ordinal	Cross-Tabulation	Not In This Class…
Interval	Mean Comparison	Correlation Coefficient

JOE BIDEN



r=-0.29

DONALD TRUMP



r=0.61

PEARSON'S R

$$r = \frac{\sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)}{n - 1}$$

- **Huh?**

OR...

Pearson Correlation Coefficient Calculator

Pearson's correlation coefficient measures the strength and direction of the relationship between two variables. To begin, you need to add your data to the text boxes below (either one value per line or as a comma delimited list). So, for example, if you were looking at the relationship between height and shoe size, you'd add your values for height into the X Values box and the values for shoes size into the Y Values box (or vice versa).

When your data is in place, and you're ready to do the calculation, just hit the "Calculate R" button, and the calculator will run various tests on your data - to make sure it is suitable for the Pearson statistic - and then spit out the correlation coefficient, together with a lot of detail about the calculation.

X Values

Y Values

OR...

Pearson Correlation Coefficient Calculator

Pearson's correlation coefficient measures the strength and direction of the relationship between two variables. To begin, you need to add your data to the text boxes below (either one value per line or as a comma delimited list). So, for example, if you were looking at the relationship between height and shoe size, you'd add your values for height into the X Values box and the values for shoes size into the Y Values box (or vice versa).

When your data is in place, and you're ready to do the calculation, just hit the "Calculate R" button, and the calculator will run various tests on your data - to make sure it is suitable for the Pearson statistic - and then spit out the correlation coefficient, together with a lot of detail about the calculation.

X Values	Y Values
4 5 8 34 24 5 -3	5 9 2 4 16 -3 4

Enter some data!

Calculate R Reset

OR...

Pearson Correlation Coefficient Calculator

The value of R is: 0.3589.

Explanation of results

As you have probably already noticed, the output of this calculator is... verbose. Although most of the information provided below is self-explanatory, there are a few things worth noting. First, the five text boxes spread across the middle of the page represent the calculations that would be required if you were to calculate the R value in stages. Second, there is more than one way to calculate the R value, but these are all mathematically equivalent, so you shouldn't worry if you don't recognize the equation used here. Third, in the "Result Details & Calculations" box, you'll find what we've called a cross-check value, which is the R value calculated using an algorithm supplied by the Meta Numerics statistical library. This should be identical to the value that we've calculated.

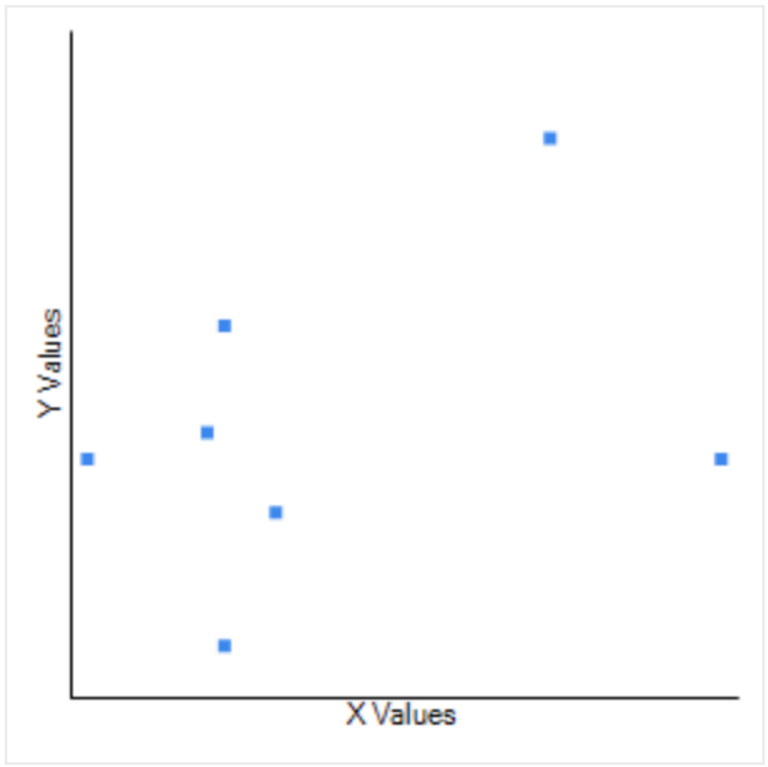
Note: If you want to calculate a P value from your R score, we have a calculator here (before clicking, remember to note your r score and record any calculation details you require).

X Values

```
4
5
8
34
24
5
-3
```

Y Values

```
5
9
2
4
16
-3
4
```

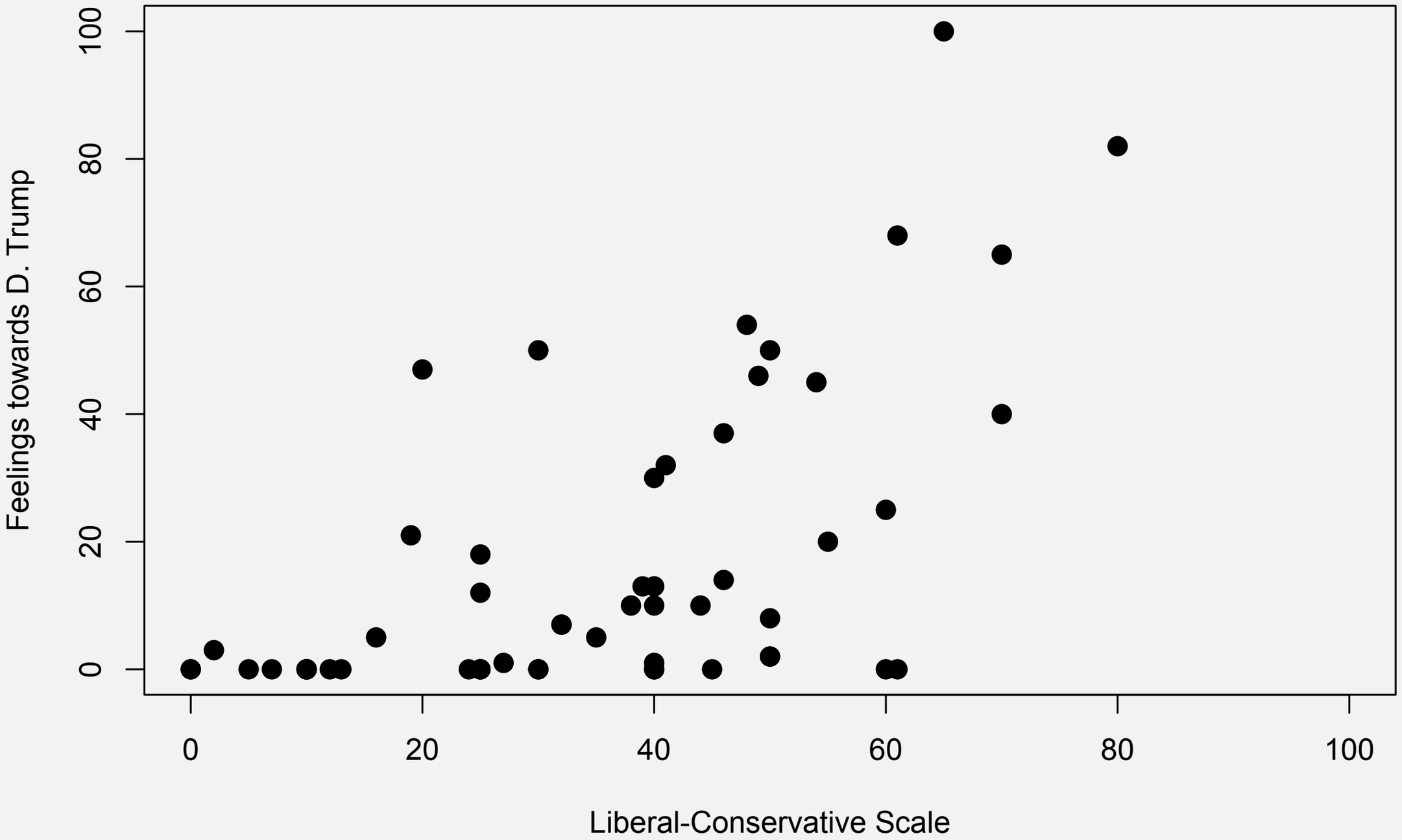


PEARSON'S R

$$r = \frac{\sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)}{n - 1}$$

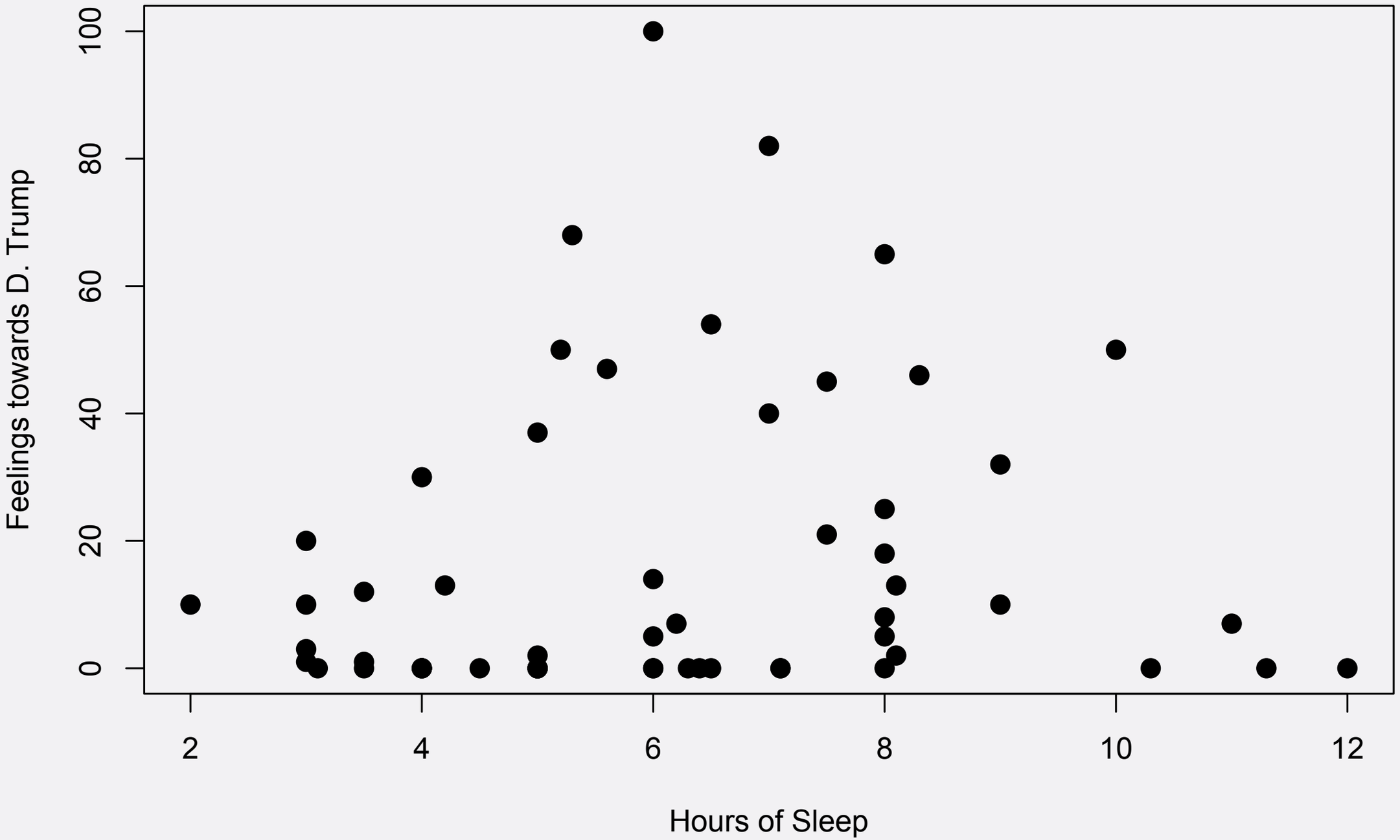
- **Intuition: Captures how much values of two variable vary together**
  - **High (positive) correlation: If X takes higher values, Y takes higher values**
  - **High (negative) correlation: If X takes higher values, Y takes lower values**
  - **Low correlation: If X takes higher values, values of Y do not move up or down**

DONALD TRUMP



r=0.61

DONALD TRUMP



r=-0.02

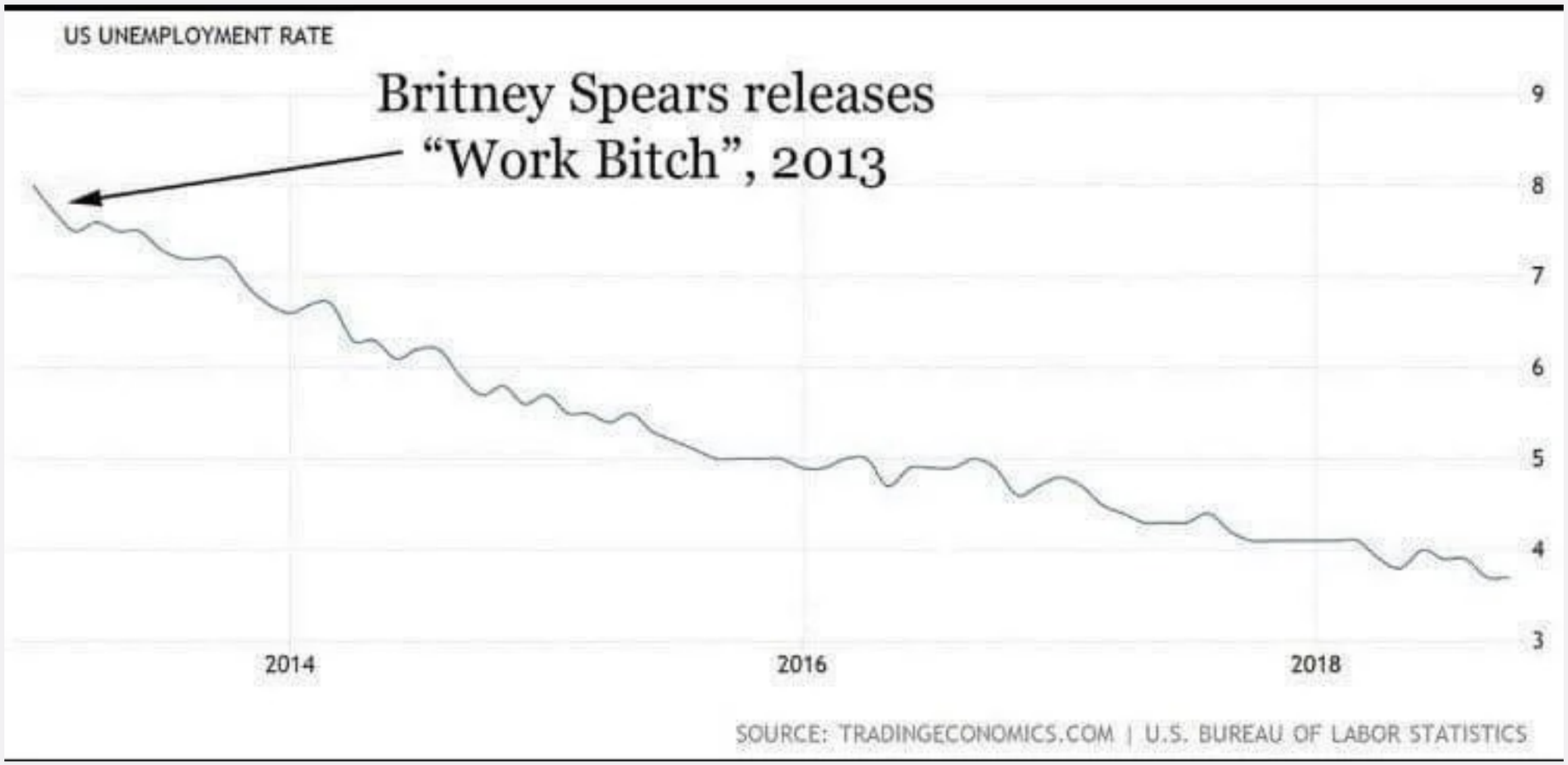
Table A.2. Correlation matrix

	PRESS	BUREAU	RULE	Log(GDP)	HUMCAP	TRADE	BLACK	ETHNIC	Corr-ICRG
PRESS	1.00								
BUREAU	−0.63	1.00							
RULE	−0.73	0.87	1.00						
Log(GDP)	−0.69	0.80	0.83	1.00					
HUMCAP	−0.60	0.69	0.64	0.79	1.00				
TRADE	−0.01	0.20	0.20	0.22	0.14	1.00			
BLACK	0.34	−0.32	−0.39	−0.45	−0.41	−0.11	1.00		
ETHNIC	0.47	−0.36	−0.41	−0.60	−0.47	−0.11	0.41	1.00	
Corr-ICRG	−0.74	0.79	0.83	0.75	0.58	0.20	−0.28	−0.43	1.00

BIVARIATE RELATIONSHIPS

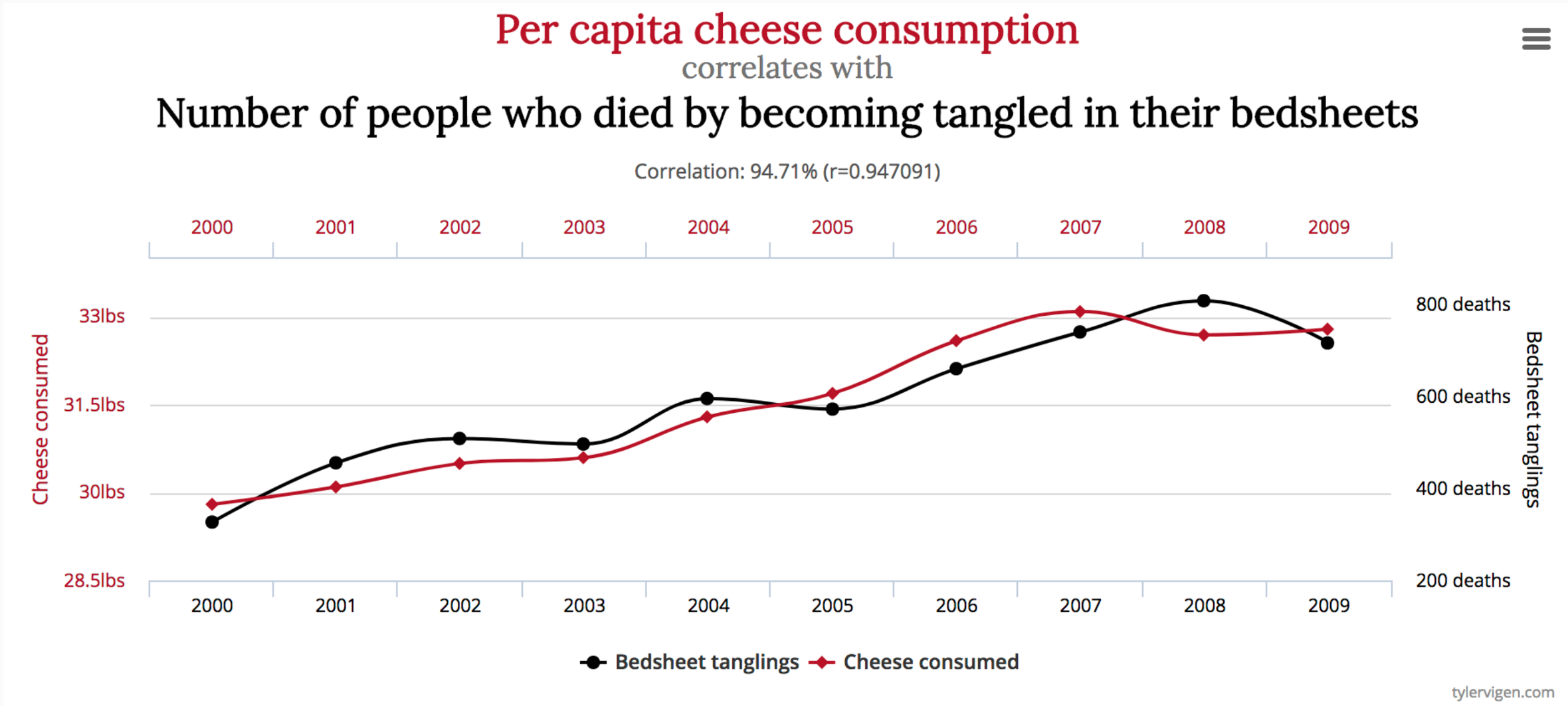
Dependent Variable \ Independent Variable	Nominal/Ordinal	Interval
Nominal/Ordinal	Cross-Tabulation	Not In This Class…
Interval	Mean Comparison	Correlation Coefficient

CAREFUL!



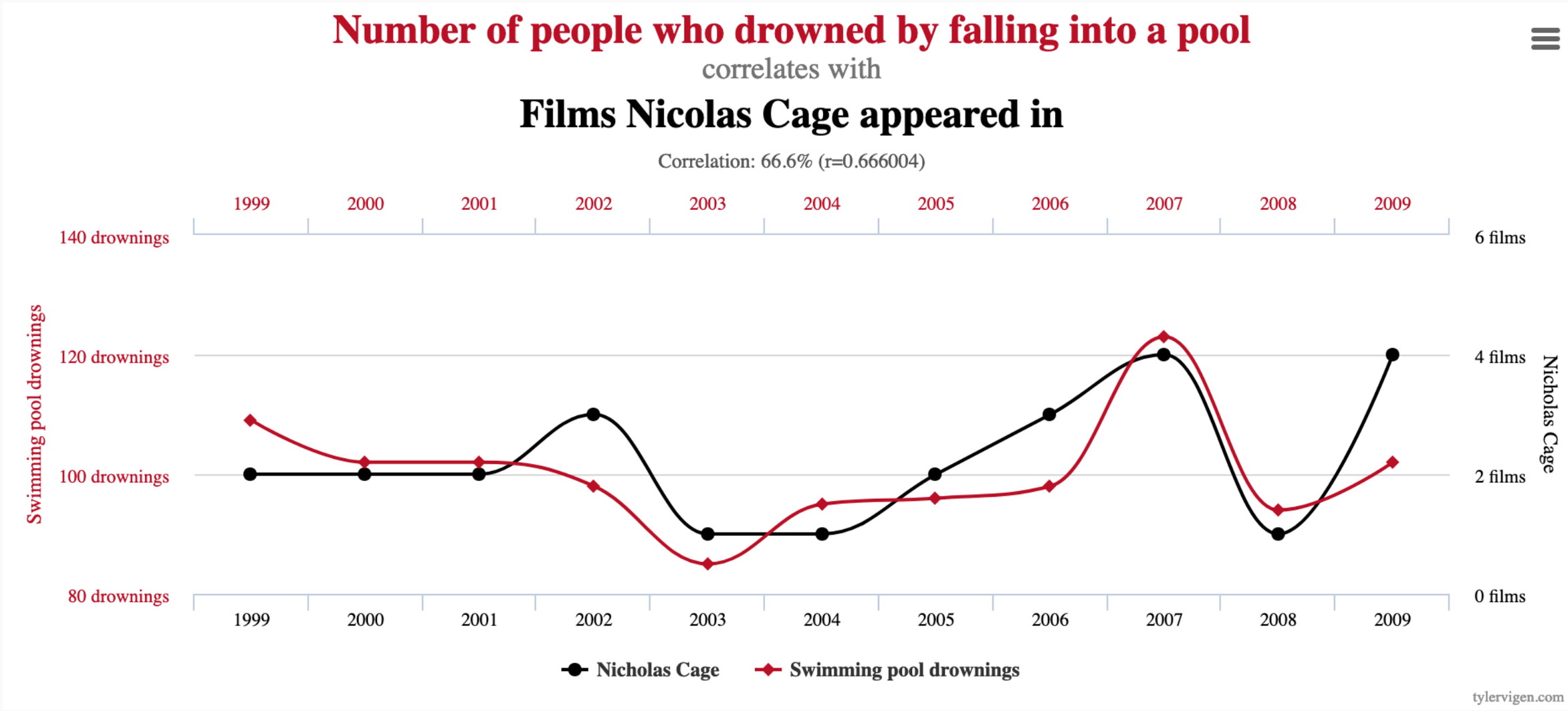
- Important: Just because we find a correlation between two variables does *not* mean that the independent variable *causes* the dependent variable

CAREFUL!



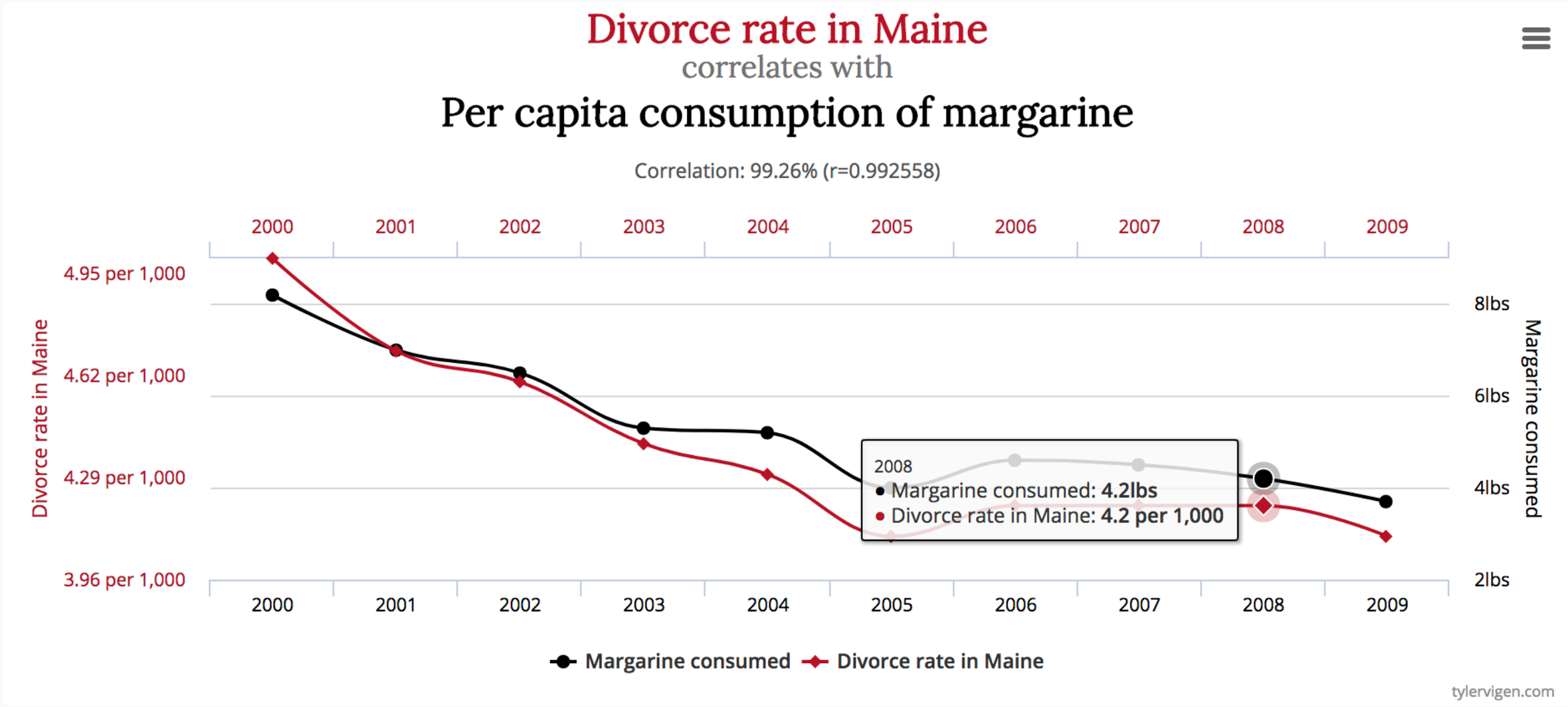
Data sources: U.S. Department of Agriculture and Centers for Disease Control & Prevention

CAREFUL!



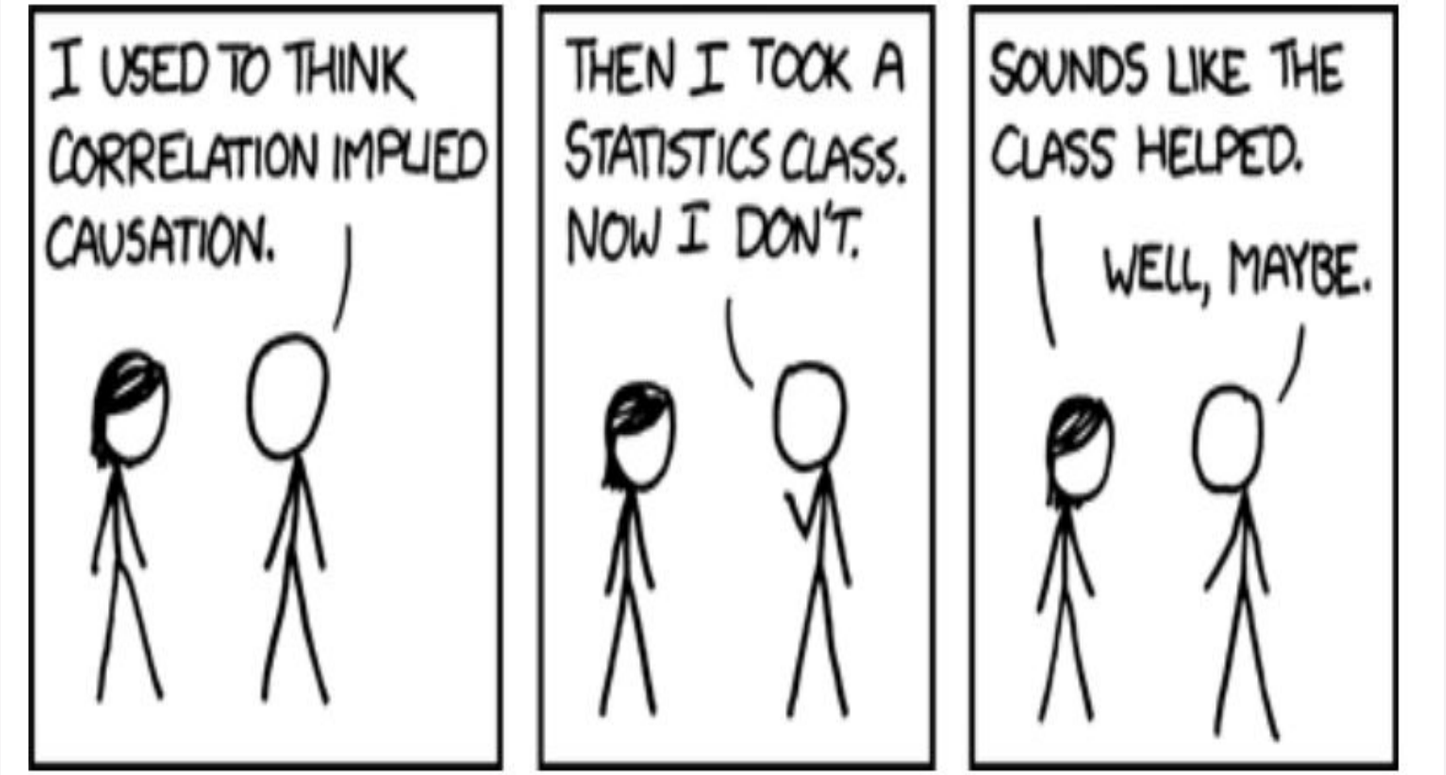
Data sources: Centers for Disease Control & Prevention and Internet Movie Database

CAREFUL!



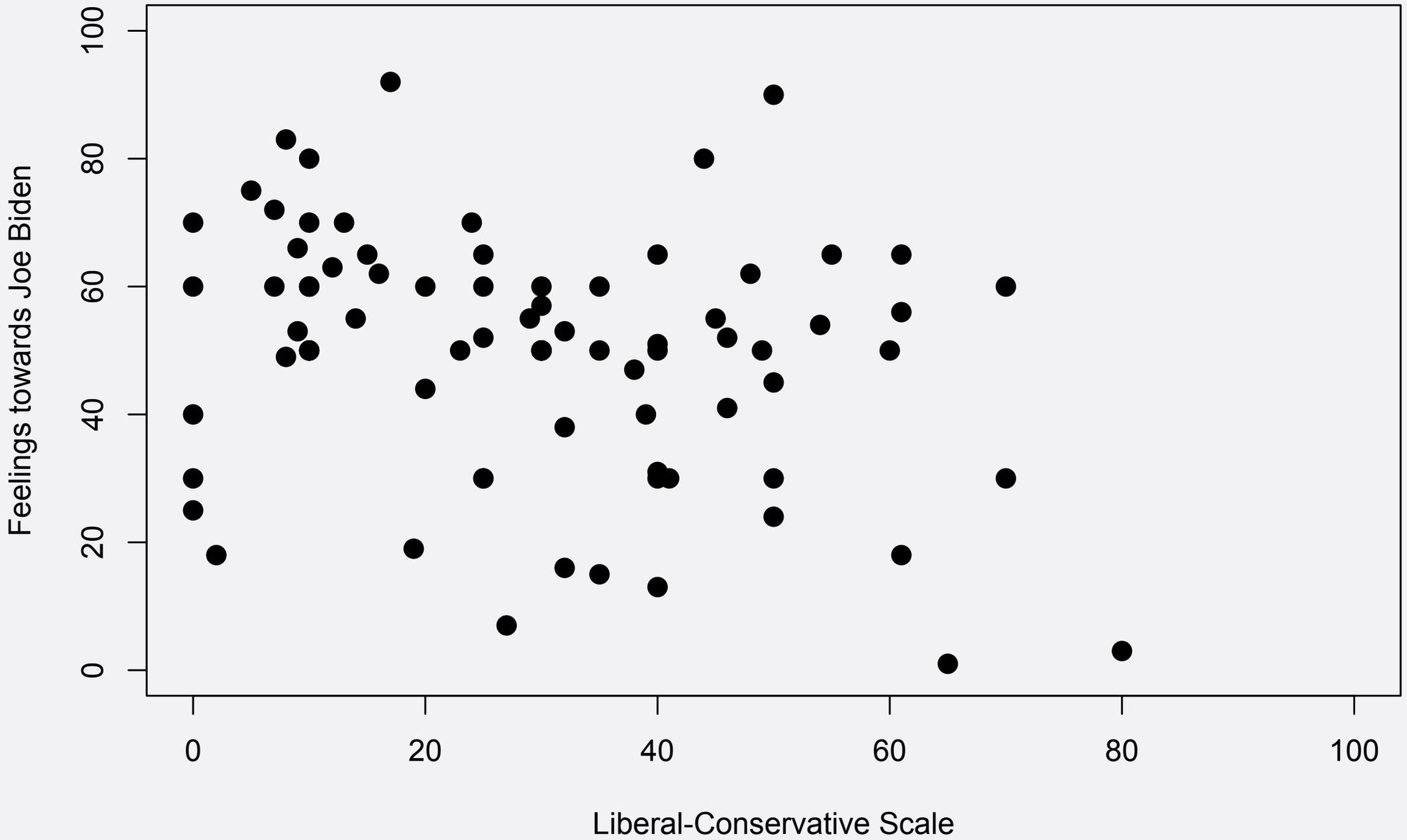
Data sources: National Vital Statistics Reports and U.S. Department of Agriculture

CAREFUL!



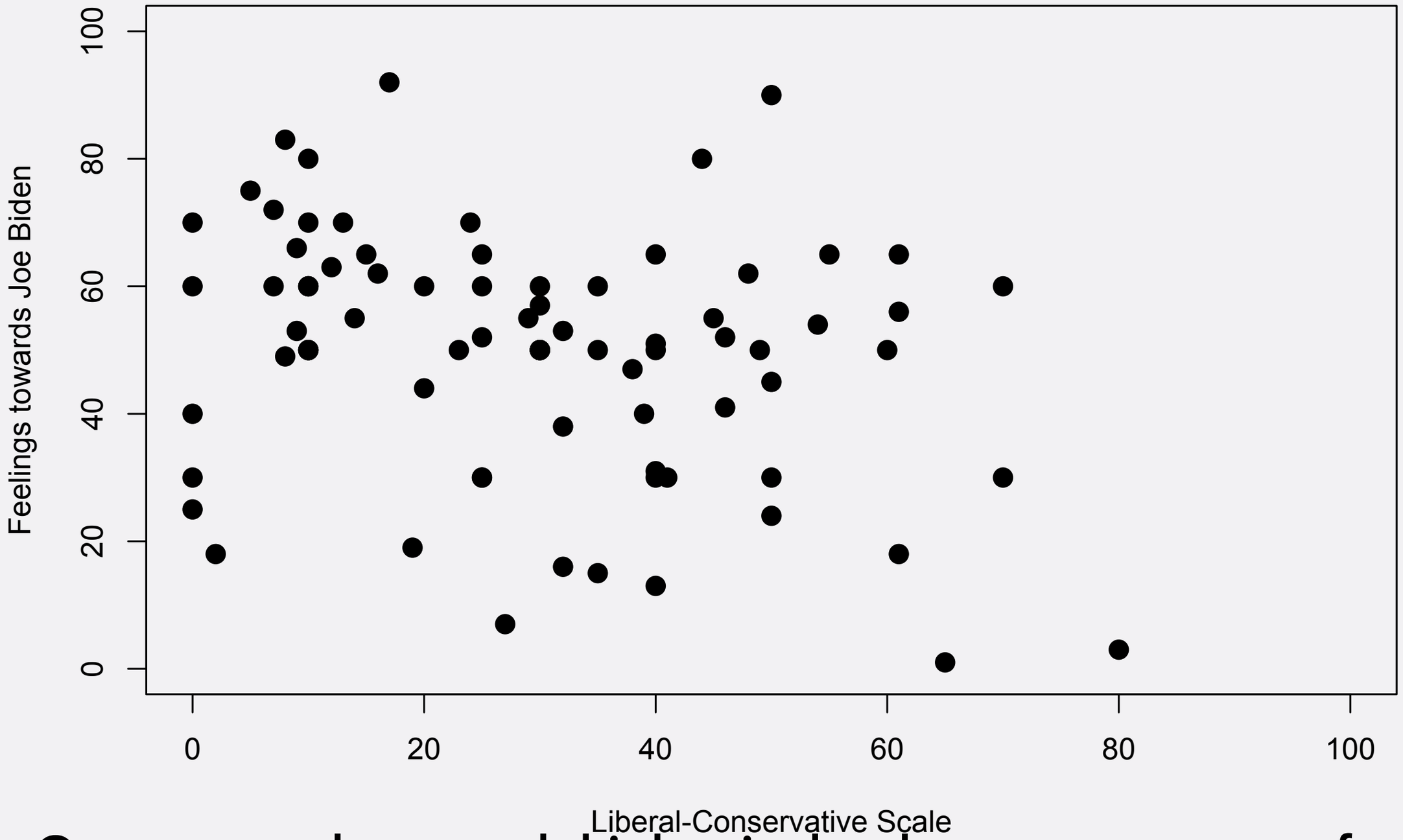
- The other hurdles to causality still apply!
  - Especially Hurdle 4: Have we controlled for all confounding variables (Z) that might make the association between X and Y spurious?
  - We'll talk about how to do this in a couple of weeks

JOE BIDEN



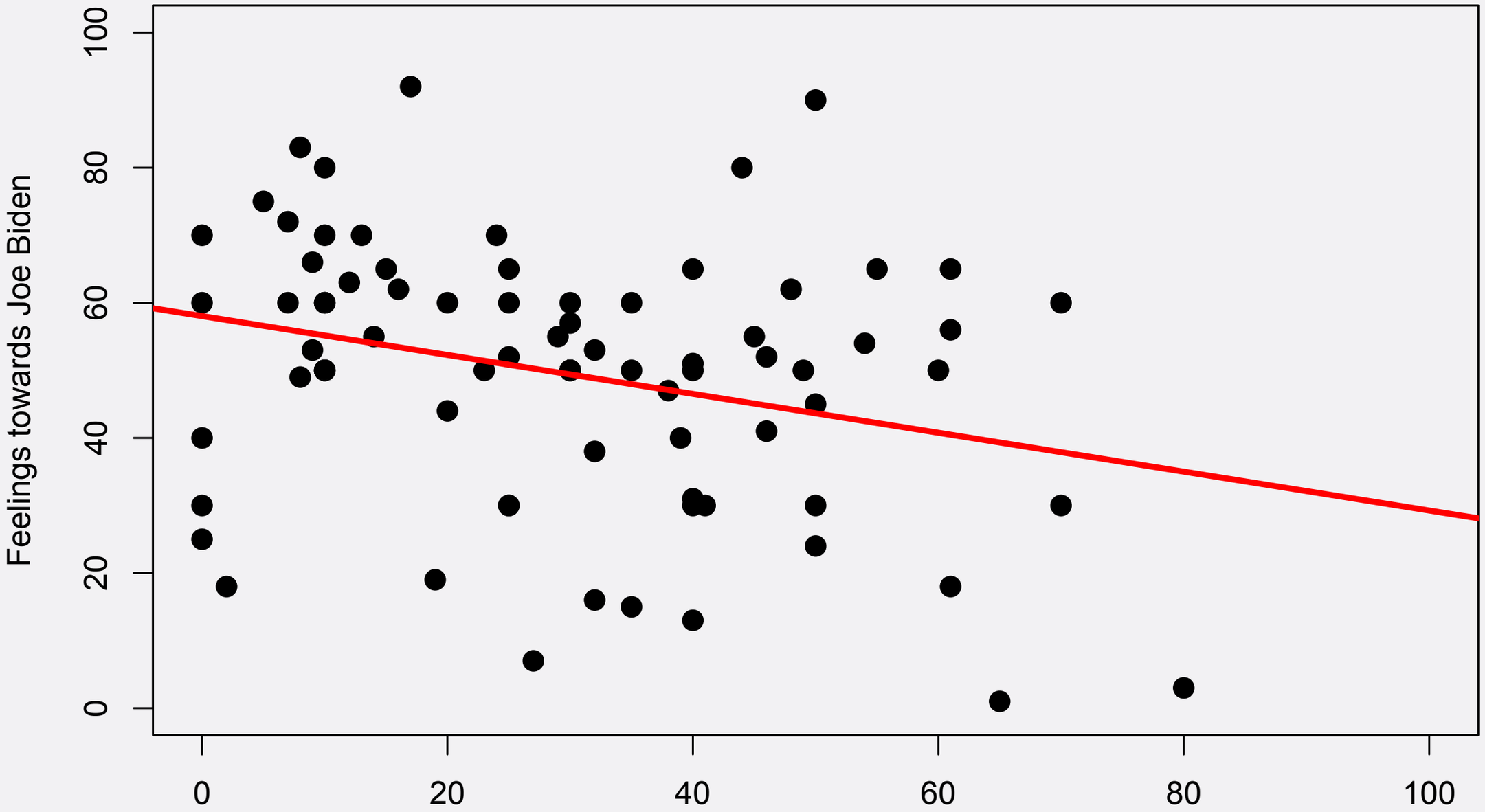
- r=-0.29
- Correlation: Direction and strength of relation, not size

JOE BIDEN



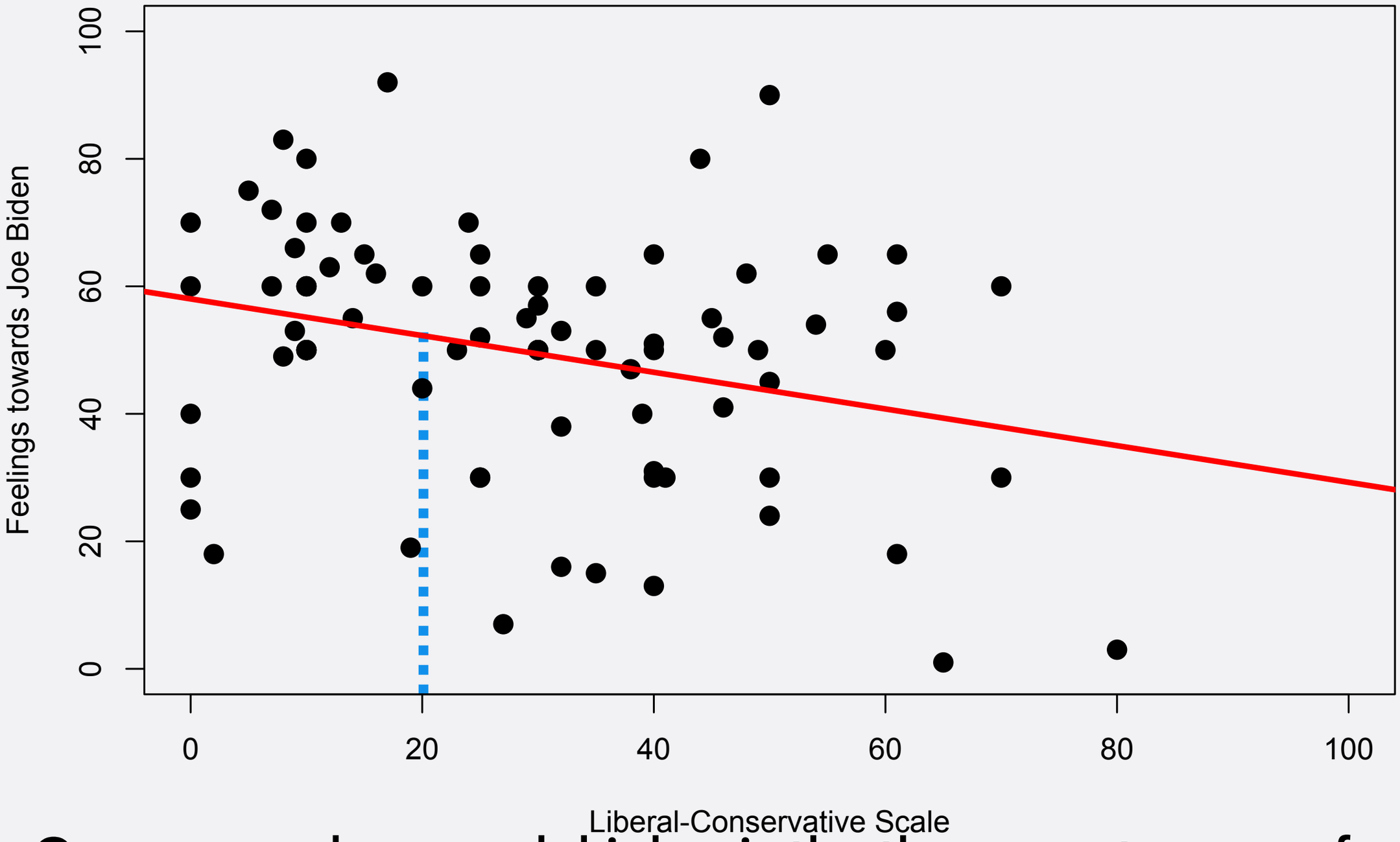
- **On average, how much higher is the thermometer score for someone who is a 20 on the liberal-conservative scale, compared to someone who is a 80?**

LINE



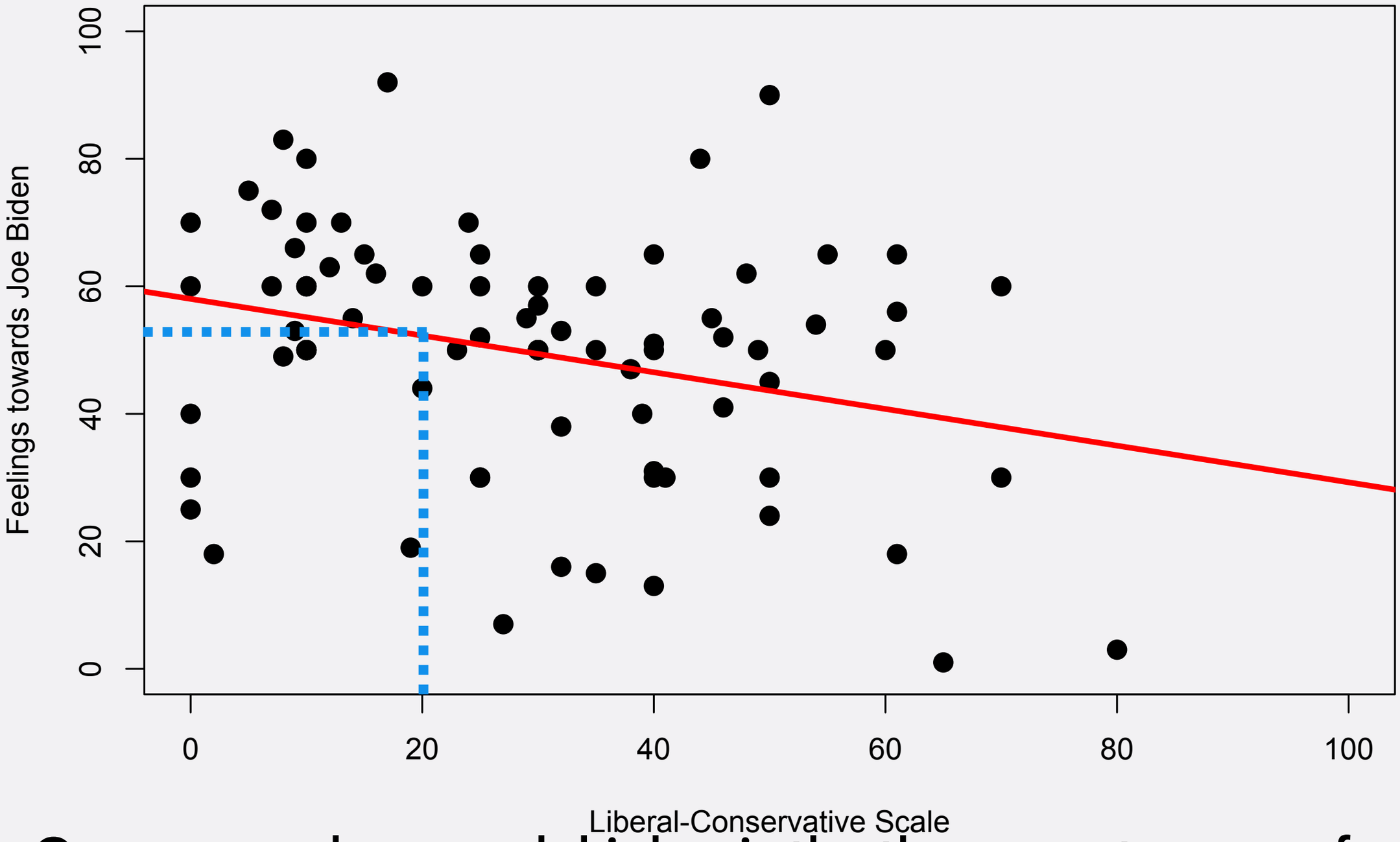
- **On average, how much higher is the thermometer score for someone who is a 20 on the liberal-conservative scale, compared to someone who is a 80?**

LINE



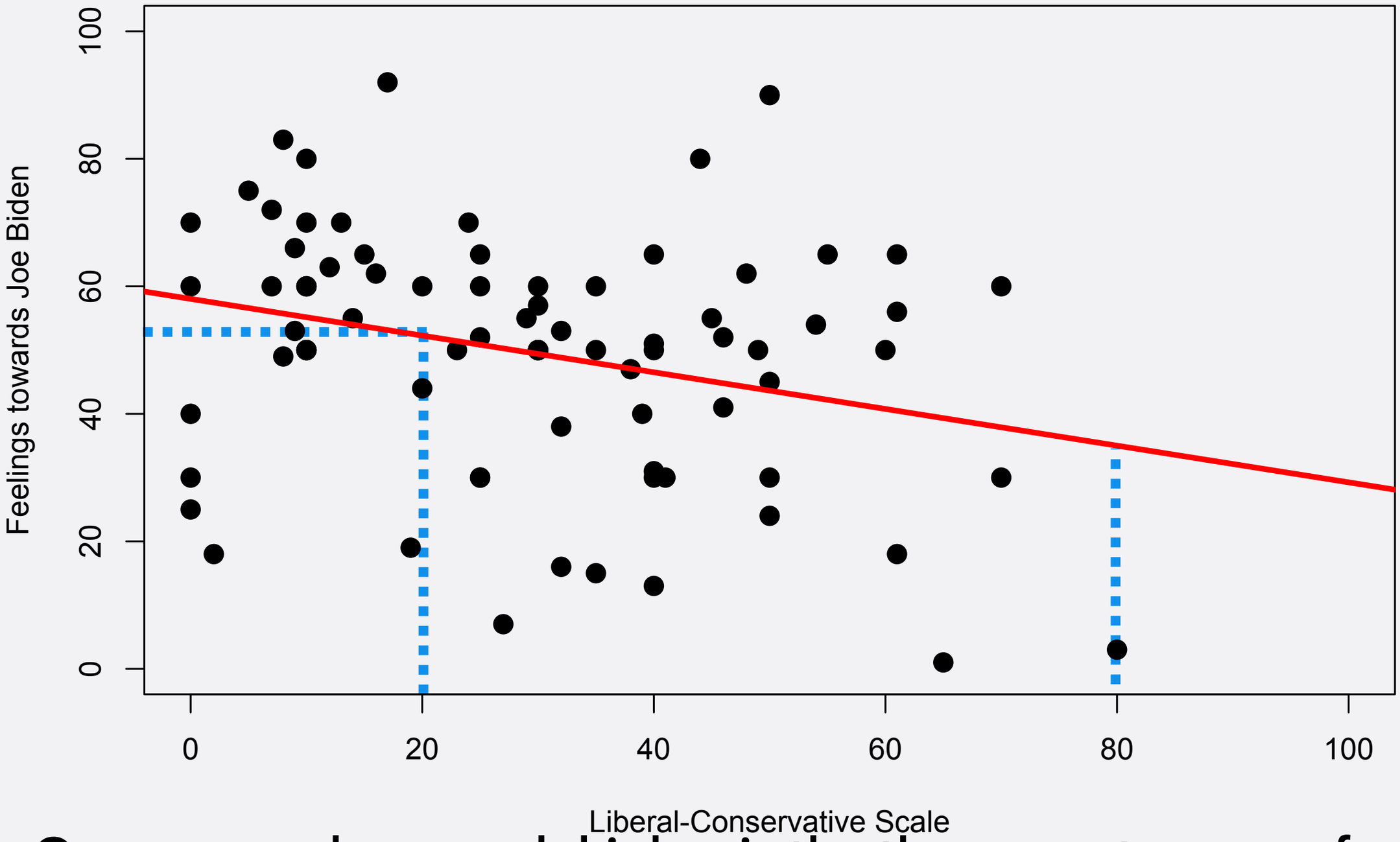
- **On average, how much higher is the thermometer score for someone who is a 20 on the liberal-conservative scale, compared to someone who is a 80?**

LINE



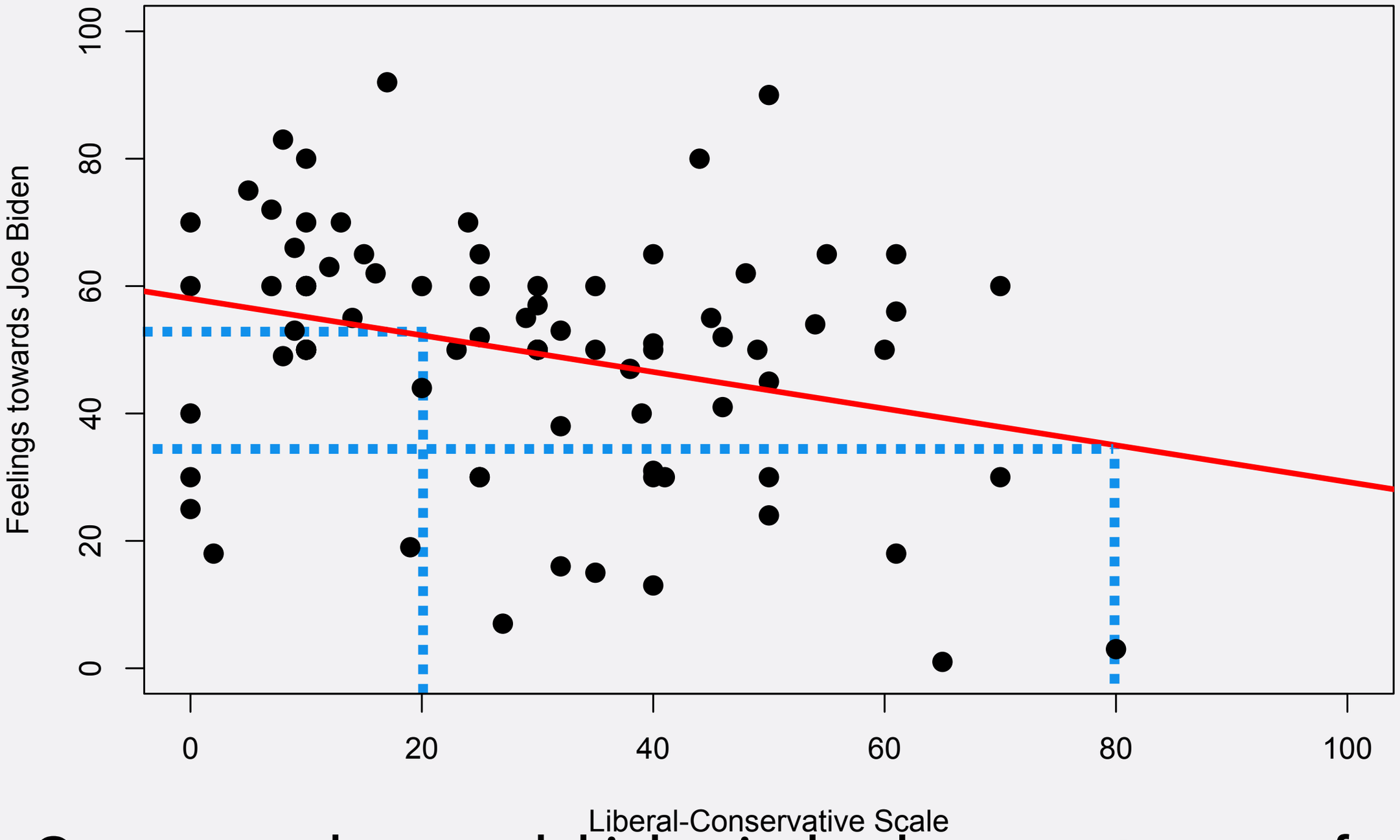
- **On average, how much higher is the thermometer score for someone who is a 20 on the liberal-conservative scale, compared to someone who is a 80?**

LINE



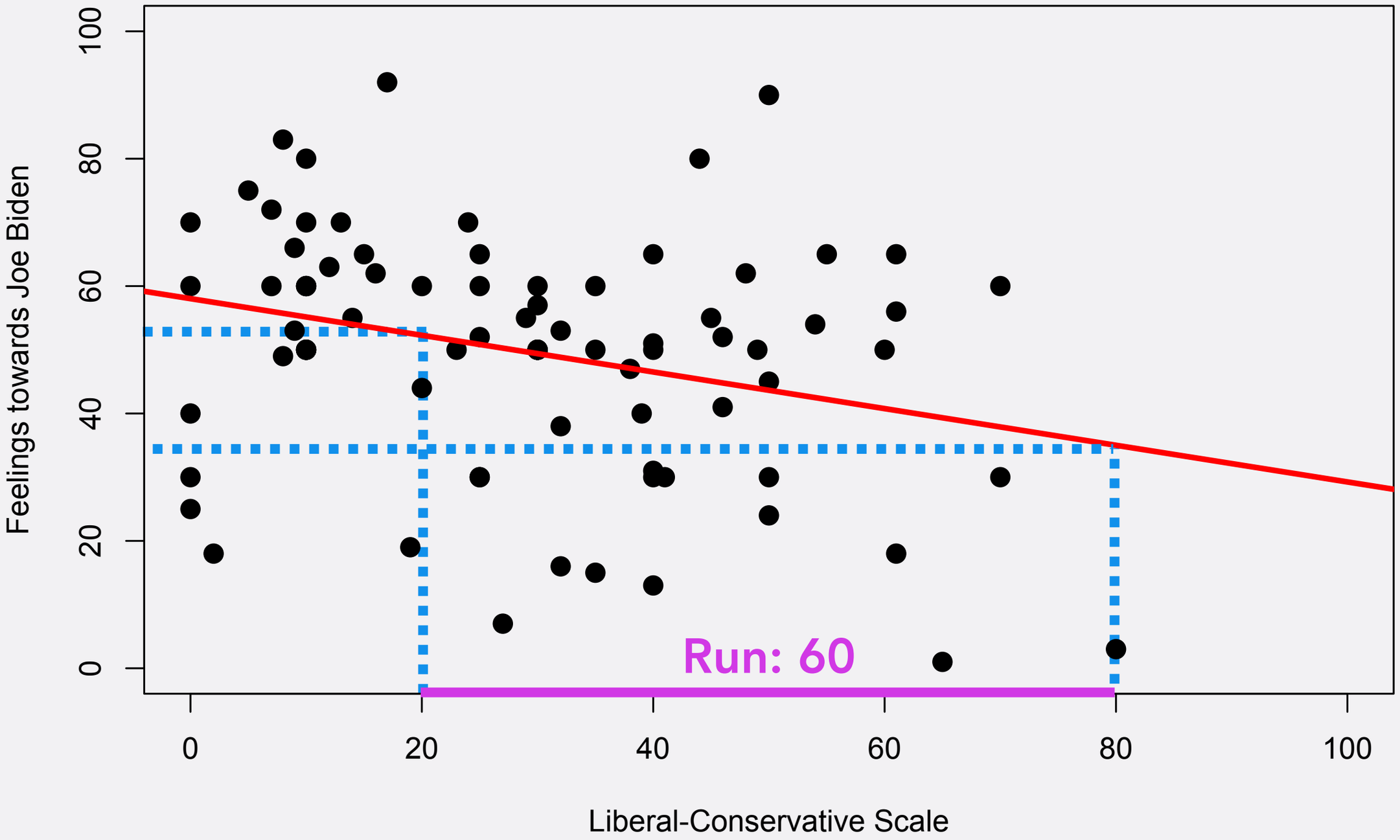
- **On average, how much higher is the thermometer score for someone who is a 20 on the liberal-conservative scale, compared to someone who is a 80?**

LINE

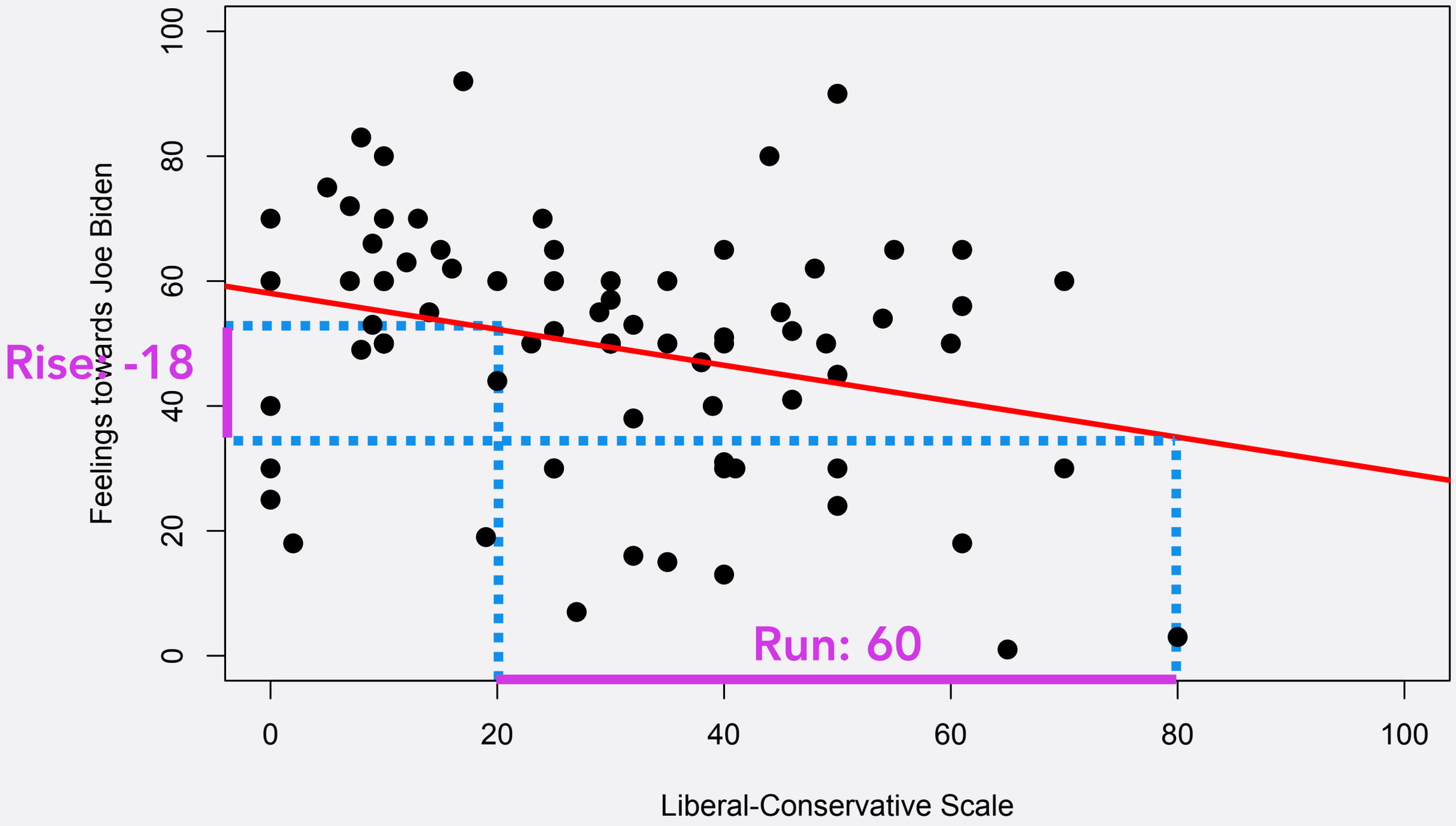


- **On average, how much higher is the thermometer score for someone who is a 20 on the liberal-conservative scale, compared to someone who is a 80?**

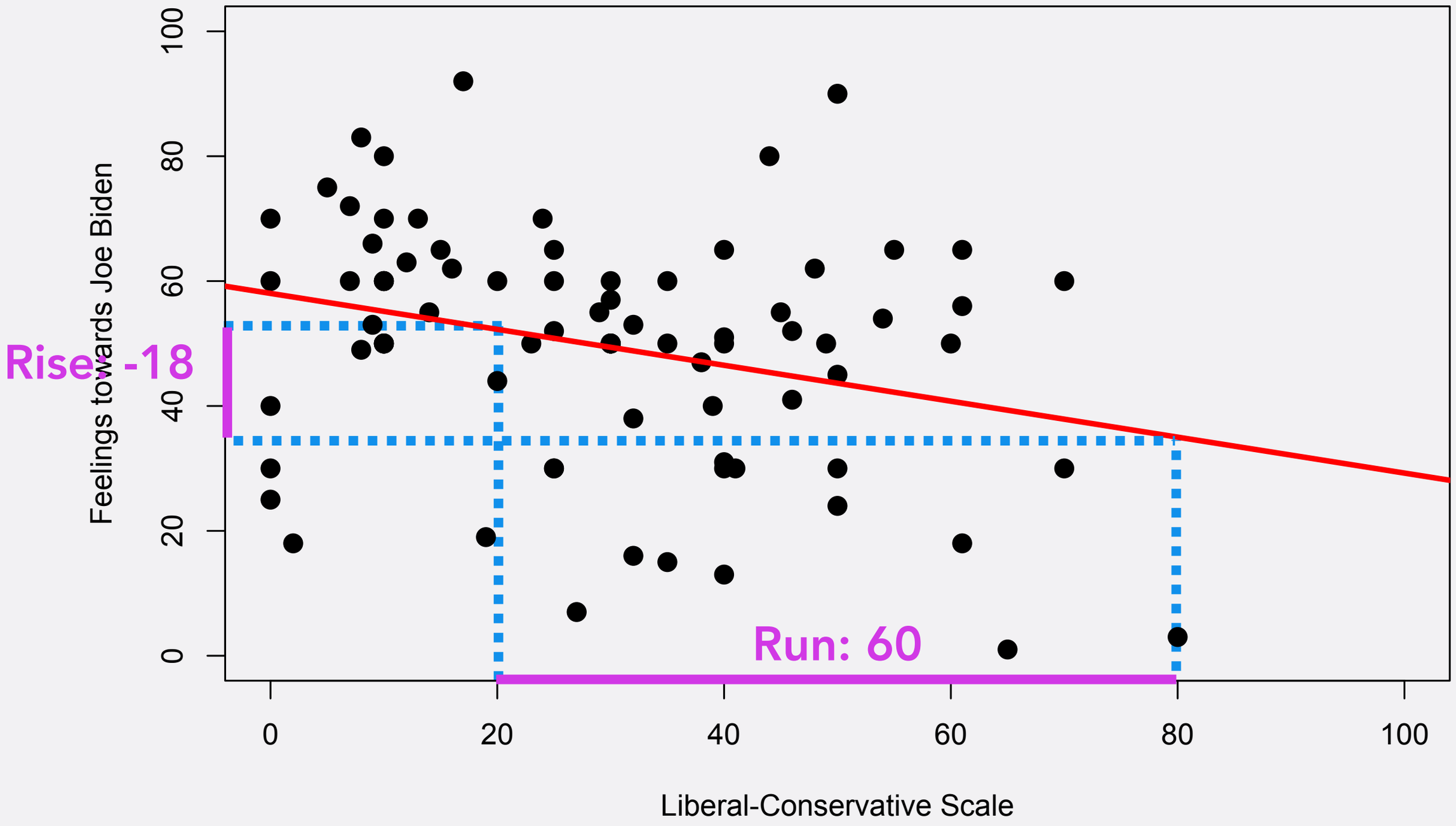
LINE


Feelings towards Joe Biden
0
20
40
60
80
100
Liberal-Conservative Scale
Run: 60

LINE


Rise: -18
Run: 60
Feelings towards Joe Biden
Liberal-Conservative Scale
0
20
40
60
80
100

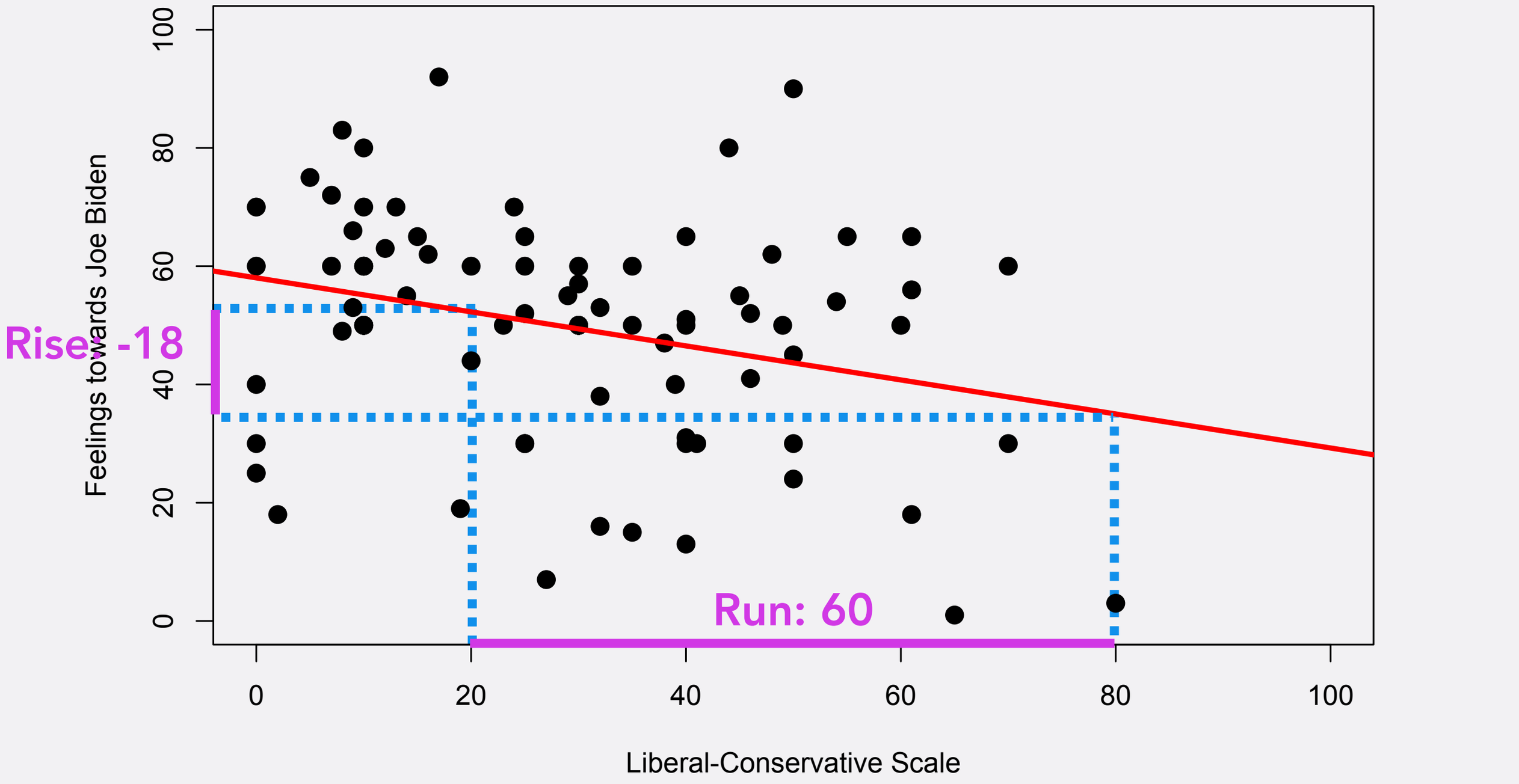
LINE



Slope=Rise over run=-18/60=-0.3

LINE

Slope=Rise over run



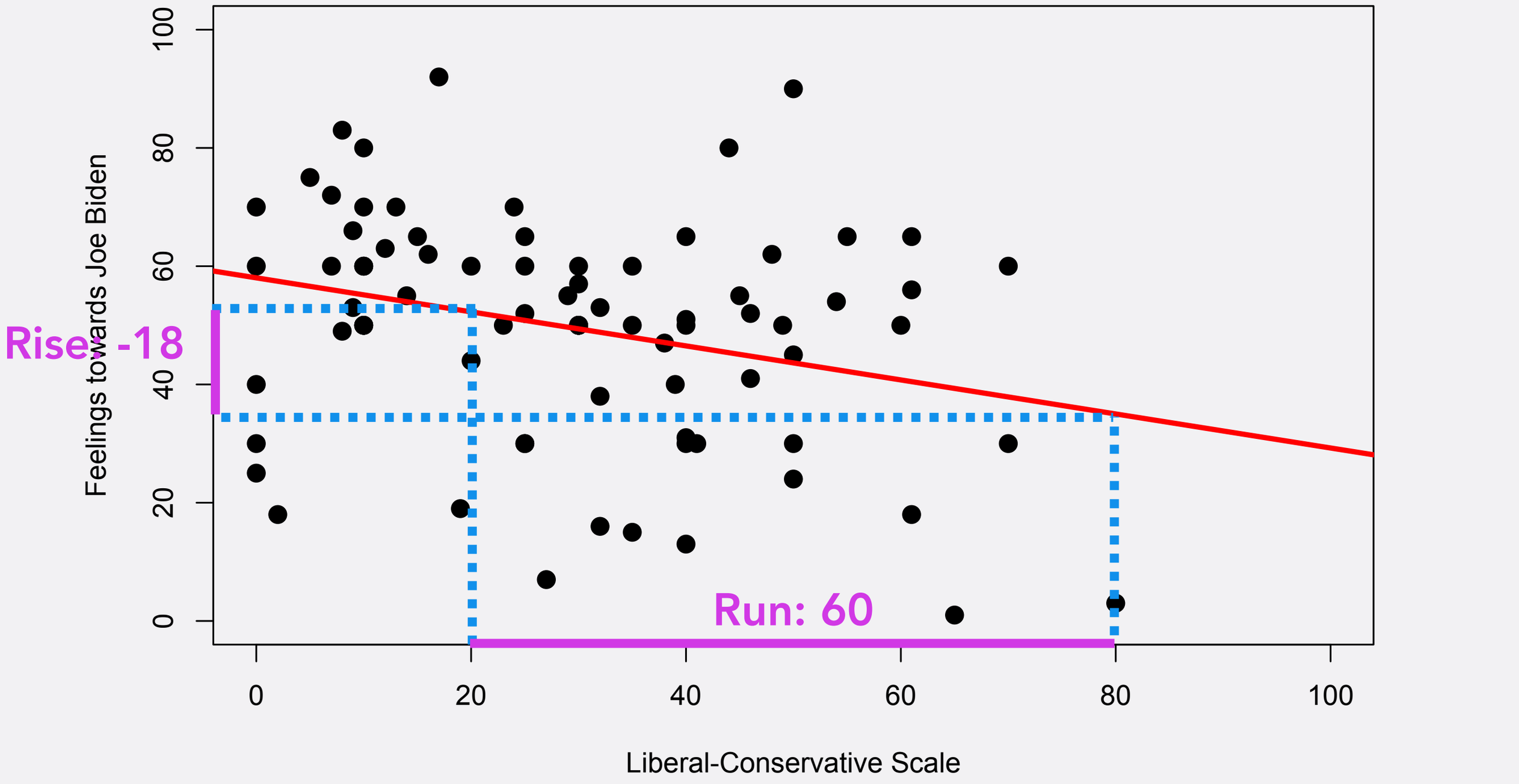
- **For each one unit increase on the liberal-conservative scale, feelings towards J. Biden go down by 0.3 points**

NOTE

- In this case, it happens to be that slope is close to correlation
- This does not need to be the case

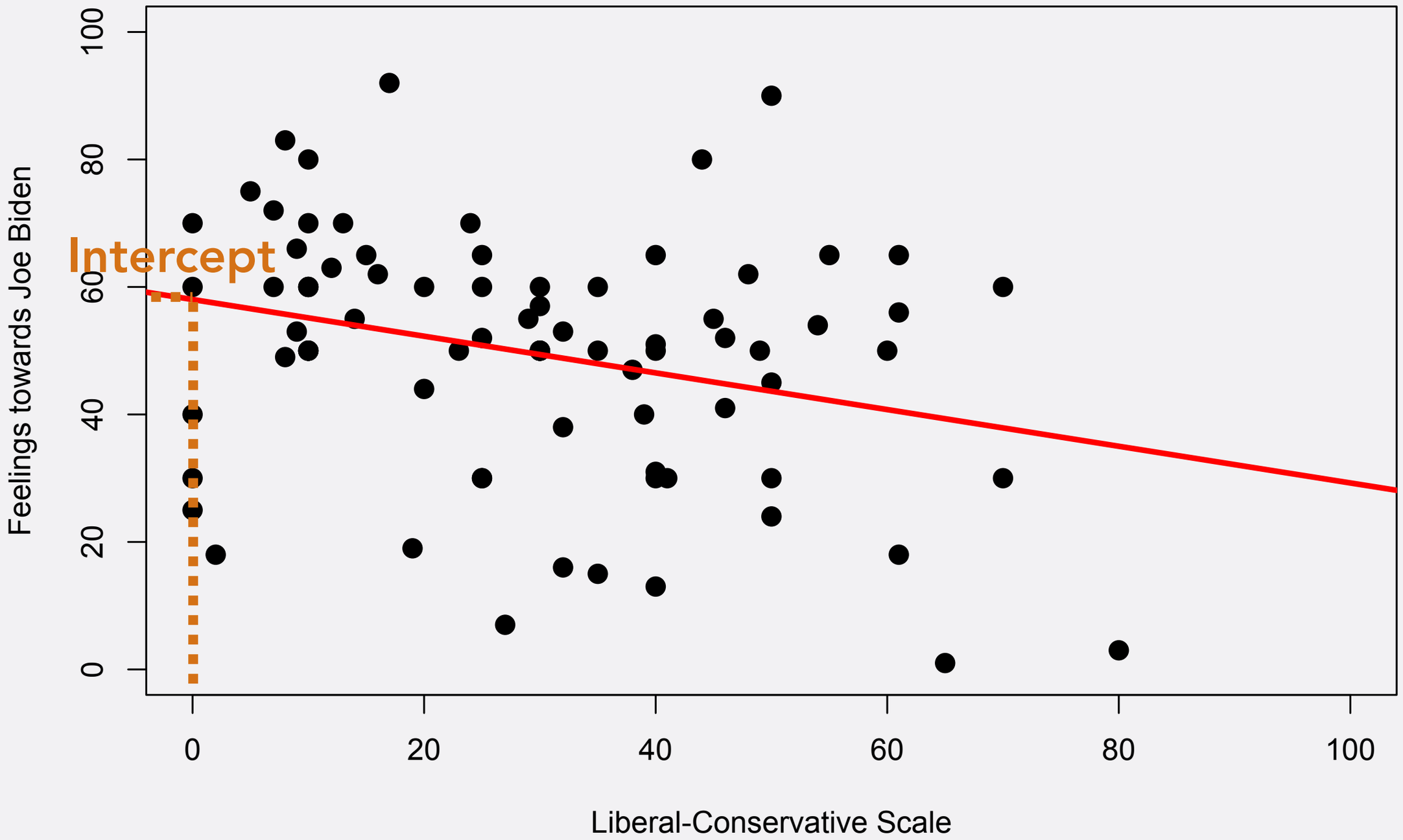
LINE

Slope=Rise over run

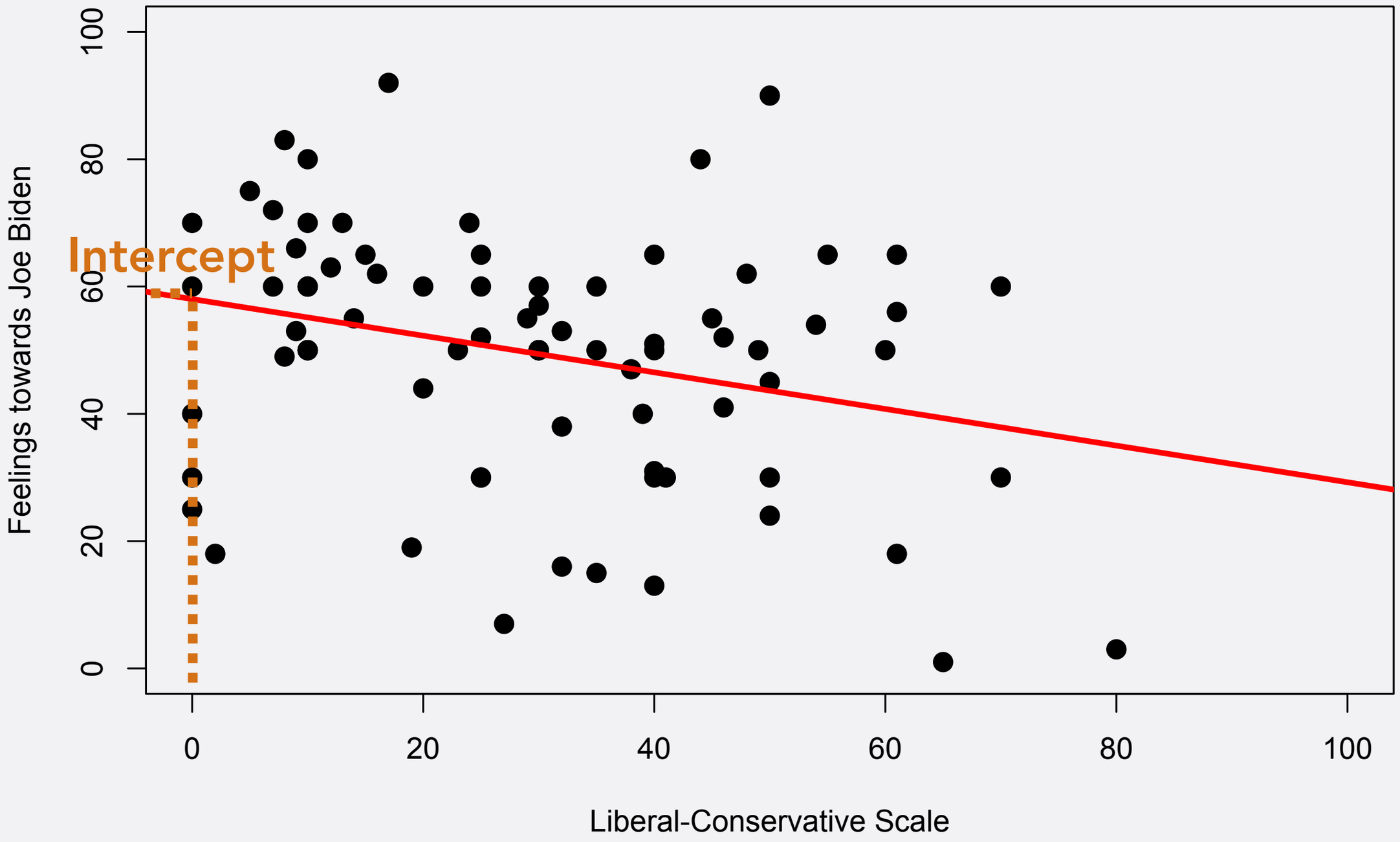


- **For each one unit increase on the liberal-conservative scale, feelings towards J. Biden go down by 0.3 points**

LINE


Intercept
Feelings towards Joe Biden
Liberal-Conservative Scale
0
20
40
60
80
100

LINE



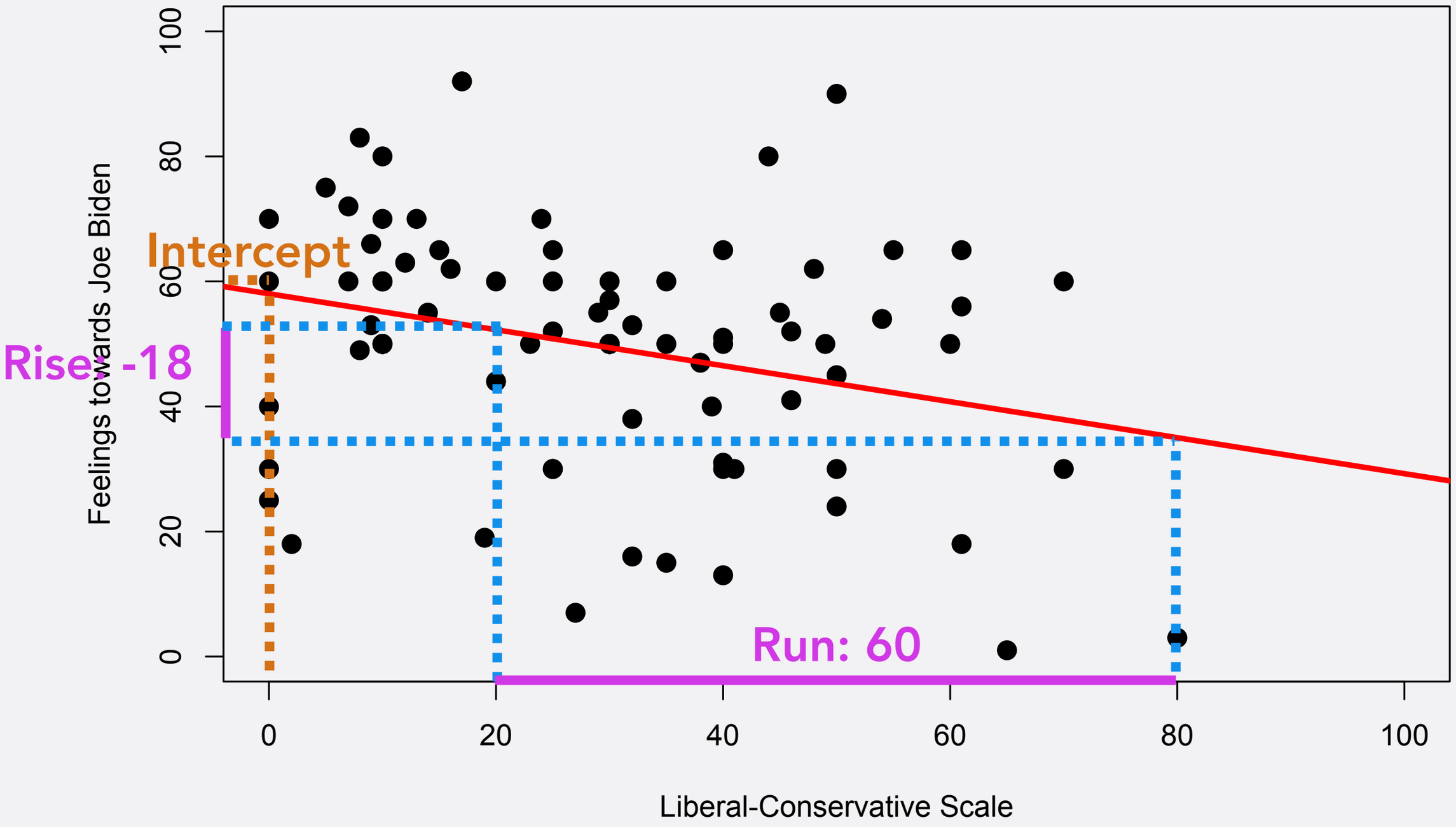
- **Students who are very liberal (score=0) are expected to have a feeling thermometer score of (about) 60.**

LINEAR REGRESSION

- Linear regression: Equation that tells us *direction* and *size* of relationship between independent variable (IV) and dependent variable (DV)
- DV = Intercept + Slope * IV

LINE

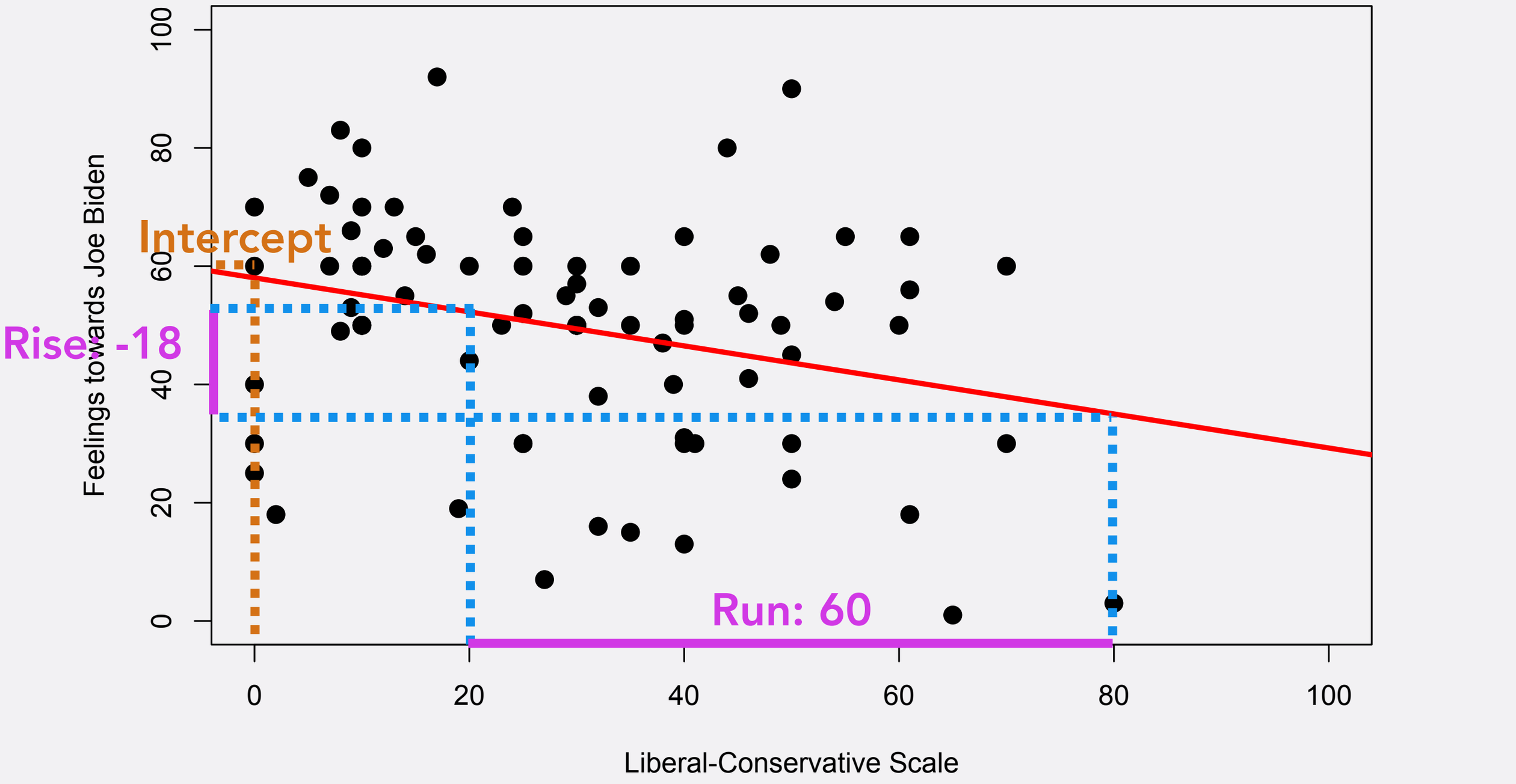
Slope=Rise over run



- Thermometer Score = Intercept + Slope * Lib/Cons

LINE

Slope=Rise over run



- Thermometer Score = 60 - 0.3 * Lib/Cons

WHAT THIS TELLS US

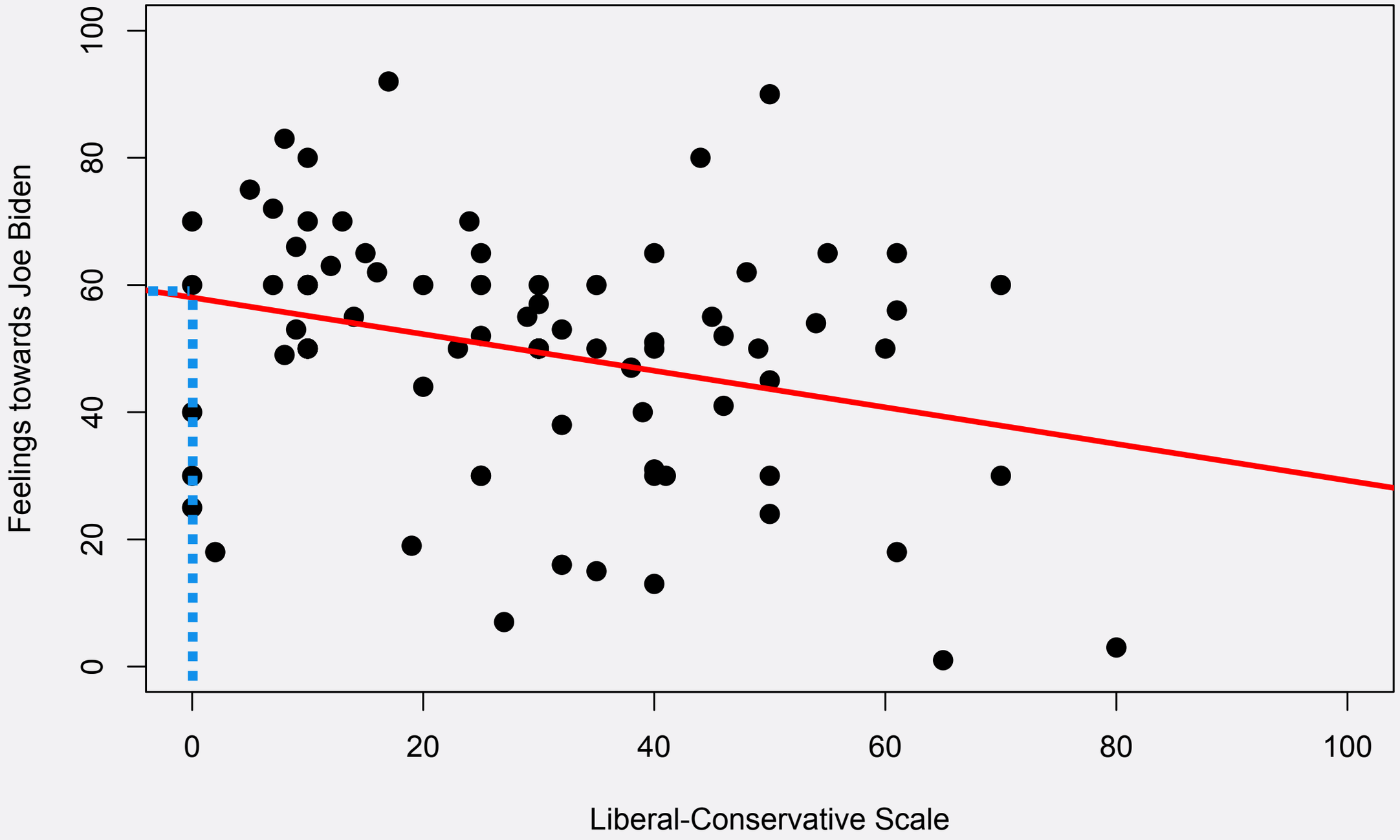
- Thermometer Score = 60 - 0.3 * Lib/Cons
- Can predict what someone's thermometer rating of Joe Biden will be, depending on where they are on liberal-conservative scale

WHAT THIS TELLS US

- Thermometer Score = 60 - 0.3 * Lib/Cons

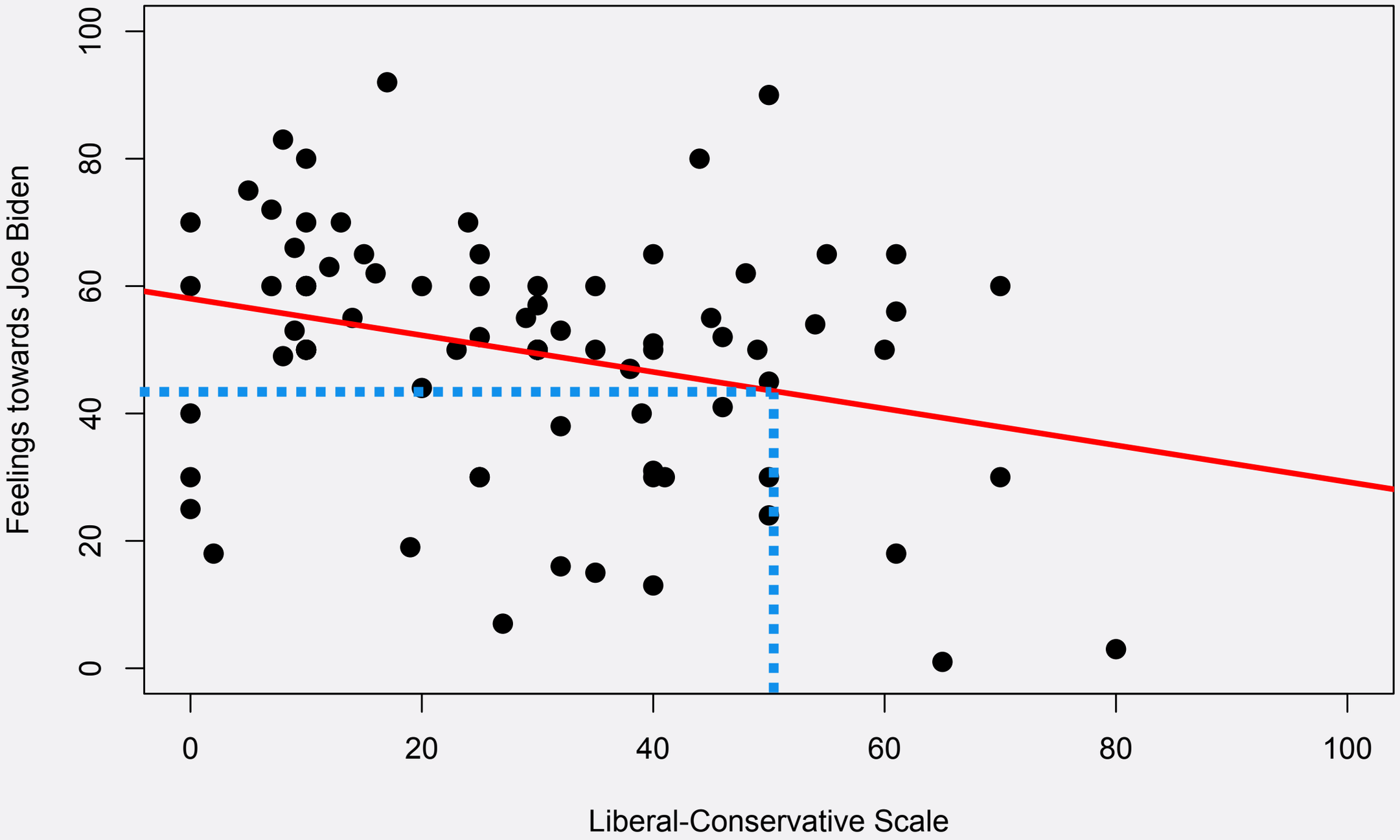
- Lib/Cons scale of 0:
  - 60 - 0.3 * 0 = 60

LINE



- 60 - 0.3 * 0 = 60

LINE

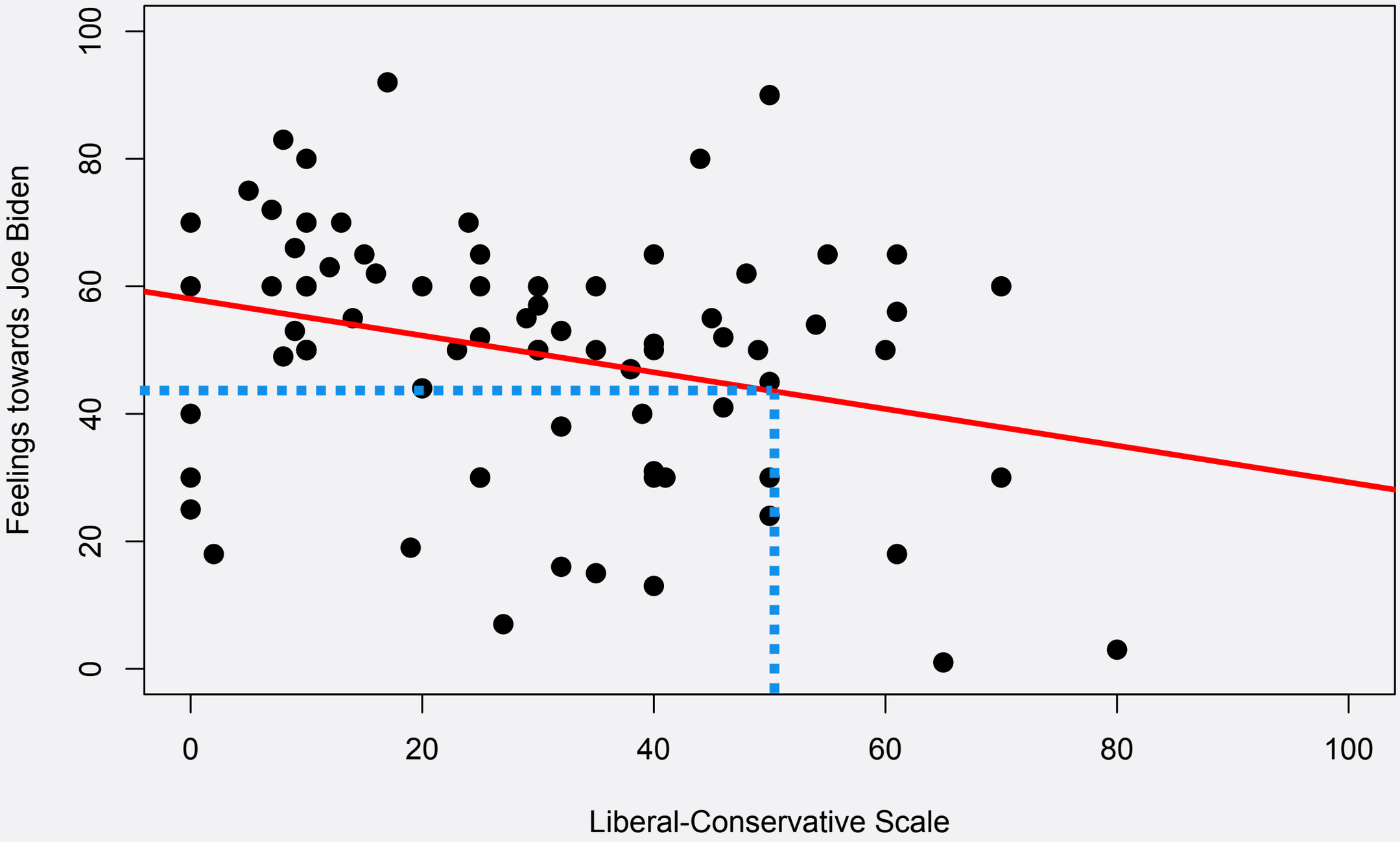

Feelings towards Joe Biden
0
20
40
60
80
100
0
20
40
60
80
100
Liberal-Conservative Scale

WHAT THIS TELLS US

- Thermometer Score = 60 - 0.3 * Lib/Cons

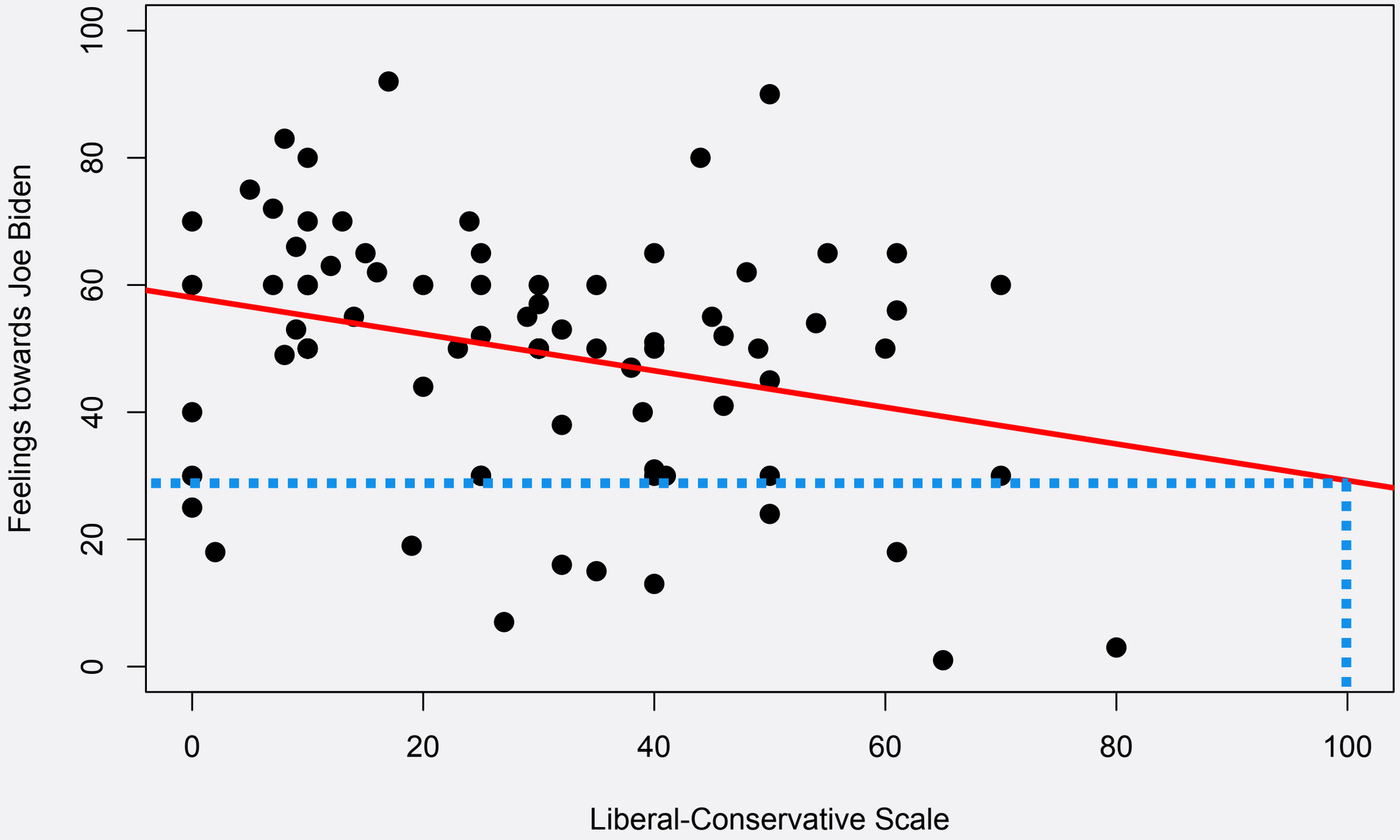
- Lib/Cons scale of 50:
  - 60 - 0.3 * 50 = 45

LINE



- 60 - 0.3 * 50 = 45

LINE



- 60 - 0.3 * 100 = 30

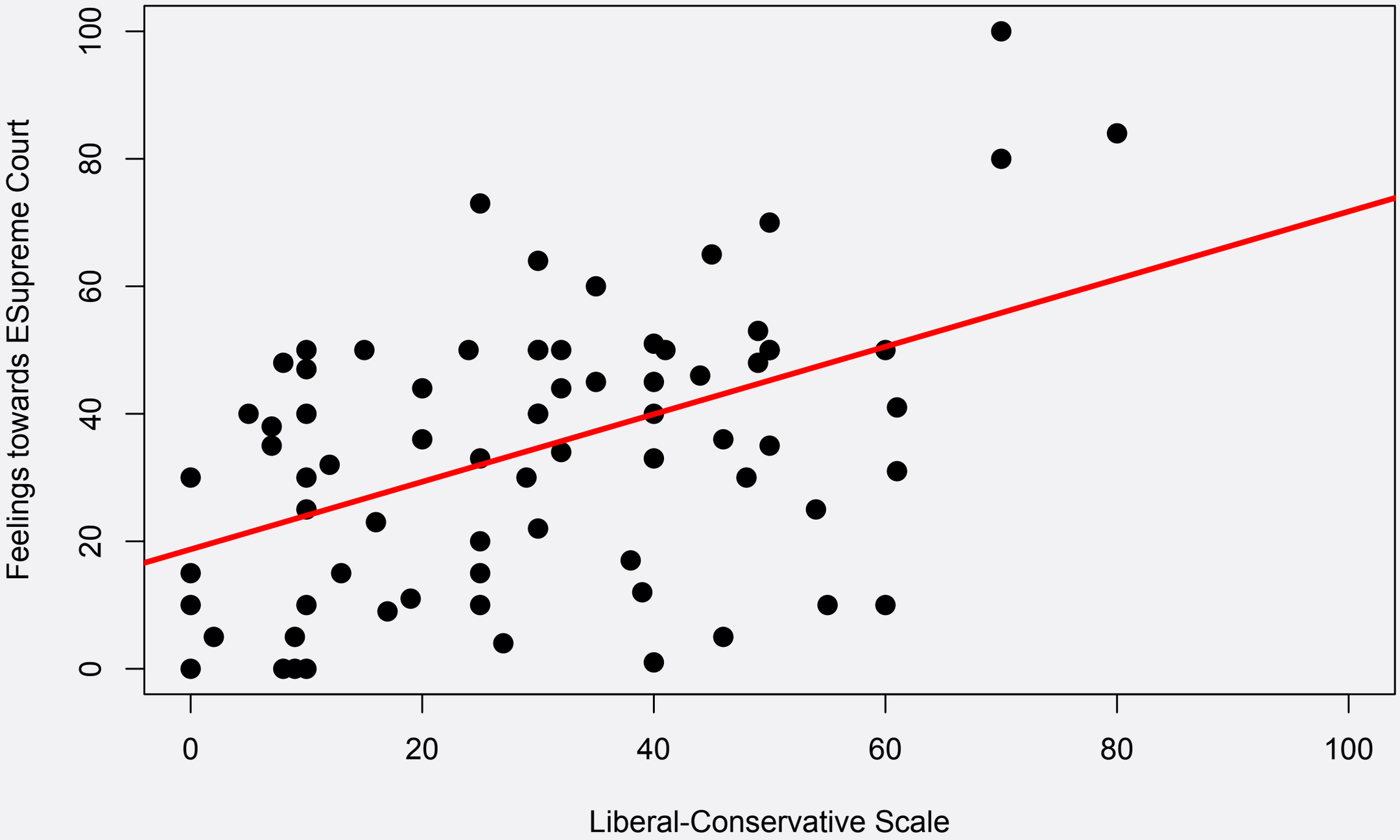
BIVARIATE RELATIONSHIPS

Dependent Variable	Independent Variable: Nominal/Ordinal	Independent Variable: Interval
Nominal/Ordinal	Cross-Tabulation	Not In This Class…
Interval	Mean Comparison	Correlation Coefficient, Linear Regression

LINEAR REGRESSION

- A tool that tells us the direction and *size* of the effect of an independent variable on a dependent variable
- both are interval-level

SUPREME COURT



- Thermometer Score = 19 + 0.53 * Lib/Cons

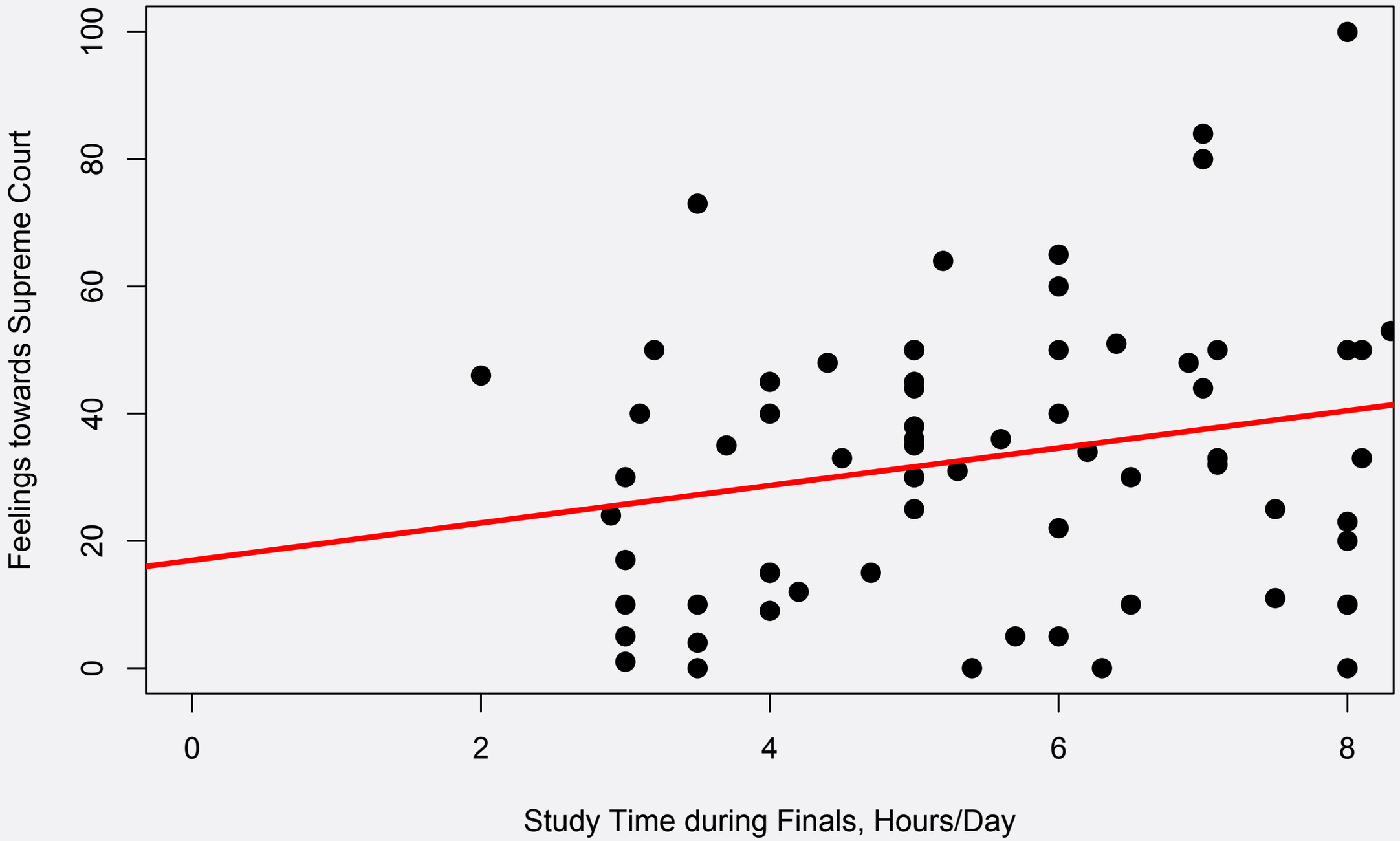
INTERPRETATION?

- Thermometer Score = 19 + 0.53 * Lib/Cons
- What does the 19 tell us?
- What does the 0.53 tell us?

INTERPRETATION?

- Thermometer Score = 19 + 0.53 * Lib/Cons
  - What does the 19 tell us?
    - A student who is 0 on the lib/cons scale has an expected thermometer score of 19
  - What does the 0.53 tell us?
    - For every one point increase in the lib/cons scale, the thermometer score is expected to increase by 0.53 points

DIFFERENT INDEPENDENT VARIABLE



- Thermometer Score = 17 + 2.9 * Hours/Day

INTERPRETATION?

- Thermometer Score = 17 + 2.9 * Hours/Day
- What does the 17 tell us?
- What does the 2.9 tell us?

INTERPRETATION?

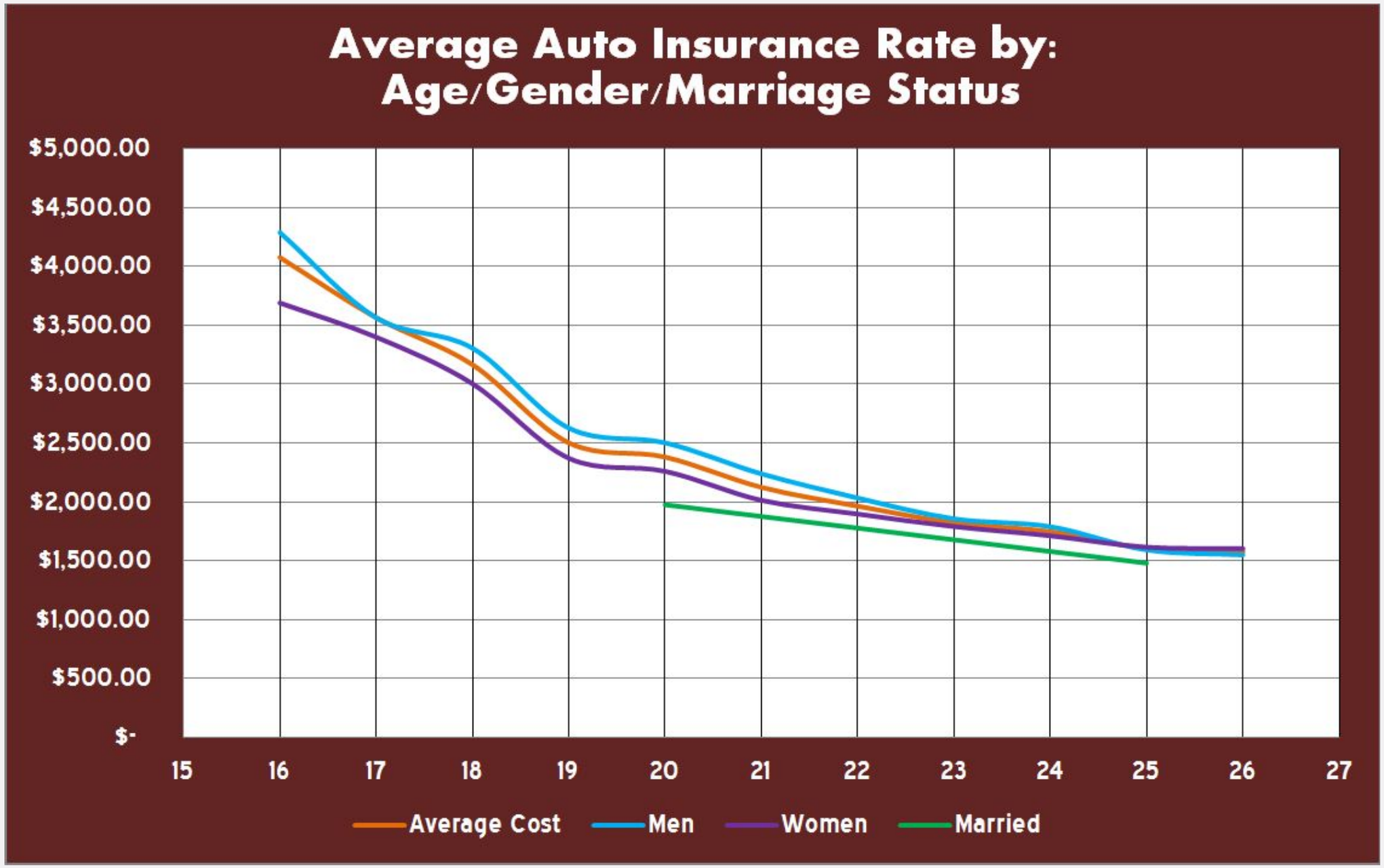
- Thermometer Score = 17 + 2.9 * Hours/Day
  - What does the 17 tell us?
    - A student who studies 0 hours per day has an expected thermometer score of 17
  - What does the 2.9 tell us?
    - For every one hour a student studies longer per day, their thermometer score is expected to increase by 2.9 points

TODAY

- How do I pick the line?
- How is linear regression useful?

HOW IS THIS USEFUL?

- Linear regression widely used in private sector



HOW IS THIS USEFUL?

- Insurance company has to decide how much to charge you
- How much to charge you depends on how much in damages they expect to have to pay for you

HOW IS THIS USEFUL?

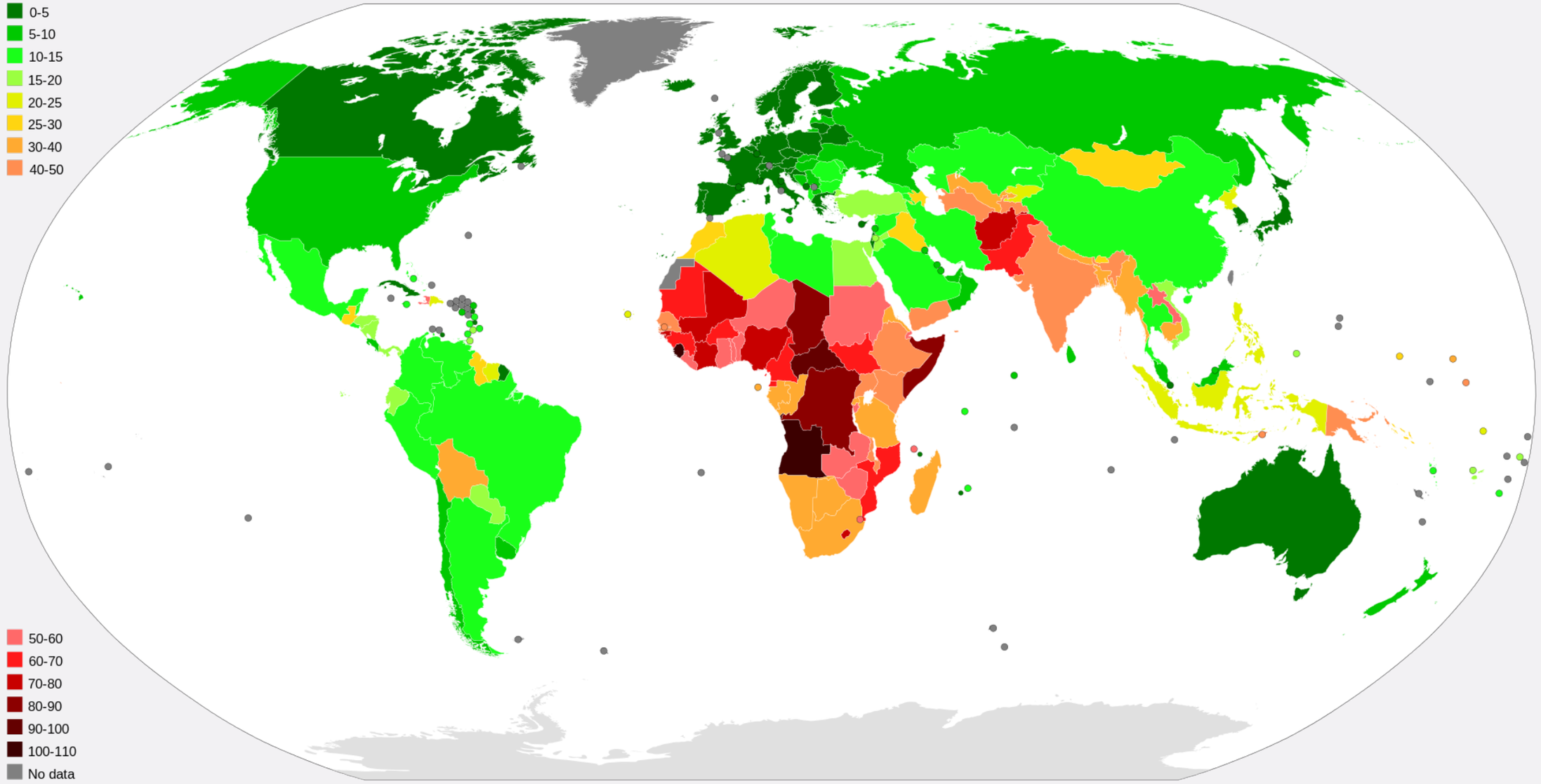
- Guessing won’t do
  - If they overestimate how much damage someone will cause, they charge too much (and the person might buy insurance elsewhere)
  - If they underestimate, they charge too little (and lose money)

HOW IS THIS USEFUL?

- They use linear regression
- Have data on how much damage other customers have caused
- Regression analysis of damages caused (Y), depending on age (X)
- Based on your age, predict how much damage you will cause
  - Damages = a + b*age
- That determines your rate

HOW IS THIS USEFUL?

- Linear regression also widely used in public policy research



- Infant mortality rates (Death under 1 year of age per 1,000 live births)

HOW IS THIS USEFUL?

- Some of these rates are appalling
  - Mali: Out of 1,000 babies born alive, 100 die before their first birthday
- If we want to lower infant mortality rates, we need to know what causes them

HOW IS THIS USEFUL?

- Infant mortality rate = 39.9 - 0.0008889*GDP per capita

HOW IS THIS USEFUL?

- Infant mortality rate = 39.9 - 0.0008889*GDP per capita
- For each dollar that GDPpc is higher, infant mortality expected to decrease by 0.0008889
- If GDPpc=0, infant mortality is expected to be 39.9

HOW IS THIS USEFUL?

- Infant mortality rate = 39.9 - 0.0008889*GDP per capita
- GDP per capita of the U.S. is $41,627
  - Expected rate: 39.9 - 0.0008889*41,627=2.90
- GDP per capita of Mexico is $11,877
  - Expected rate: 39.9 - 0.0008889*11,877=29.34

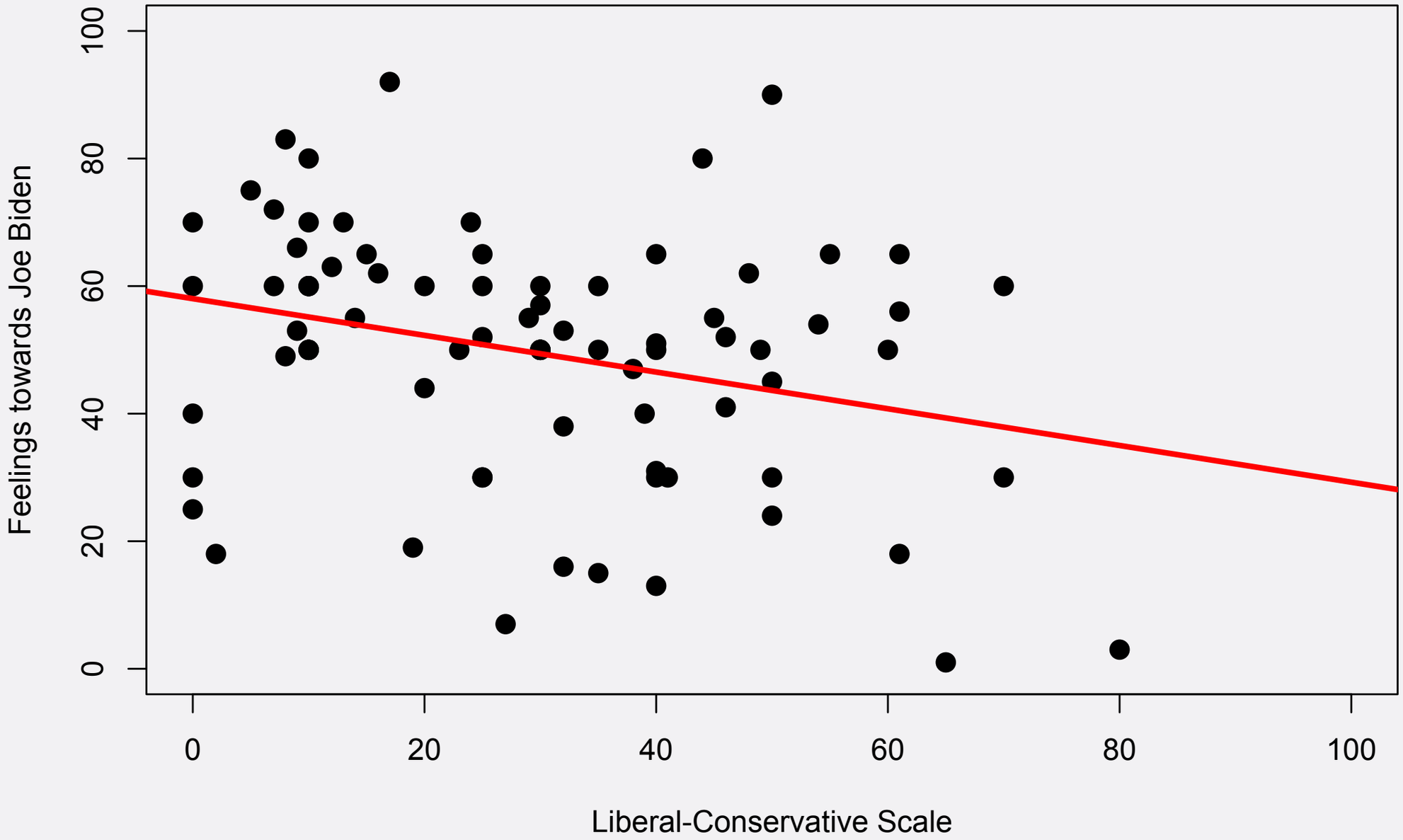
BIVARIATE RELATIONSHIPS

Dependent Variable	Independent Variable: Nominal/Ordinal	Independent Variable: Interval
Nominal/Ordinal	Cross-Tabulation	Not In This Class…
Interval	Mean Comparison	Correlation Coefficient, Linear Regression

WHAT WE CAN DO

- We can now estimate how much an independent variable X affects a dependent variable Y

NEXT TIME



- **Is the effect of lib/cons on ratings of J. Biden real?**
- **Or is it only something that we found in our sample, but lib/cons actually has no effect in the population?**