PSC 202
SYRACUSE UNIVERSITY

INTRODUCTION TO POLITICAL ANALYSIS

BIVARIATE HYPOTHESIS TESTING PART 2

EXAM #2

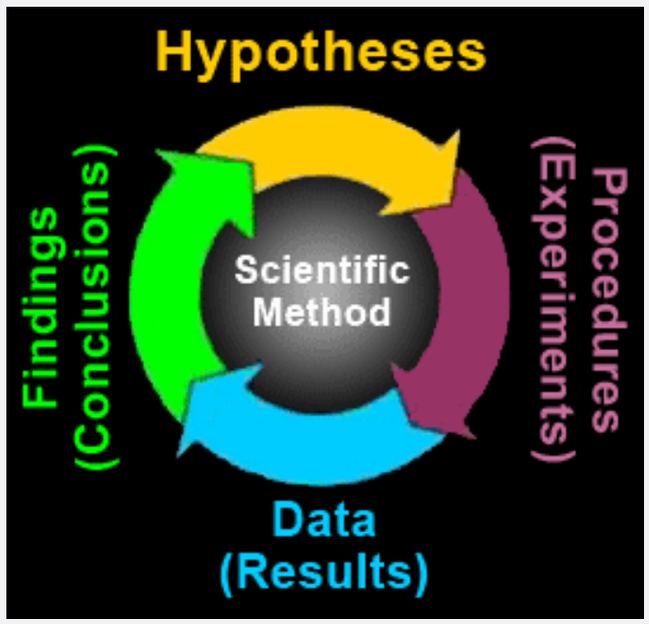
- Originally scheduled for March 27 (next Monday)
- Moved to April 3 (one week later)

PROBLEM SETS

- Problem set 6 due on Friday

WHERE WE ARE

- Formulate research question
- Propose explanation/theory, hypotheses
- Data collection process
- Use data to evaluate hypotheses
- Reassess explanation



HURDLES

- 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

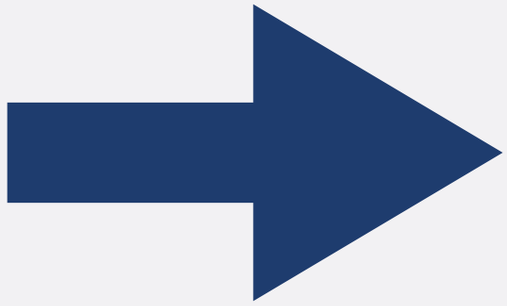
Dependent Variable	Independent Variable: Nominal/Ordinal	Independent Variable: Interval
Nominal/Ordinal	Cross-Tabulation	Not In This Class…
Interval	Mean Comparison	?

BIVARIATE RELATIONSHIPS

Dependent Variable	Independent Variable: Nominal/Ordinal	Independent Variable: Interval
Nominal/Ordinal	Cross-Tabulation	Not In This Class…
Interval	Mean Comparison	?

BIVARIATE RELATIONSHIP

Gender



Approval of
J. Biden

CROSS-TABULATIONS

Approve of Biden	Gender: Male	Gender: Female	Total
Approve	43.2% (19)	60.3% (38)	53.2% (57)
Do Not Approve	56.8% (25)	39.7% (25)	46.7% (50)
Total	100% (44)	100% (63)	100% (107)

COVARIATION

- Covariation between gender and approval: Proportion of women who approve is larger than proportion of men who approve

TERMINOLOGY

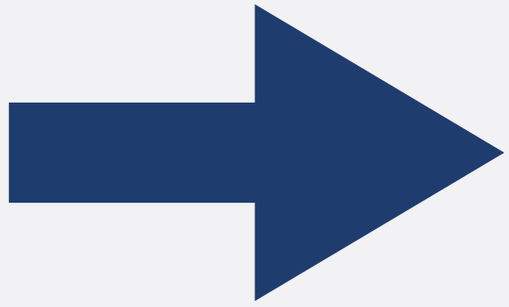
- Zero-order relationship: relationship between two variables, without controlling for any other factors
- Women are 17.1 percentage points more likely to approve of Biden than men (60.3% vs. 43.2%)

BIVARIATE RELATIONSHIPS

Dependent Variable	Independent Variable: Nominal/Ordinal	Independent Variable: Interval
Nominal/Ordinal	Cross-Tabulation	Not In This Class…
Interval	Mean Comparison	?

BIVARIATE RELATIONSHIP

Gender



Feelings towards Democratic Party

DEMOCRATIC PARTY

	Mean Thermometer Score	Frequency
Female	62.7	79
Male	43.9	50
Total	55.1	129

ZERO-ORDER RELATIONSHIP

	Mean Thermometer Score	Frequency
Female	62.7	79
Male	43.9	50
Total	55.1	129

18.8

BIVARIATE RELATIONSHIPS

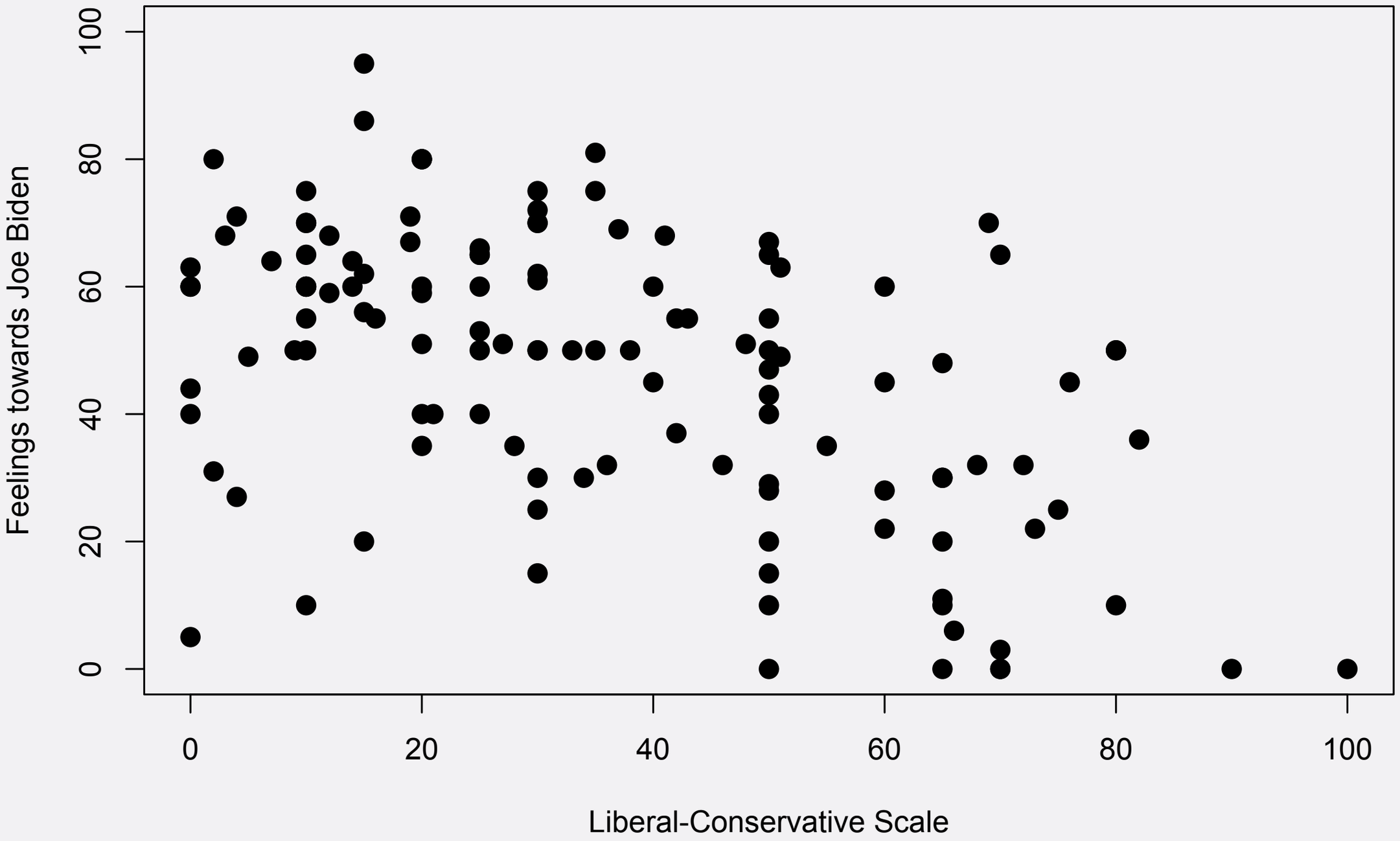
Dependent Variable	Independent Variable: Nominal/Ordinal	Independent Variable: Interval
Nominal/Ordinal	Cross-Tabulation	Not In This Class…
Interval	Mean Comparison	?

BIVARIATE RELATIONSHIP

Liberal-Conservative → Feelings towards Joe Biden

- DV: Feelings towards Joe Biden (0-100)
- IV: Liberal-Conservative (0-100)

JOE BIDEN

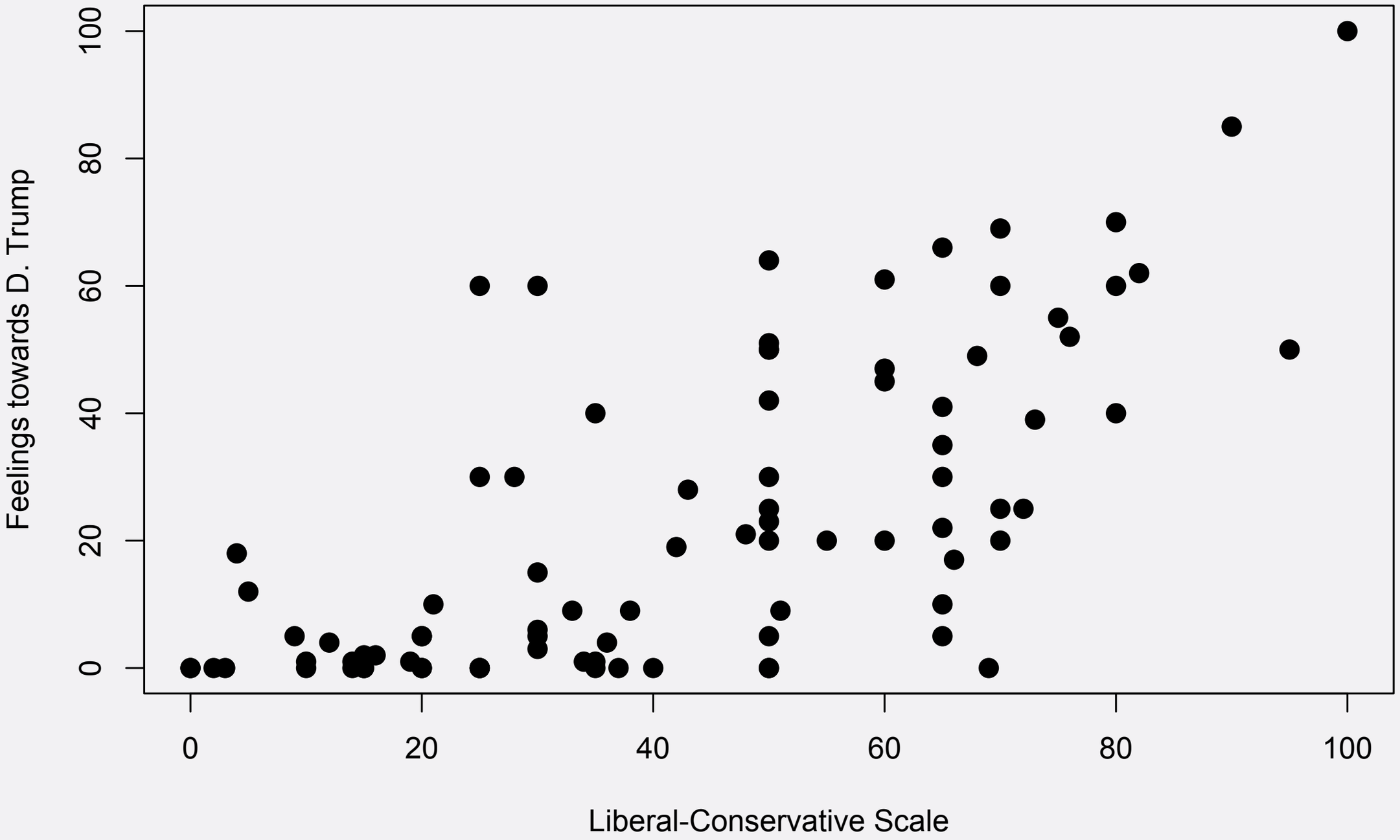

Feelings towards Joe Biden
0
20
40
60
80
100
0
20
40
60
80
100
Liberal-Conservative Scale

BIVARIATE RELATIONSHIP

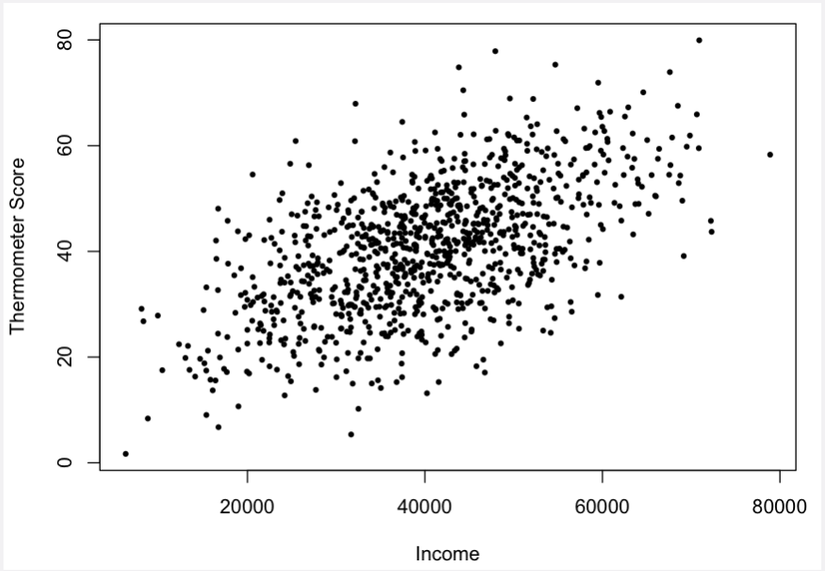
Liberal-Conservative → Feelings towards Donald Trump

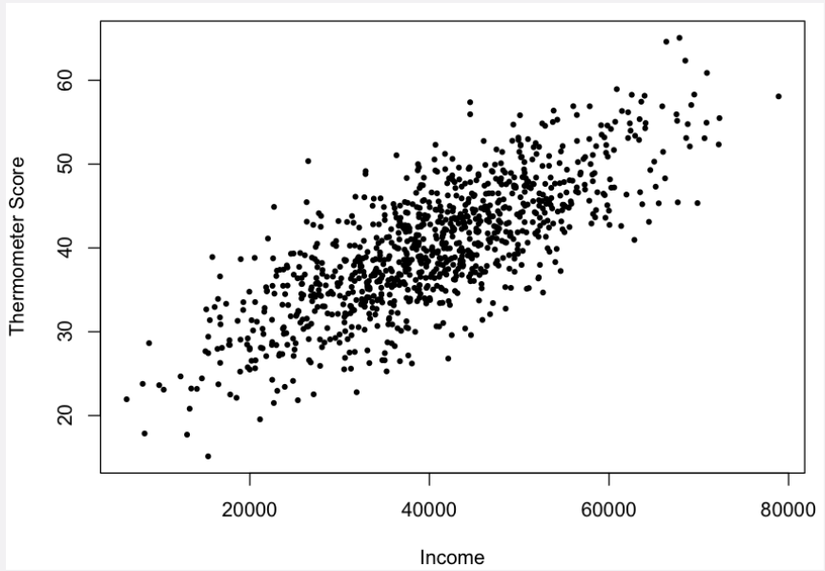
- DV: Feelings towards Donald Trump (0-100)
- IV: Liberal-Conservative (0-100)

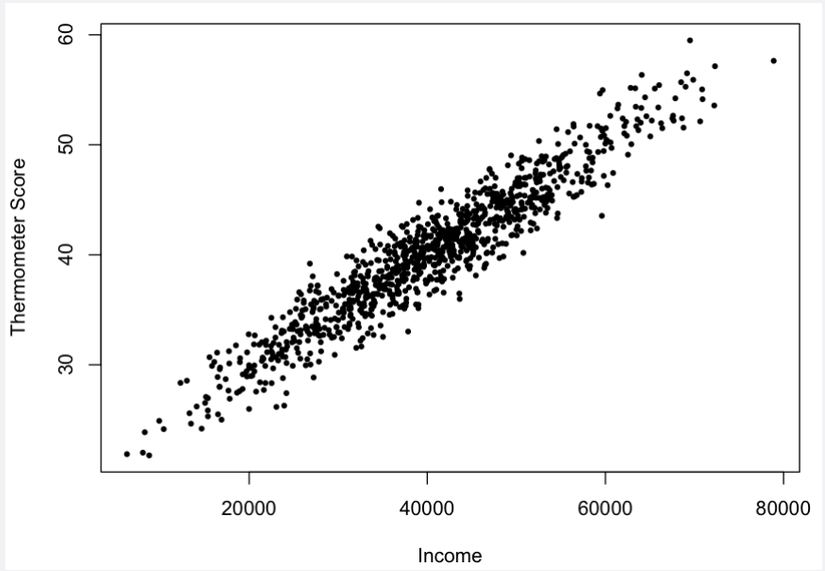
DONALD TRUMP

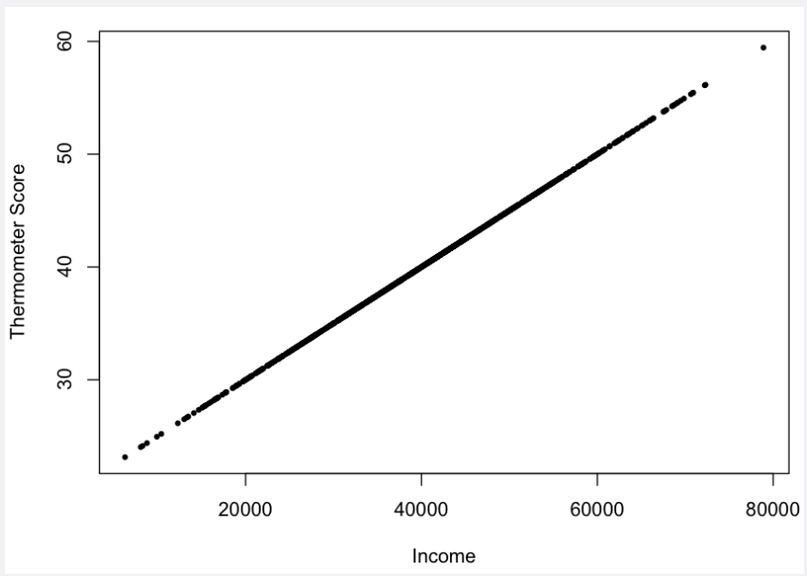


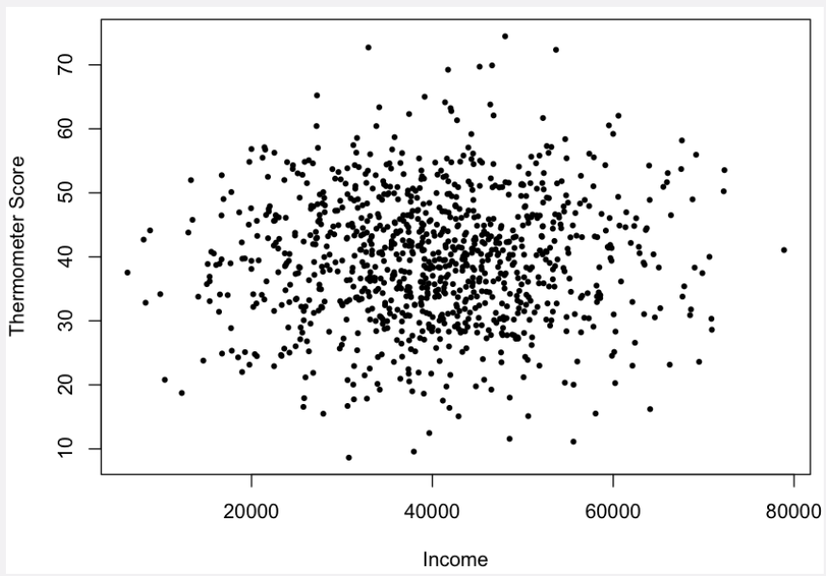
SCATTERPLOTS

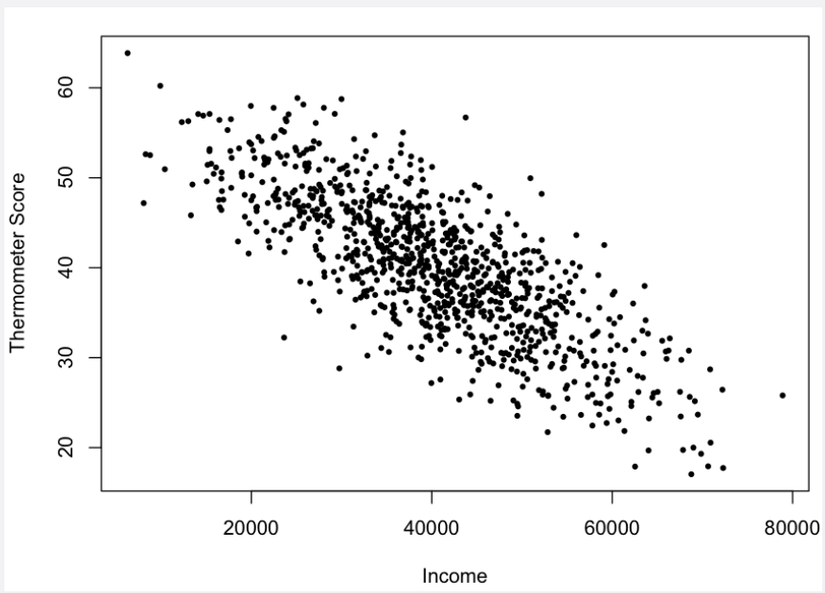












SCATTERPLOT

- Some relations are positive, some negative, others show no relation
- Some are more positive than others
- We want to make statements that are more precise than "Some are more positive than others"

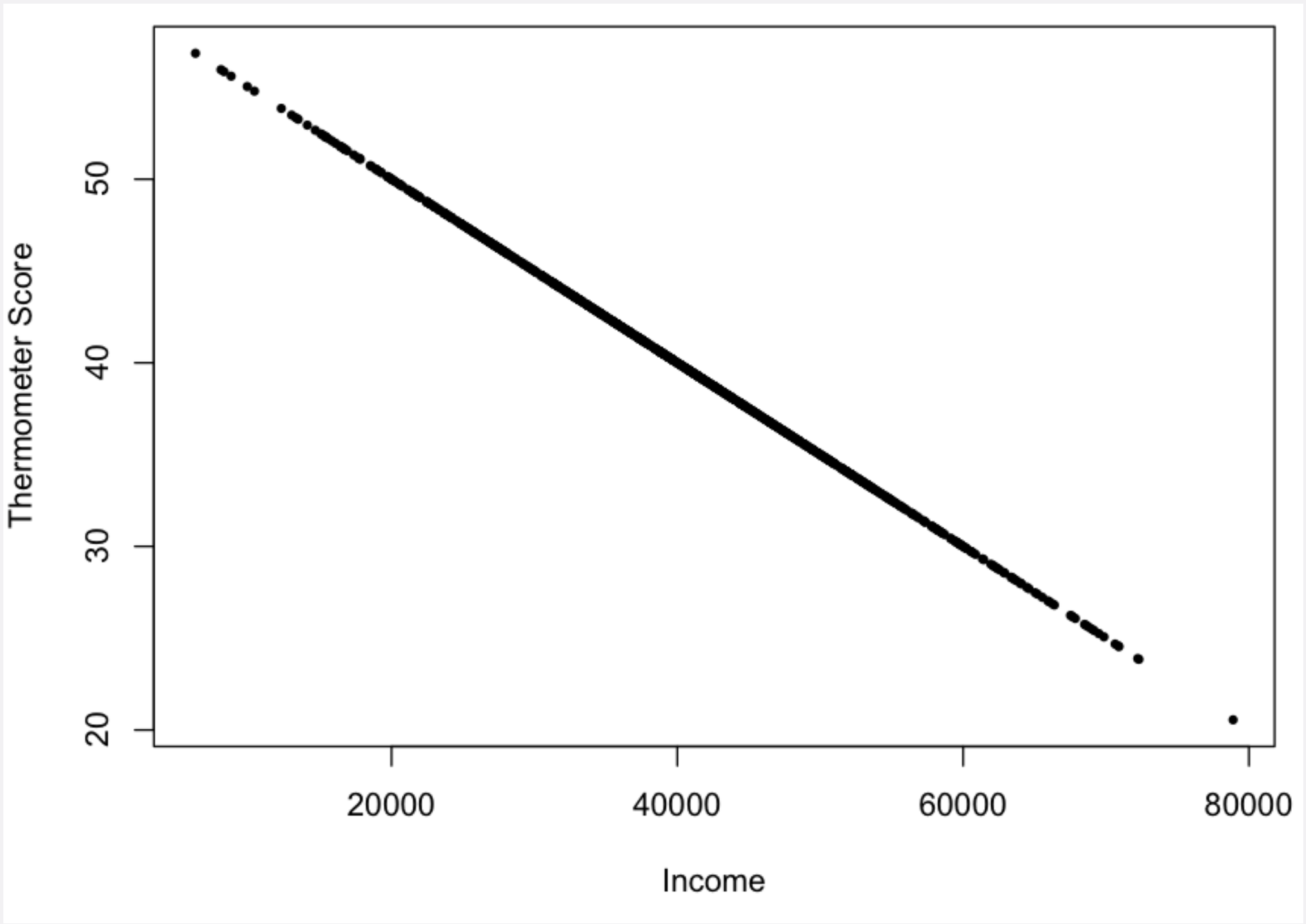
CORRELATION

- Systematic way to measure the *direction* and *linearity* of the relationship between two variables
- Pearson correlation *r*
- Ranges between -1 and +1

PEARSON'S R

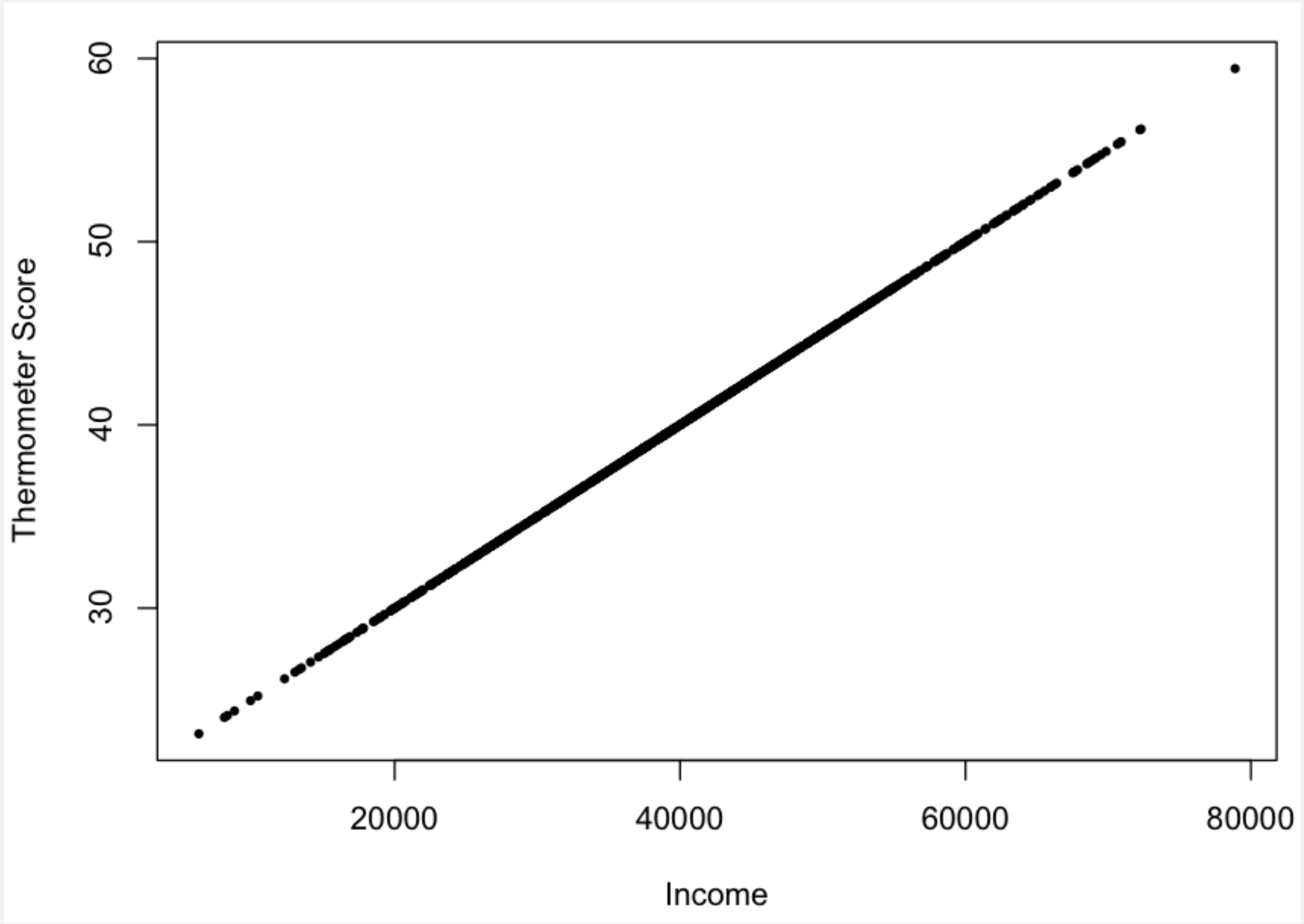
- If the coefficient is 0, no bivariate relationship exists
  - A positive coefficient means that a positive relationship exists
  - A negative coefficient means that a negative relationship exists
- Correlation of -1 or +1 means that relation between X and Y is perfectly linear

CORRELATION



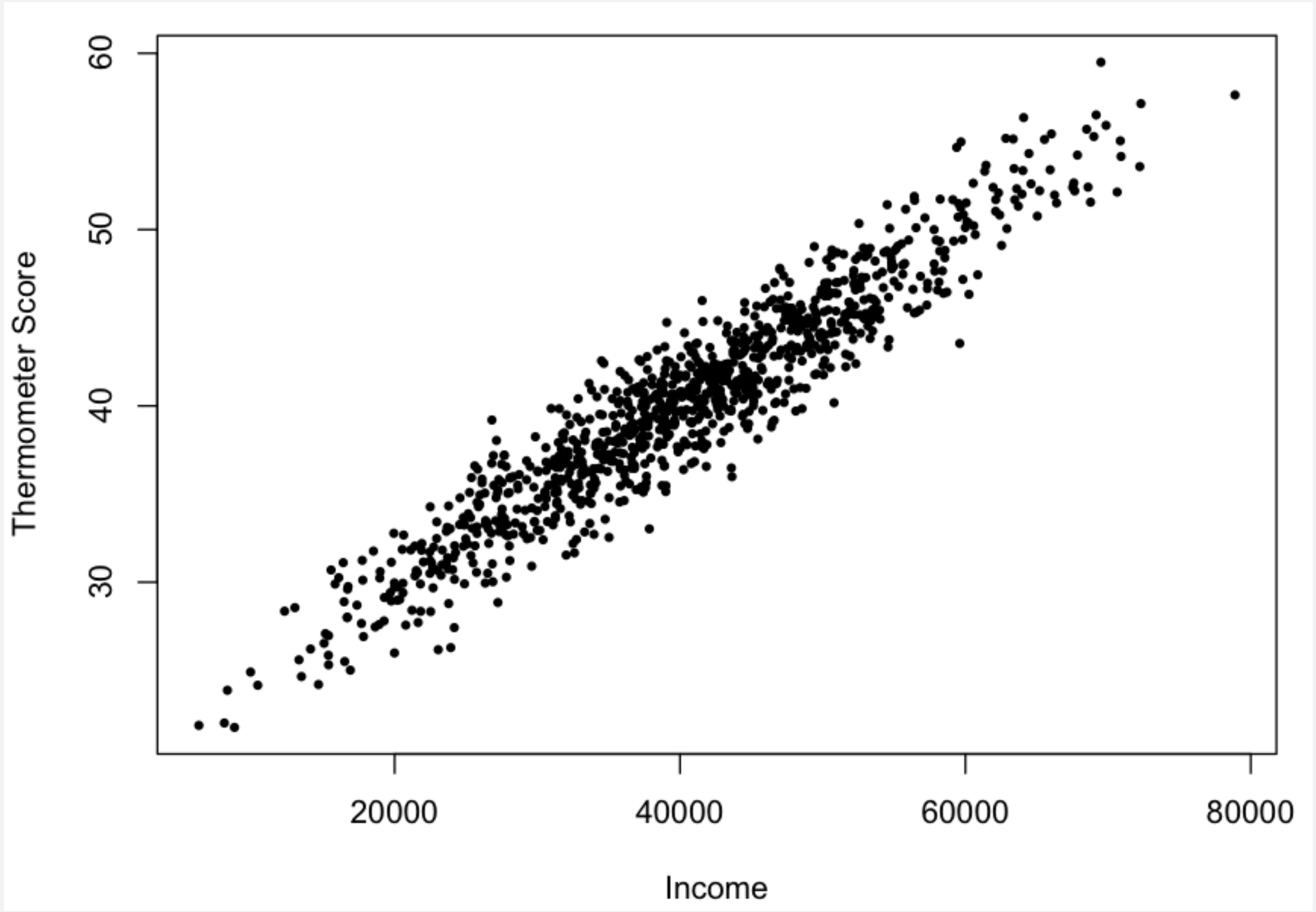
r=-1

CORRELATION



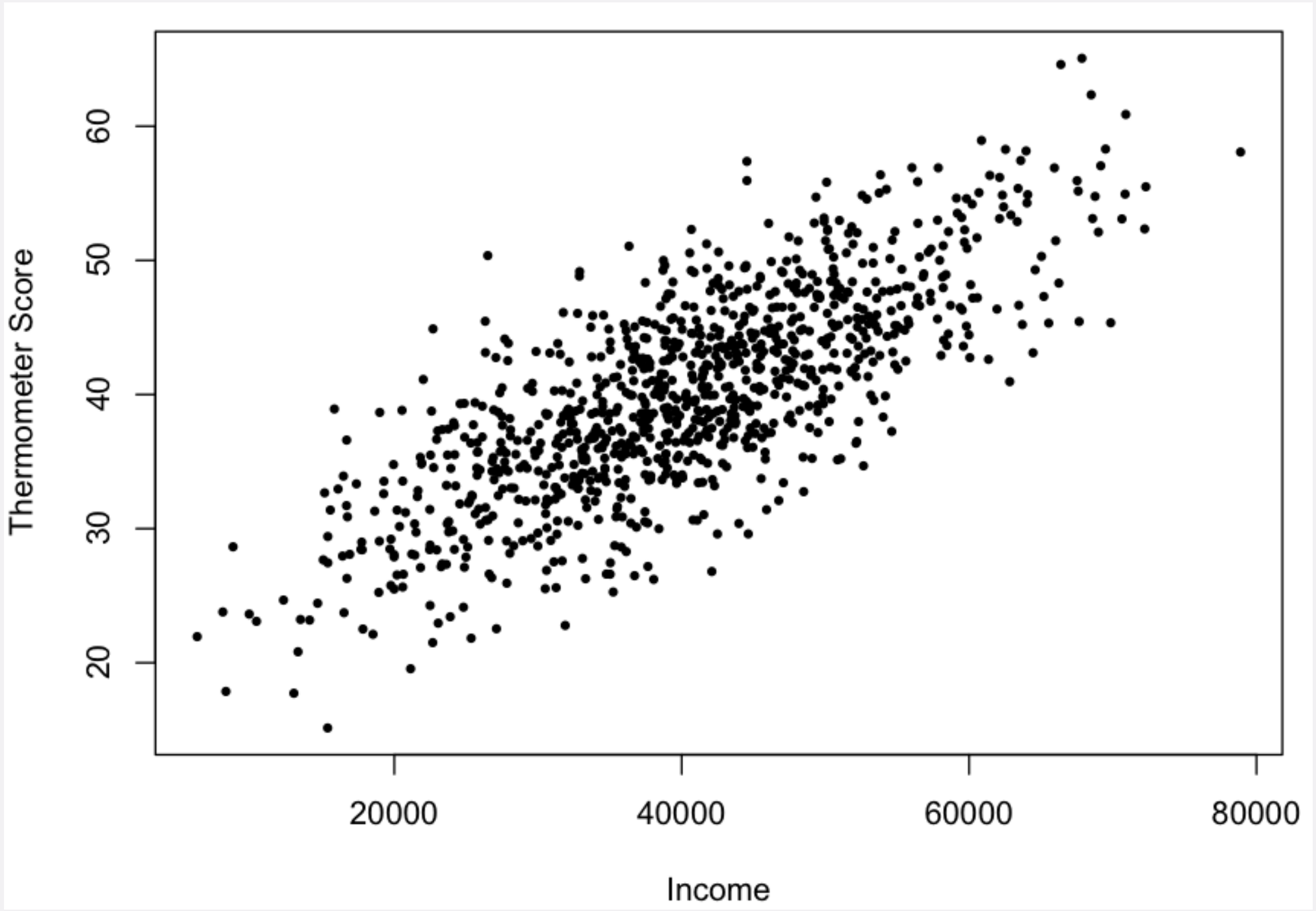
r=1

CORRELATION



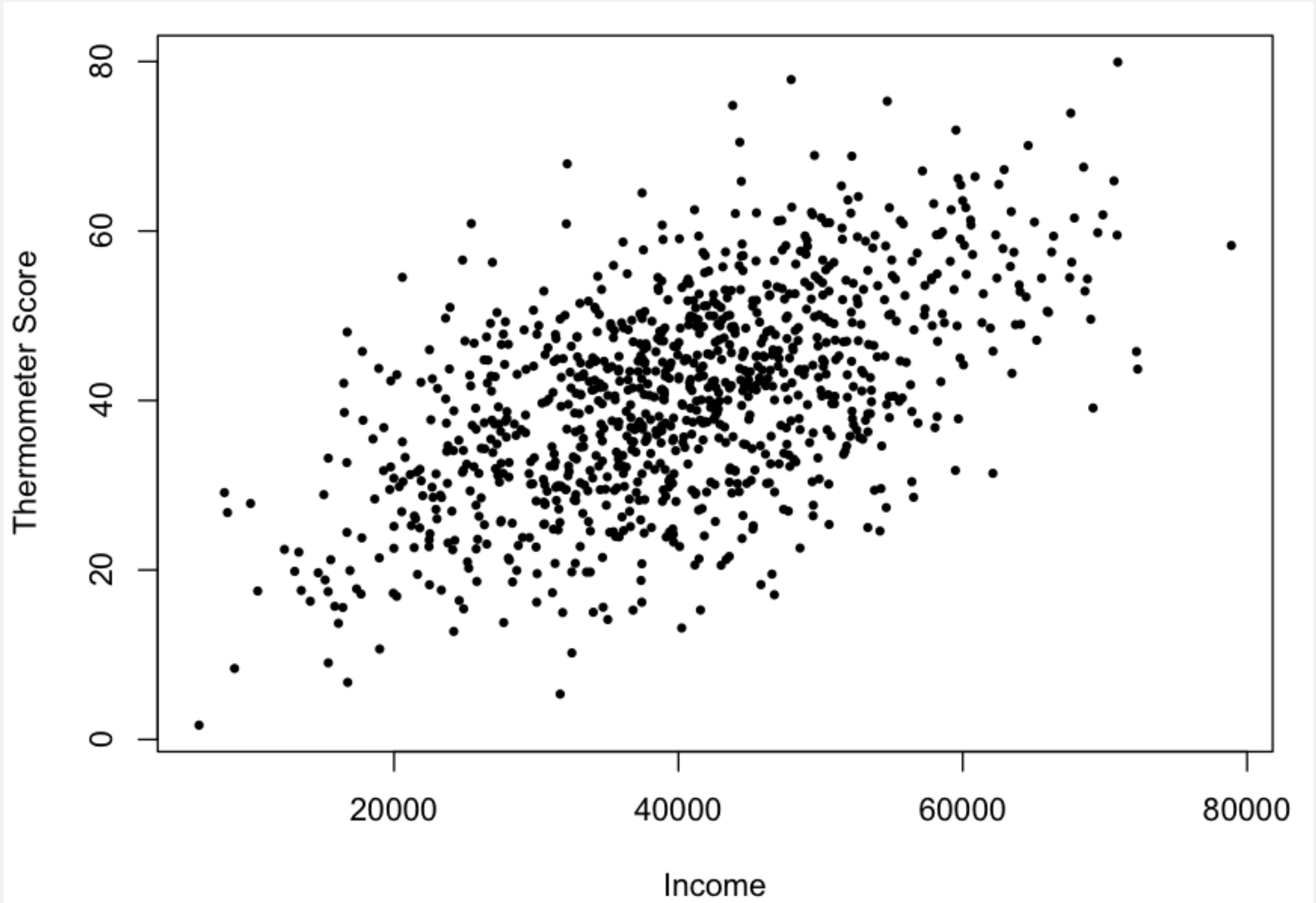
r=0.95

CORRELATION



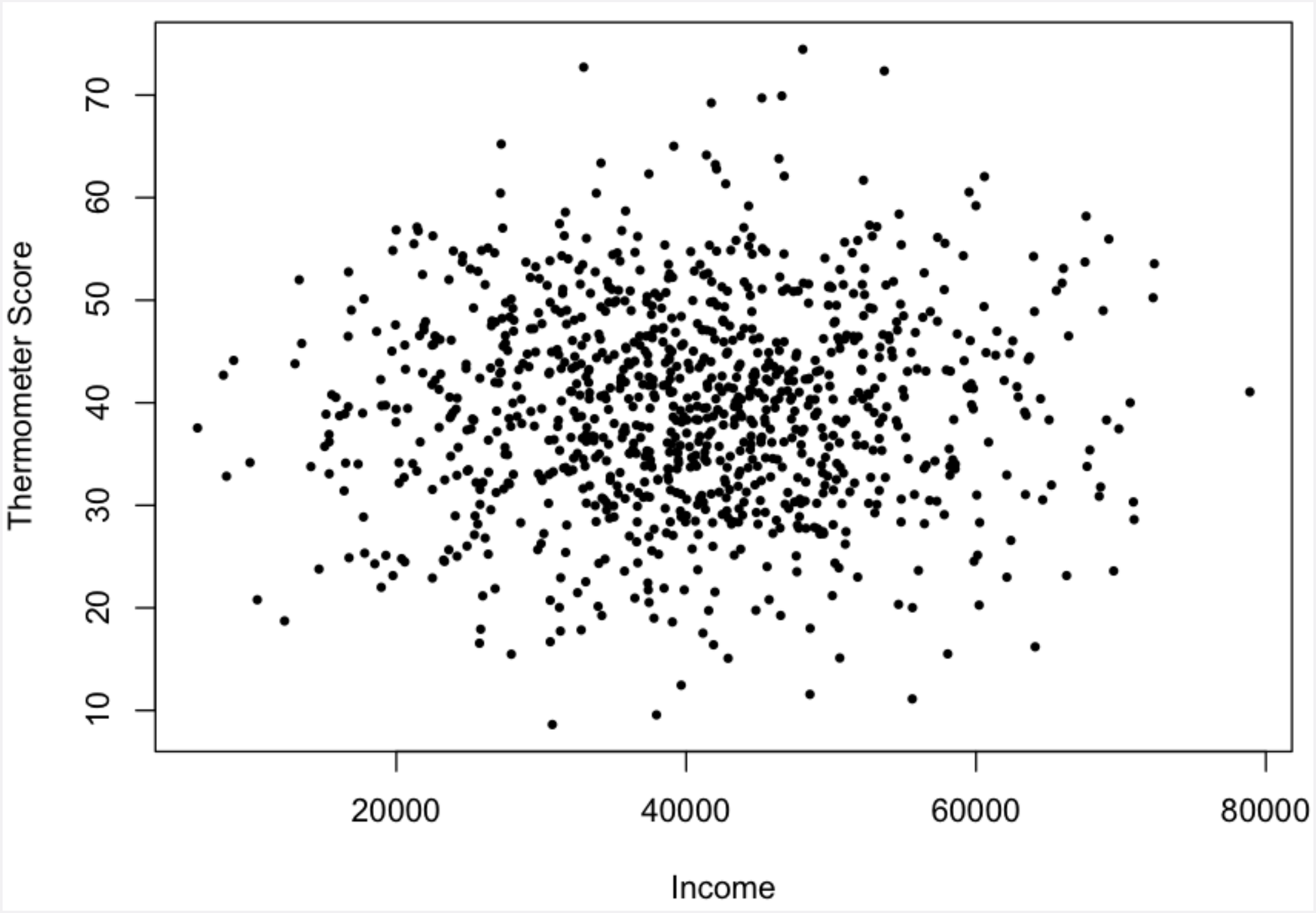
r=0.77

CORRELATION



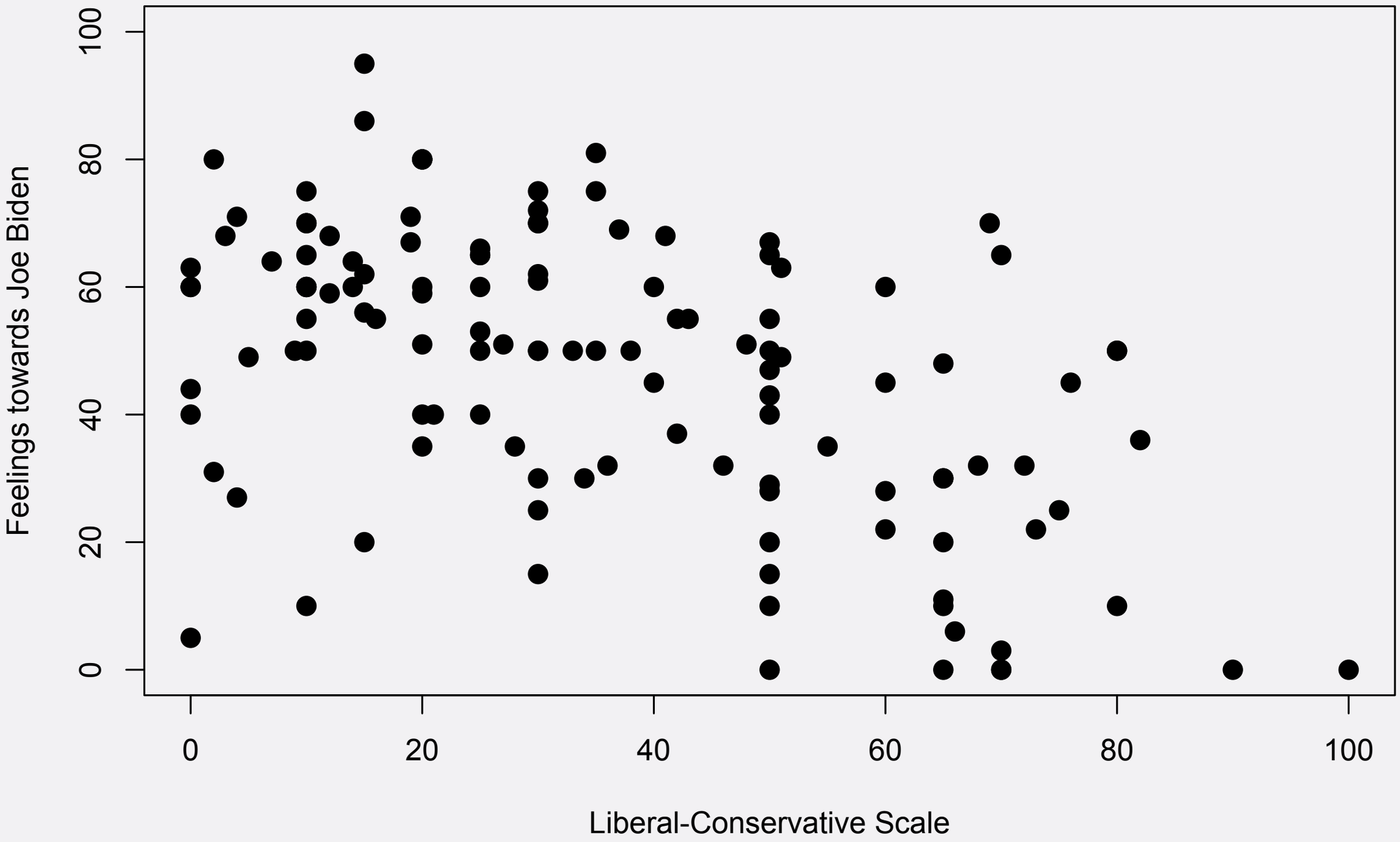
r=0.53

CORRELATION



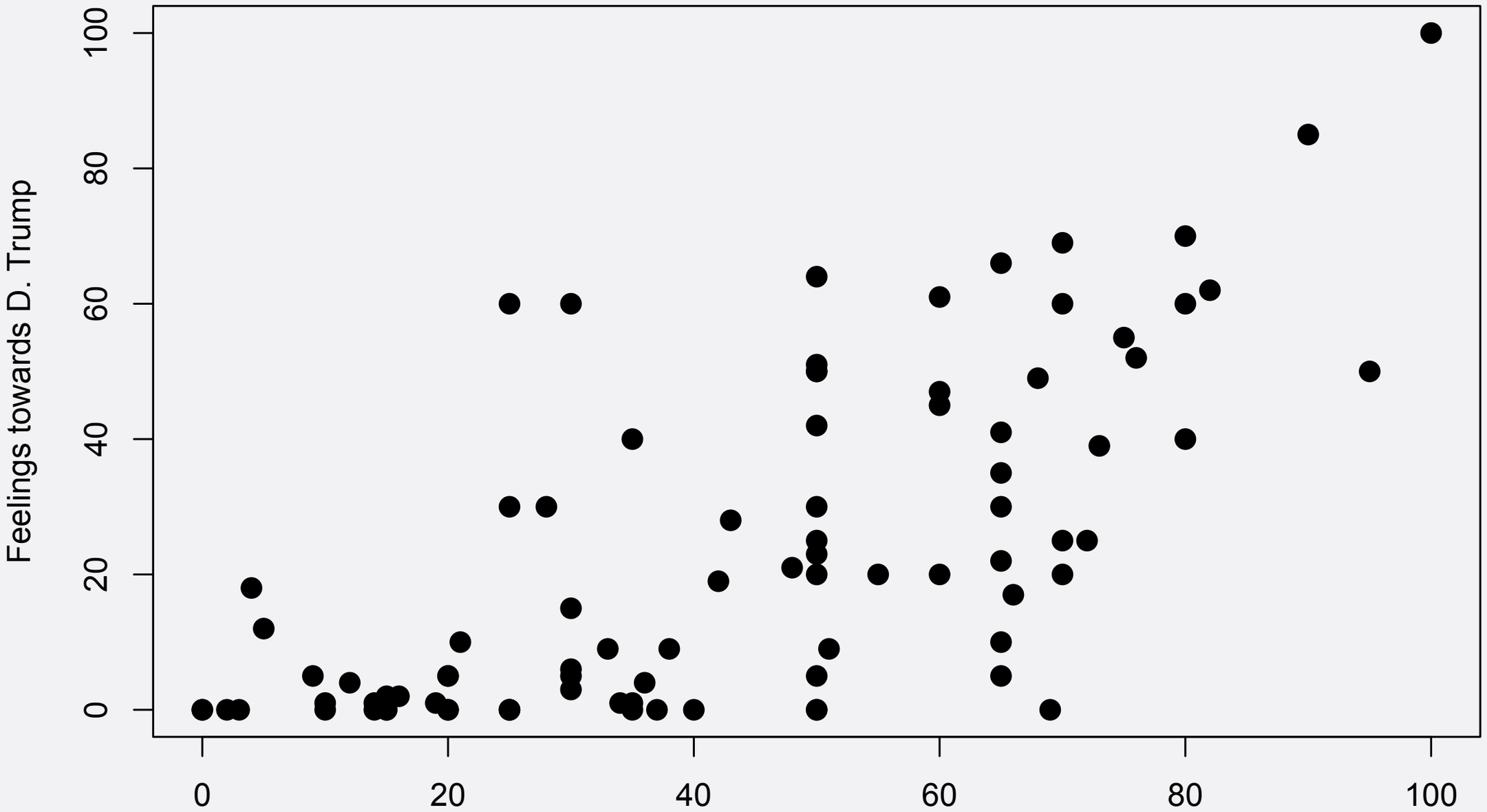
r=0

JOE BIDEN



r=-0.51

DONALD TRUMP



r=0.70

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.

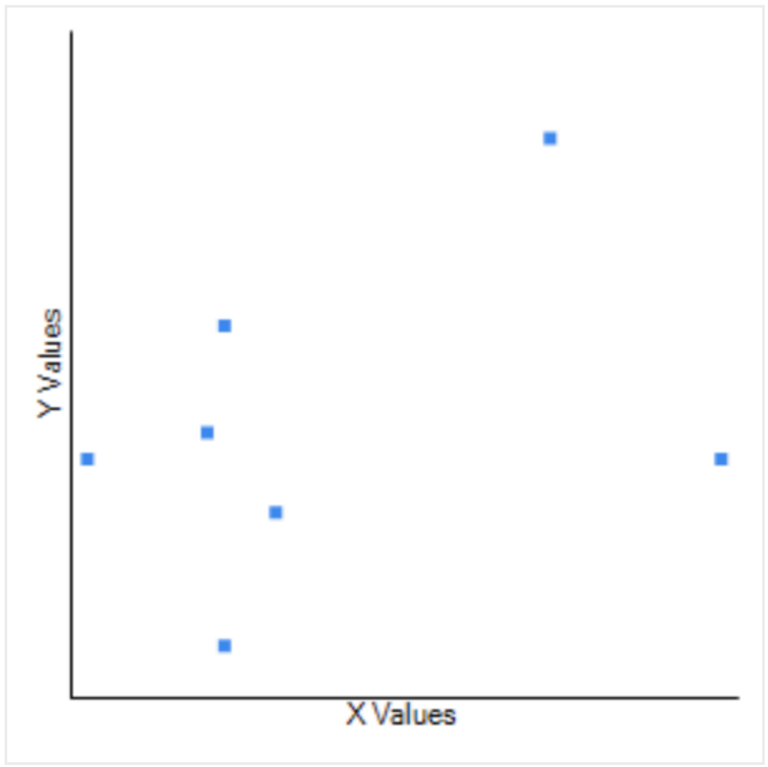
Note: If you want to calculate a P value from your R score, we have a calculator here (before clicking, remember to note your r score and record any calculation details you require).

X Values

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

Y Values

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

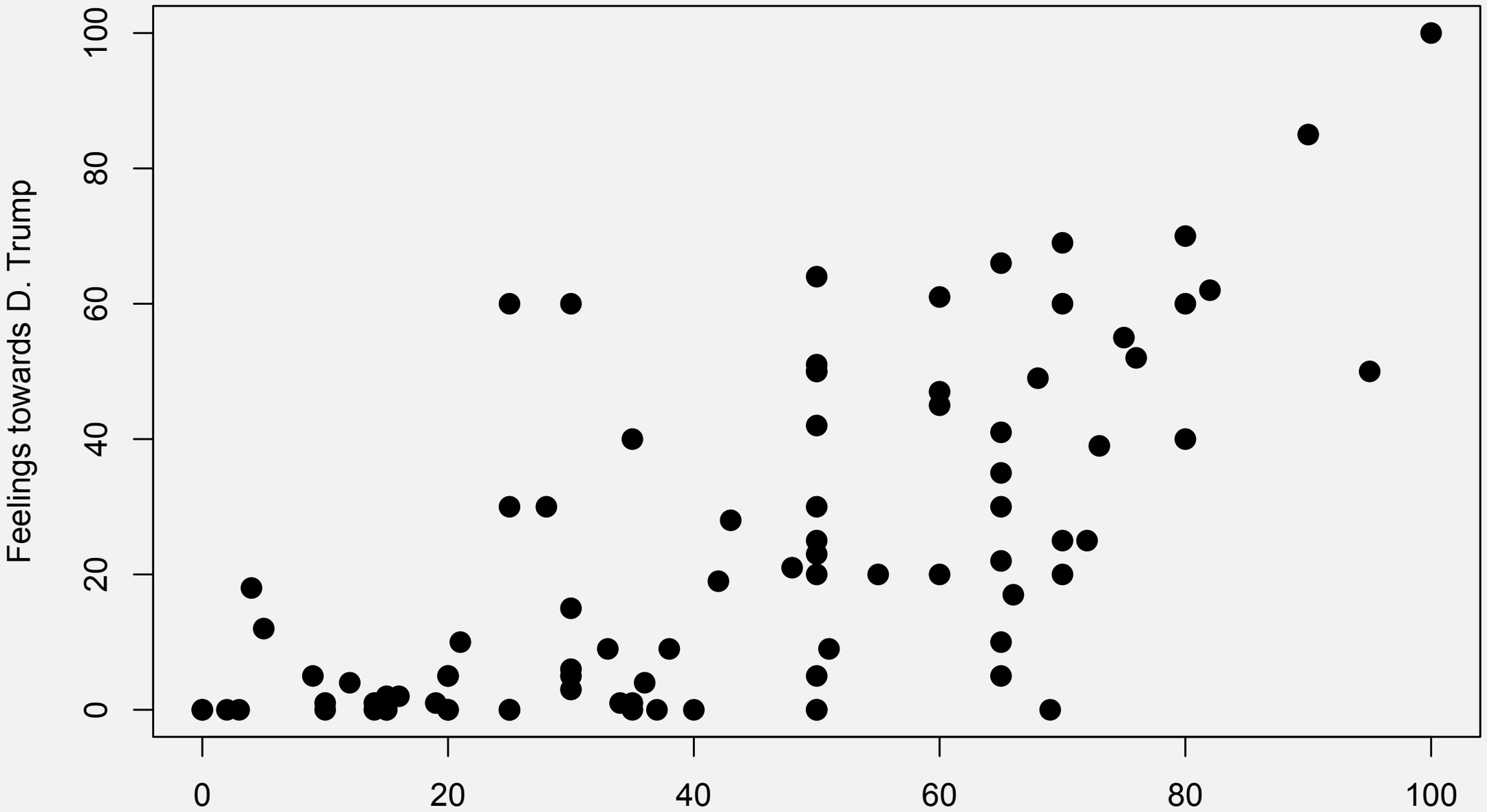


PEARSON'S R

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

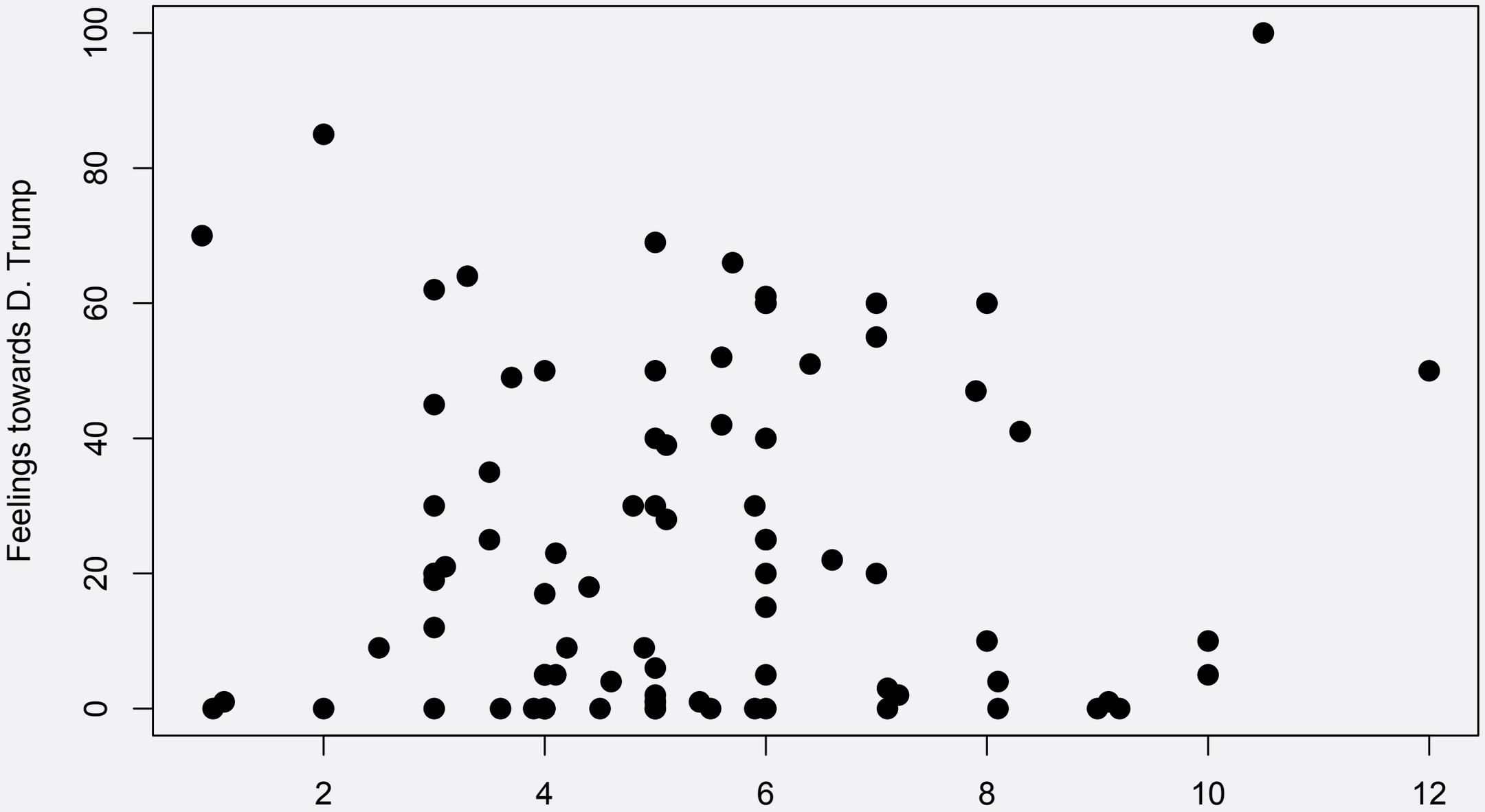
- **Intuition: Captures how much values of two variable vary together**
  - **High (positive) correlation: If X takes higher values, Y takes higher values**
  - **High (negative) correlation: If X takes higher values, Y takes lower values**
  - **Low correlation: If X takes higher values, values of Y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

DONALD TRUMP



r=0.70

DONALD TRUMP



r=0.07

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

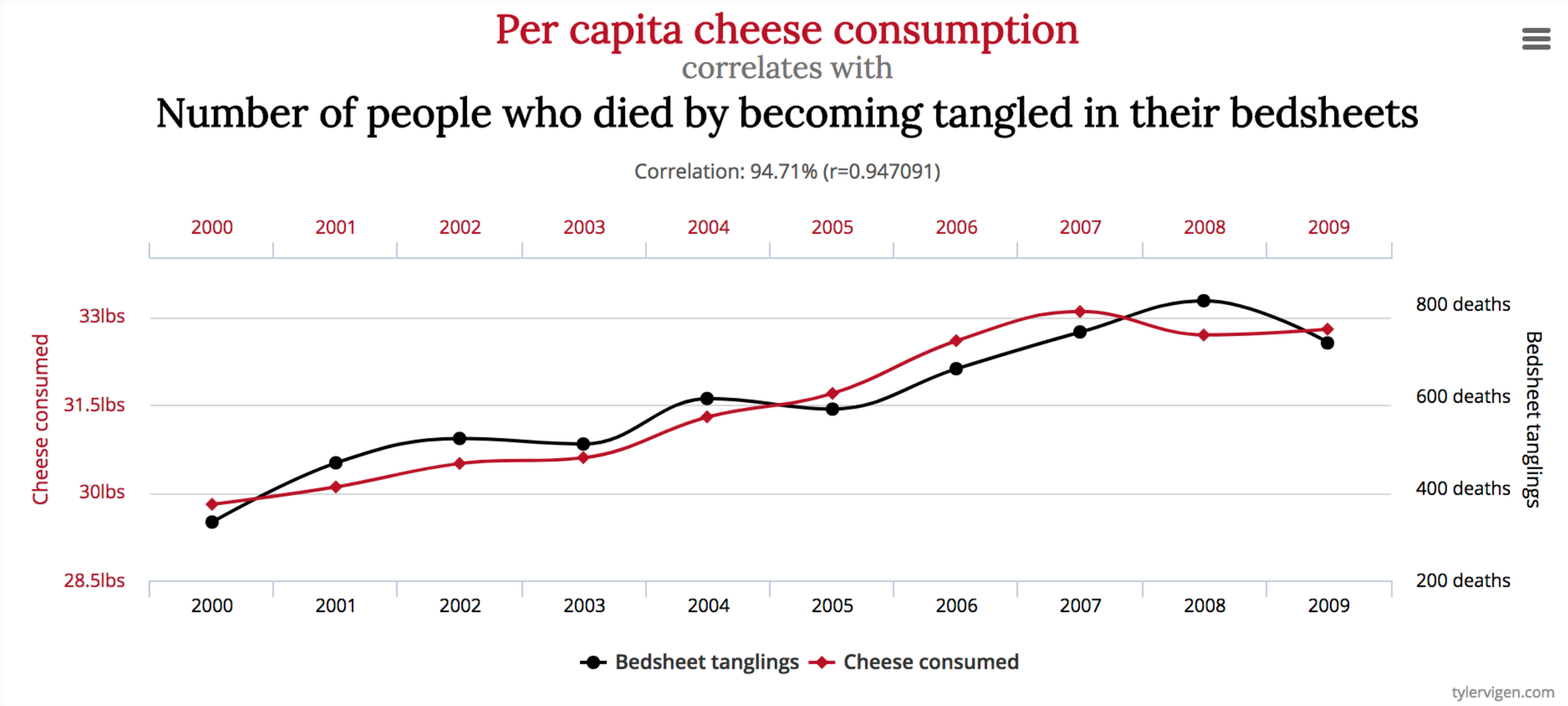
BIVARIATE RELATIONSHIPS

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

CAREFUL!

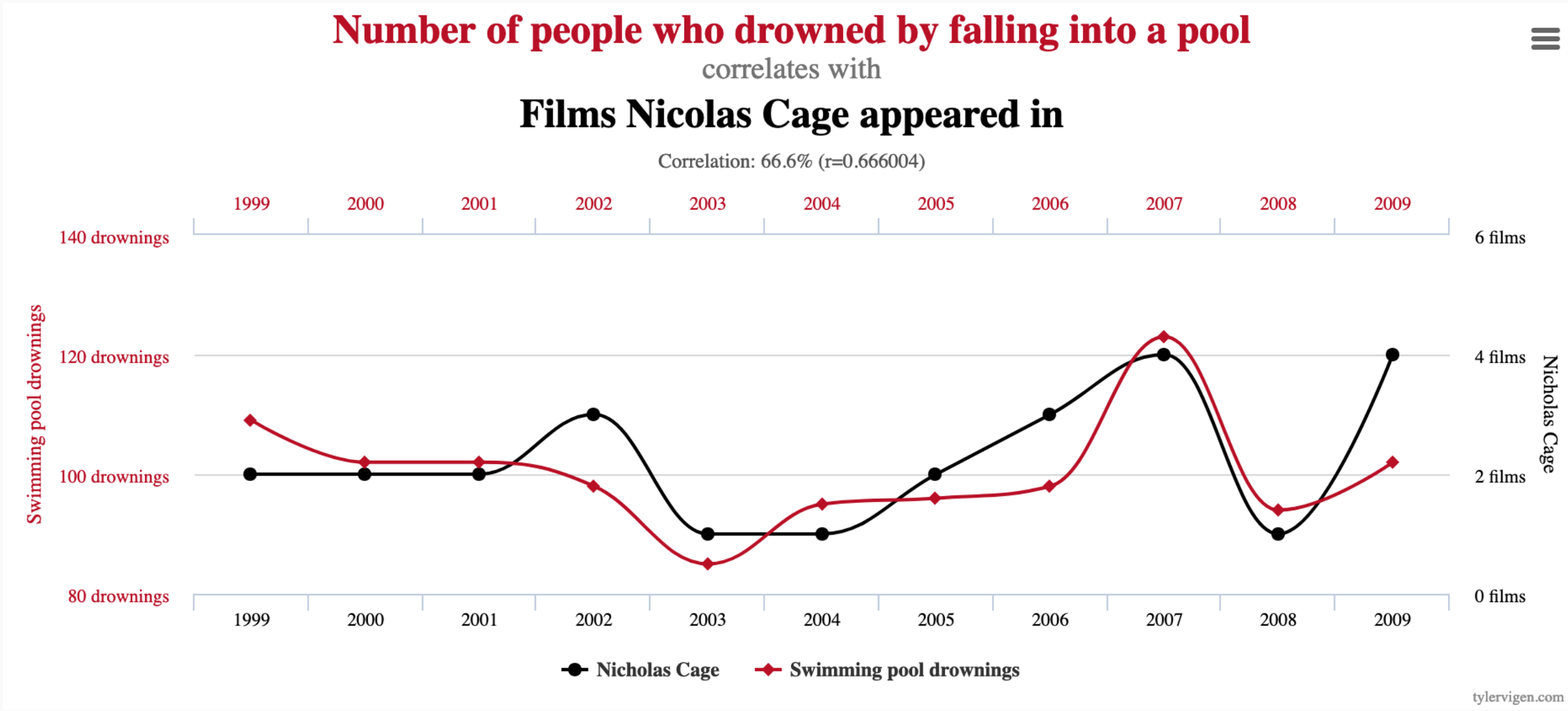
- Important: Just because we find a correlation between two variables does *_not_* mean that the independent variable *causes* the dependent variable
- The other hurdles to causality still apply!

CAREFUL!



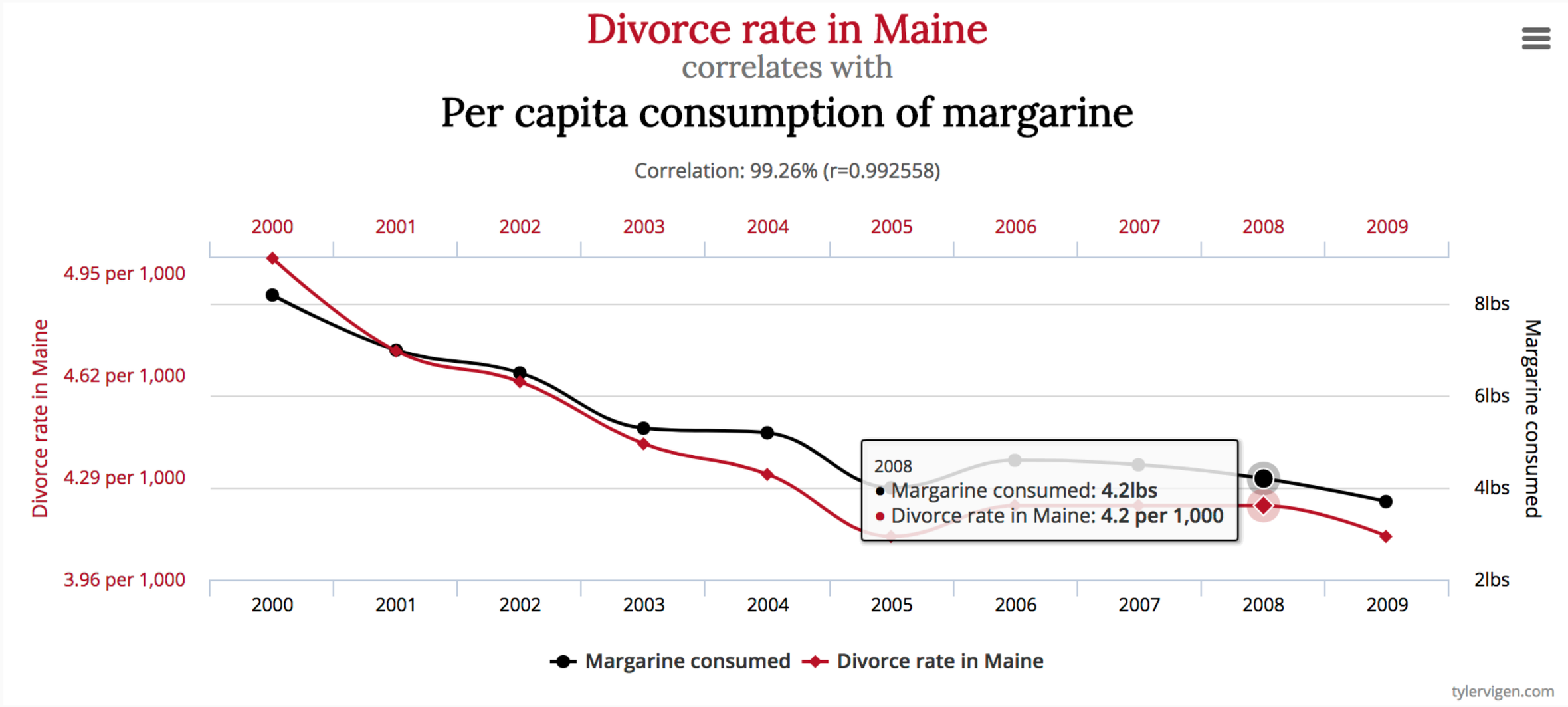
Data sources: U.S. Department of Agriculture and Centers for Disease Control & Prevention

CAREFUL!



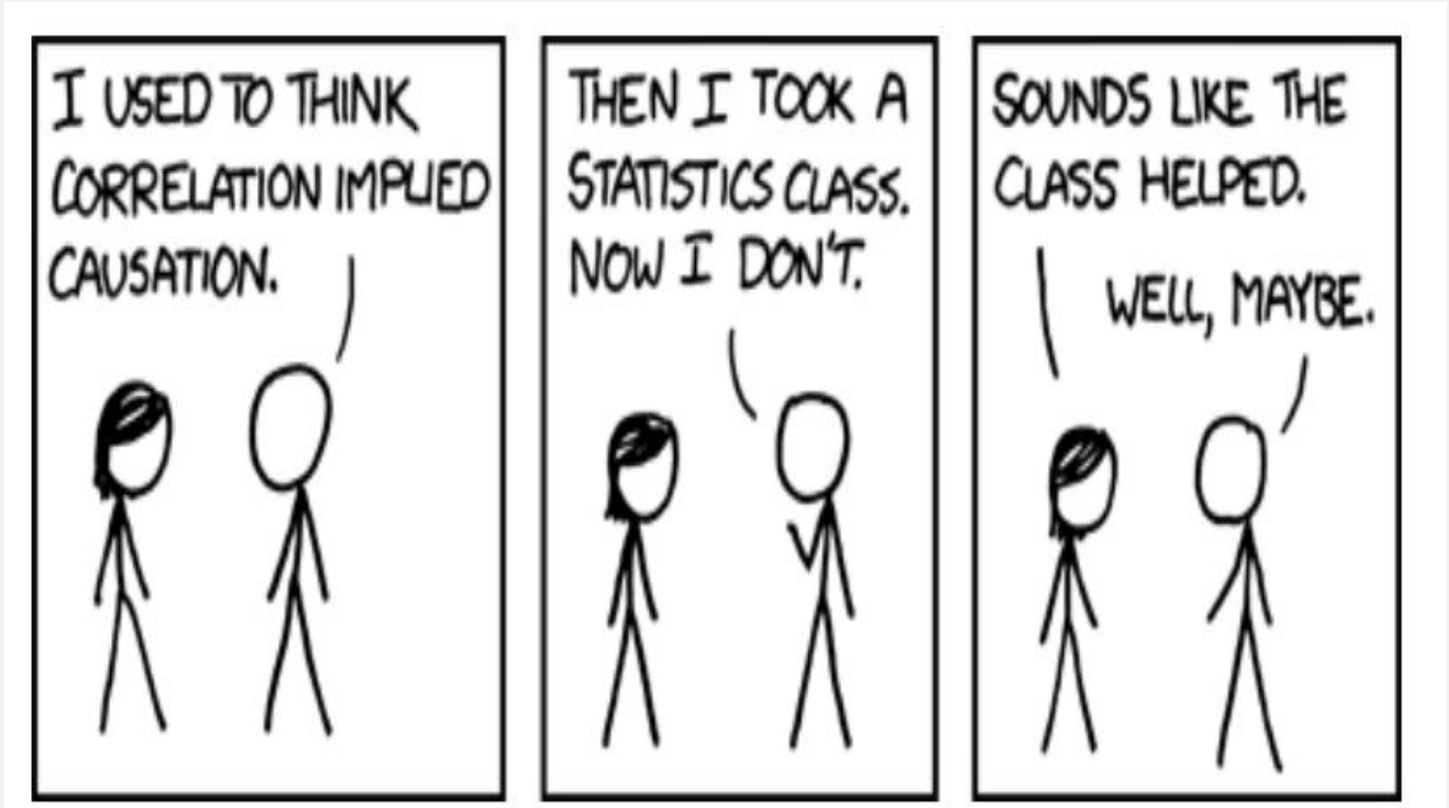
Data sources: Centers for Disease Control & Prevention and Internet Movie Database

CAREFUL!



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

CAREFUL!

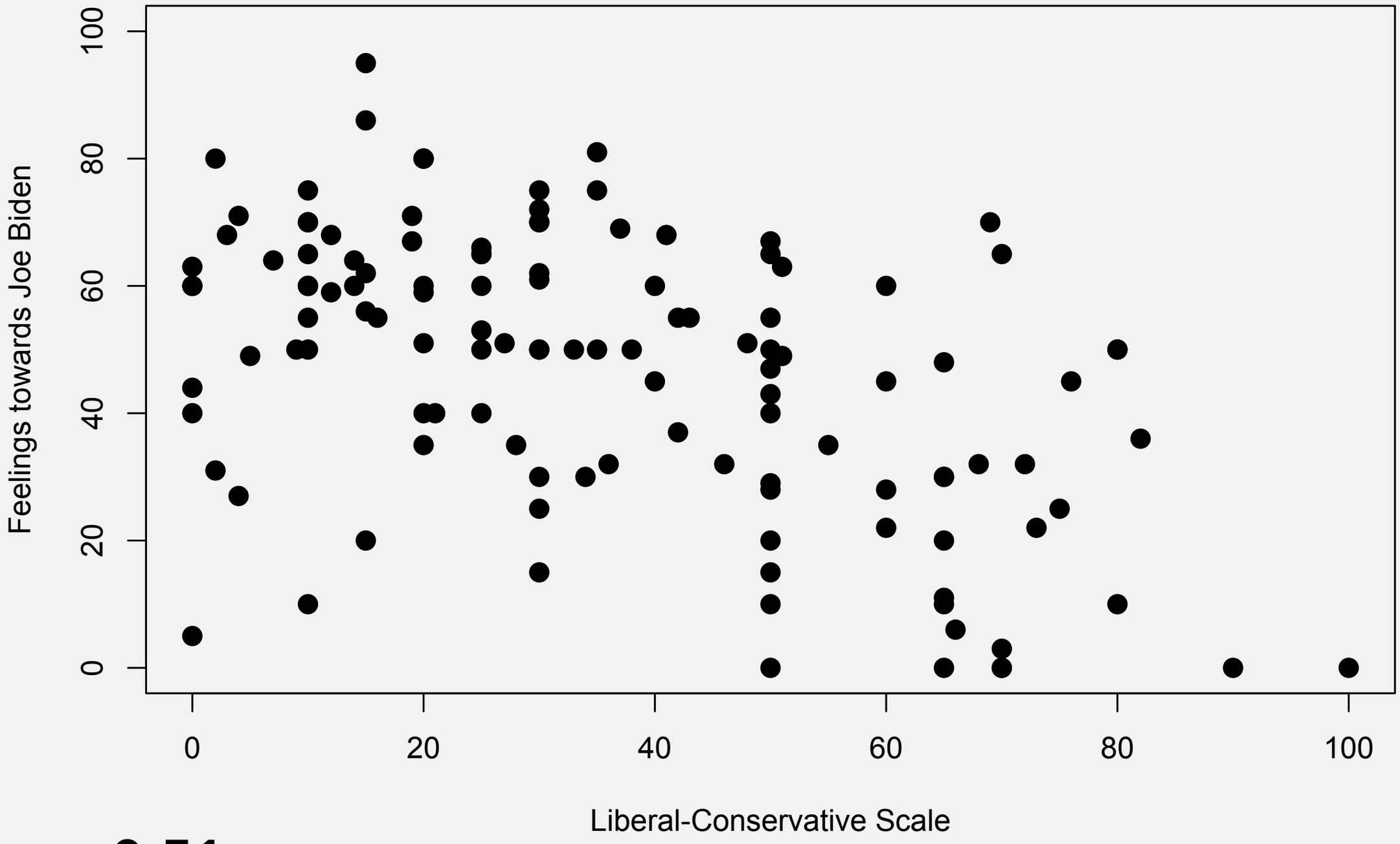


- **Important: Just because we find a correlation between two variables does **not** mean that the independent variable *causes* the dependent variable**
  - **The other hurdles to causality still apply!**

BIVARIATE RELATIONSHIPS

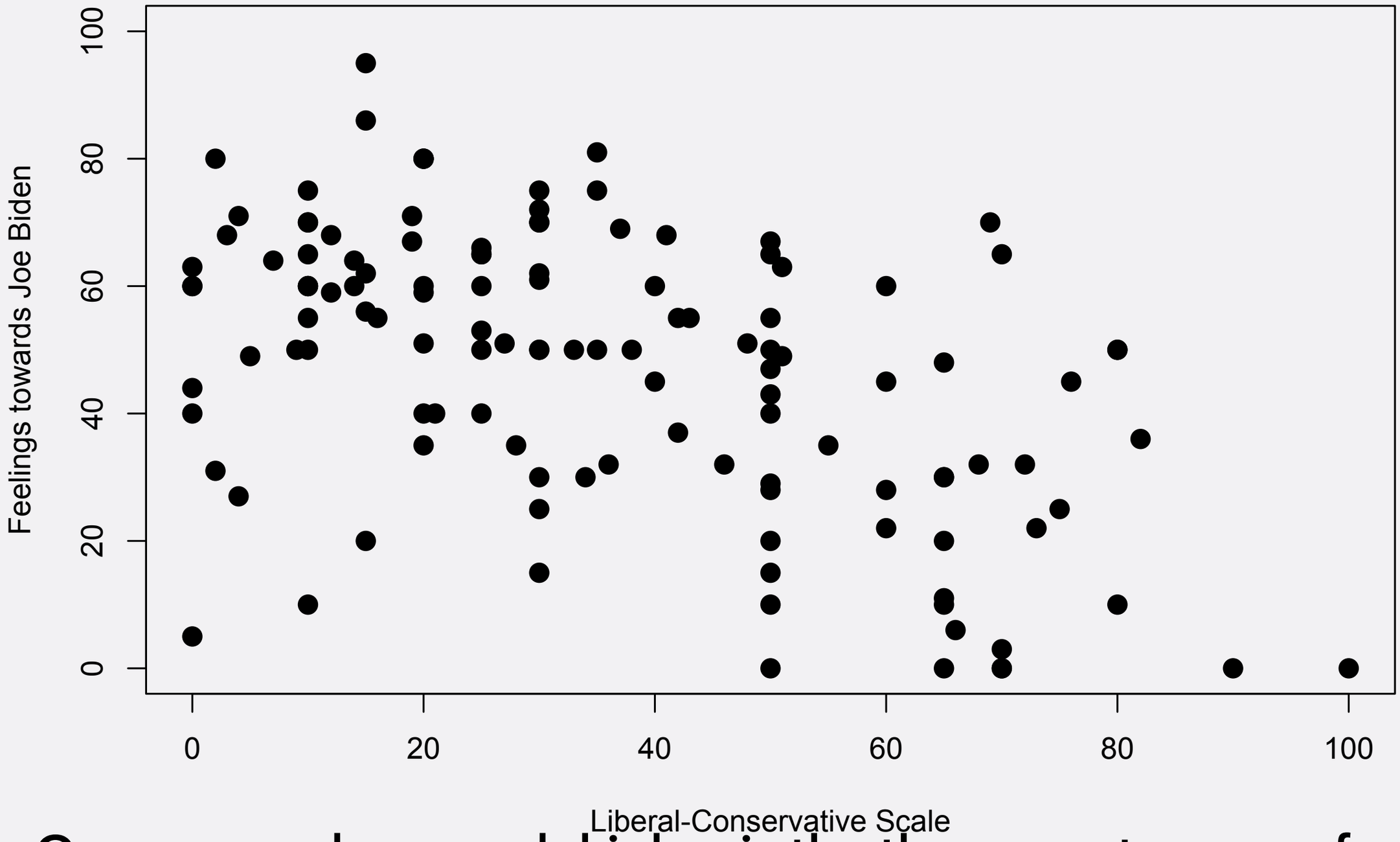
Dependent Variable	Independent Variable: Nominal/Ordinal	Independent Variable: Interval
Nominal/Ordinal	Cross-Tabulation	Not In This Class…
Interval	Mean Comparison	Correlation Coefficient

JOE BIDEN



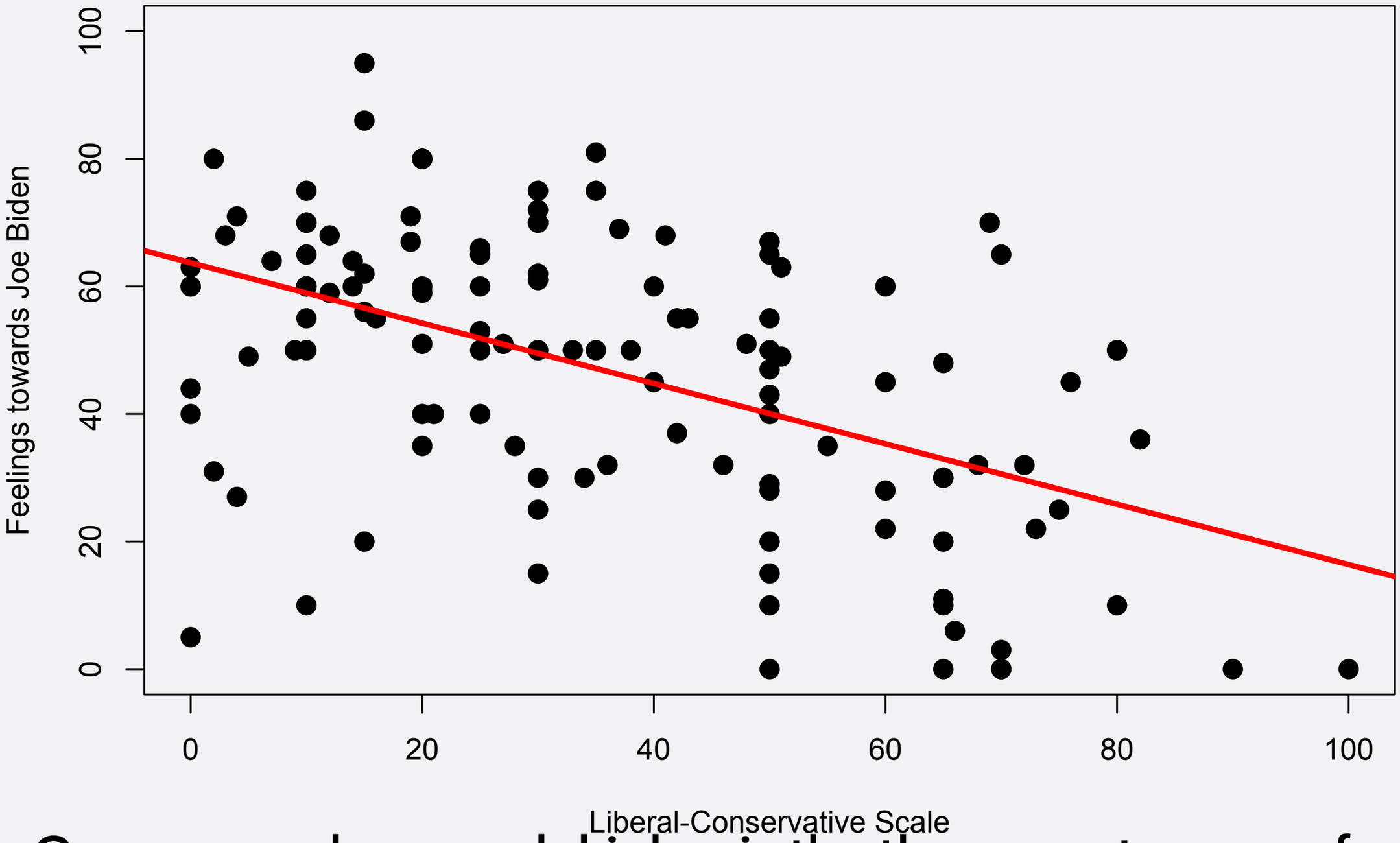
- r=-0.51
- Correlation: Direction and strength of relation, not size

JOE BIDEN



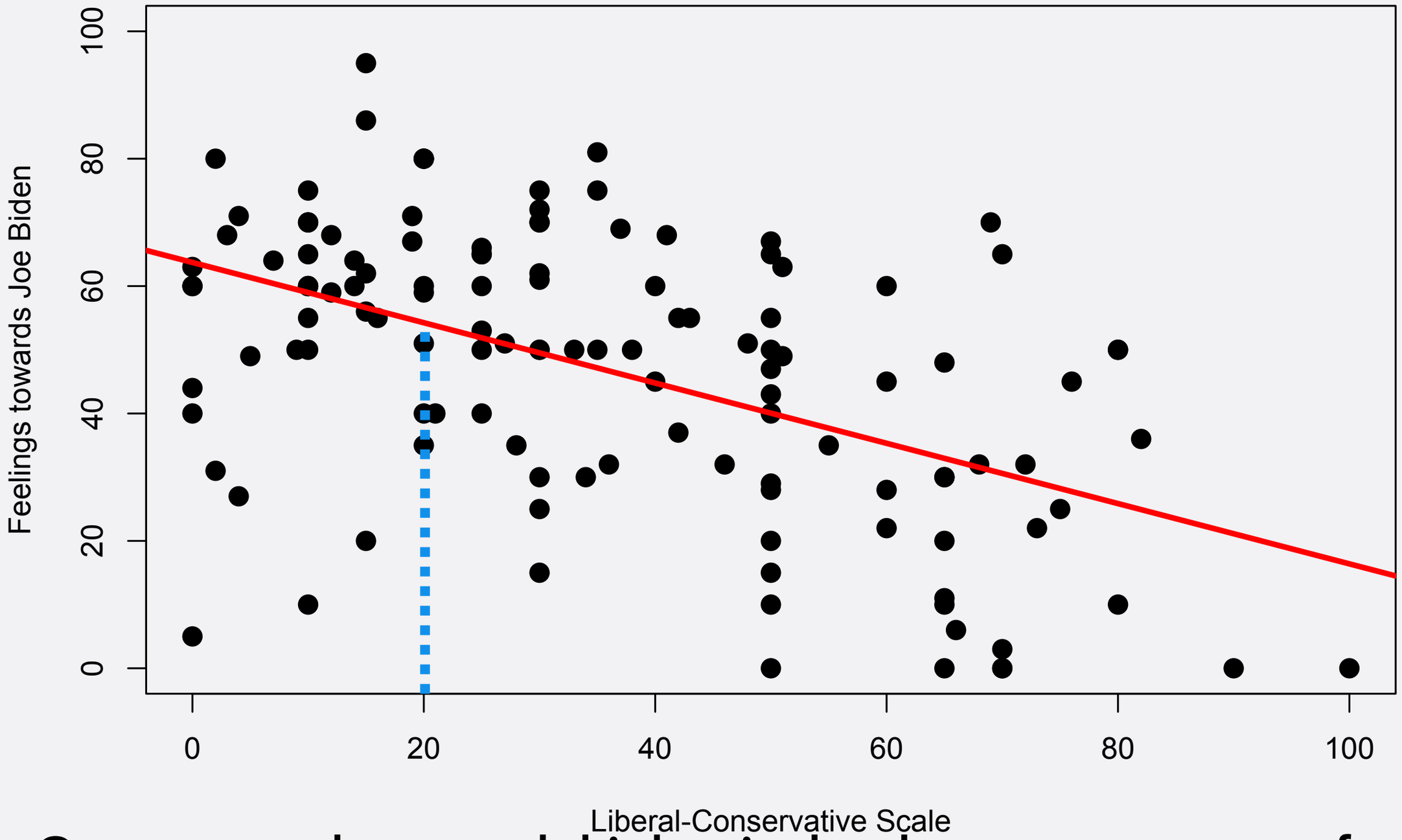
- **On average, how much higher is the thermometer score for someone who is a 20 on the liberal-conservative scale, compared to someone who is a 80?**

LINE



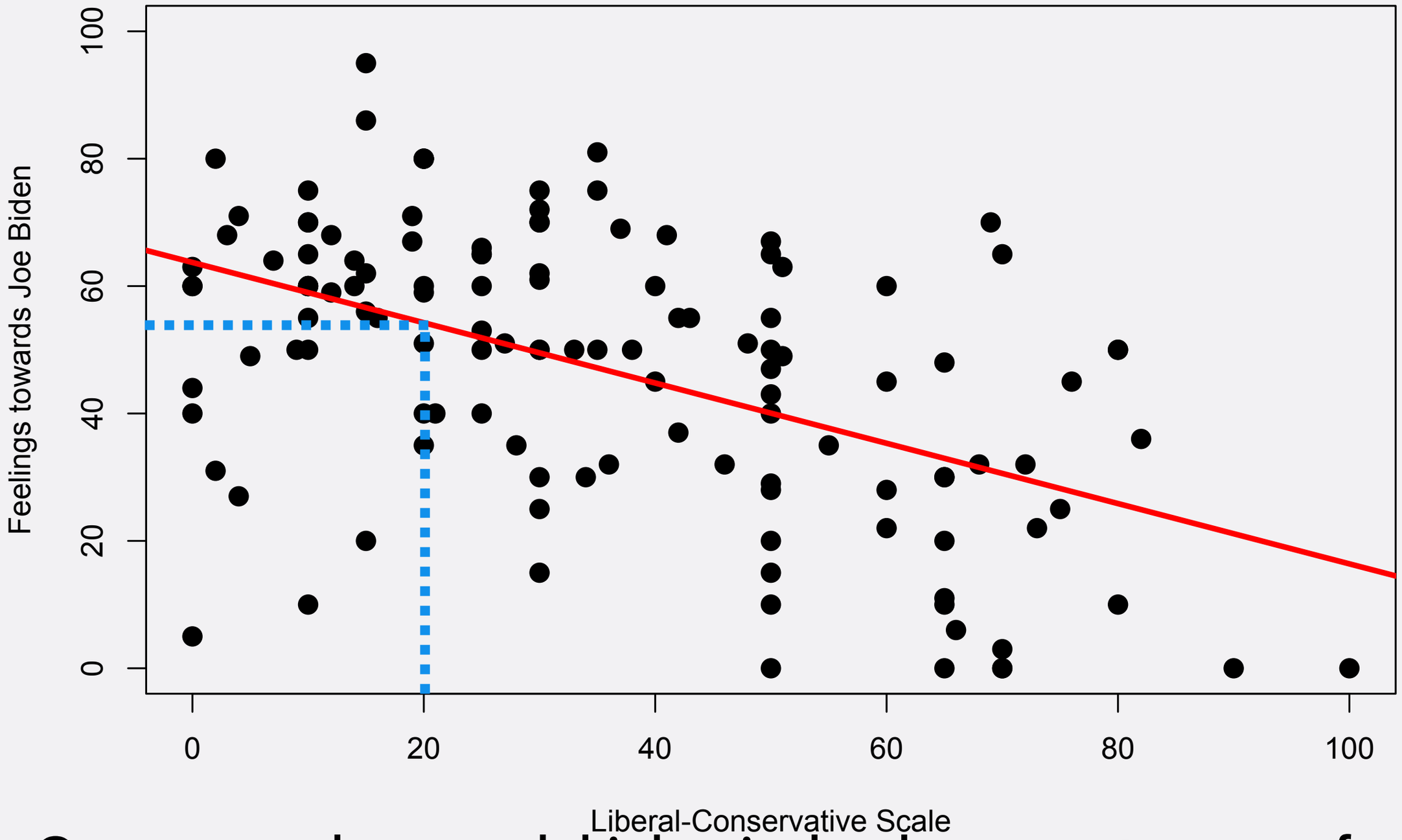
- **On average, how much higher is the thermometer score for someone who is a 20 on the liberal-conservative scale, compared to someone who is a 80?**

LINE



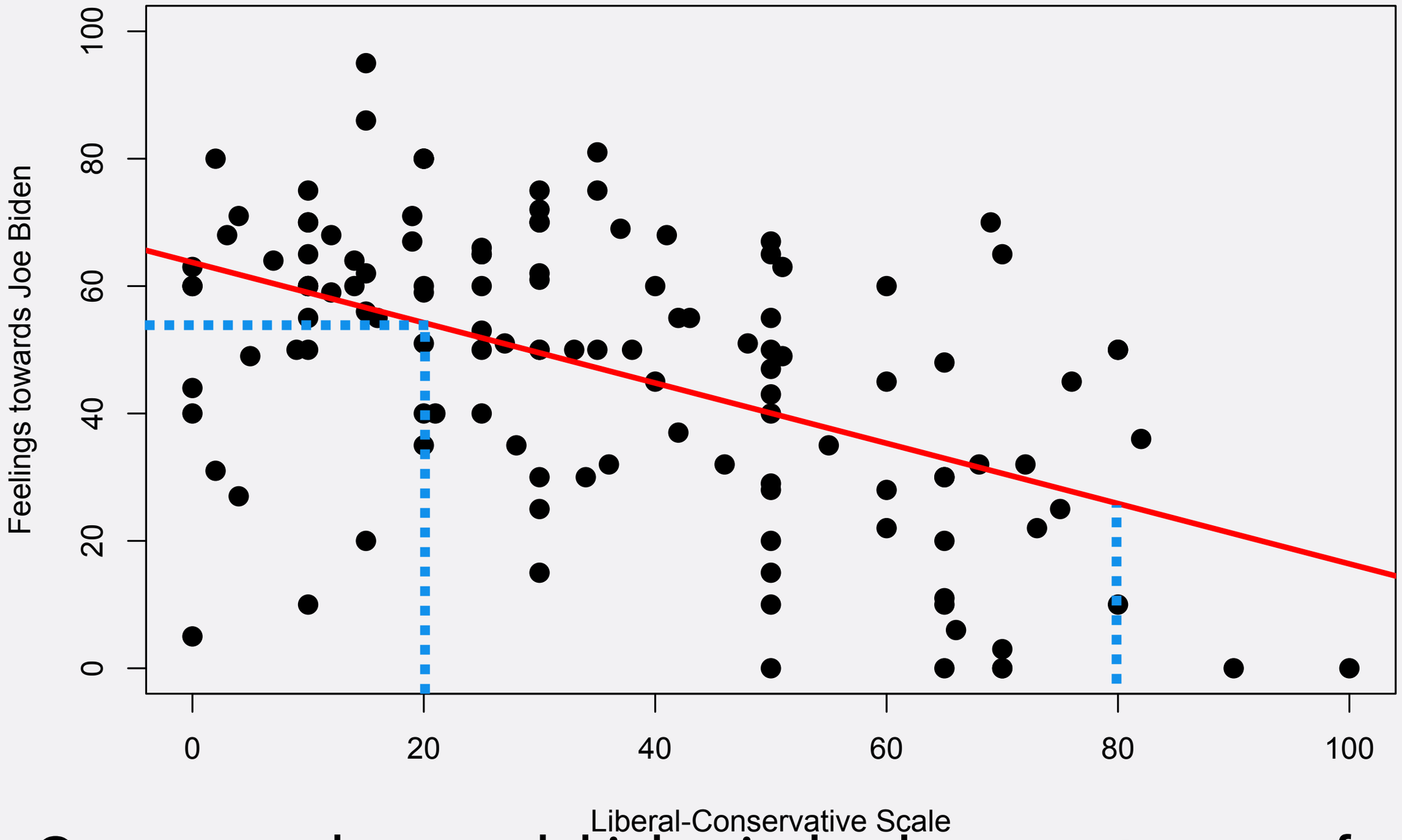
- **On average, how much higher is the thermometer score for someone who is a 20 on the liberal-conservative scale, compared to someone who is a 80?**

LINE



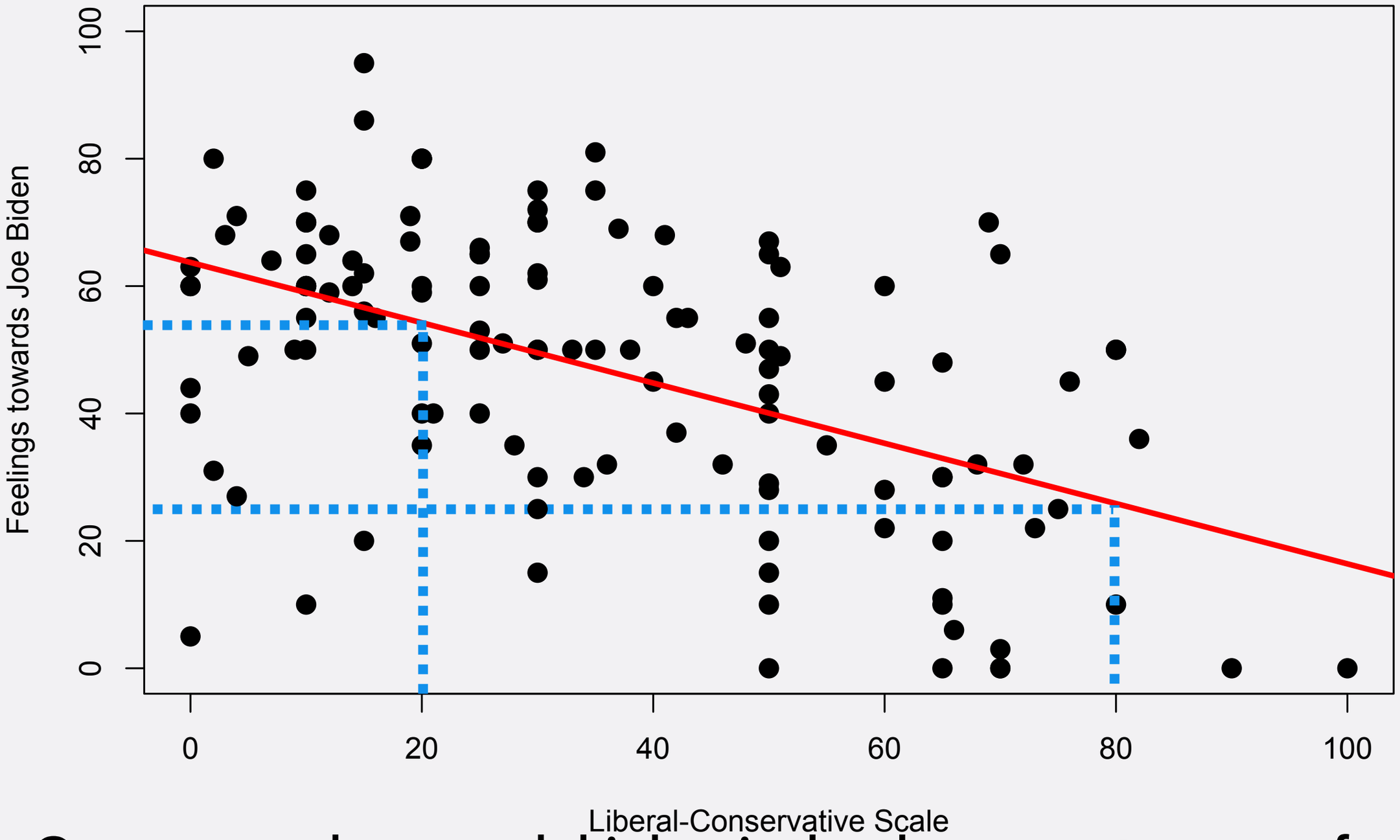
- **On average, how much higher is the thermometer score for someone who is a 20 on the liberal-conservative scale, compared to someone who is a 80?**

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?**

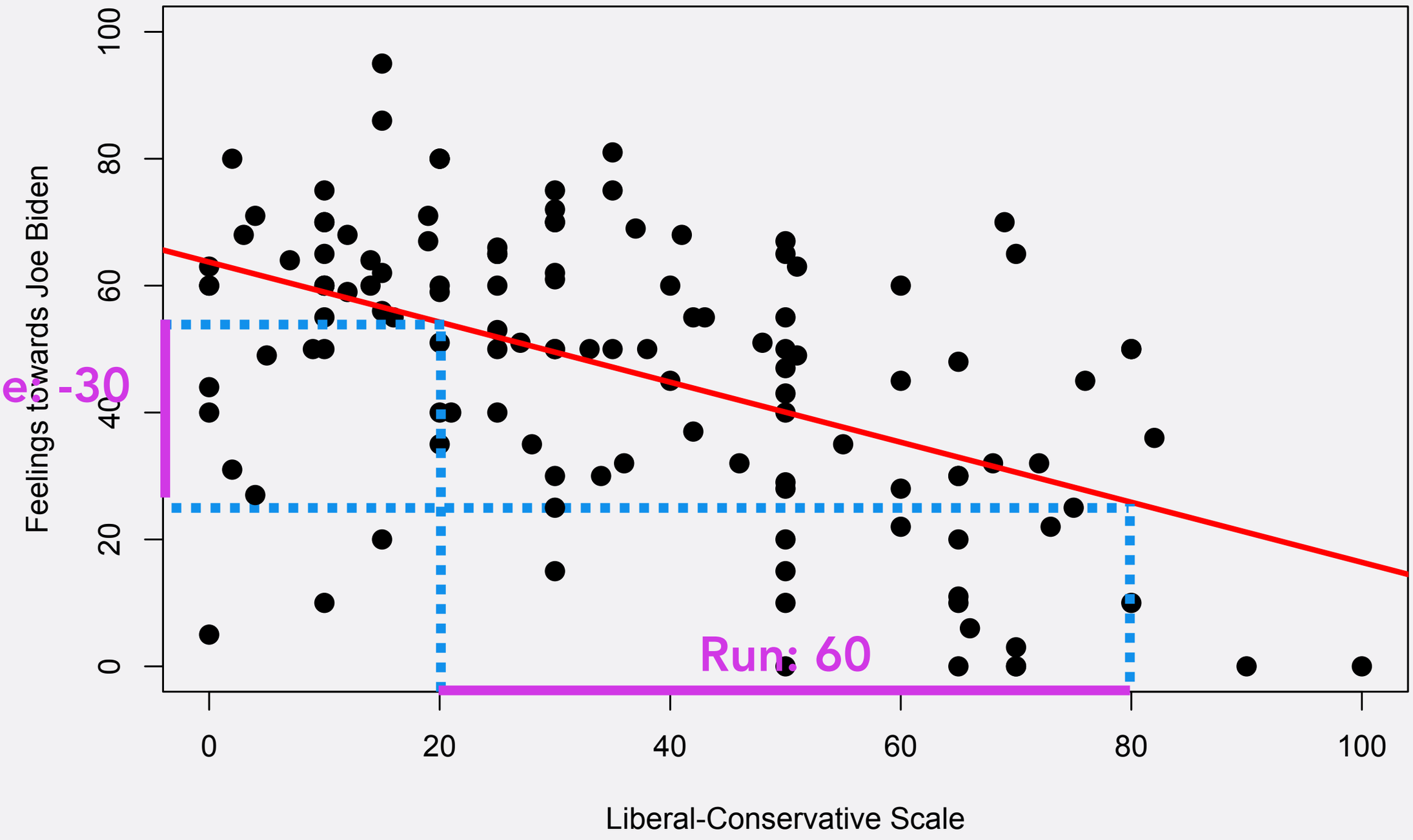
LINE



- **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

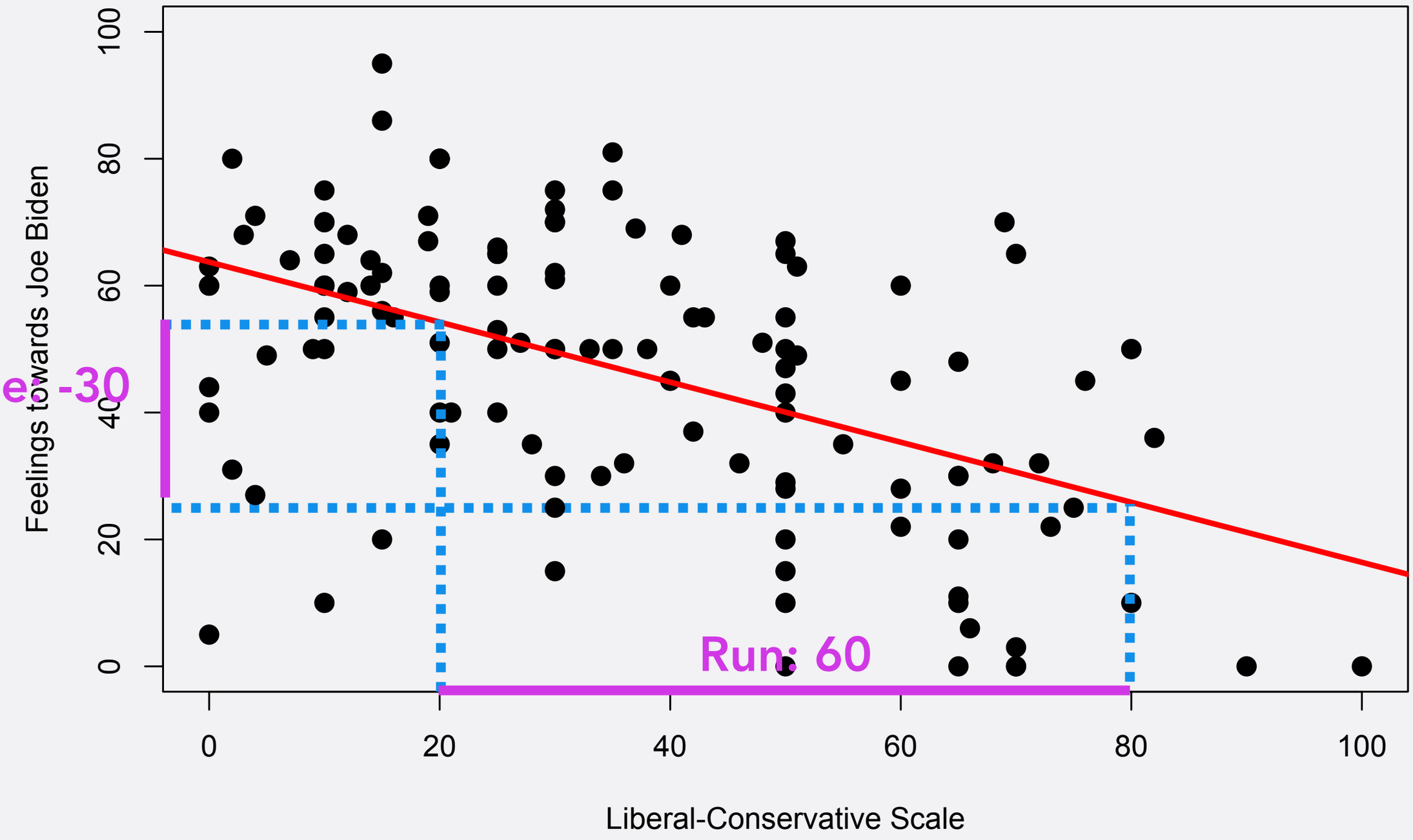
Slope=Rise over run



Slope=Rise over run=-30/60=-0.5

LINE

Slope=Rise over run



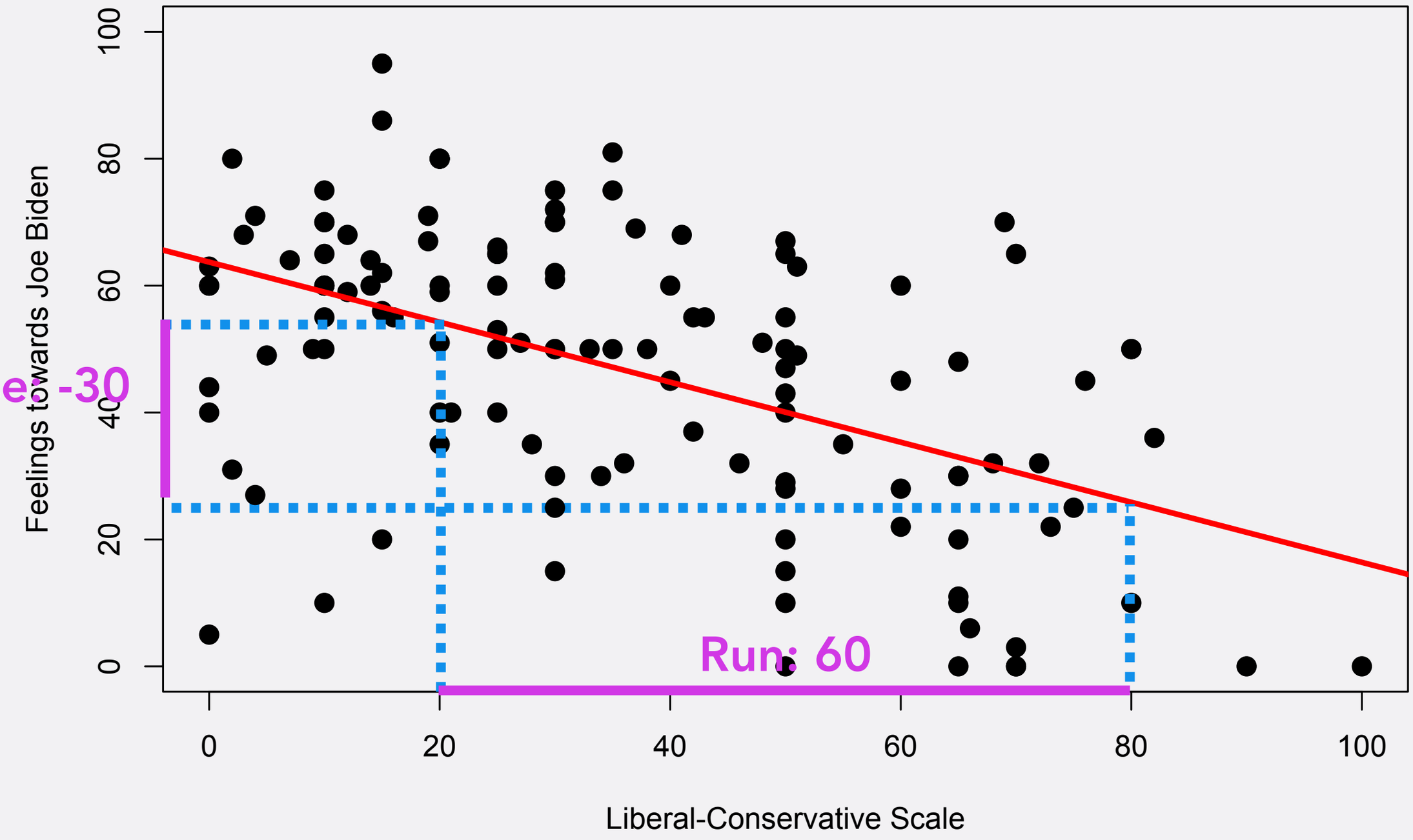
- **For each one unit increase on the liberal-conservative scale, feelings towards J. Biden go down by 0.5 points**

NOTE

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

LINE

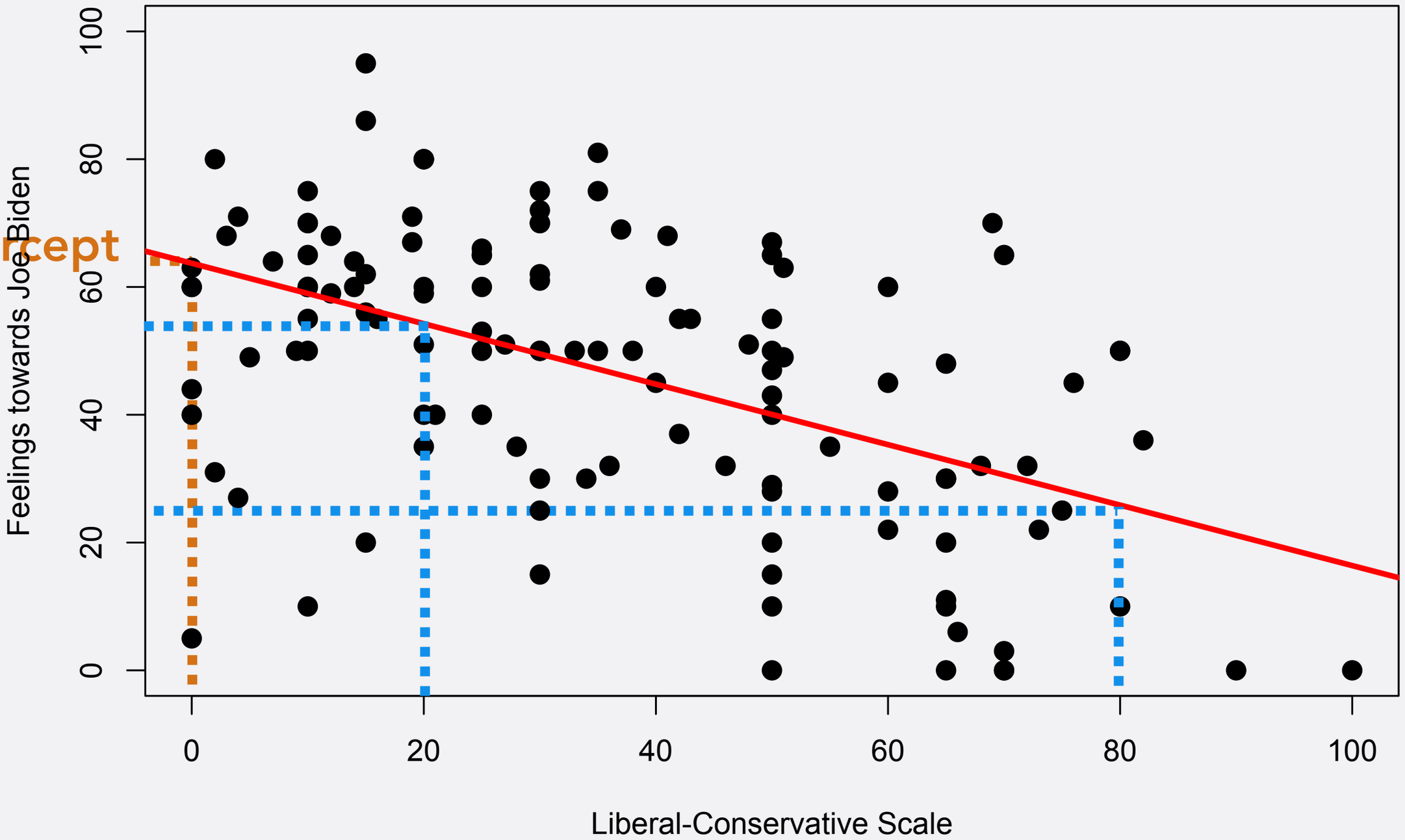
Slope=Rise over run



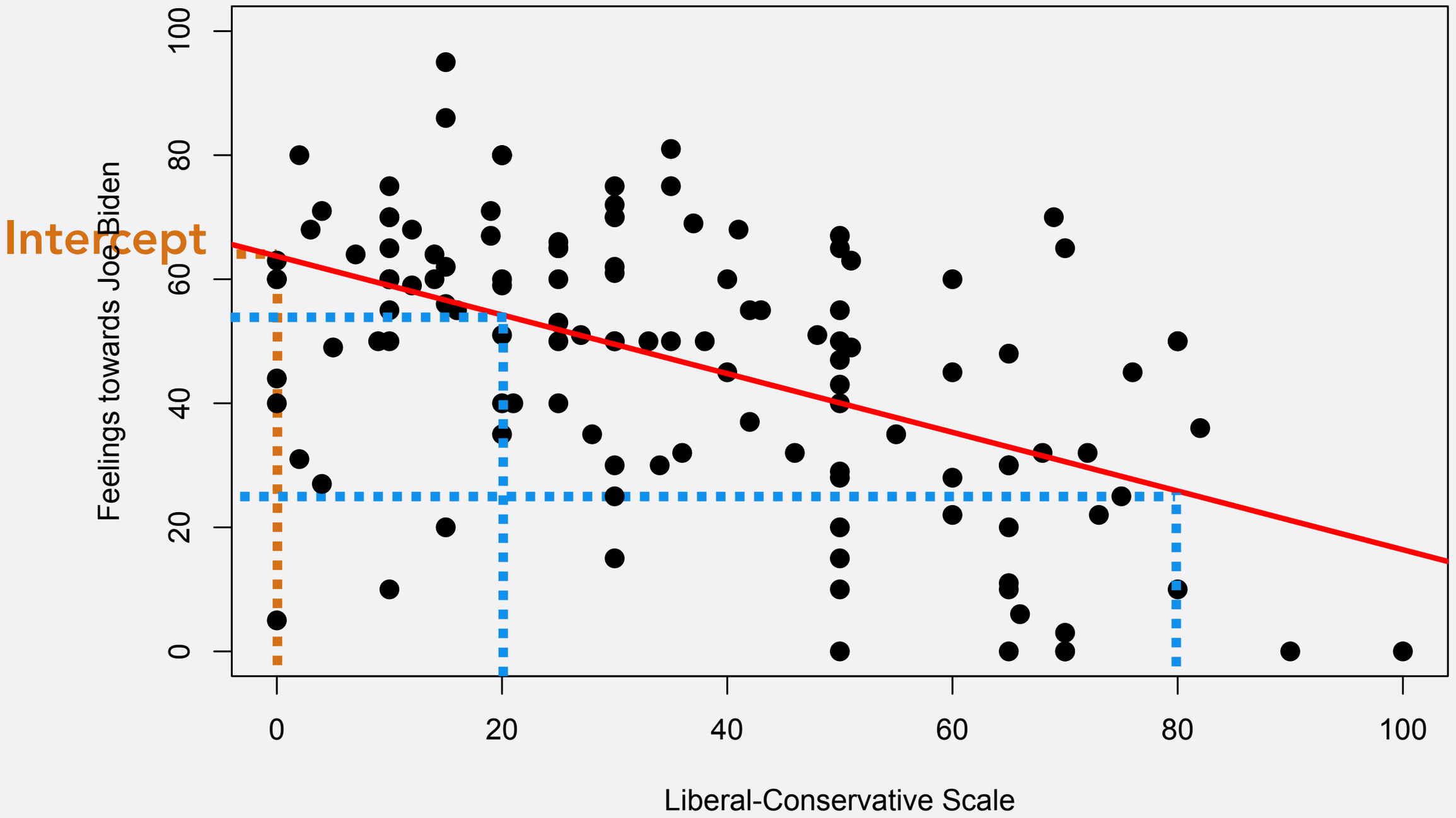
- **For each one unit increase on the liberal-conservative scale, feelings towards J. Biden go down by 0.5 points**

LINE

Intercept



LINE



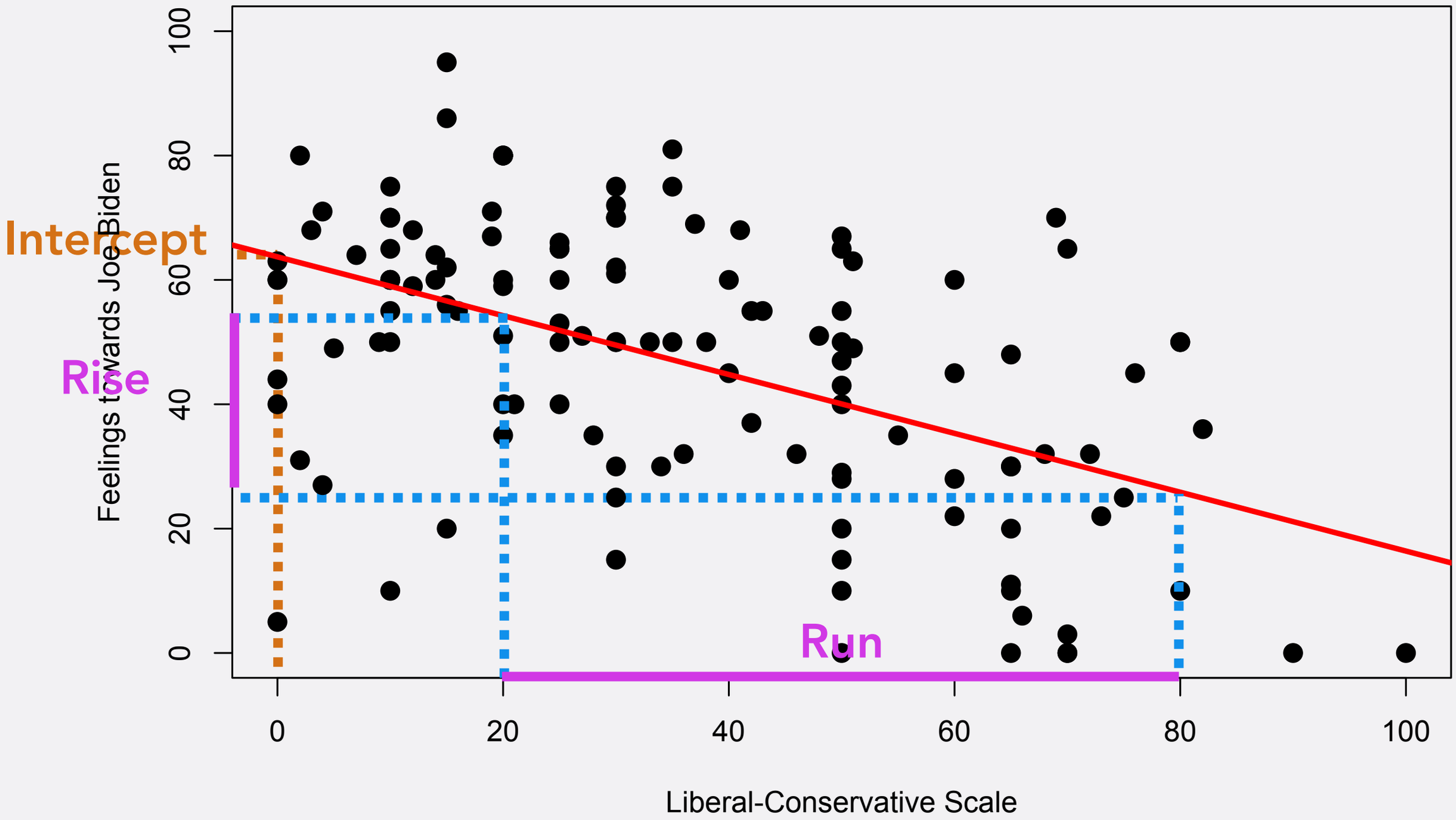
- **Students who are very liberal (score=0) are expected to have a feeling thermometer score of 64.**

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

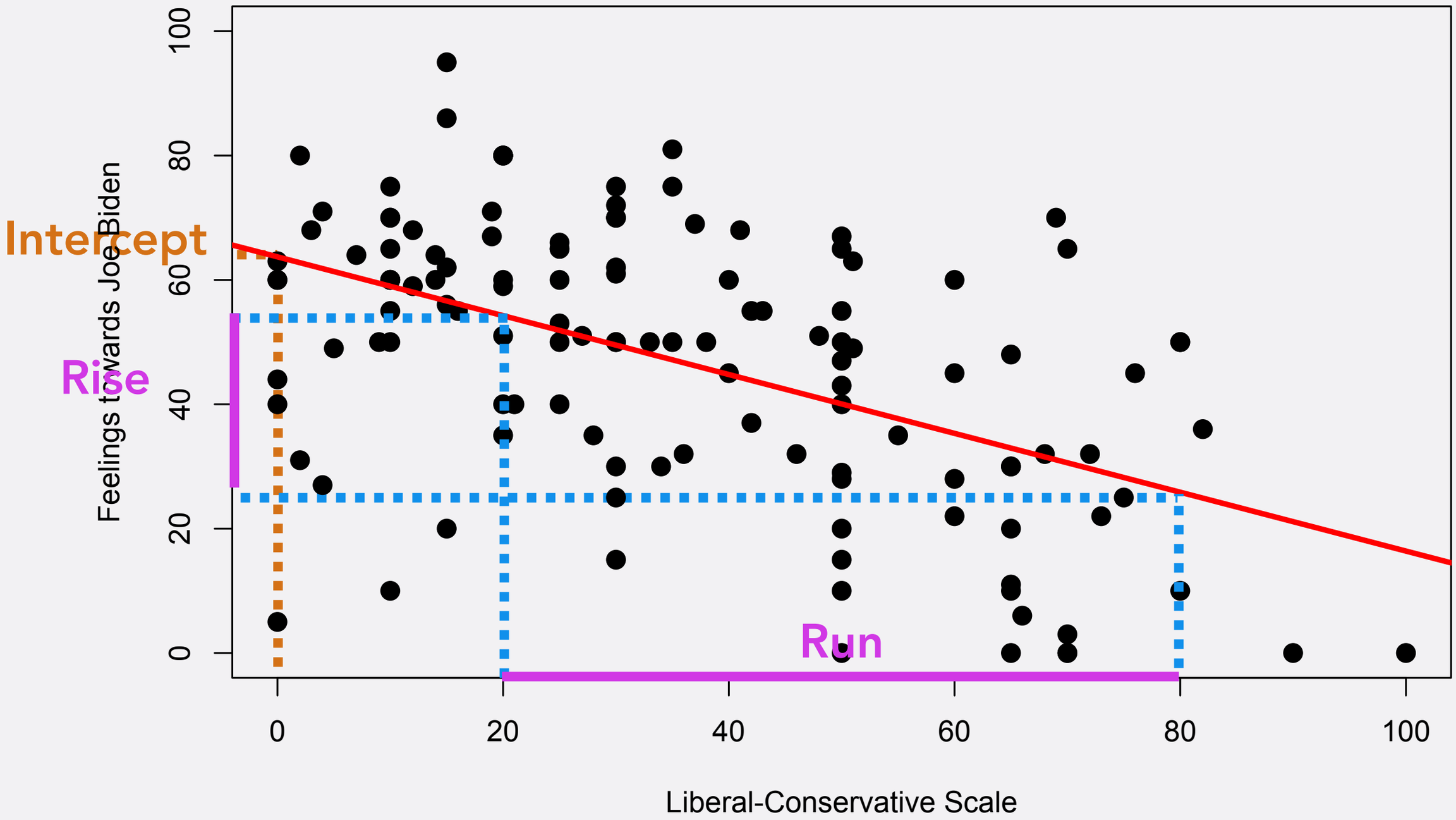
Slope=Rise over run



- Thermometer Score = Intercept + Slope * Lib/Cons

LINE

Slope=Rise over run



- Thermometer Score = 64 - 0.5 * Lib/Cons

WHAT THIS TELLS US

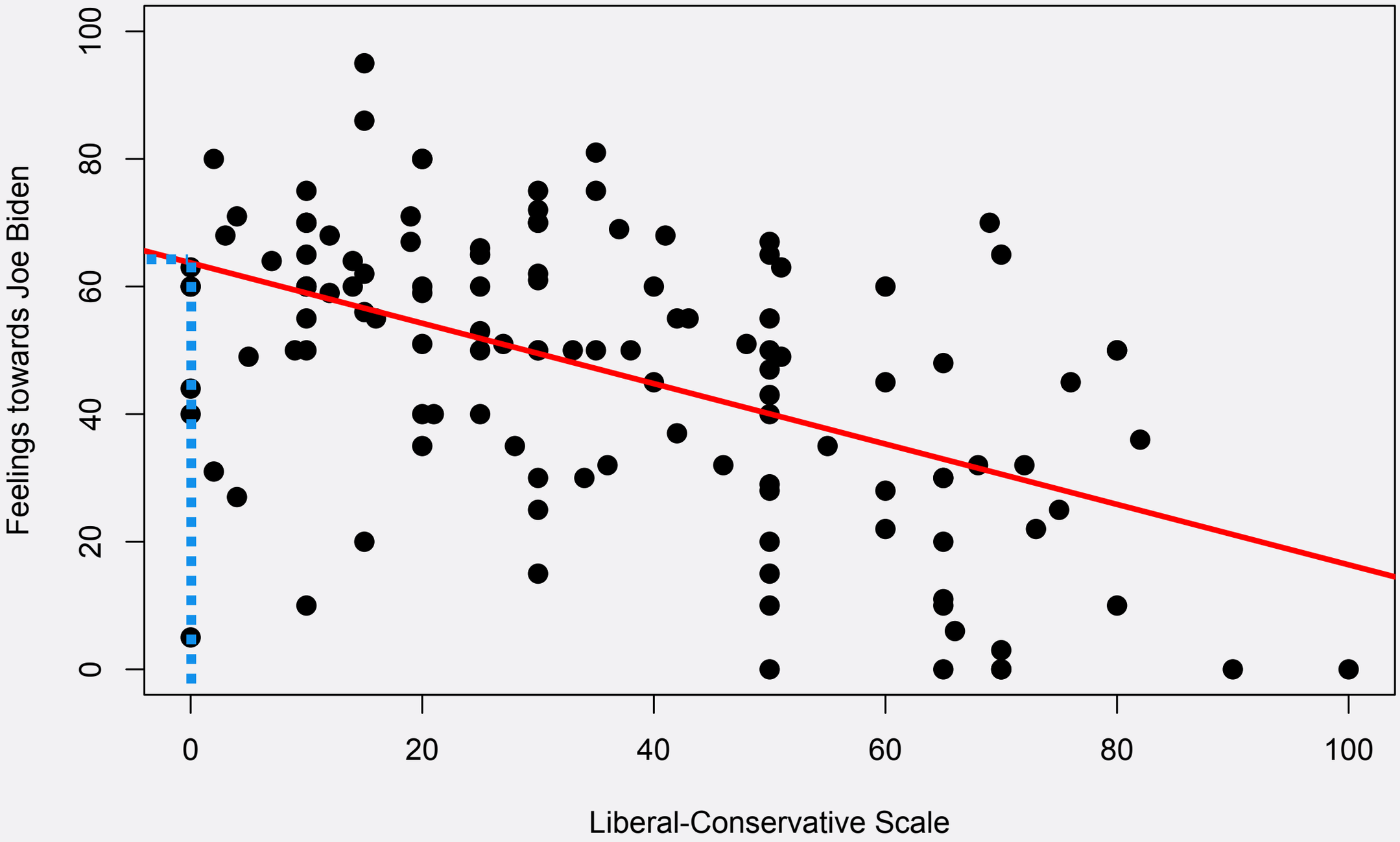
- Thermometer Score = 64 - 0.5 * Lib/Cons
- Can predict what someone's thermometer rating of Joe Biden will be, depending on where they are on liberal-conservative scale

WHAT THIS TELLS US

- Thermometer Score = 64 - 0.5 * Lib/Cons

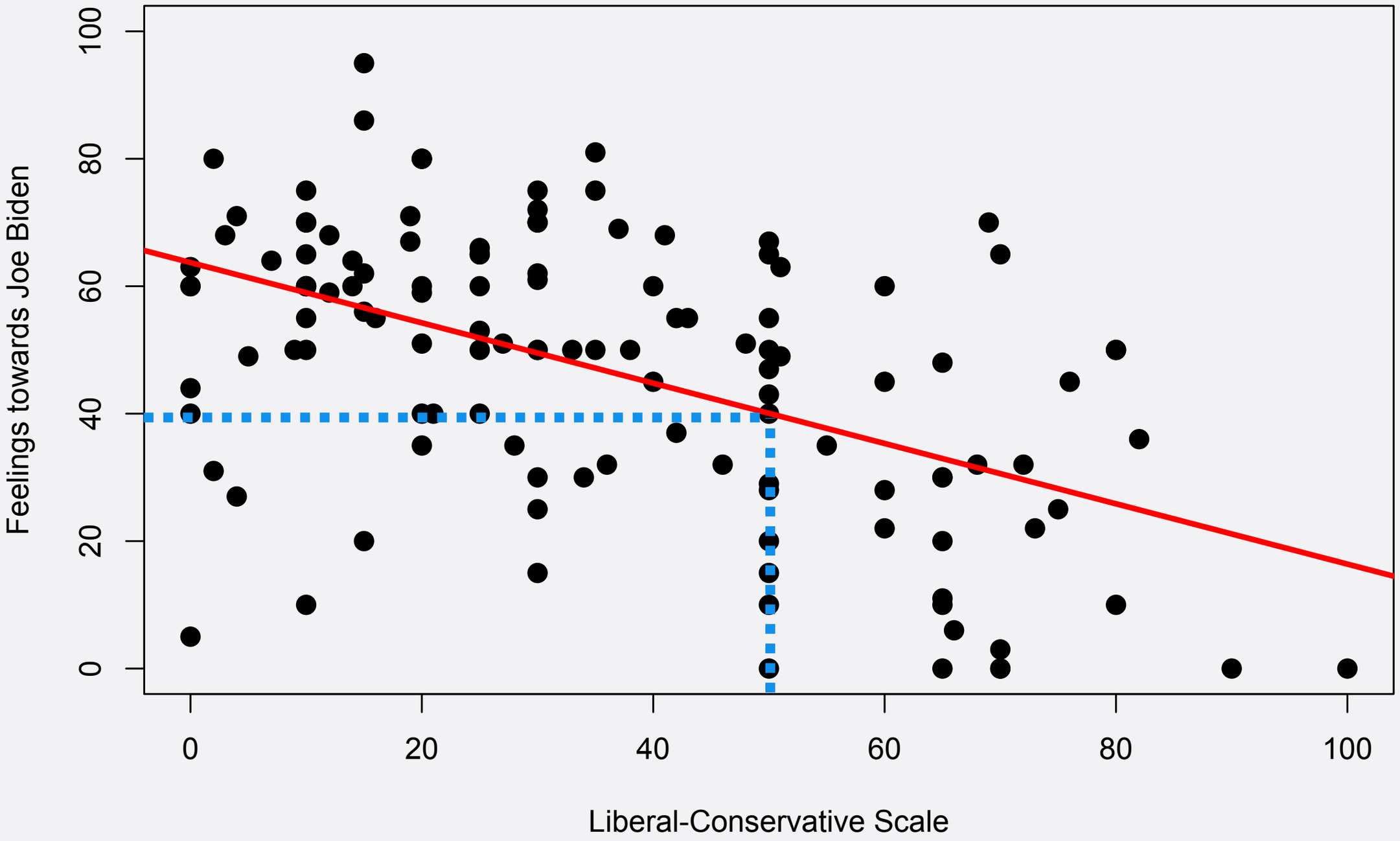
- Lib/Cons scale of 0:
  - 64 - 0.5 * 0 = 64

LINE



- 64 - 0.33 * 0 = 64

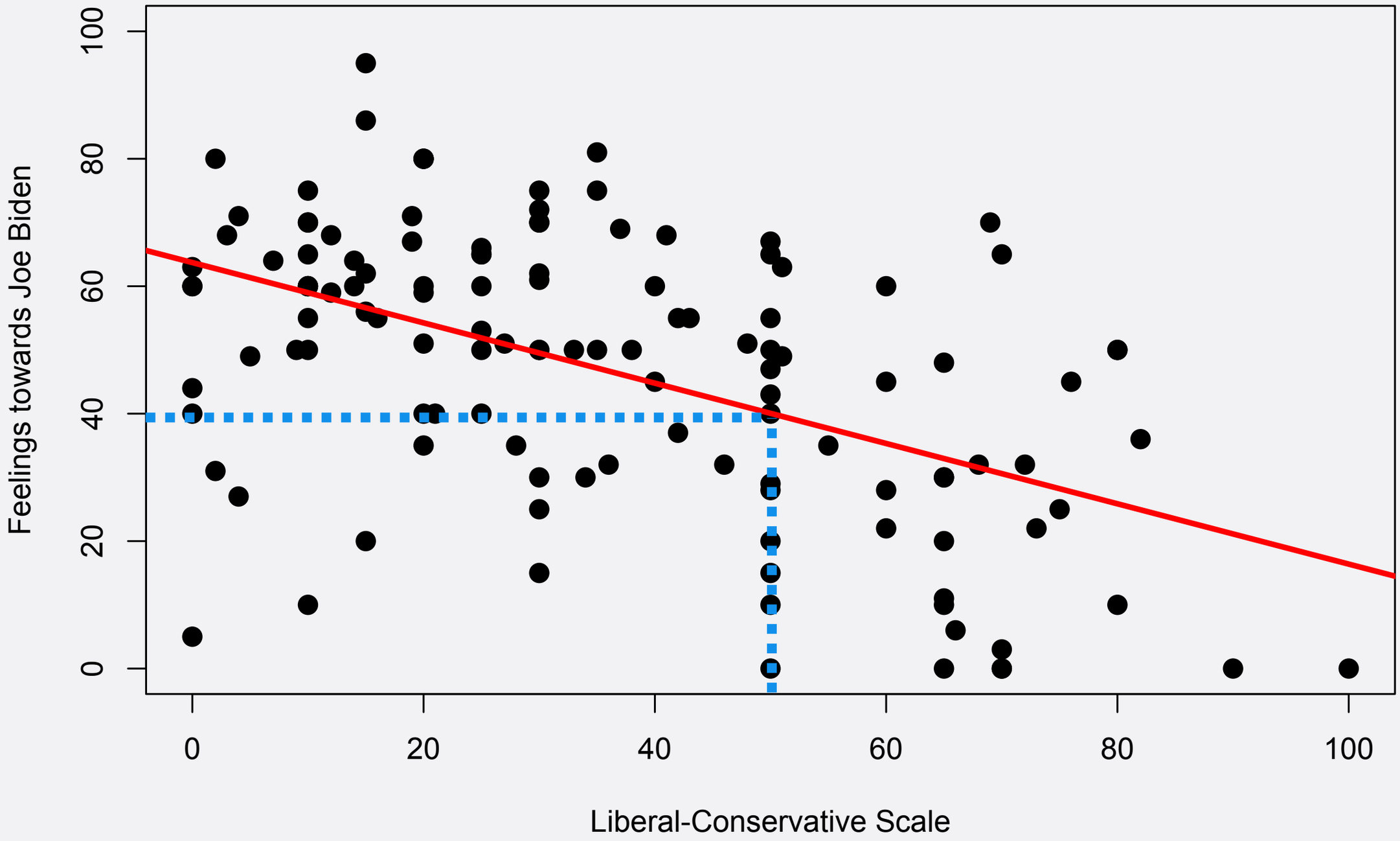
LINE


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

WHAT THIS TELLS US

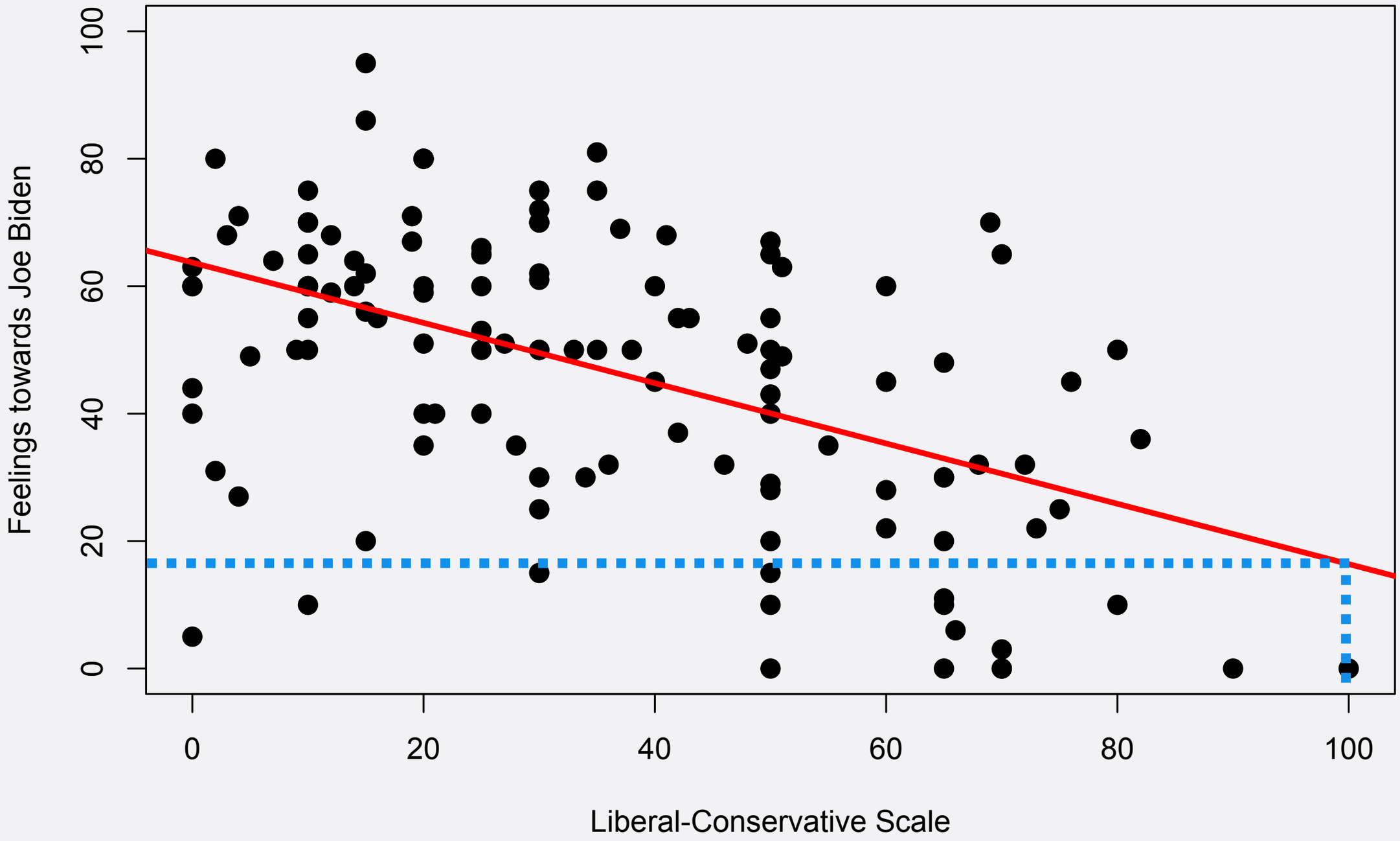
- Thermometer Score = 64 - 0.5 * Lib/Cons
- Lib/Cons scale of 50:
  - 64 - 0.5 * 50 = 39

LINE



- 64 - 0.5 * 50 = 39

LINE



- 64 - 0.5 * 100 = 14

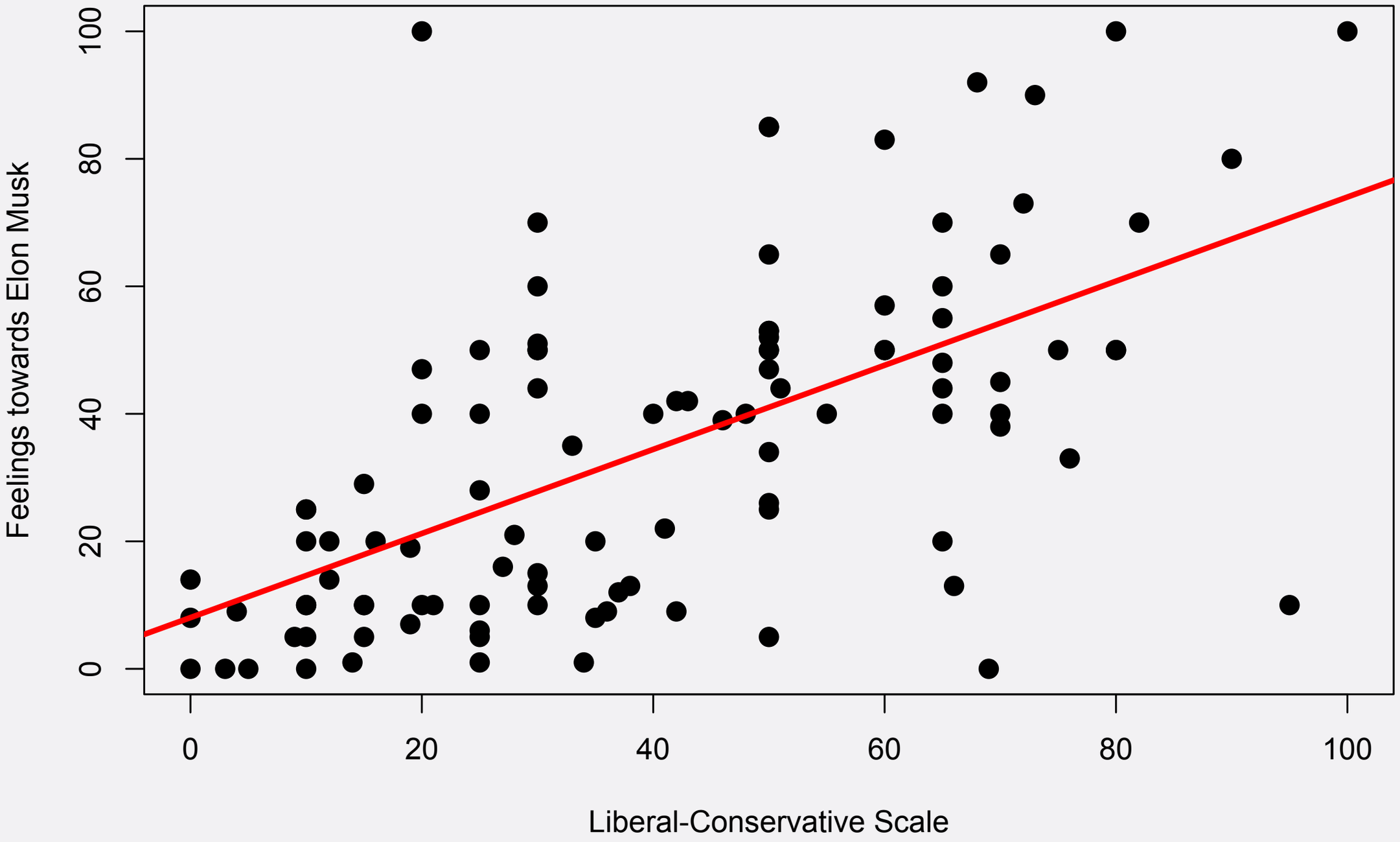
BIVARIATE RELATIONSHIPS

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

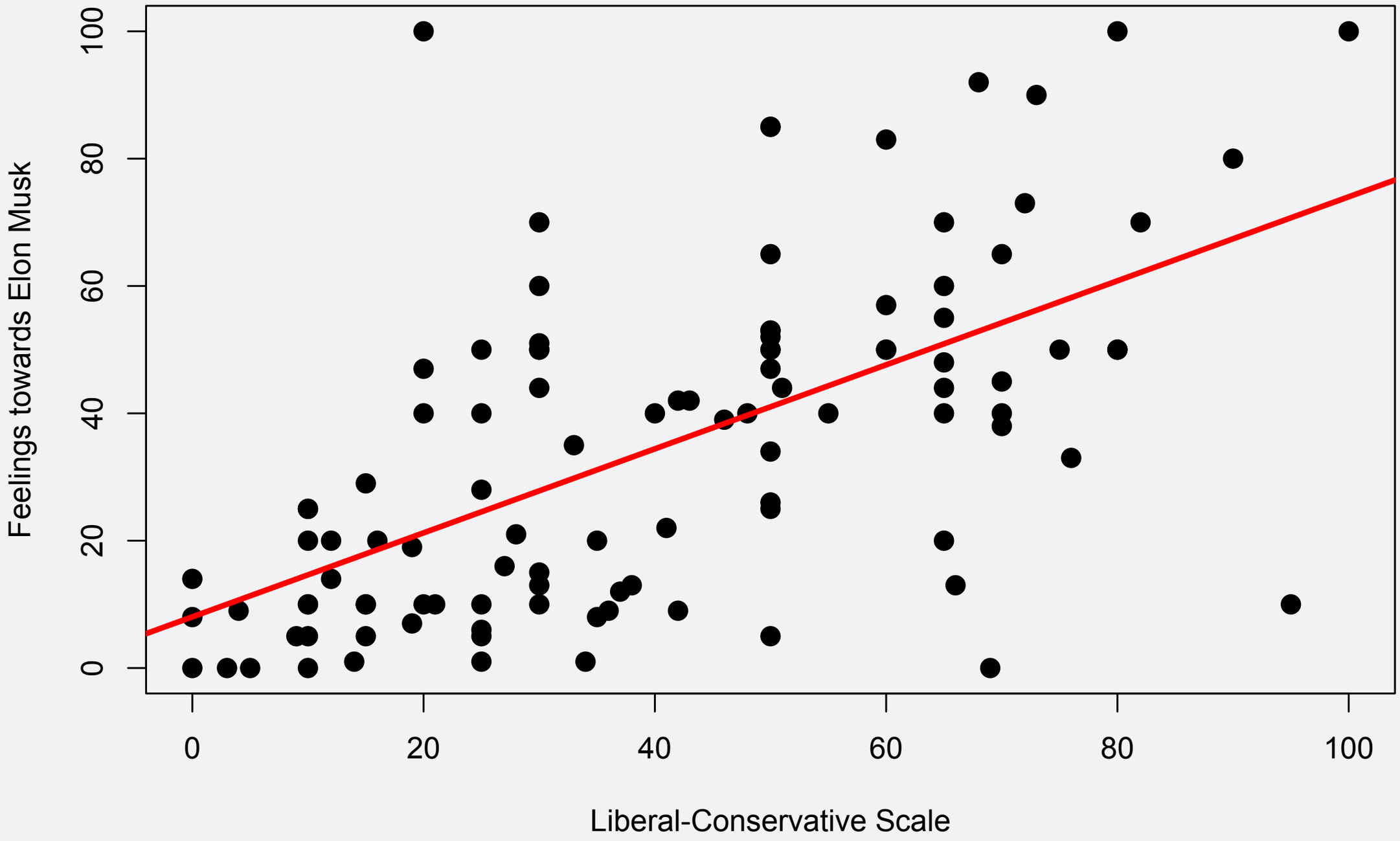
LINEAR REGRESSION

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

ELON MUSK


Feelings towards Elon Musk
0
20
40
60
80
100
0
20
40
60
80
100
Liberal-Conservative Scale

ELON MUSK



- Thermometer Score = 8 + 0.66 * Lib/Cons

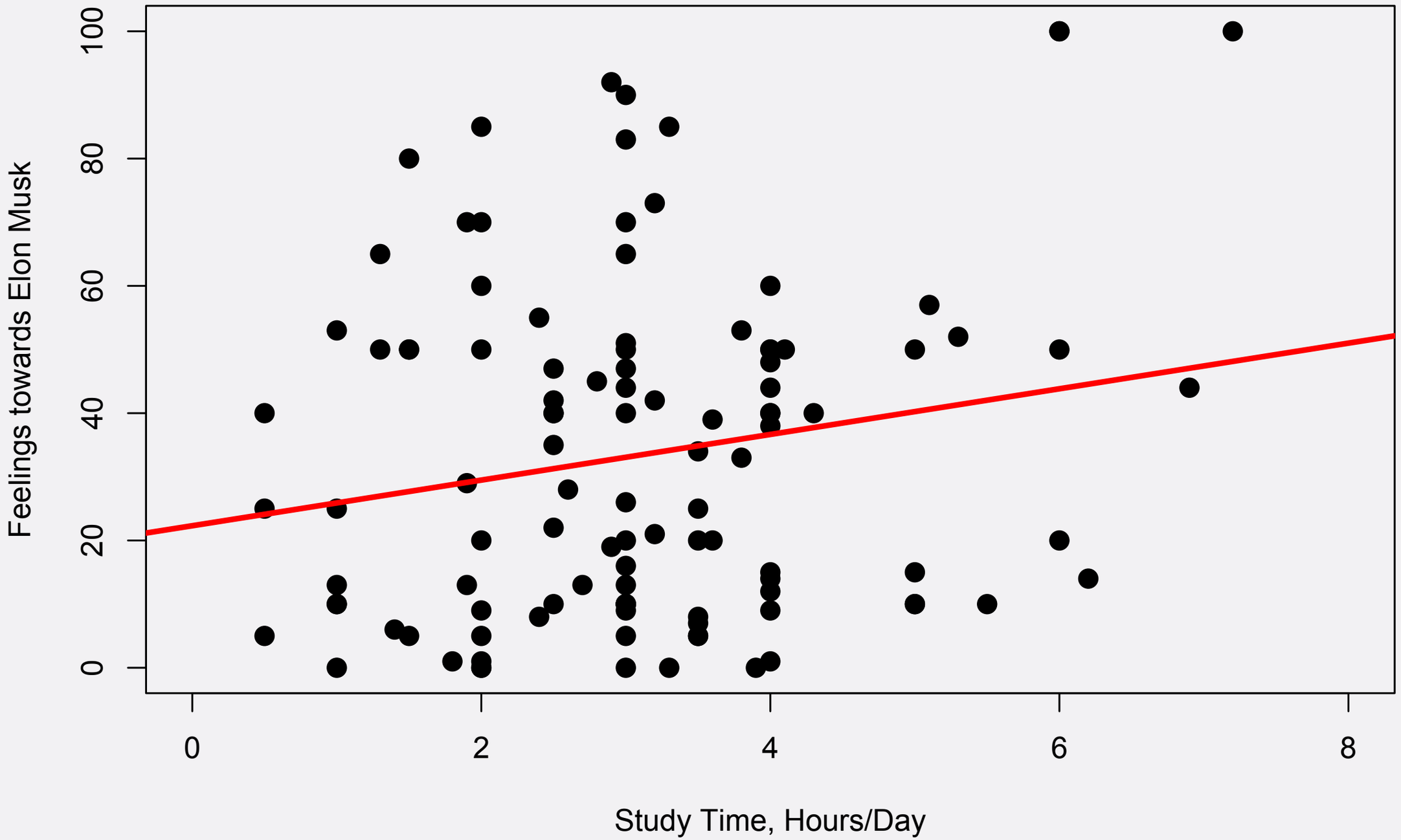
INTERPRETATION?

- Thermometer Score = 8 + 0.66 * Lib/Cons
- What does the 8 tell us?
- What does the 0.66 tell us?

INTERPRETATION?

- Thermometer Score = 8 + 0.66 * Lib/Cons
  - What does the 8 tell us?
    - A student who has a 0 on the liberal-conservative scale has an expected thermometer score of 8
  - What does the 0.66 tell us?
    - For every one point increase on the liberal-conservative scale, their thermometer score is expected to increase by 0.66 points

DIFFERENT INDEPENDENT VARIABLE



- Thermometer Score = 22 + 3.6 * Hours/Day

INTERPRETATION?

- Thermometer Score = 22 + 3.6 * Hours/Day
- What does the 22 tell us?
- What does the 3.6 tell us?

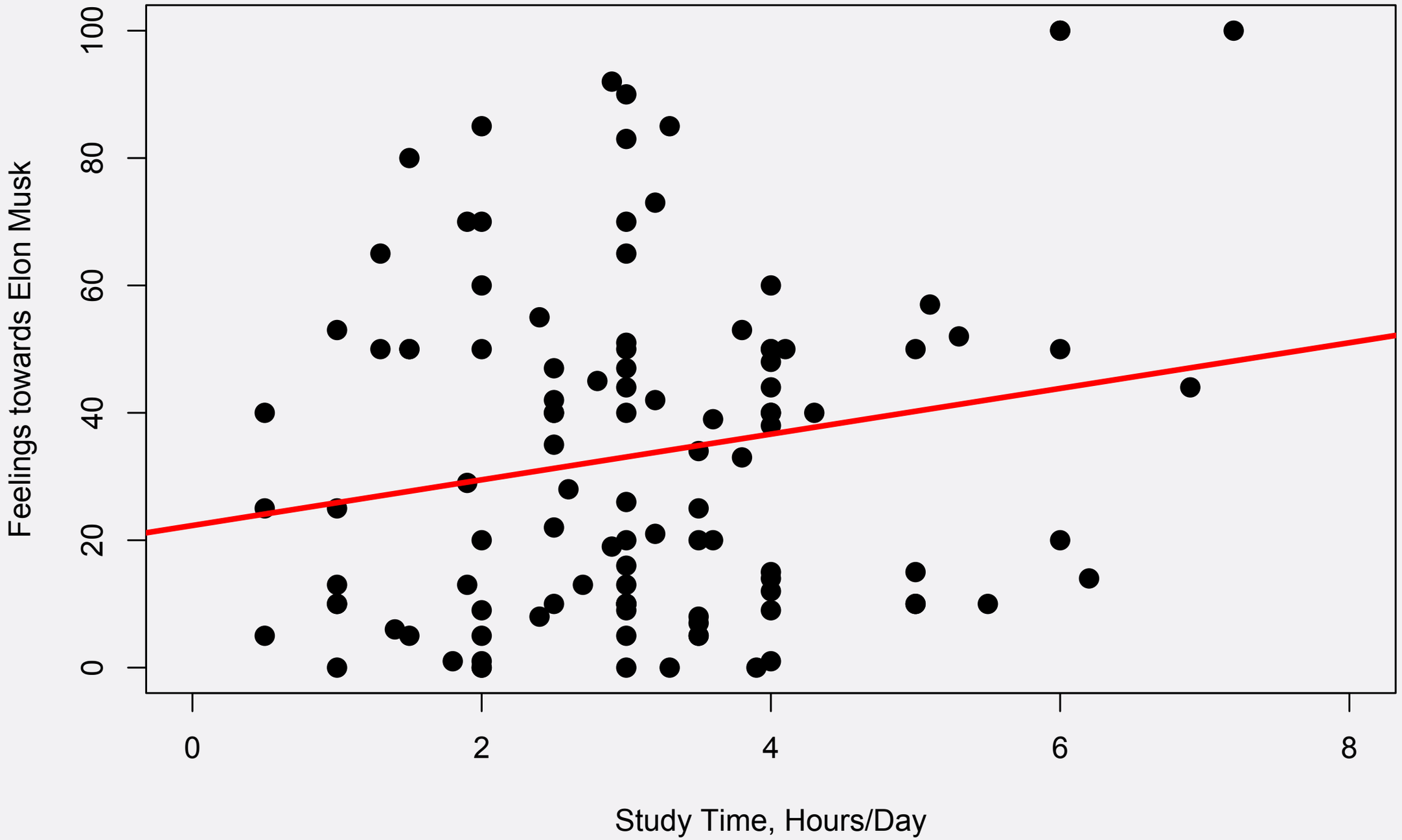
INTERPRETATION?

- Thermometer Score = 22 + 3.6 * Hours/Day
  - What does the 22 tell us?
    - A student who studies 0 hours per day has an expected thermometer score of 22
  - What does the 3.6 tell us?
    - For every one hour a student studies longer per day, their thermometer score is expected to increase by 3.6 points

INTERPRETATION?

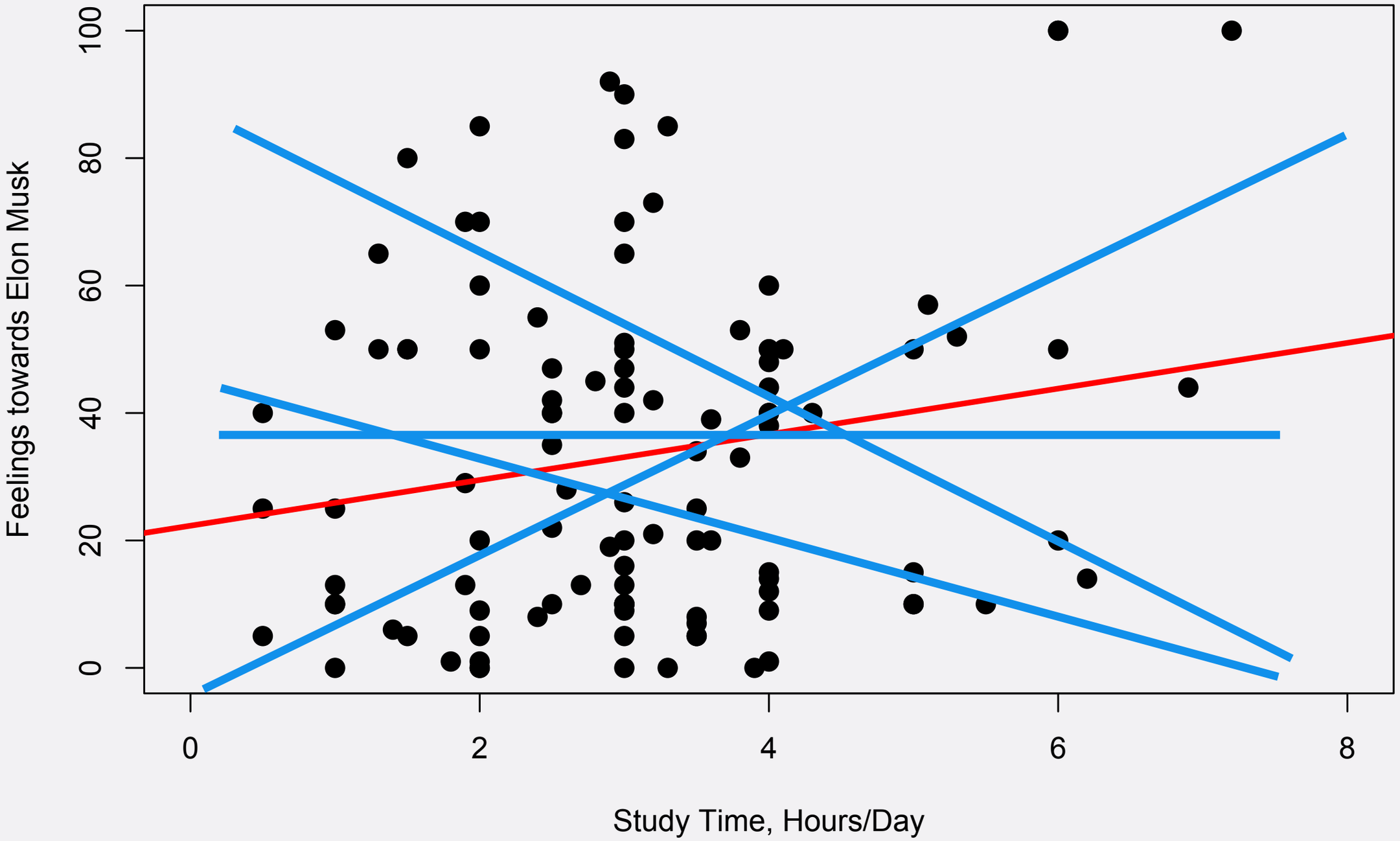
- Earlier we had:
  - Thermometer Score = 8 + 0.66 * Lib/Cons
- Now we have:
  - Thermometer Score = 22 + 3.6 * Hours/Day
- Does this mean that the effect of hours of study is larger than of how liberal-conservative students are?

HOW TO PICK THE LINE



- **Why this line?**

HOW TO PICK THE LINE



- **Why not these?**