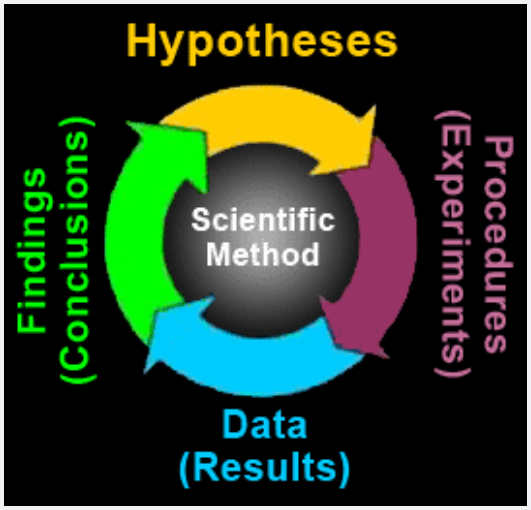
PSC 202
SYRACUSE UNIVERSITY

INTRODUCTION TO POLITICAL ANALYSIS

BIVARIATE HYPOTHESIS TESTING PART 1

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?

CAUSALITY

- What is the *causal* effect of attending Syracuse University on (future) income?
  - As opposed to attending e.g. a public university

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?

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?

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?

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?

CAUSALITY

- Income if attending SU - Income if attending public university
  - C*ausal* effect of attending Syracuse University
  - Problem?

CAUSALITY

- Income if attending SU - Income if attending public university
  - Can either observe person's income after attending SU

CAUSALITY

- Income if attending SU - Income if attending public university
  - Or person's income after attending public university

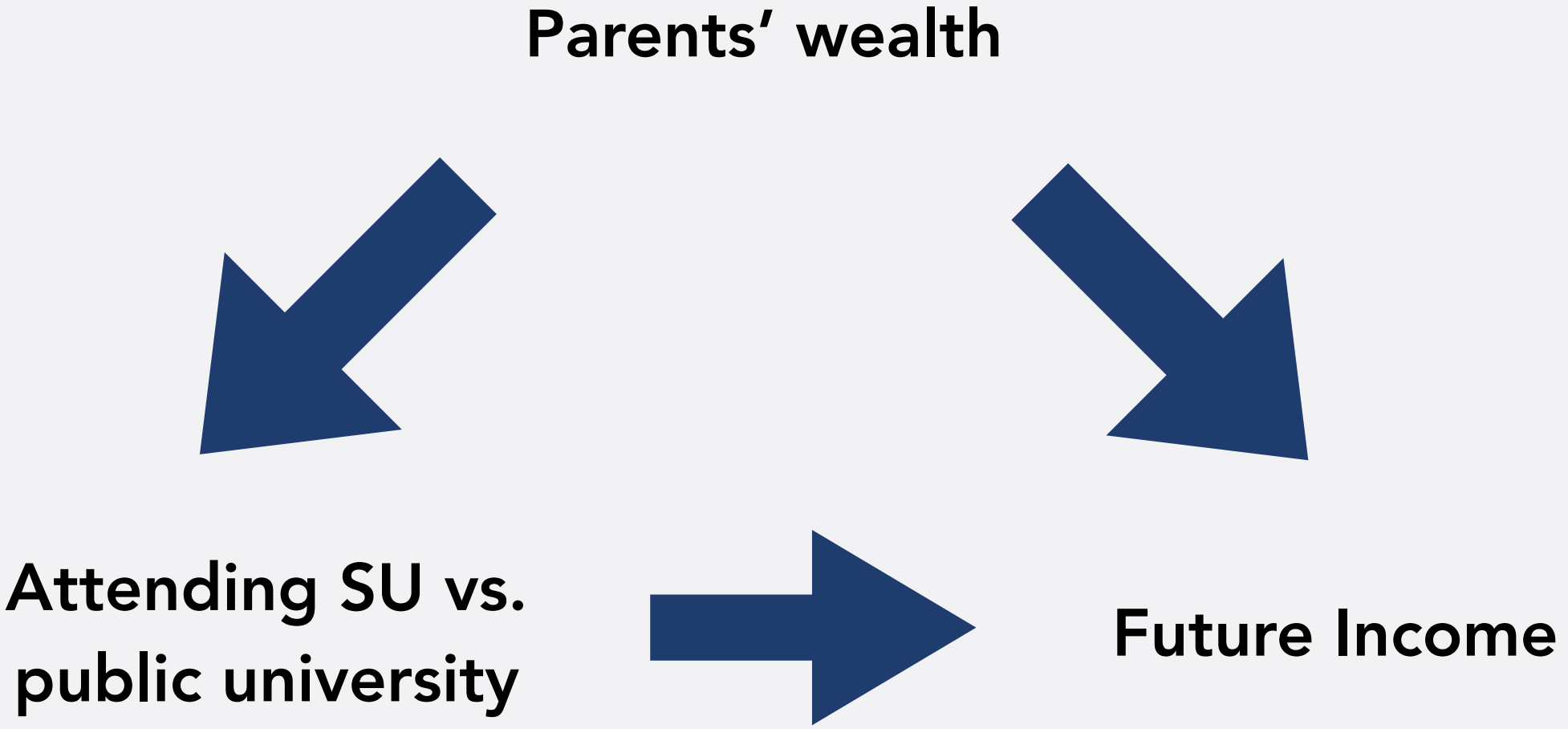
CAUSALITY

- Income if attending SU - Income if attending public university
  - But not both!
  - “Fundamental problem of causal inference”: We can’t observe alternate reality in which you didn’t attend SU!

CAUSALITY

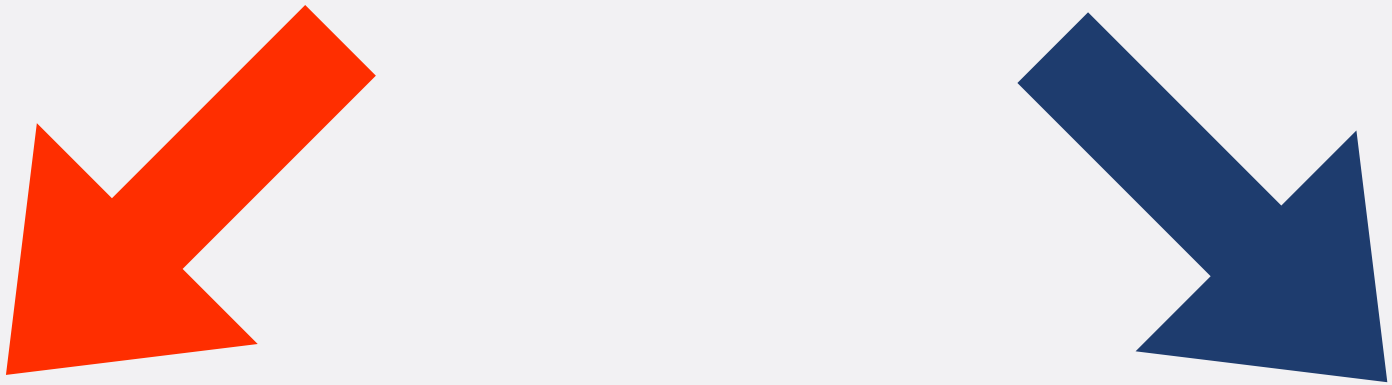
- Income of people attending SU - Income of people attending public university
  - This we can compute
  - But: Students who *choose* to attend SU are likely different from students who *choose* to attend public university
  - These differences potentially affect our ability to compute the causal effect of attending SU

HURDLE 4


Parents' wealth
Attending SU vs. public university
Future Income

HURDLE 4

Parents' wealth



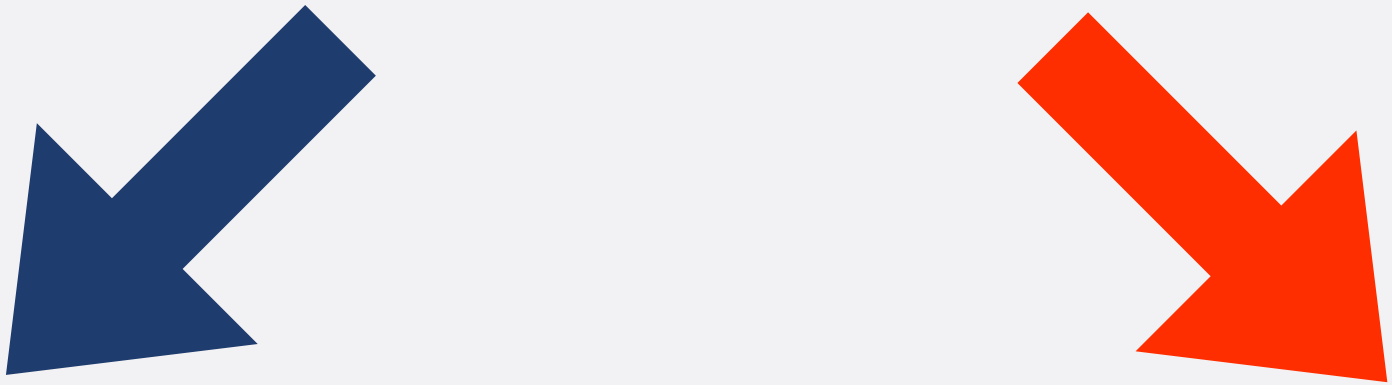
Attending SU vs. public university

Future Income

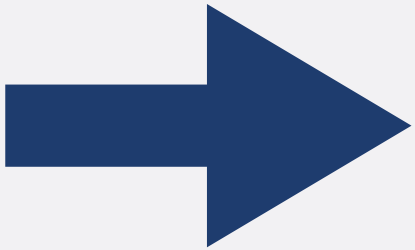
- SU is more expensive than public university
  - Students with rich parents more likely to attend SU
  - Students without rich parents more likely to attend public university

HURDLE 4

Parents' wealth



Attending SU vs. public university



Future Income

- Getting a well-paying job is (in part) about connections
  - Students with rich parents have better connections to companies with well-paying jobs
  - Students without rich parents have fewer connections

HURDLE 4

- So: income differences can be due to:
  - Causal effect of SU, and/or
  - Differences in parents' wealth
- If we want to estimate causal effect of SU, we need to "control for" differences in parents' wealth
  - Want to compare students who did and did not attend SU with similar parental wealth

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?

LARGE N AND SMALL N

- Qualitative studies (small n)
- Quantitative studies (large n)

HURDLE 3

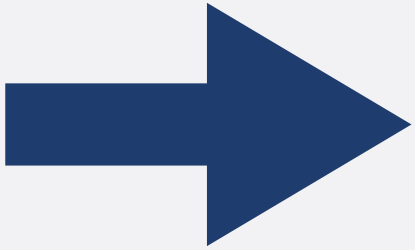
- Is there a credible causal mechanism that connects X to Y?
- Can we rule out the possibility that Y could cause X?
- Is there covariation between X and Y?
- Have we controlled for all confounding variables (Z) that might make the association between X and Y spurious?

SURVEY

- Do you approve or disapprove of the way Joe Biden is handling his job as President?

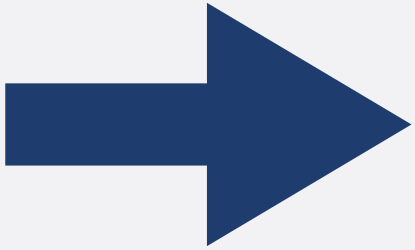
- Excluding students who said “Don’t know”

BIVARIATE RELATIONSHIP

?  Approval of J. Biden

- What explains why some of you approve, while others don't?

BIVARIATE RELATIONSHIP

Gender  Approval of J. Biden

- If gender has an effect on approval, what would we expect to see?
- How could we show it?

BIVARIATE RELATIONSHIP

- Male
  - Approve: 11
  - Do not approve: 14
- Female
  - Approve: 22
  - Do not approve: 20

- Excluding students who said “Don’t know”

CROSS-TABULATIONS

Approve of Biden	Gender: Male	Gender: Female
Approve	11	22
Do Not Approve	14	20

CROSS-TABULATIONS

Approve of Biden	Gender: Male	Gender: Female	Total
Approve	11	22	33
Do Not Approve	14	20	34
Total	25	42	67

CROSS-TABULATIONS

Gender

Approve of Biden	Male	Female	Total
Approve	11	22	33
Do Not Approve	14	20	34
Total	25	42	67

Total number approving

CROSS-TABULATIONS

Gender

Approve of Biden	Male	Female	Total
Approve	11	22	33
Do Not Approve	14	20	34
Total	25	42	67

Total number Men

CROSS-TABULATIONS

Gender

Approve of Biden	Male	Female	Total
Approve	11	22	33
Do Not Approve	14	20	34
Total	25	42	67

Total number

CROSS-TABULATIONS

Approve of Biden	Gender: Male	Gender: Female	Total
Approve	11	22	33
Do Not Approve	14	20	34
Total	25	42	67

CROSS-TABULATIONS

Approve of Biden	Gender: Male	Gender: Female	Total
Approve	% (11)	% (22)	% (33)
Do Not Approve	% (14)	% (20)	% (34)
Total	% (25)	% (42)	100% (67)

CROSS-TABULATIONS

Gender

Approve of Biden	Male	Female	Total
Approve	44.0% (11)	% (22)	% (33)
Do Not Approve	56.0% (14)	% (20)	% (34)
Total	100% (25)	% (42)	100% (67)

CROSS-TABULATIONS

Approve of Biden	Gender: Male	Gender: Female	Total
Approve	44.0% (11)	52.4% (22)	% (33)
Do Not Approve	56.0% (14)	47.6% (20)	% (34)
Total	100% (25)	100% (42)	100% (67)

CROSS-TABULATIONS

Gender

Approve of Biden	Male	Female	Total
Approve	44.0% (11)	52.4% (22)	49.2% (33)
Do Not Approve	56.0% (14)	47.6% (20)	50.8% (34)
Total	100% (25)	100% (42)	100% (67)

CROSS-TABULATIONS

Approve of Biden	Gender: Male	Gender: Female	Total
Approve	44.0% (11)	52.4% (22)	49.2% (33)
Do Not Approve	56.0% (14)	47.6% (20)	50.8% (34)
Total	100% (25)	100% (42)	100% (67)

% Men who approve

CROSS-TABULATIONS

Gender

Approve of Biden	Male	Female	Total
Approve	44.0% (11)	52.4% (22)	49.2% (33)
Do Not Approve	56.0% (14)	47.6% (20)	50.8% (34)
Total	100% (25)	100% (42)	100% (67)

% Women who approve

CROSS-TABULATIONS

Gender

Approve of Biden	Male	Female	Total
			% who approve
Approve	44.0% (11)	52.4% (22)	49.2% (33)
Do Not Approve	56.0% (14)	47.6% (20)	50.8% (34)
Total	100% (25)	100% (42)	100% (67)

TEMPLATE

Dependent Variable	Independent Variable: I V Value 1	Independent Variable: I V Value 2	Total
D V Value 1	% In Column (# Cases)	% In Column (# Cases)	% Of Total (# In Row)
D V Value 2	% In Column (# Cases)	% In Column (# Cases)	% Of Total (# In Row)
Total	100% (# In Column)	100% (# In Column)	100% (# Total)

CROSS-TABULATIONS

Gender

Approve of Biden	Male	Female	Total
Approve	44.0% (11)	52.4% (22)	49.2% (33)
Do Not Approve	56.0% (14)	47.6% (20)	50.8% (34)
Total	100% (25)	100% (42)	100% (67)

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 8.4 percentage points more likely to approve of Biden than men (52.4% vs. 44.0%)

EXERCISE

- In fraternity/sorority
  - Approve: 14
  - Do not approve: 11
- Not in fraternity/sorority
  - Approve: 19
  - Do not approve: 23

- What is the zero-order relationship between being in a fraternity/sorority and Biden approval?

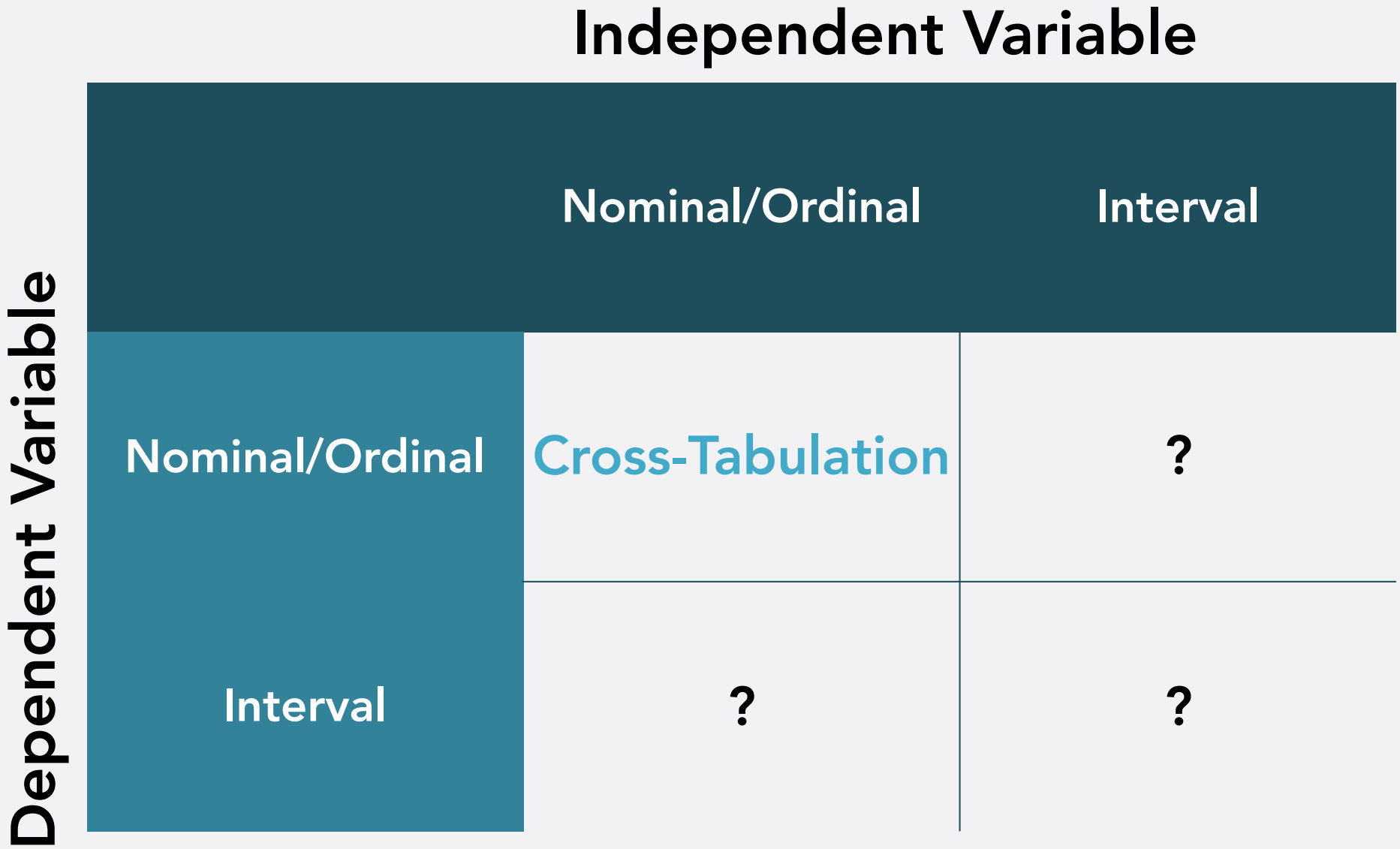
CROSS-TABULATIONS

Sorority/Fraternity

Approve of Biden	Member	Not A Member	Total
Yes	56.0% (14)	45.2% (19)	49.2% (33)
No	44.0% (11)	54.8% (23)	50.8% (34)
Total	100% 25	100% (42)	100% (67)

- Zero-order relationship: Greek members are 10.8 percentage points more likely to approve of Biden than non-members

BIVARIATE RELATIONSHIPS



BIVARIATE RELATIONSHIPS

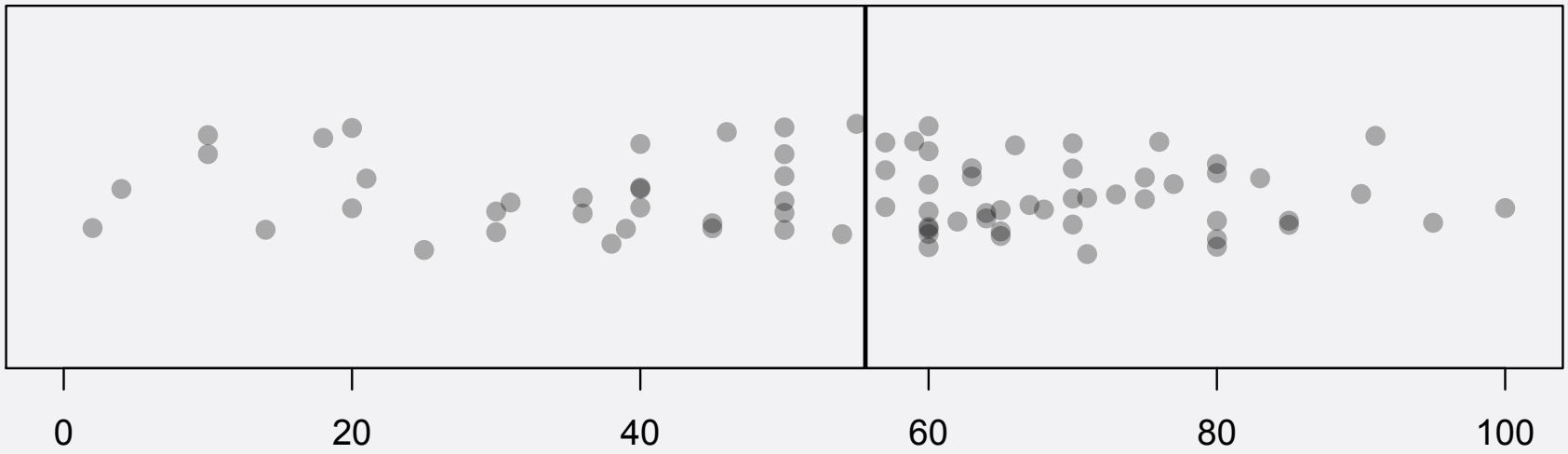
Dependent Variable \ Independent Variable	Nominal/Ordinal	Interval
Nominal/Ordinal	Cross-Tabulation	Not In This Class…
Interval	?	?

BIVARIATE RELATIONSHIPS

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

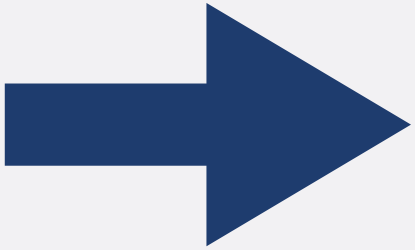
SURVEY

- Feelings towards the Democratic Party



BIVARIATE RELATIONSHIP

?

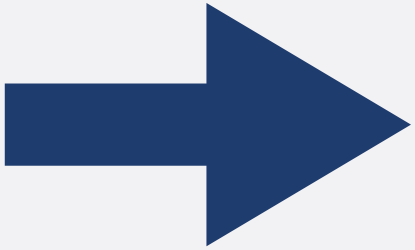


Feelings towards Democratic Party

- What explains variation in feelings towards the Democratic Party?

BIVARIATE RELATIONSHIP

Gender



Feelings towards Democratic Party

- If gender has an effect on feelings towards Democratic Party, what would we expect to see?
- How could we show it?

DEMOCRATIC PARTY

	Mean Thermometer Score	Frequency
Female	57.9	54
Male	50.0	27
Total	55.6	81

ZERO-ORDER RELATIONSHIP

	Mean Thermometer Score	Frequency
Female	57.9	54
Male	50.0	27
Total	55.6	81

7.9

ZERO-ORDER RELATIONSHIP

- There is covariation between gender and feelings towards Democratic Party
  - Women's feelings towards the party are on average 7.9 points higher than men's

MEAN COMPARISON TABLE

	Average of DV	Frequency
IV Value 1	Mean of DV for IV Value 1	# Cases IV Value 1
IV Value 2	Mean of DV for IV Value 2	# Cases IV Value 2
Total	Mean of DV overall	# Cases overall

- DV: Dependent variable; IV: Independent variable

REPUBLICAN PARTY

	Mean Thermometer Score	Frequency
Female	20.6	54
Male	31.4	27
Total	24.2	81

-10.8

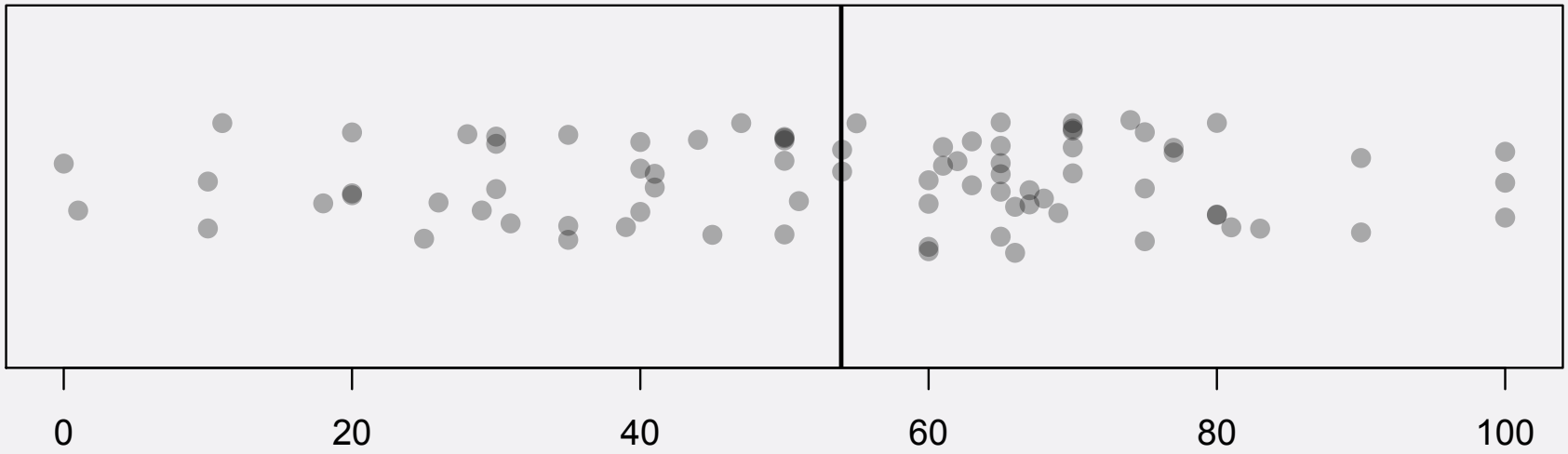
BIVARIATE RELATIONSHIPS

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

BIVARIATE RELATIONSHIPS

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

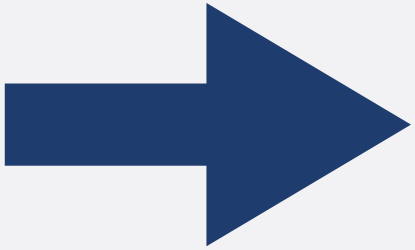
BIVARIATE RELATIONSHIP



?

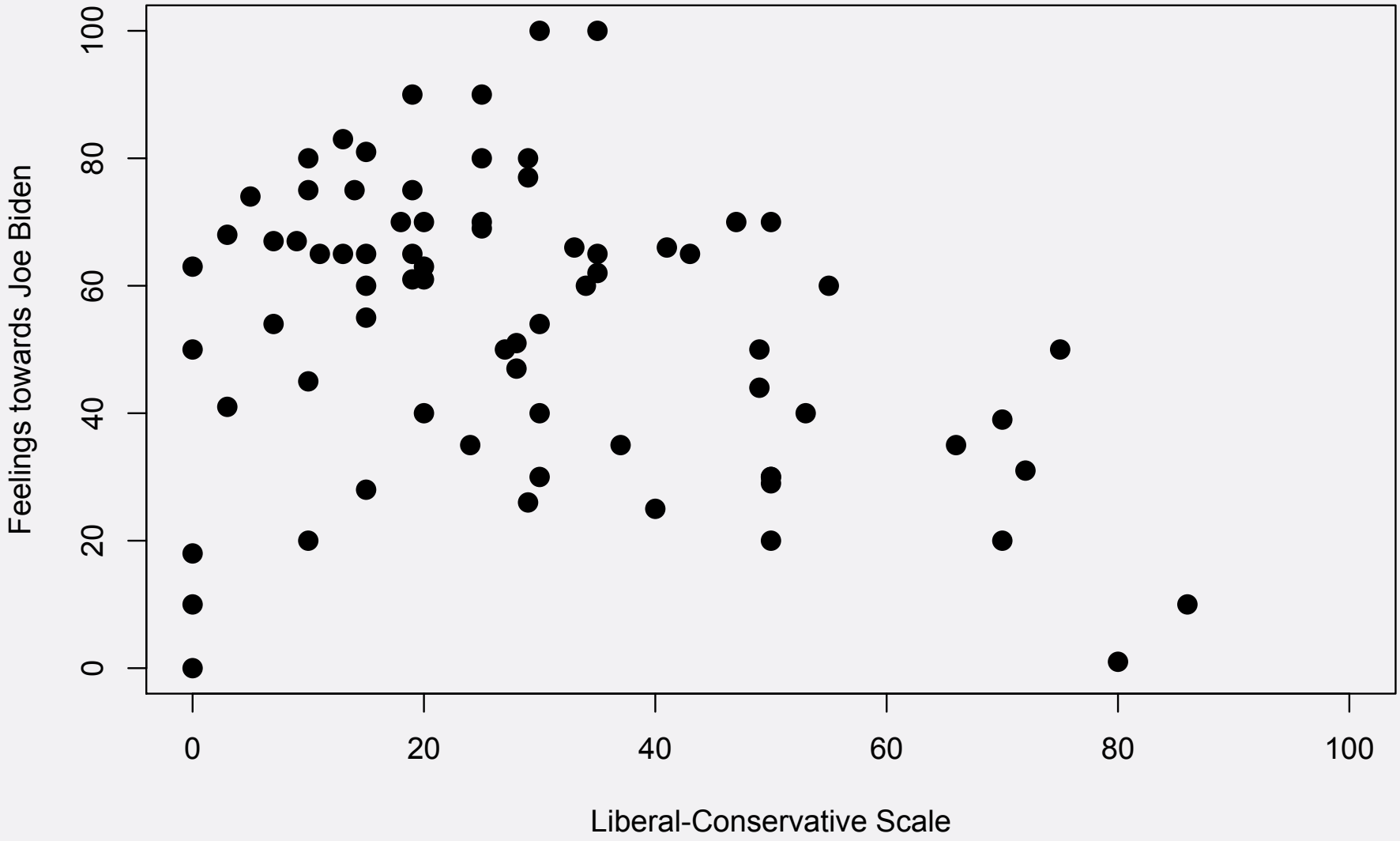
Feelings towards Joe Biden

BIVARIATE RELATIONSHIP

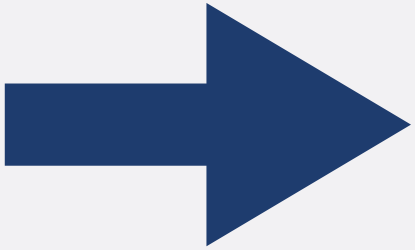
Liberal-Conservative  Feelings towards Joe Biden

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

JOE BIDEN

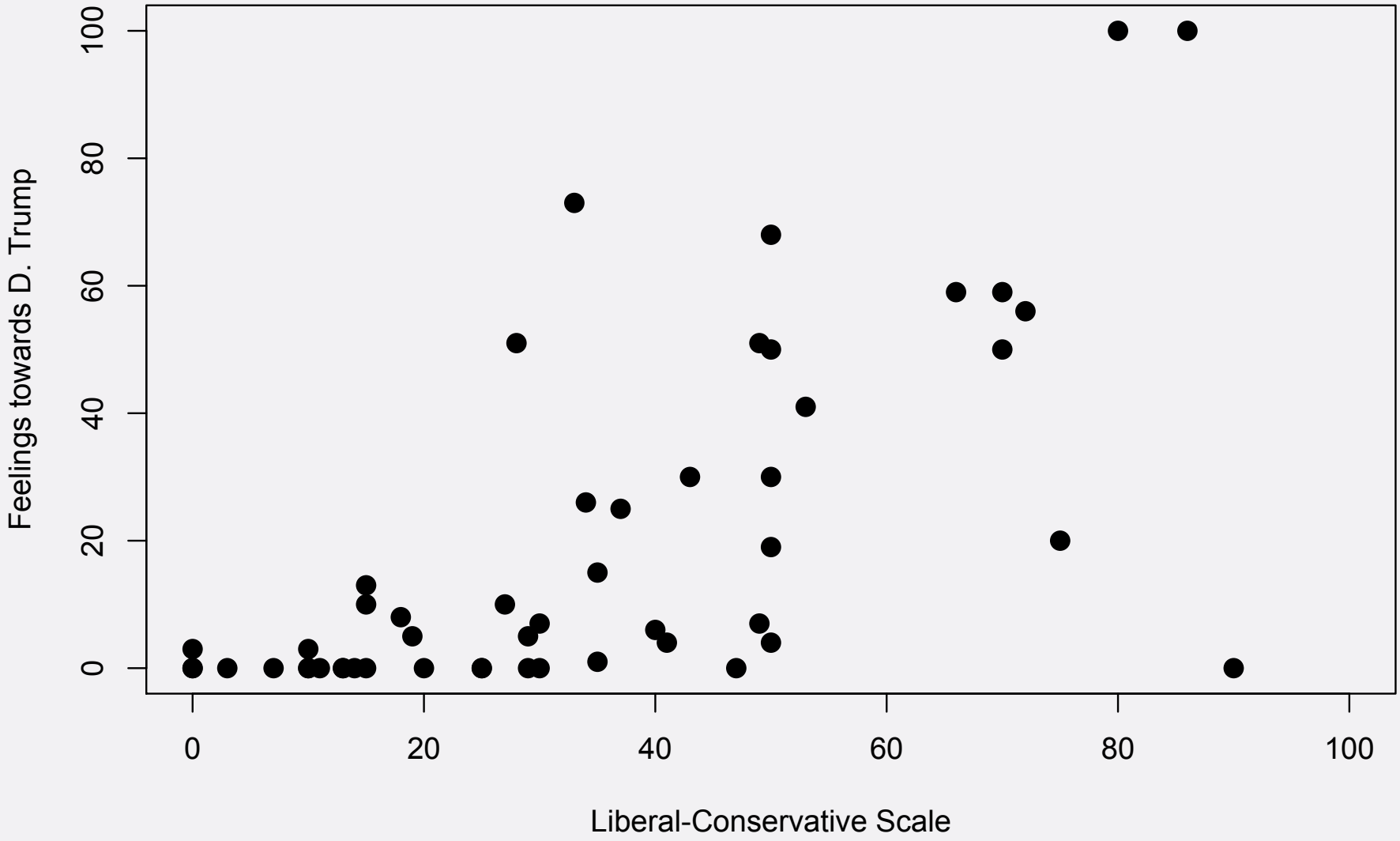

Feelings towards Joe Biden
0
20
40
60
80
100
0
20
40
60
80
100
Liberal-Conservative Scale

BIVARIATE RELATIONSHIP

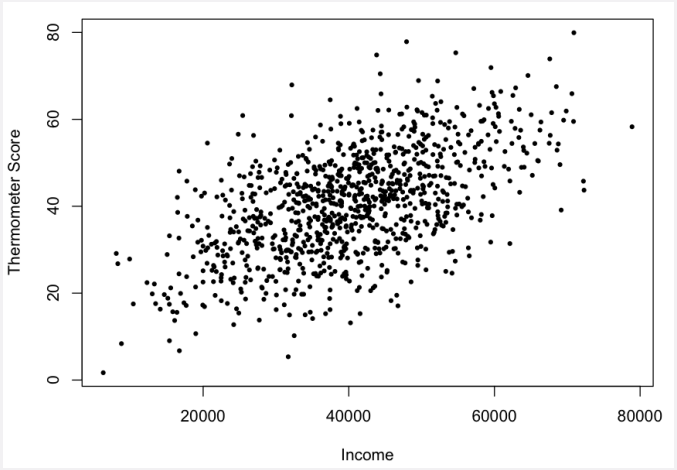
Liberal-Conservative  Feelings towards Donald Trump

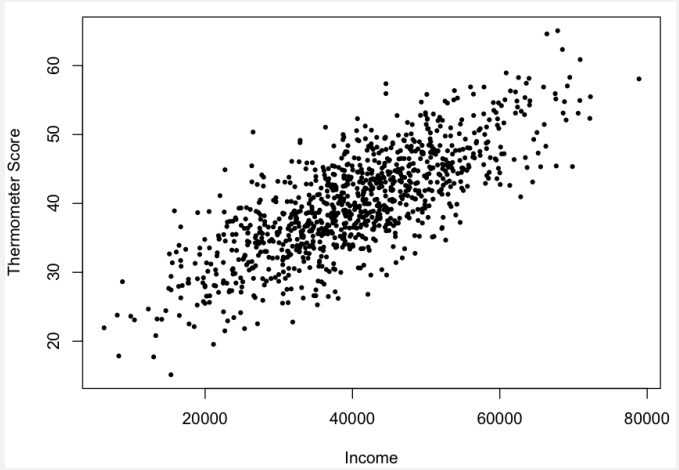
- DV: Feelings towards Donald Trump (0-100)
- IV: Liberal-Conservative (0-100)

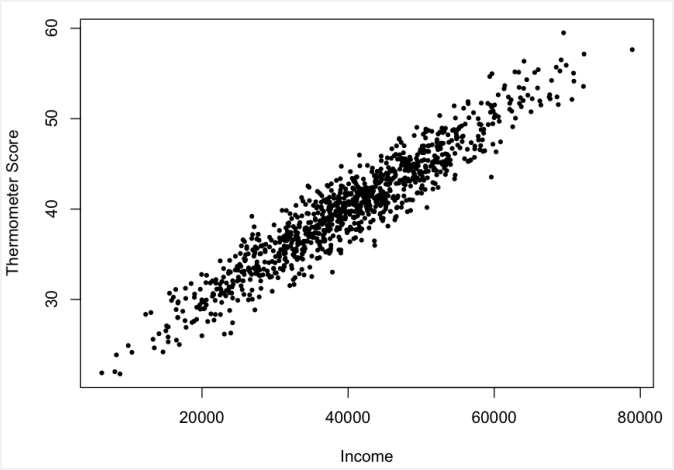
DONALD TRUMP

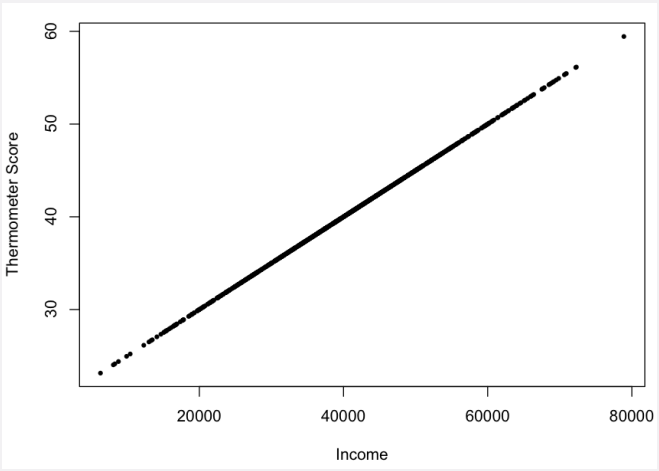


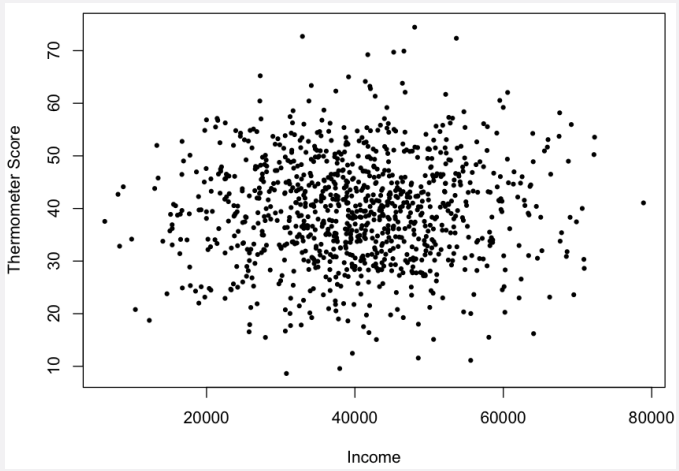
SCATTERPLOTS

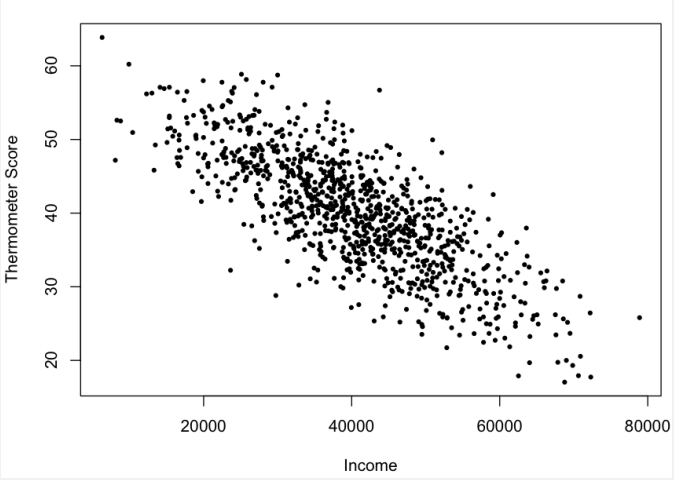












SCATTERPLOT

- 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”

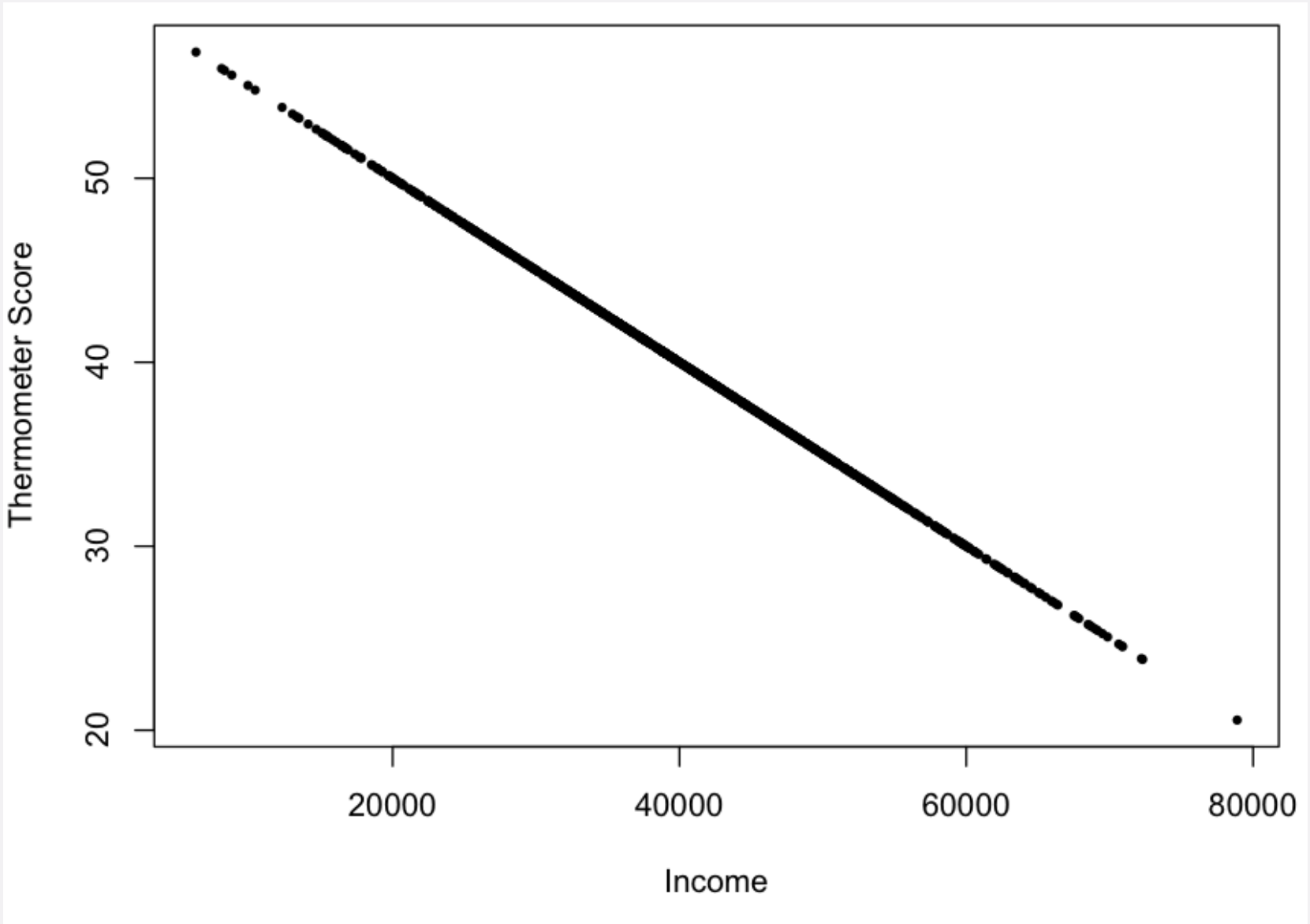
CORRELATION

- 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

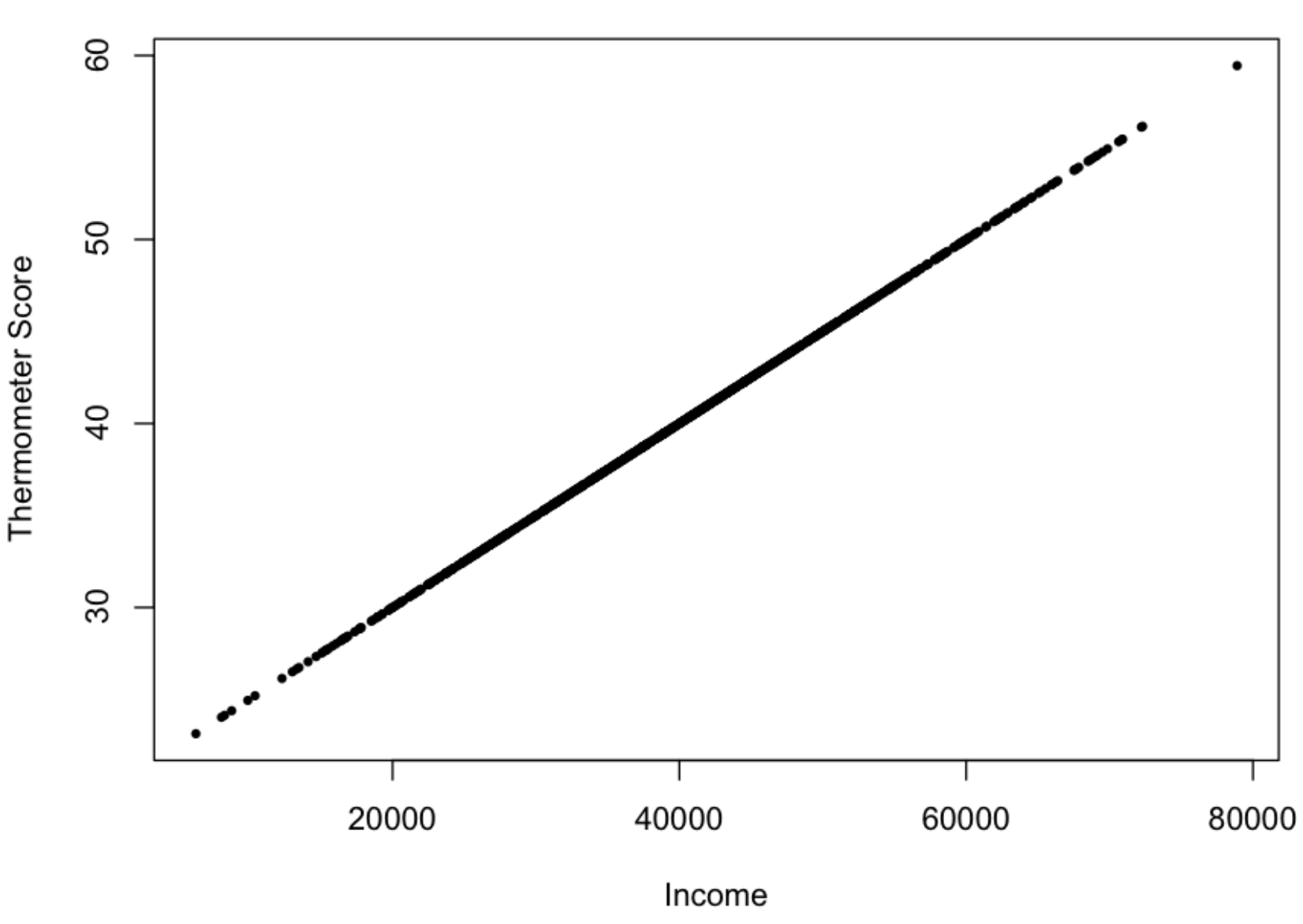
- If the coefficient is 0, no bivariate relationship exists
  - A positive coefficient means that a positive relationship exists
  - A negative coefficient means that a negative relationship exists
- Correlation of -1 or +1 means that relation between X and Y is perfectly linear

CORRELATION



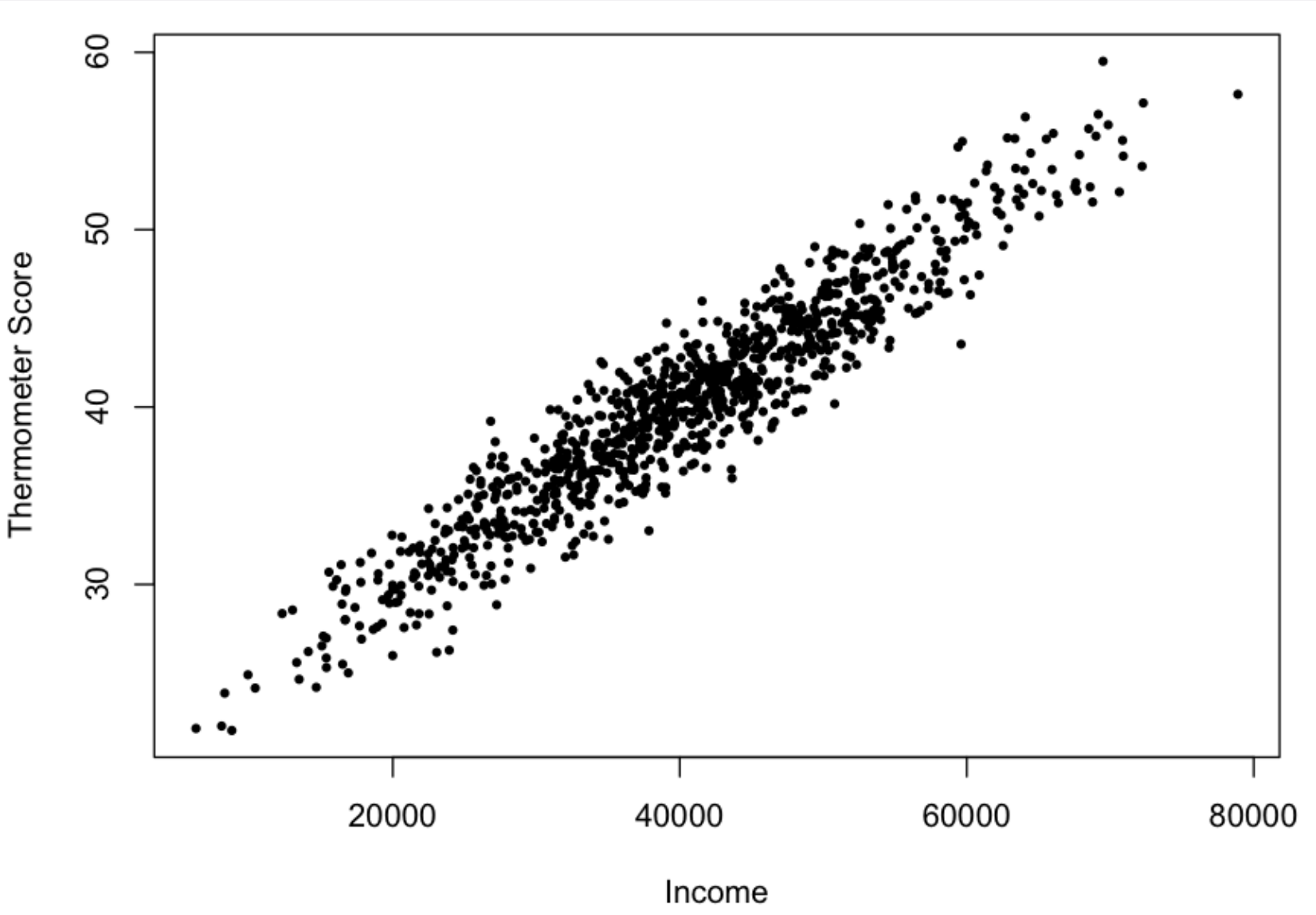
r=-1

CORRELATION



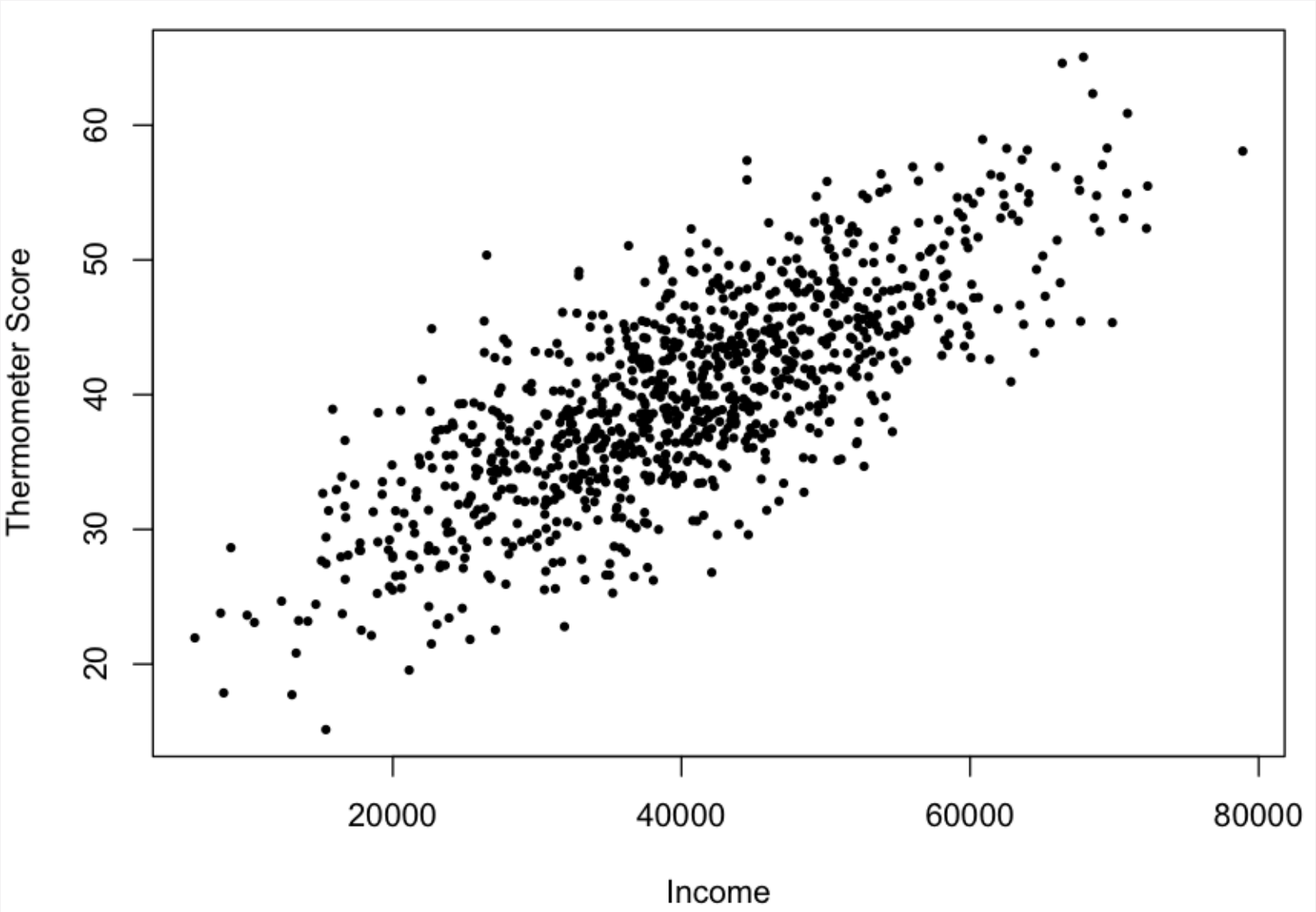
r=1

CORRELATION



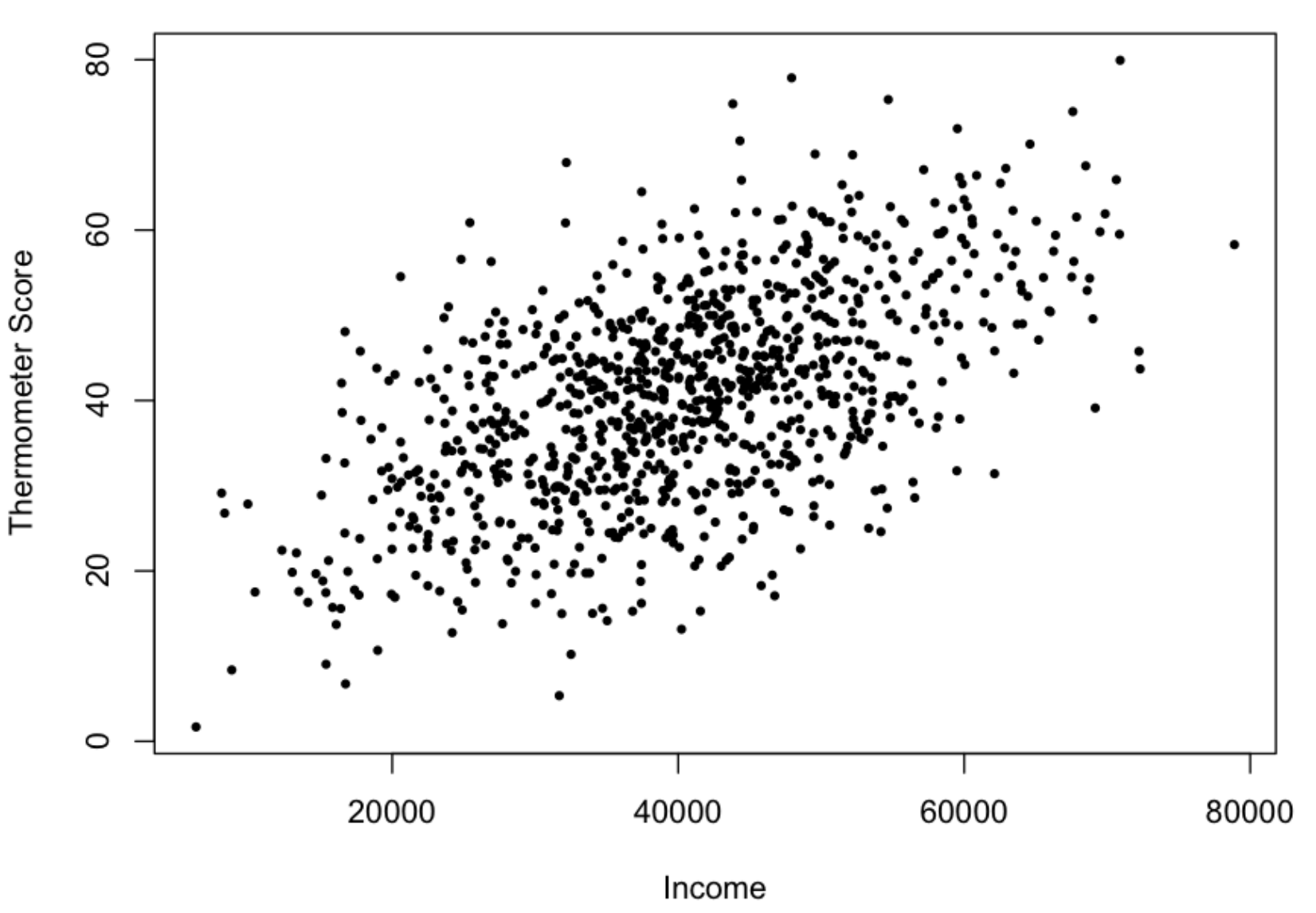
r=0.95

CORRELATION



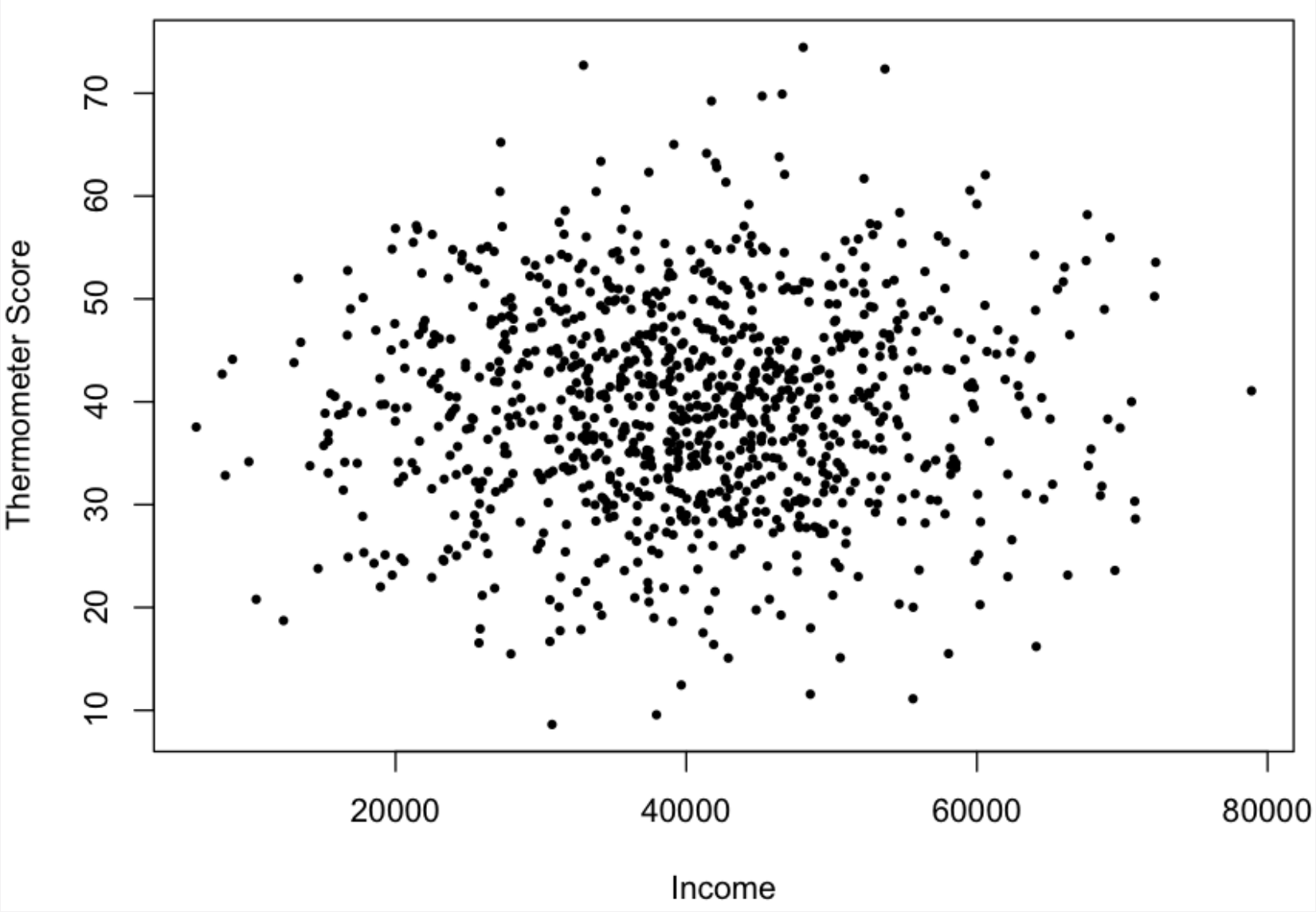
r=0.77

CORRELATION



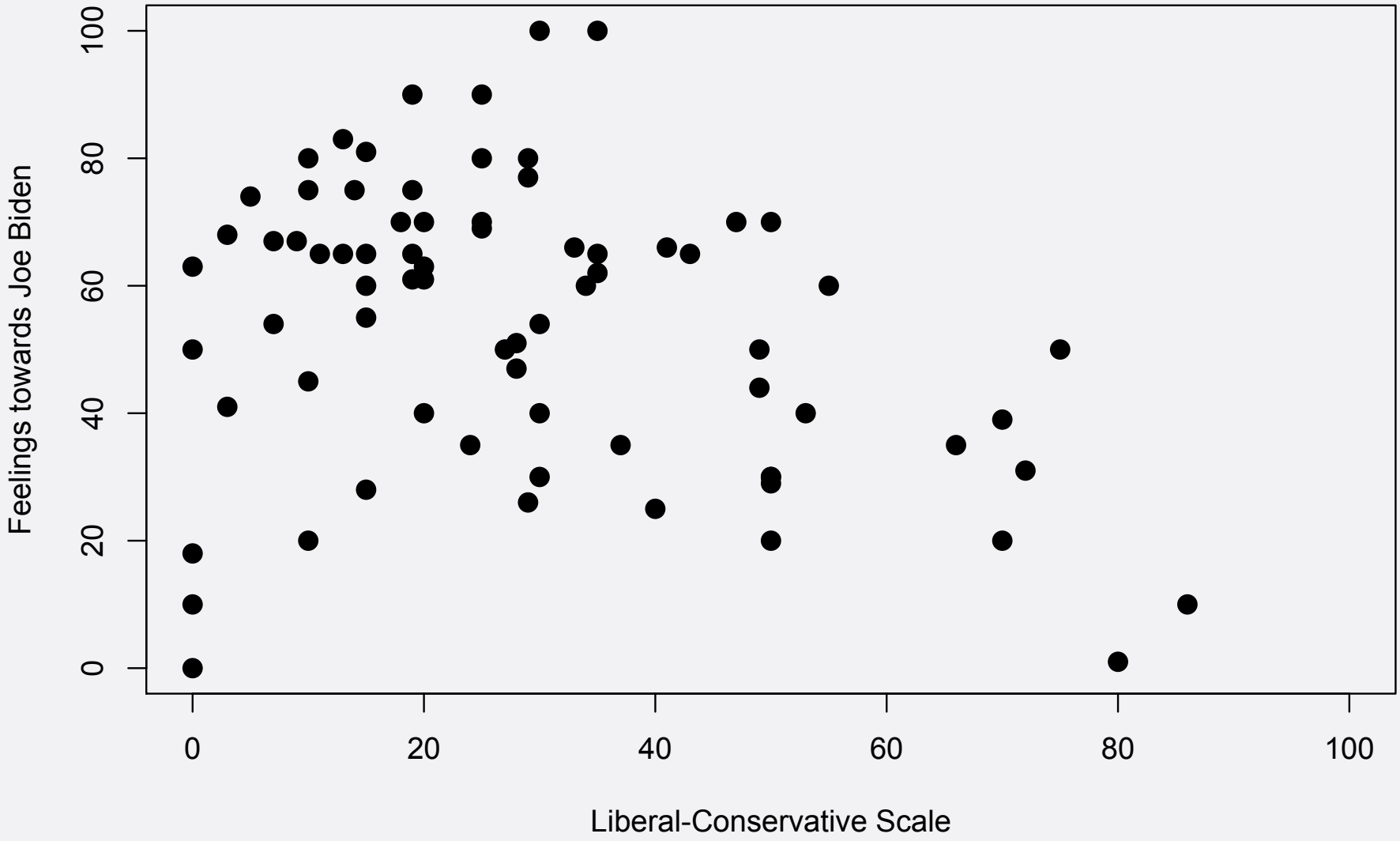
r=0.53

CORRELATION



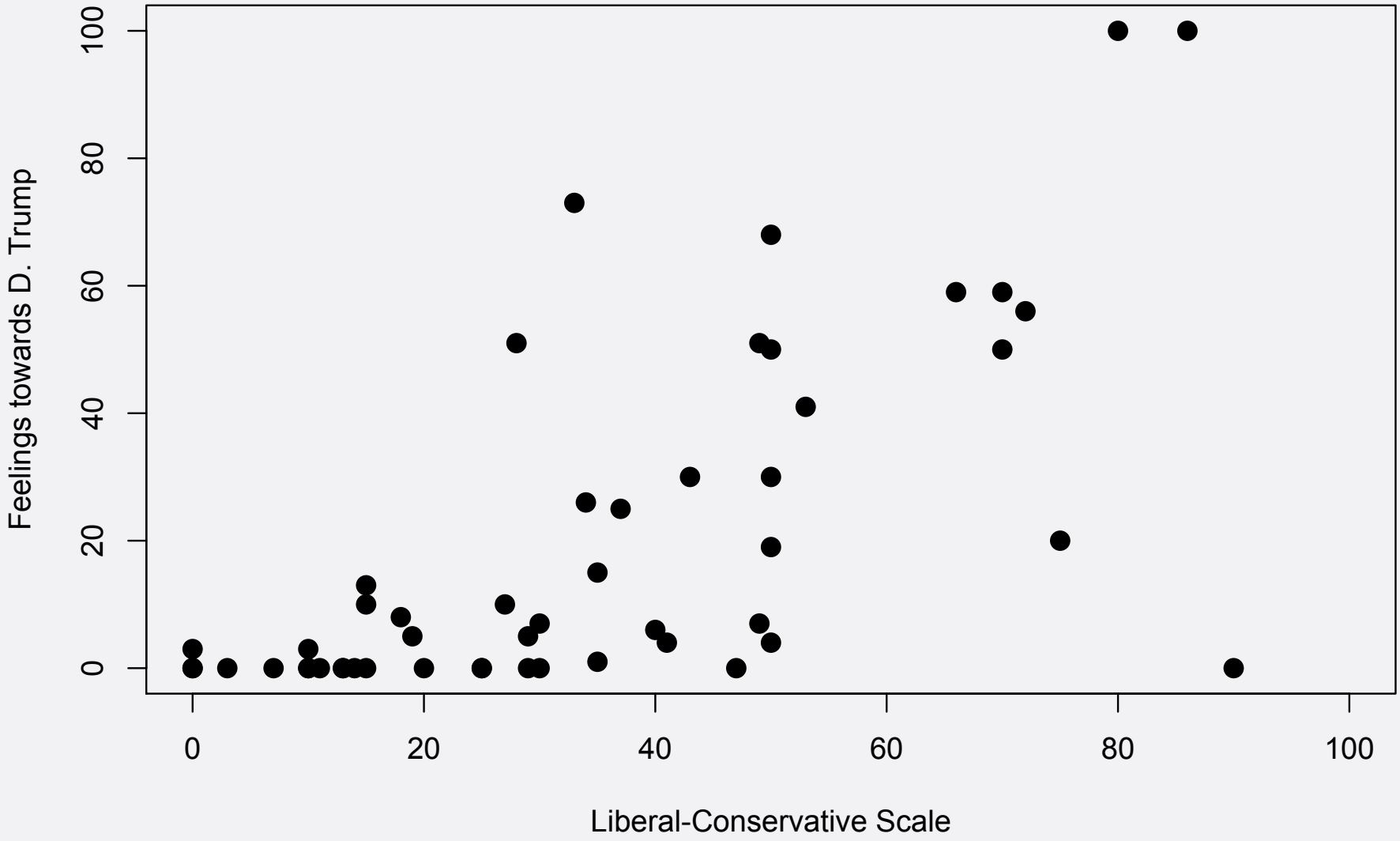
r=0

JOE BIDEN



r=-0.32

DONALD TRUMP



r=0.67

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

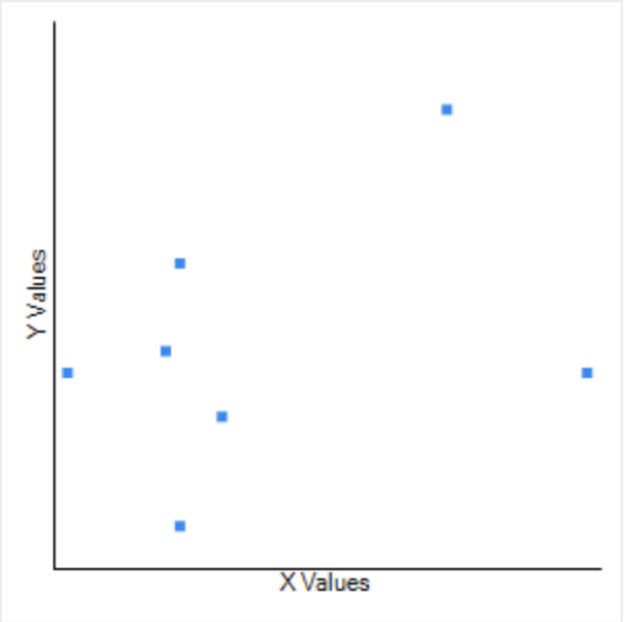
The value of R is: 0.3589.

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

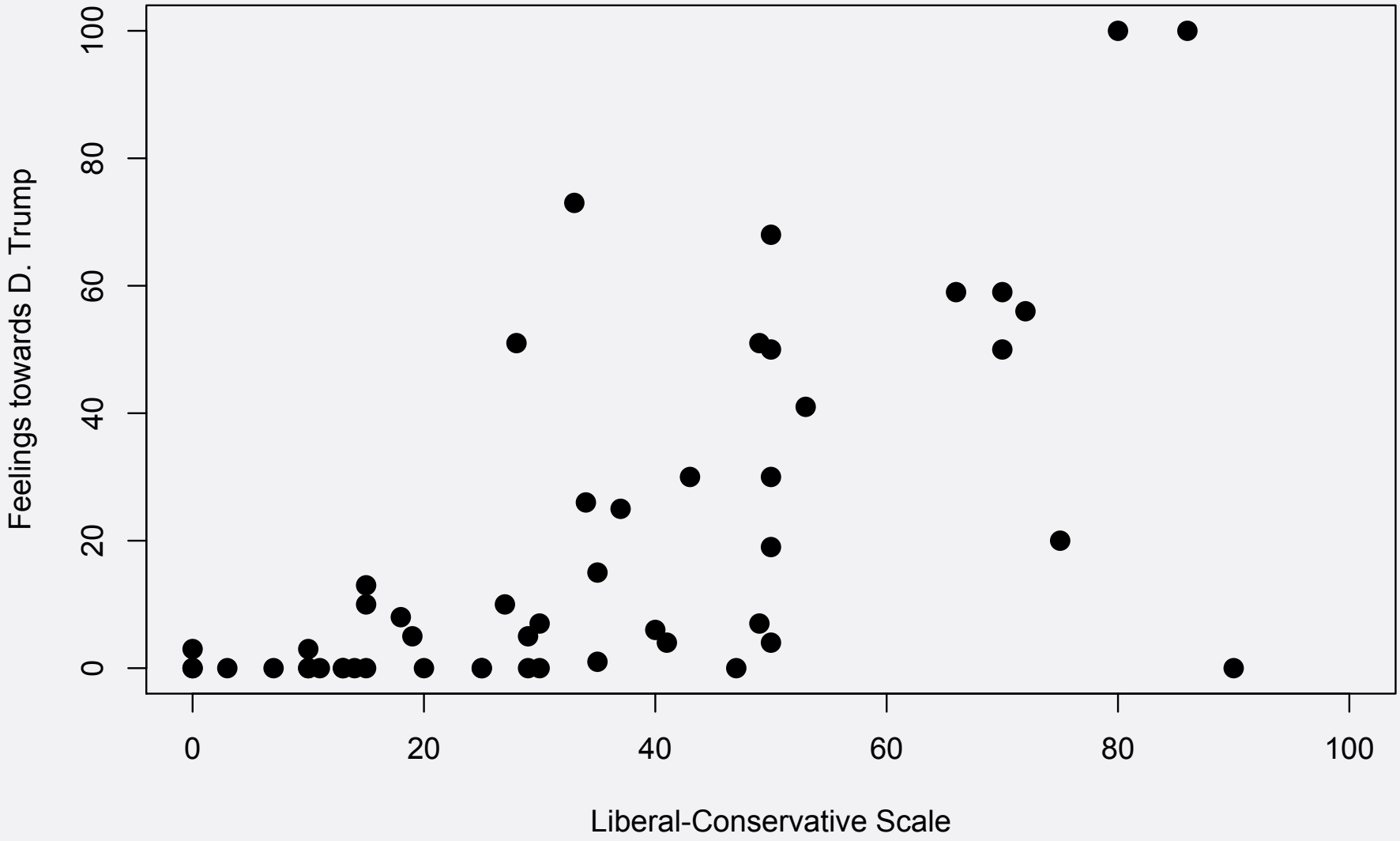


PEARSON'S R

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

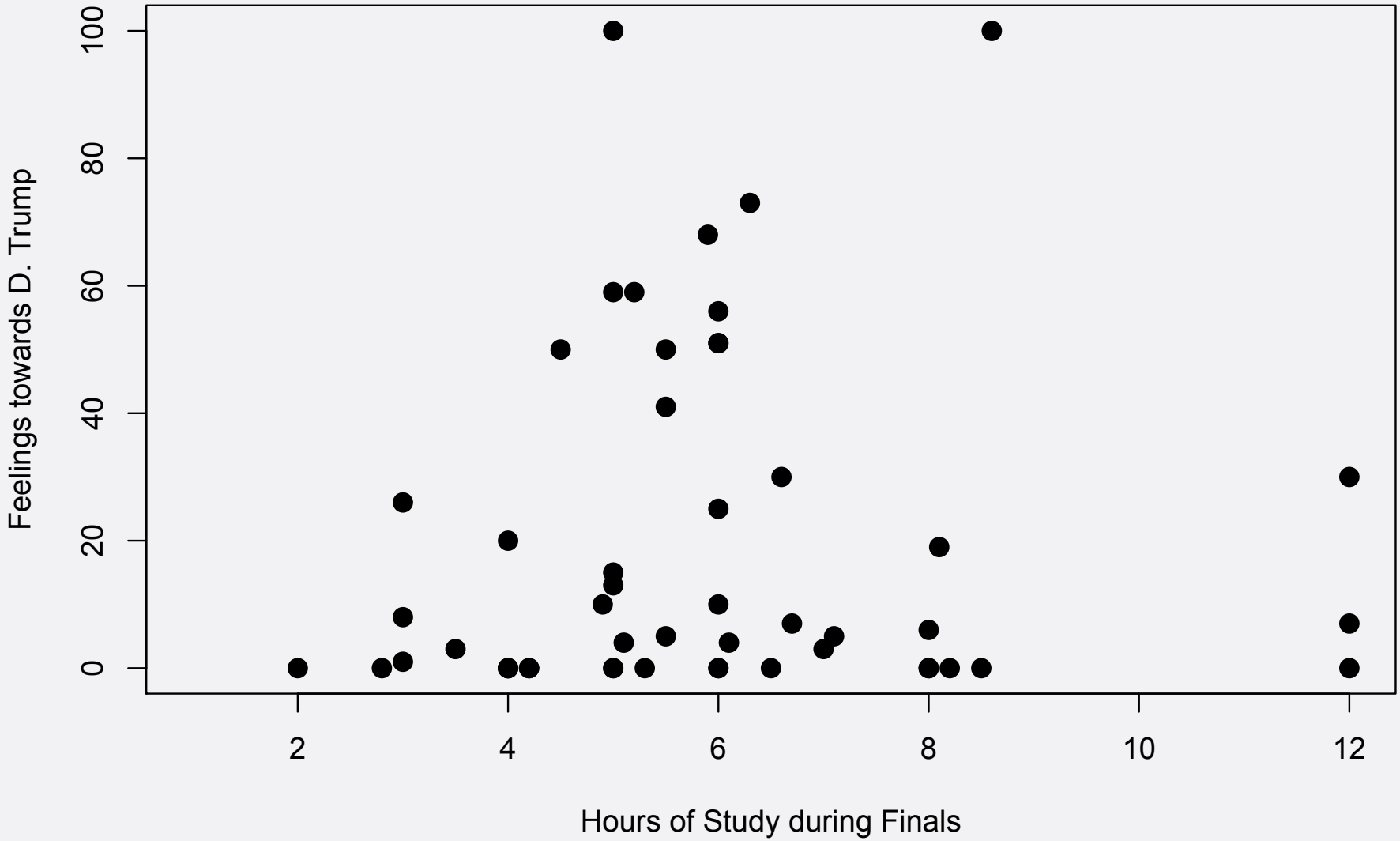
- **Intuition: Captures how much values of two variable vary together**
  - **High (positive) correlation: If X takes higher values, Y takes higher values**
  - **High (negative) correlation: If X takes higher values, Y takes lower values**
  - **Low correlation: If X takes higher values, values of Y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

DONALD TRUMP



r=0.67

DONALD TRUMP



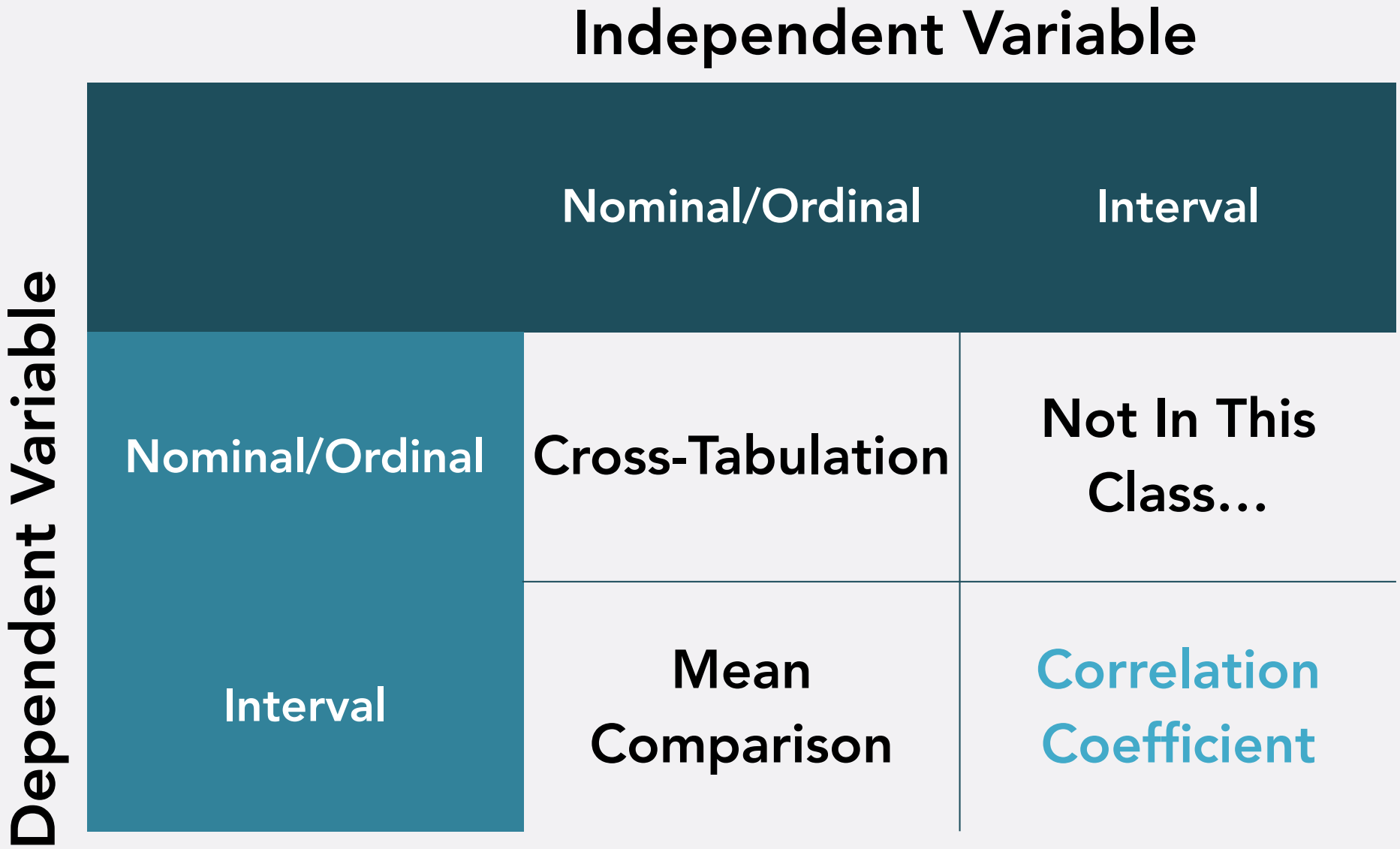
r=0.05

ACTUAL POLITICAL SCIENCE

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



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!