PSC 202

SYRACUSE UNIVERSITY

INTRODUCTION TO POLITICAL ANALYSIS

SAMPLING AND SURVEYS

SURVEY

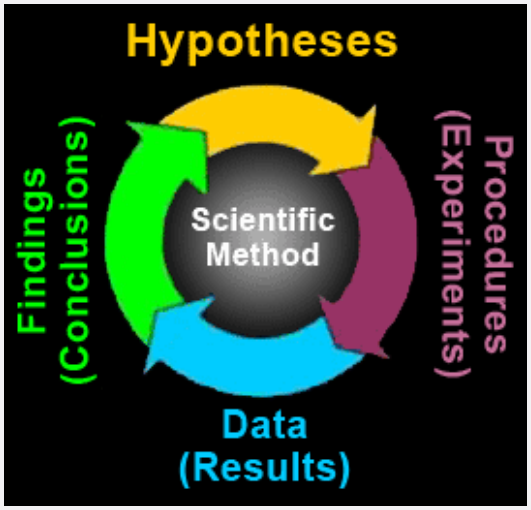
- Take it if you haven’t yet!
- Response rate: 68%
  - Need 85% to get extra credit for whole class
- rebrand.ly/202survey

EXAM

- Next week Monday: Exam #1
  - Bring a calculator (no phone etc.)
- Wednesday: Review
  - Email questions etc. by tomorrow evening
  - Formulas will be given on exam, no need to memorize
- If you take exam at ODS, please sign up now!

WHERE WE ARE

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



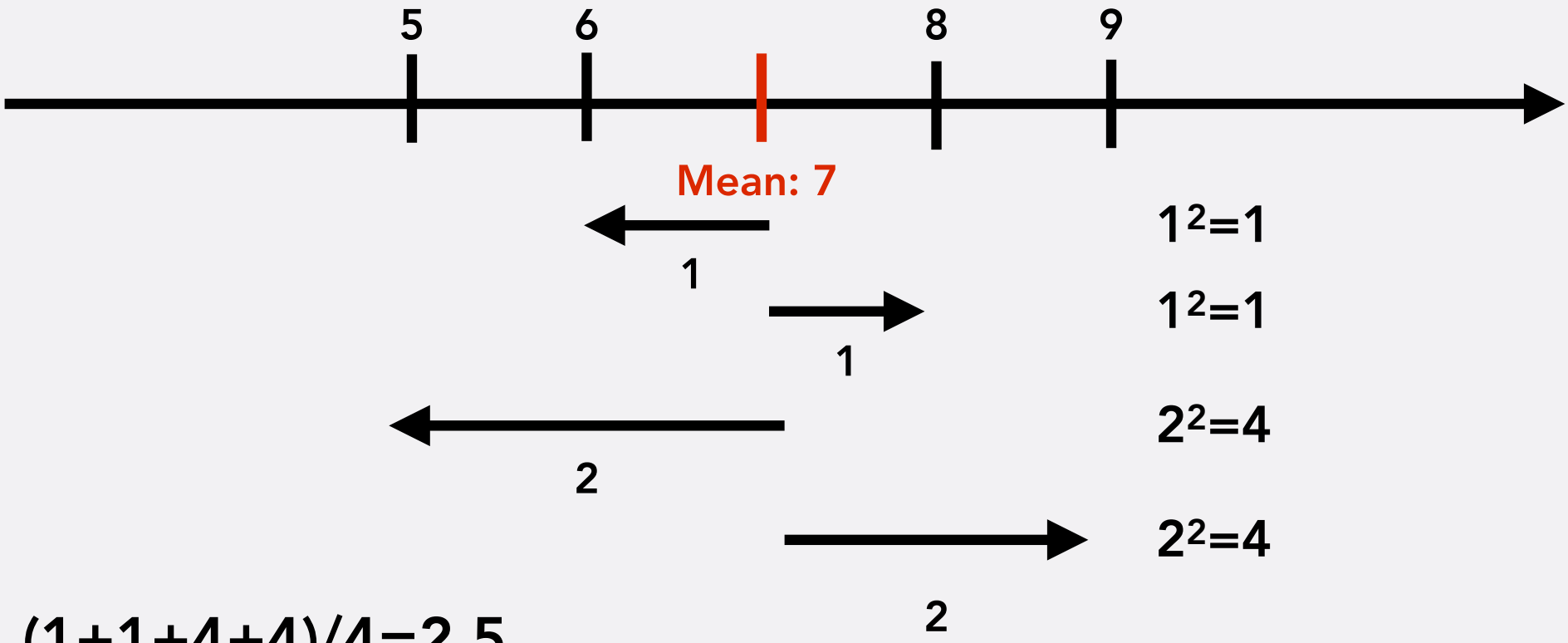
DESCRIBING VARIABLES

- Central tendency
- Frequency tables
- Dispersion

STANDARD DEVIATION

1. Calculate each value's deviation from mean
2. Square each deviation
3. Calculate the average of the sum of the squared deviations ("variance")
4. Take the square root of the variance ("standard deviation")

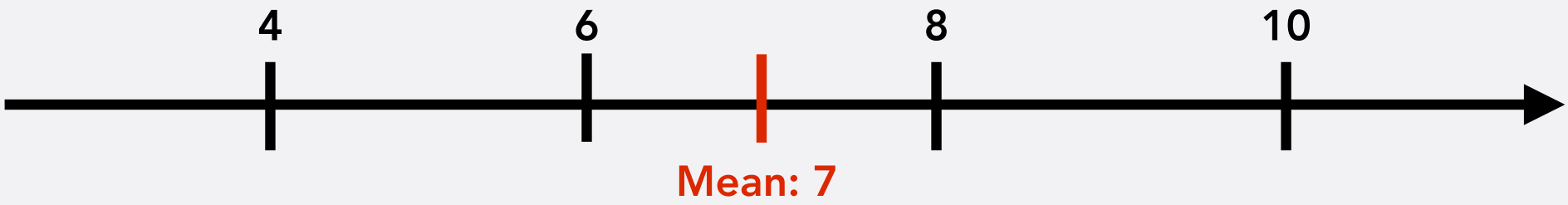
STANDARD DEVIATION



- (1+1+4+4)/4=2.5
- √2.5=1.58

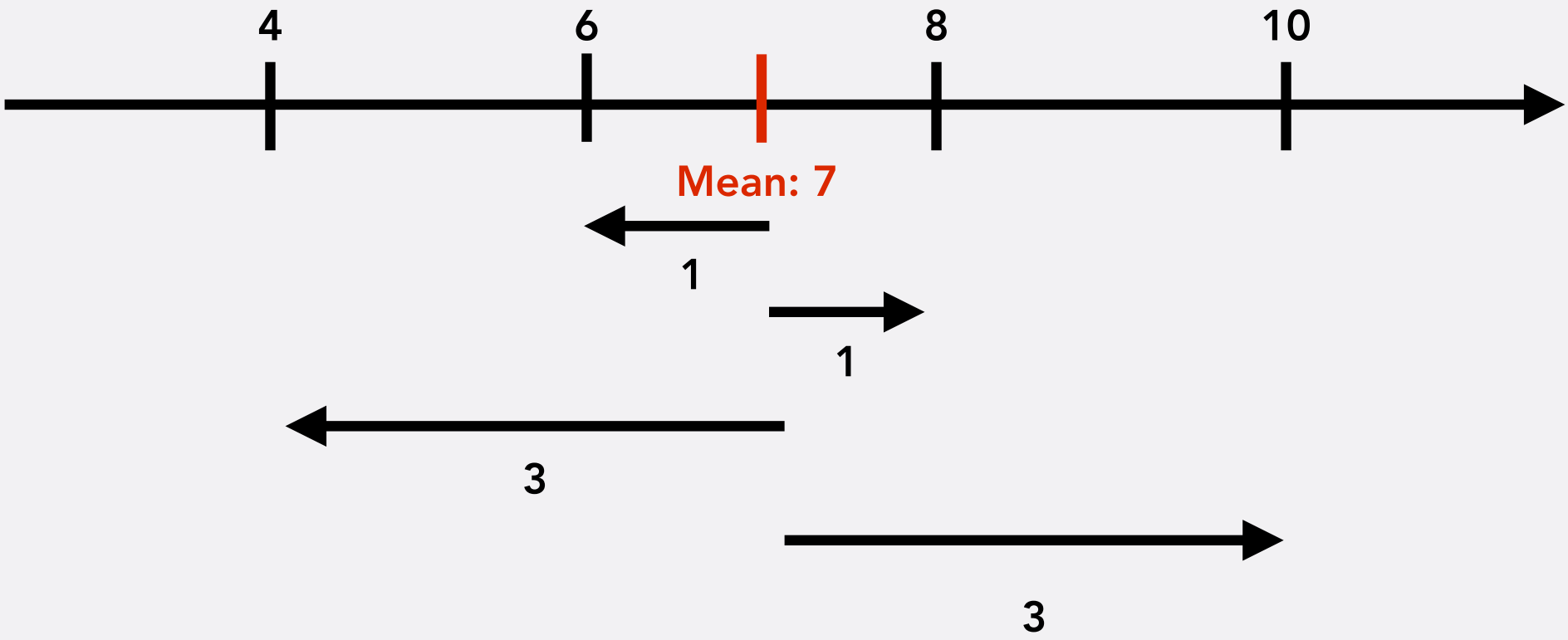
STANDARD DEVIATION

- Example with greater dispersion



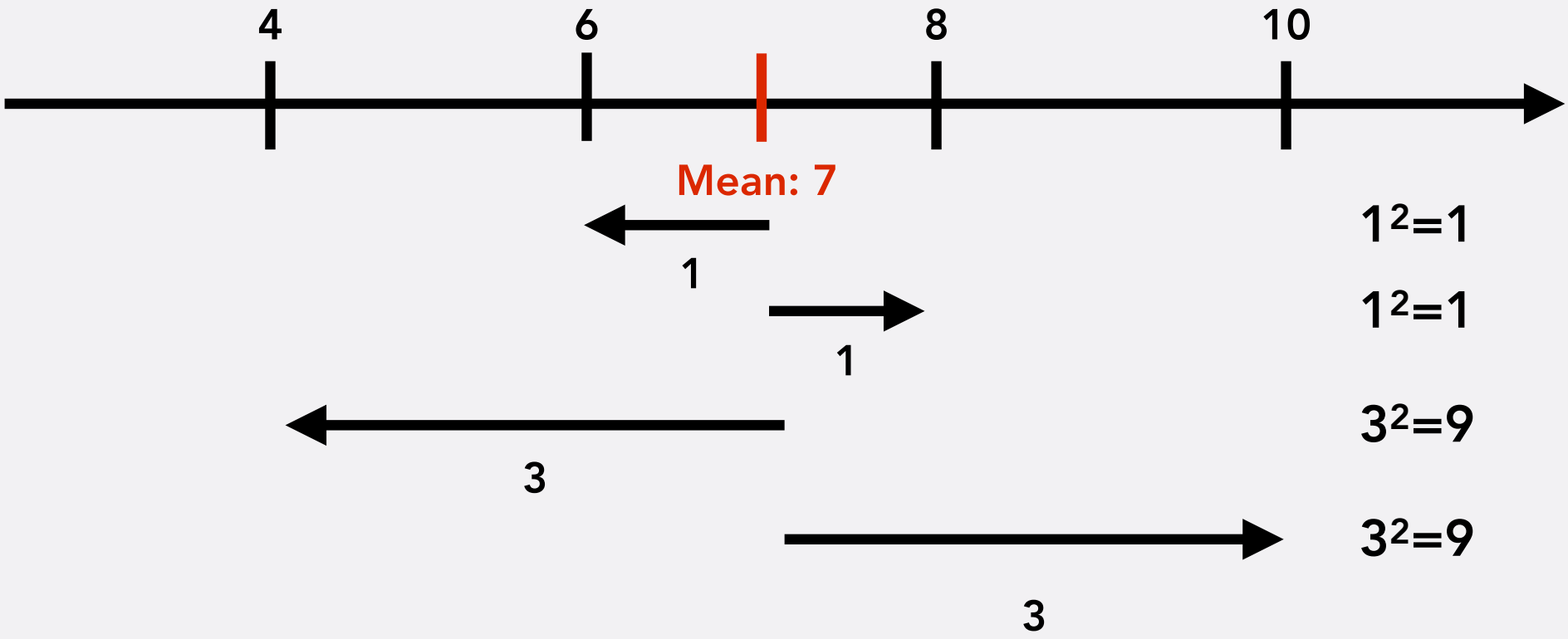
STANDARD DEVIATION

1. Calculate each value's deviation from mean



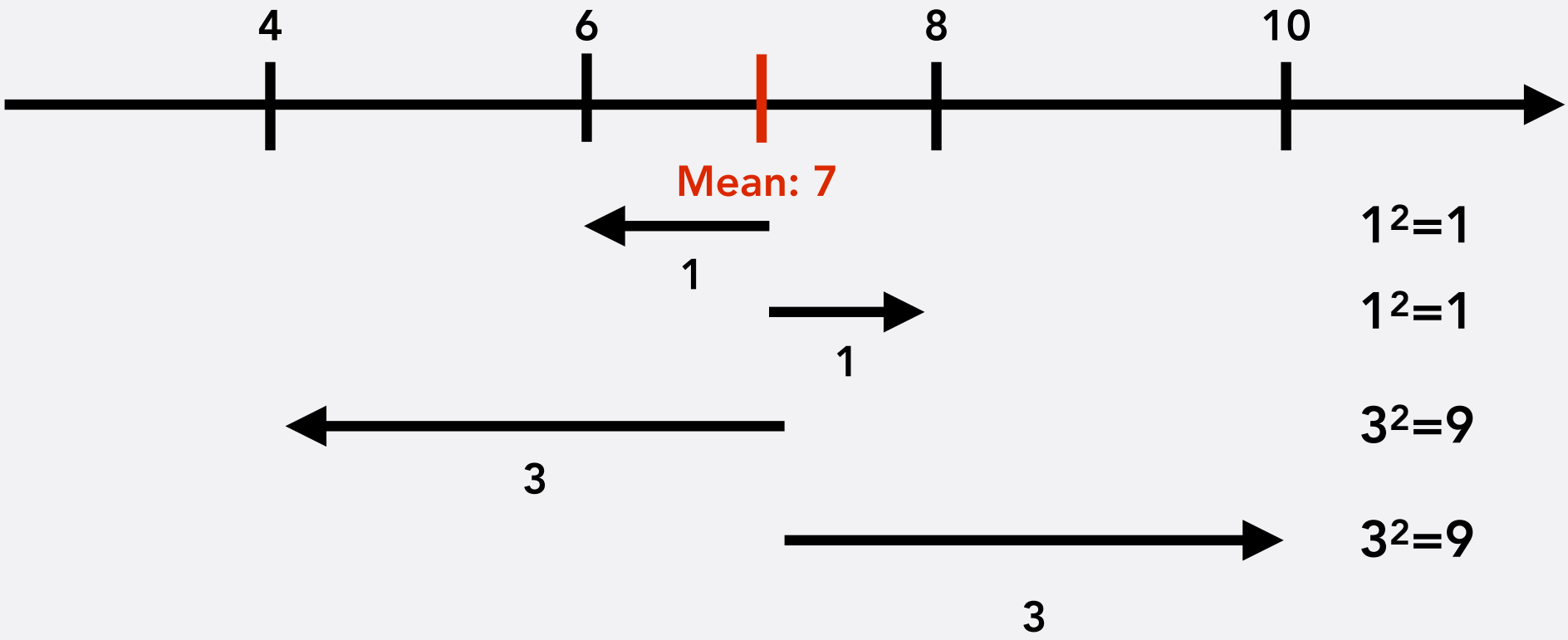
STANDARD DEVIATION

2. Square each deviation



STANDARD DEVIATION

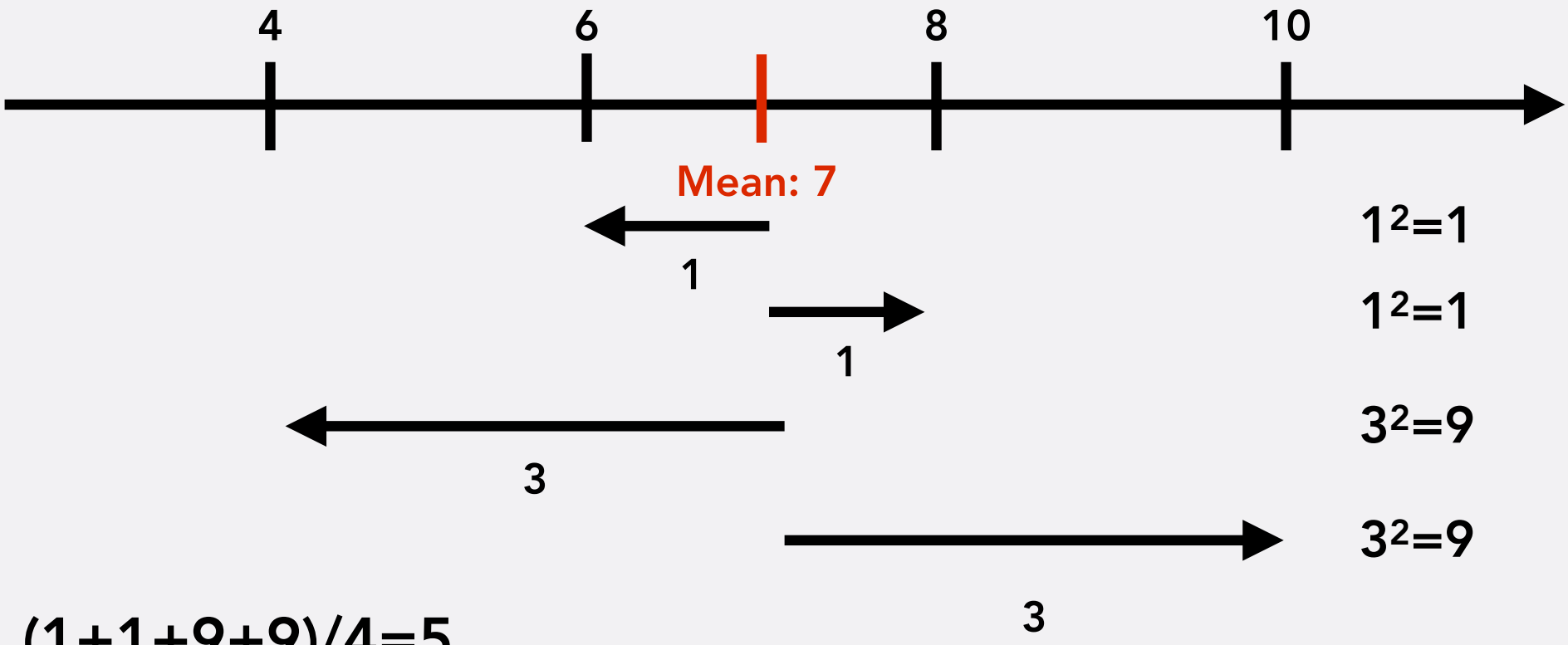
3. Calculate the average of the sum of the squared deviations ("variance")



- (1+1+9+9)/4=5

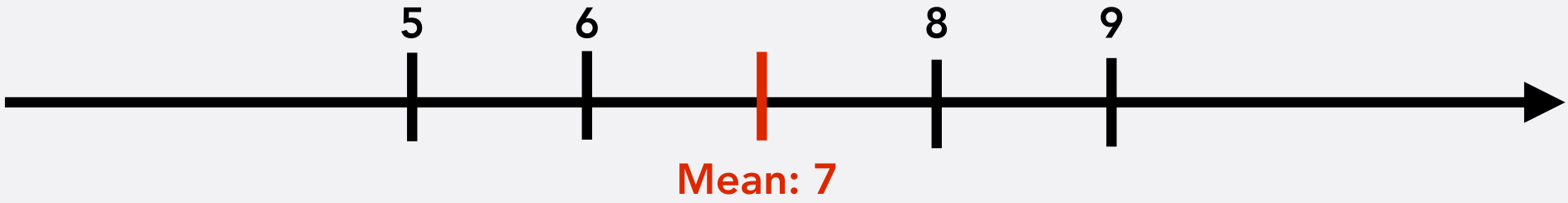
STANDARD DEVIATION

4. Take the square root of the variance ("standard deviation")

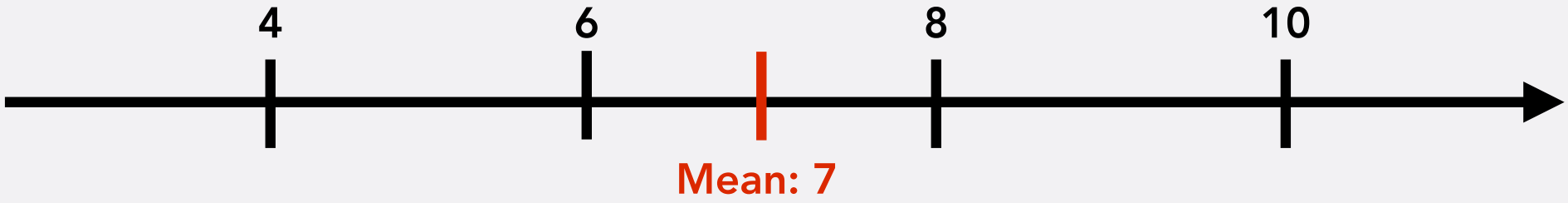


- (1+1+9+9)/4=5
- √5=2.24

COMPARISON



- Standard deviation: 1.58



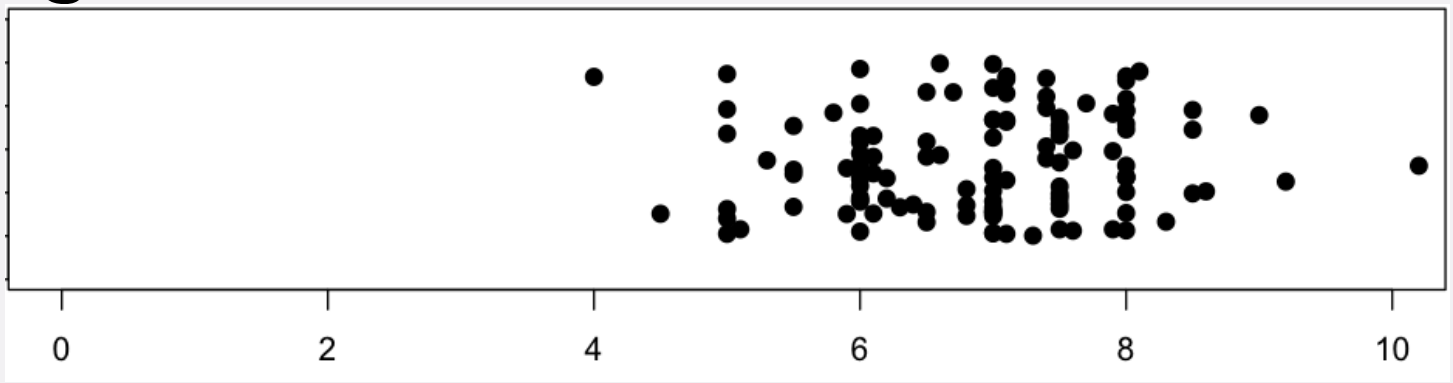
- Standard deviation: 2.24

STANDARD DEVIATION

- Standard deviation is helpful when comparing samples
  - Two countries with average income of 50k
  - One has a standard deviation of 5k, the other one a standard deviation of 15k

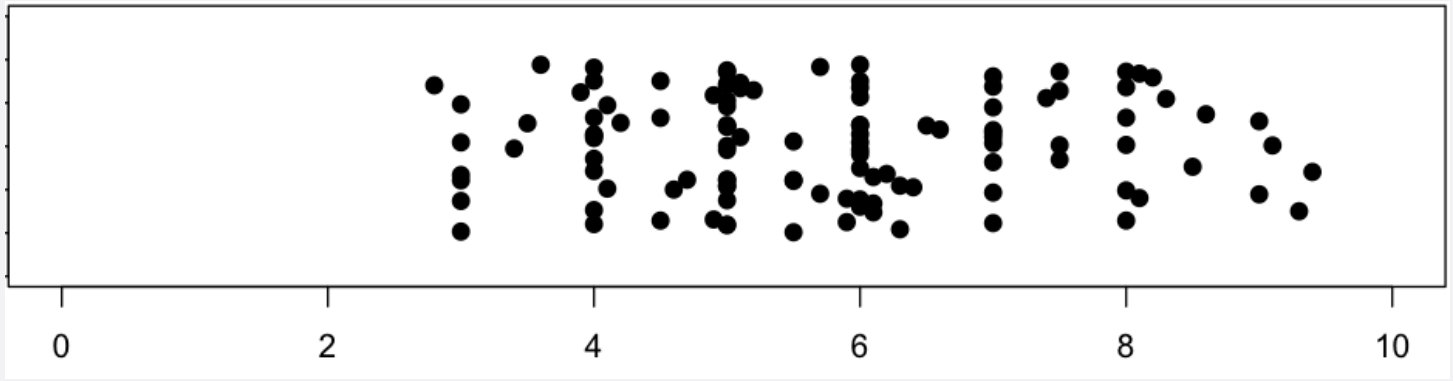
STANDARD DEVIATION

- How many hours to you sleep at night?
  - Regular week:



Mean: 6.9; Standard deviation: 1.06

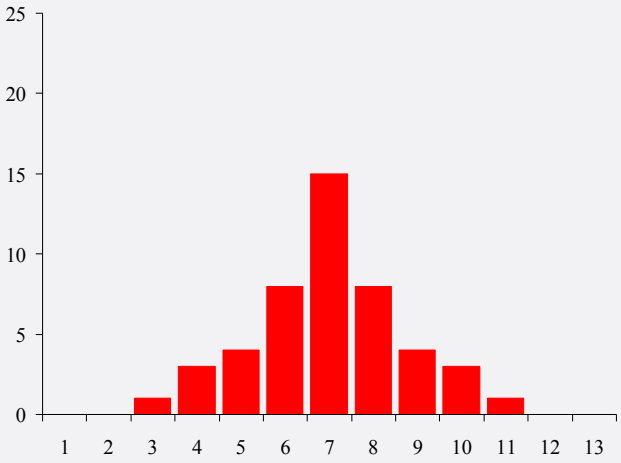
  - Finals week:



Mean: 5.8; Standard deviation: 1.59

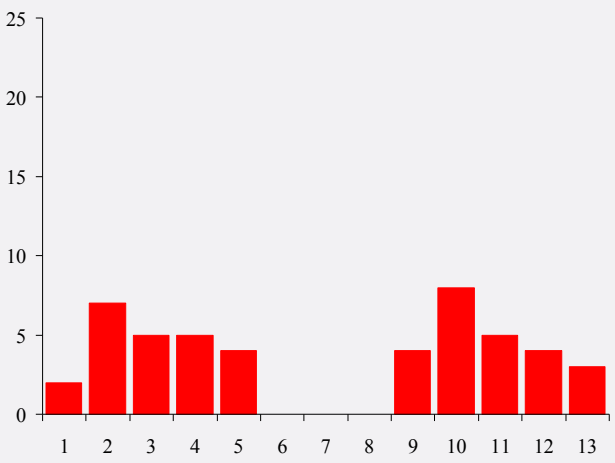
STANDARD DEVIATION

- Mean=7, median=7



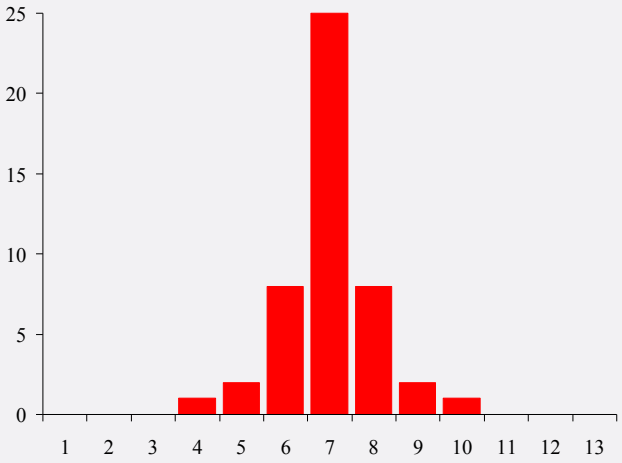
SD=1.67

Many students perform mediocre, some do well, others not



SD=4.01

One group does very well, one group does not



SD=1.02

All students perform relatively similarly (mediocre)

WHAT YOU CAN DO NOW

- Describe a variable
  - central tendency
  - dispersion
  - frequency table

TODAY: SURVEYS AND SAMPLING

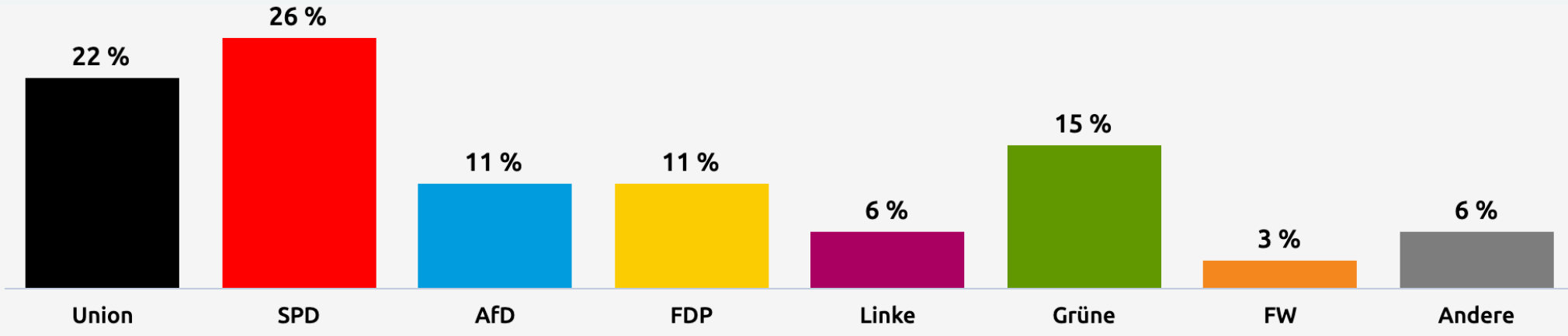
POLITICS SEPTEMBER 22, 2021

Biden's Approval Rating Hits New Low of 43%; Harris' Is 49%

Results for this Gallup poll are based on telephone interviews conducted Sept. 1-17, 2021, with a random sample of 1,005 adults, aged 18 and older, living in all 50 U.S. states and the District of Columbia. For results based on the total sample of national adults, the margin of sampling error is ±4 percentage points at the 95% confidence level. All reported margins of sampling error include computed design effects for weighting.

EXAMPLE

16.09.2021



WHAT THIS DOES AND DOESN'T TELL US

- What this tells us:
  - Biden's approval rating is 43% among 1,005 people who were interviewed
  - 26% plan on voting for SPD in sample of 1,512 Germans who were interviewed

WHAT THIS DOES AND DOESN'T TELL US

- What we are *really* interested in
  - Approval rating for Biden among *all* American voters
  - Preference for SPD among *all* German voters

TODAY AND NEXT CLASS

- How confident can we be that the 43% approval rating among 1,005 respondents is close to the approval rating of *all* American voters?

HOW IS THIS POSSIBLE?

Triumph of the Nerds: Nate Silver Wins in 50 States

Share on Facebook | Share on Twitter | +

Numbers nerd Nate Silver's forecasts prove all right on election night

FiveThirtyEight blogger predicted the outcome in all 50 states, assuming Barack Obama's Florida victory is confirmed

KEY DISTINCTION

- Population: the entire universe of objects to which our hypothesis applies (citizens, voters, countries, etc.)
  - Size of population: N
  - e.g. registered American voters, N=168,308,000

KEY DISTINCTION

- Sample: the subset of the population that we study in order to make inferences about the full population
  - Size of sample: n
  - e.g. respondents in the poll, n= 1,005

HOW TO SAMPLE

- How do we get a sample from a population?

HOW TO SAMPLE

- 1936 Presidential Election: Franklin D. Roosevelt (D) vs. Alf Landon (R)
- Who will win?
- *Literary Digest* magazine poll
  - Sent 10 million mock ballots to people, addresses from subscriber data, phone records, automobile registrations
  - 2.5 million people returned ballot
- Prediction: Roosevelt 43%, Landon 57%

HOW TO SAMPLE

Straw Vote Fight Arouses Interest

Literary Digest and American Institute Are Far Apart In Pre-election Forecast—Roosevelt, Landon Both Get Around 56 Per Cent

- *American Institute for Public Opinion* poll
  - Interviewed 50,000 randomly sampled people
- Prediction: Roosevelt 56%, Landon 44%

HOW TO SAMPLE

The Pittsburgh Press

WEATHER—Rain changing to snow and colder tonight. Thursday fair.

(Copyright, 1936, by Post Publishing Co. All Rights Reserved)

FINAL STOCKS
CLOSING PRICES
PRICE THREE CENTS

VOLUME 53; No. 134 48 PAGES PITTSBURGH, PA., WEDNESDAY, NOVEMBER 4, 1936

ROOSEVELT WINS 46 STATES

Landslide Gives Him Pennsylvania By 600,000 And County By 188,000

Nominee	**Franklin D. Roosevelt**	Alf Landon
Party	Democratic	Republican
Home state	New York	Kansas
Running mate	**John Nance Garner**	Frank Knox
Electoral vote	**523**	8
States carried	**46**	2
Popular vote	**27,747,636**	16,679,543
Percentage	**60.8%**	36.5%

WHAT WENT WRONG FOR LD?

WHAT WENT WRONG FOR LD?

- *Literary Digest* magazine poll
  - People who subscribe to magazines, own phones/ cars are wealthy
  - Therefore more likely to vote for Republican
  - People who are less wealthy (and are more likely to for Democrat) were not in LD's sample
  - LD's sample was not representative of the population

WHAT WENT WRONG FOR LD?

- *American Institute* poll
  - Randomly selected people to interview
  - Its sample was representative of the population
  - It sampled poor and rich voters roughly in the same proportion as they are present in the population
  - Much more accurate prediction with 50,000 respondents than with 2.5 million of LD

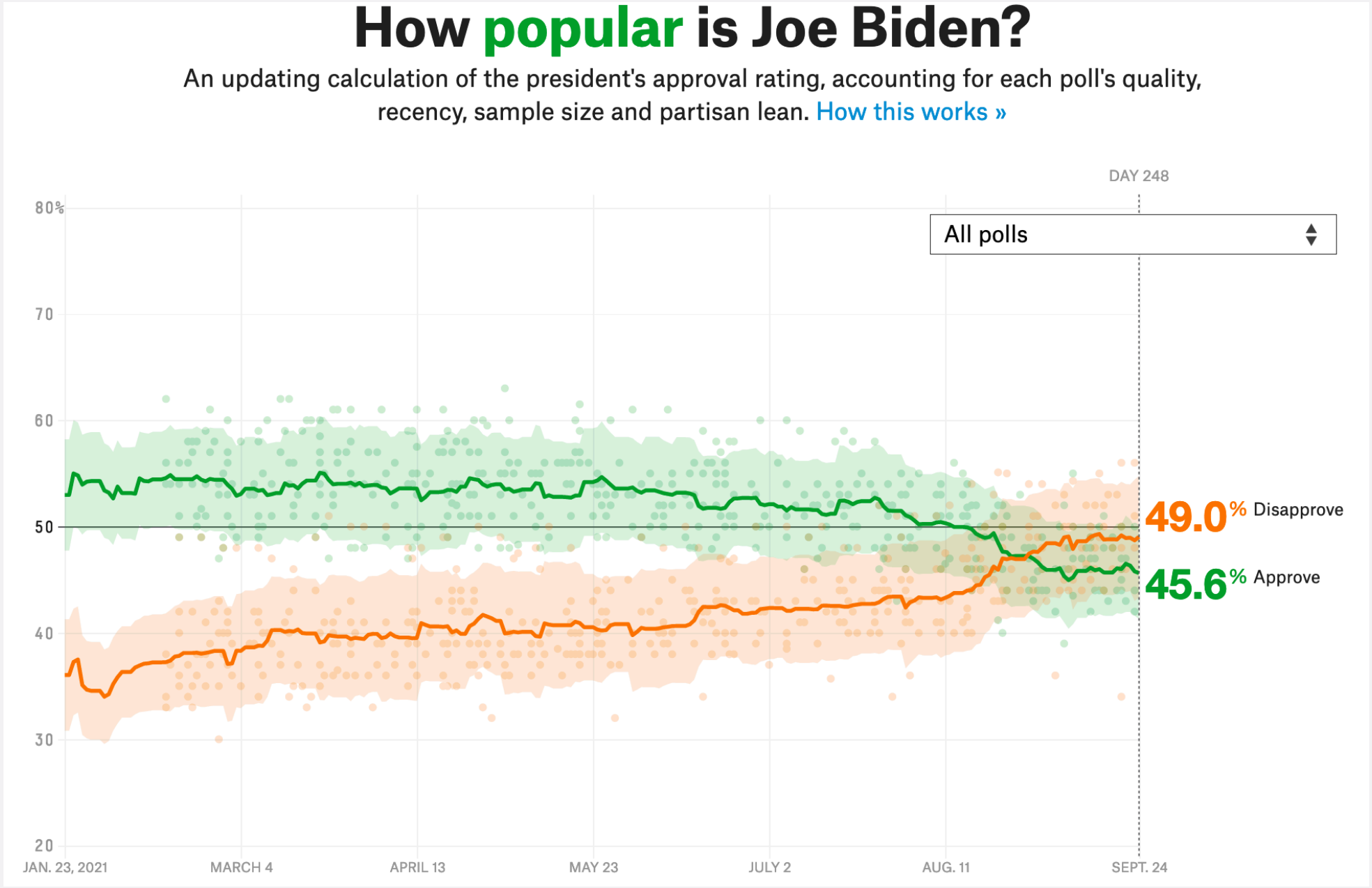
RANDOM SAMPLING

- A random sample of the population avoids *systematic* sampling error
- If we use random sampling, we can use our sample's characteristics to estimate the population's characteristics
  - e.g. can use 1,005 randomly selected survey respondents to infer approval rating of J. Biden in American population

RANDOM SAMPLING ERROR

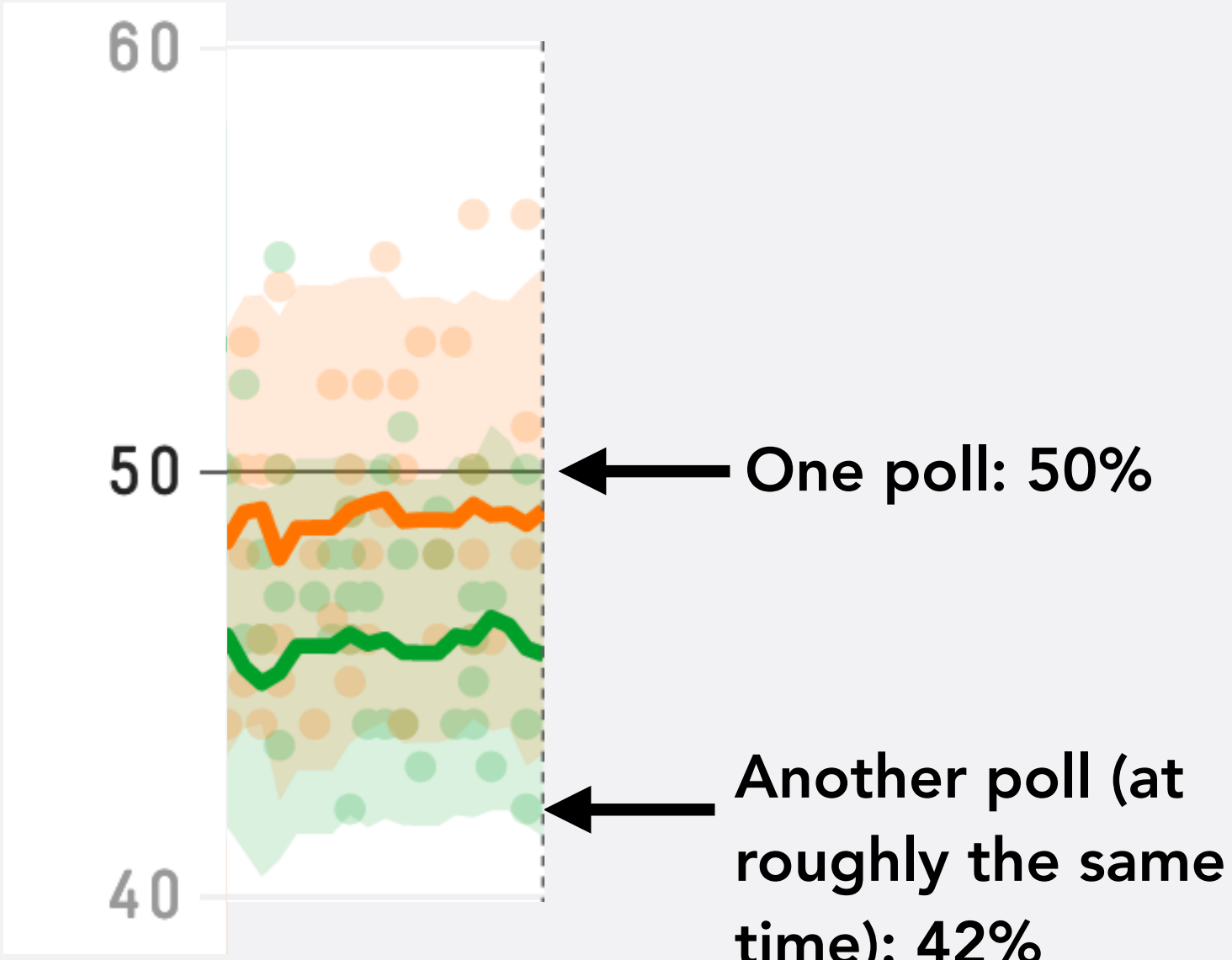
- But: random sampling introduces *random* sampling error
  - It is unlikely that our random sample looks *exactly* like the American population
  - e.g. by chance, we might draw more people that approve of Biden than is the case in the population
  - Or we might draw more people that disapprove of his performance than in the population

RANDOM SAMPLING ERROR



- https://projects.fivethirtyeight.com/biden-approval-rating

RANDOM SAMPLING ERROR



RANDOM SAMPLING ERROR

- Random sampling introduces *random* sampling error
  - Example: Flipping a coin
  - For a fair coin, we know that Heads=50%, Tails=50%
  - We flip a coin 10 times:
    - We may get HHTHTTHTHT (5H, 5T)
    - We might also get HHHHHTHHHT (8H, 2T)
    - Or TTTHTTTTHT (2H, 8T)

THE PROBLEM

- Population parameter = Sample statistic + random sampling error

THE PROBLEM

Unknown: Approval rating in population

Known: Approval rating in survey

- Population parameter = Sample statistic + random sampling error

Also unknown

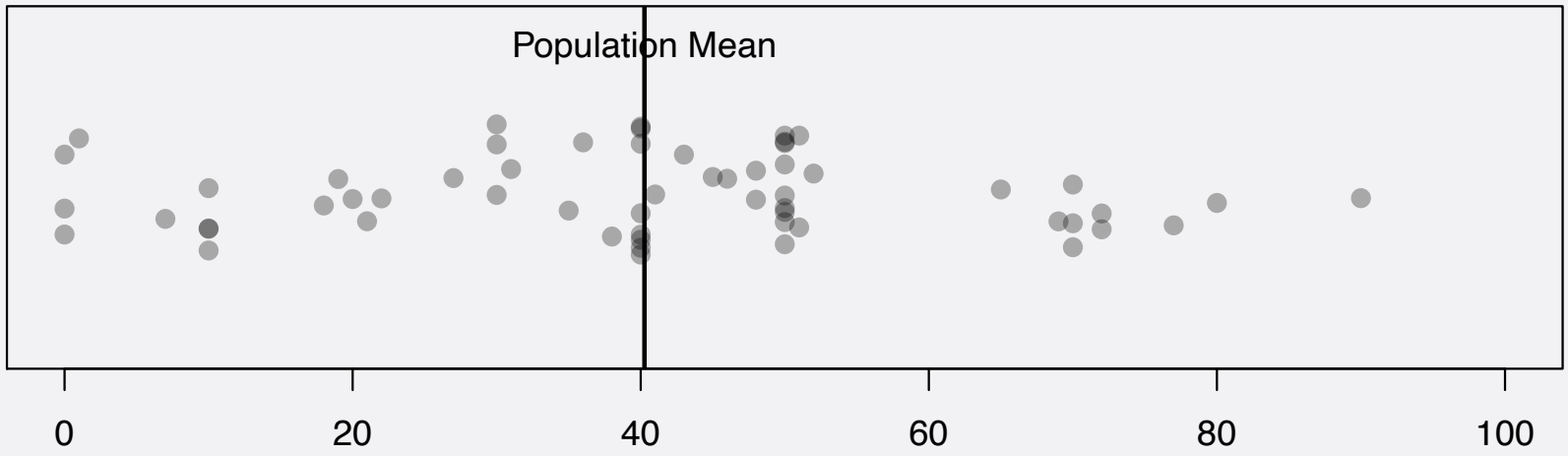
GOOD NEWS

- We can figure out how large the random sampling error is

OUR SURVEY

- Feelings towards Kent Syverud among population of PSC 202 students
  - On a scale from 0 to 100

OUR SURVEY



- Mean: 40.3
- Standard deviation: 21.7

POPULATION

- This is the view of among the *population* of PSC 202 students
  - Usually: survey of sample, not of population
  - Here: we *do* have the population
  - What we can do: See what would happen if we only had a random sample of PSC 202 students, and compare that to the population
  - This will give us an idea of how large the sampling error is

THE PROBLEM

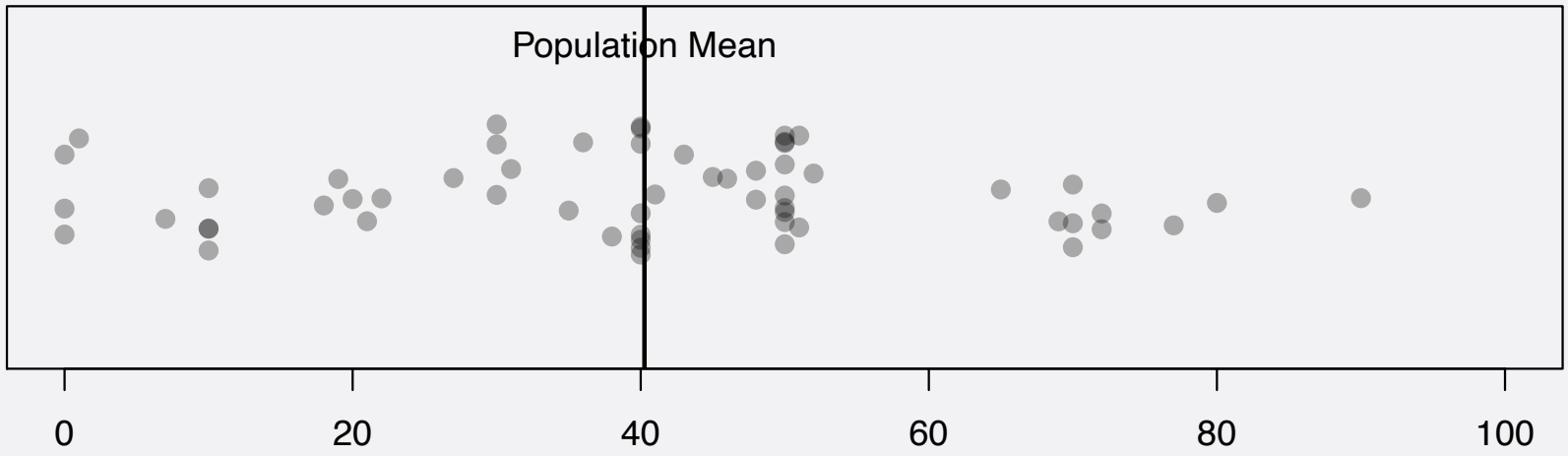
Known (here): approval rating in population

Also known: approval rating in random sample

- Population parameter = Sample statistic + random sampling error

We can figure this out

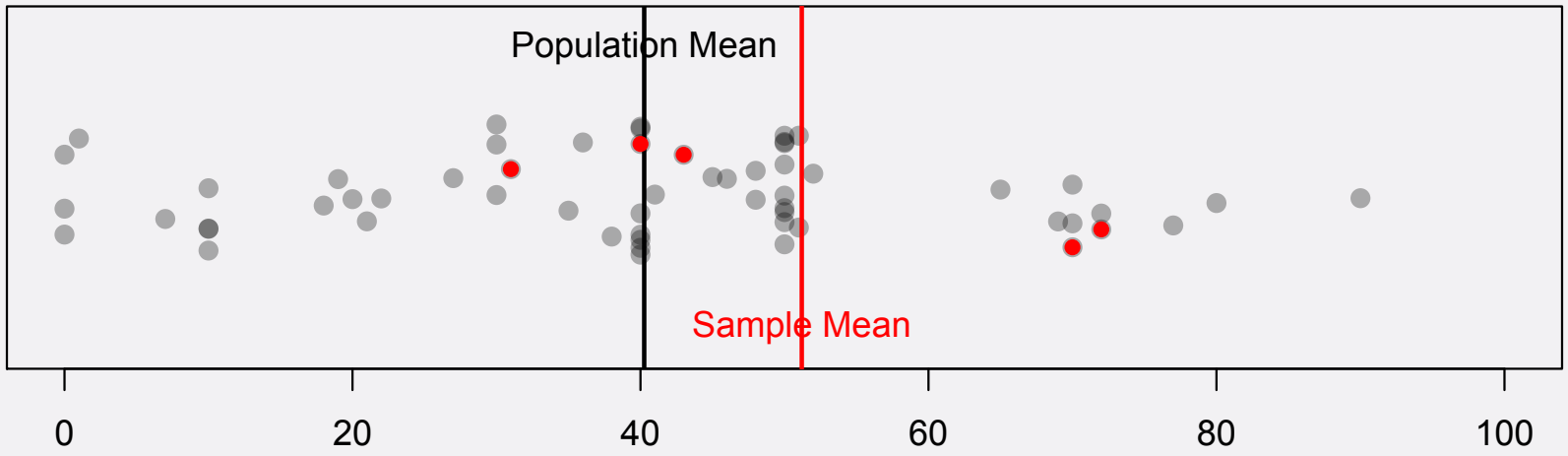
OUR SURVEY



- Mean: 40.3
- Standard deviation: 21.7

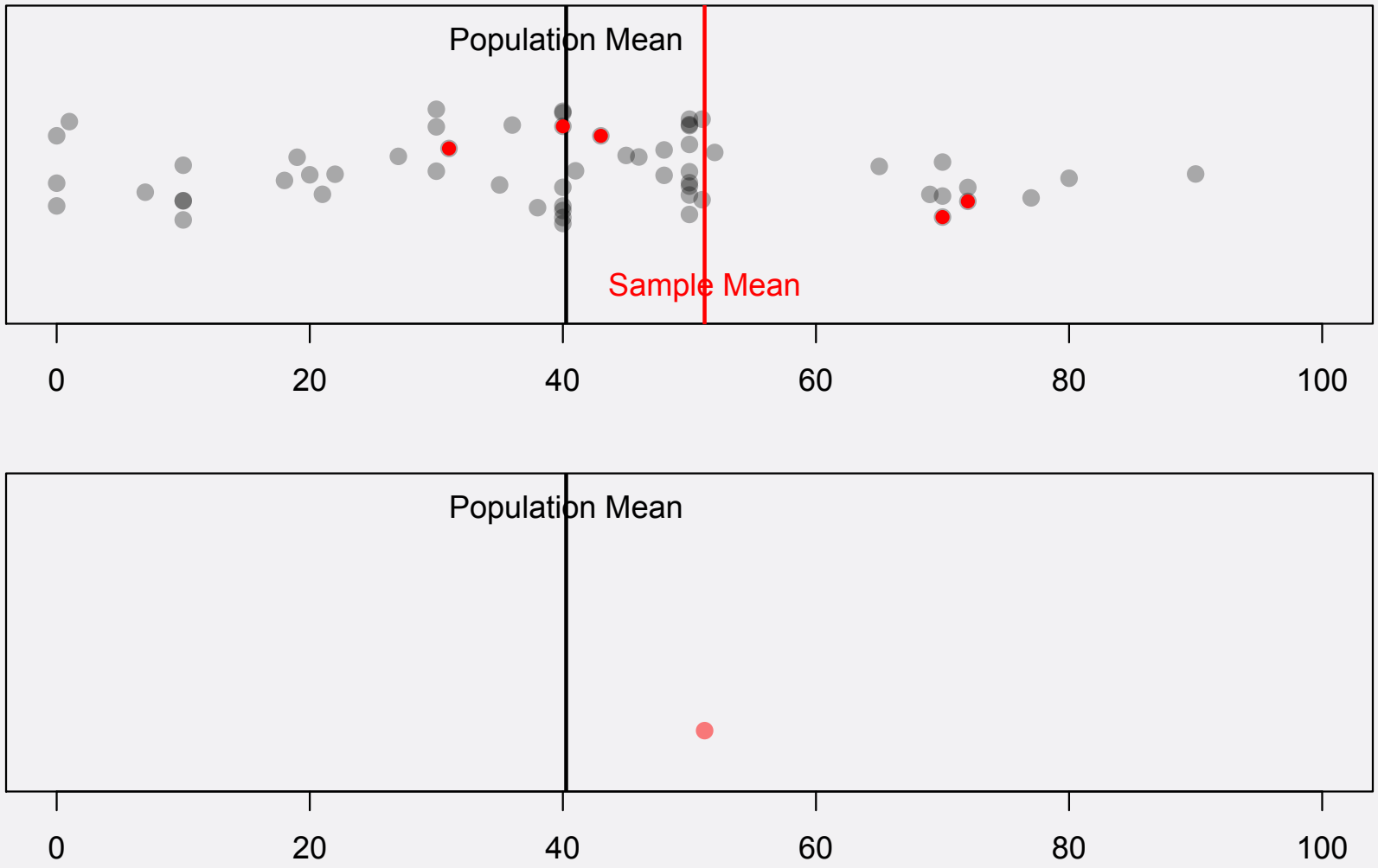
- Let's say we could only survey a random sample of 5 students of PSC 202

OUR SURVEY

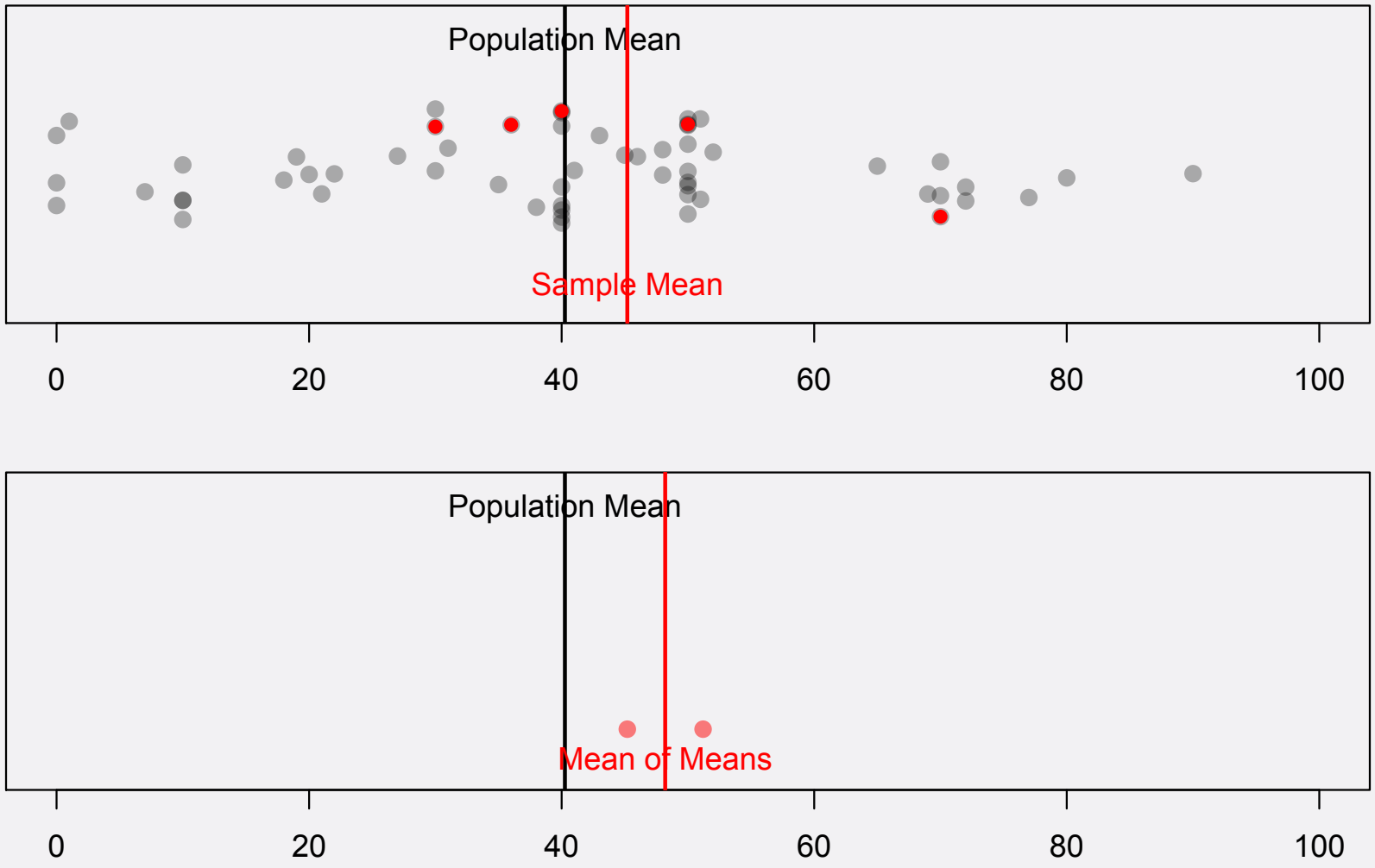


- Mean of this sample: a lot higher than the population mean

OUR SURVEY

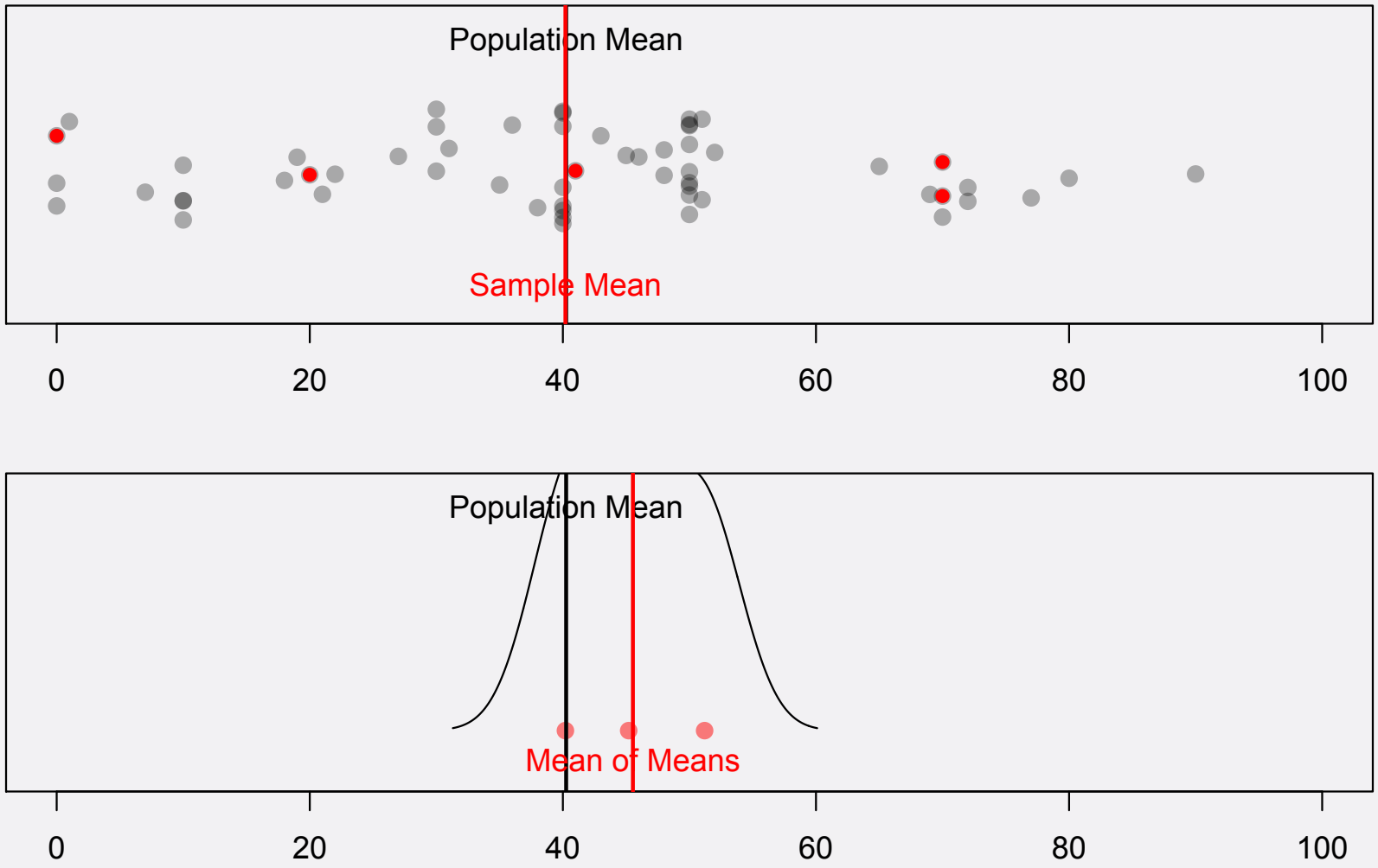

Population Mean
Sample Mean
0
20
40
60
80
100
Population Mean
0
20
40
60
80
100

OUR SURVEY

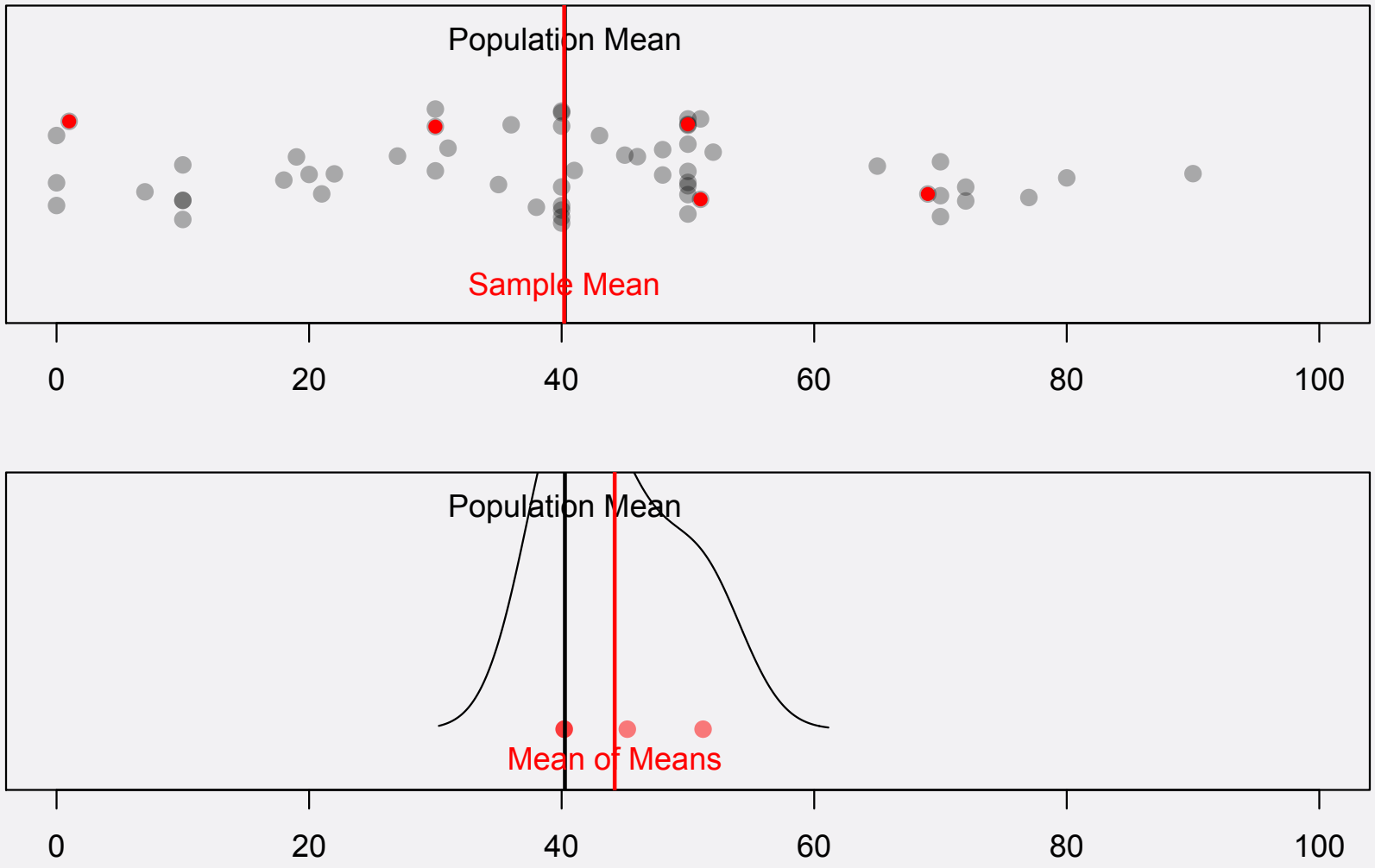


- Let's do this again…another random sample of 5

OUR SURVEY


Population Mean
Sample Mean
0
20
40
60
80
100
Population Mean
Mean of Means
0
20
40
60
80
100

OUR SURVEY



- **We actually need to do this a lot of times…**

OUR SURVEY

- This was 500 random samples from the population
  - Sample mean jump around
  - Some are far away from population mean, but most are quite close
  - Also: Mean of sample means gets pretty close to the population mean

THAT'S HOW THIS IS POSSIBLE

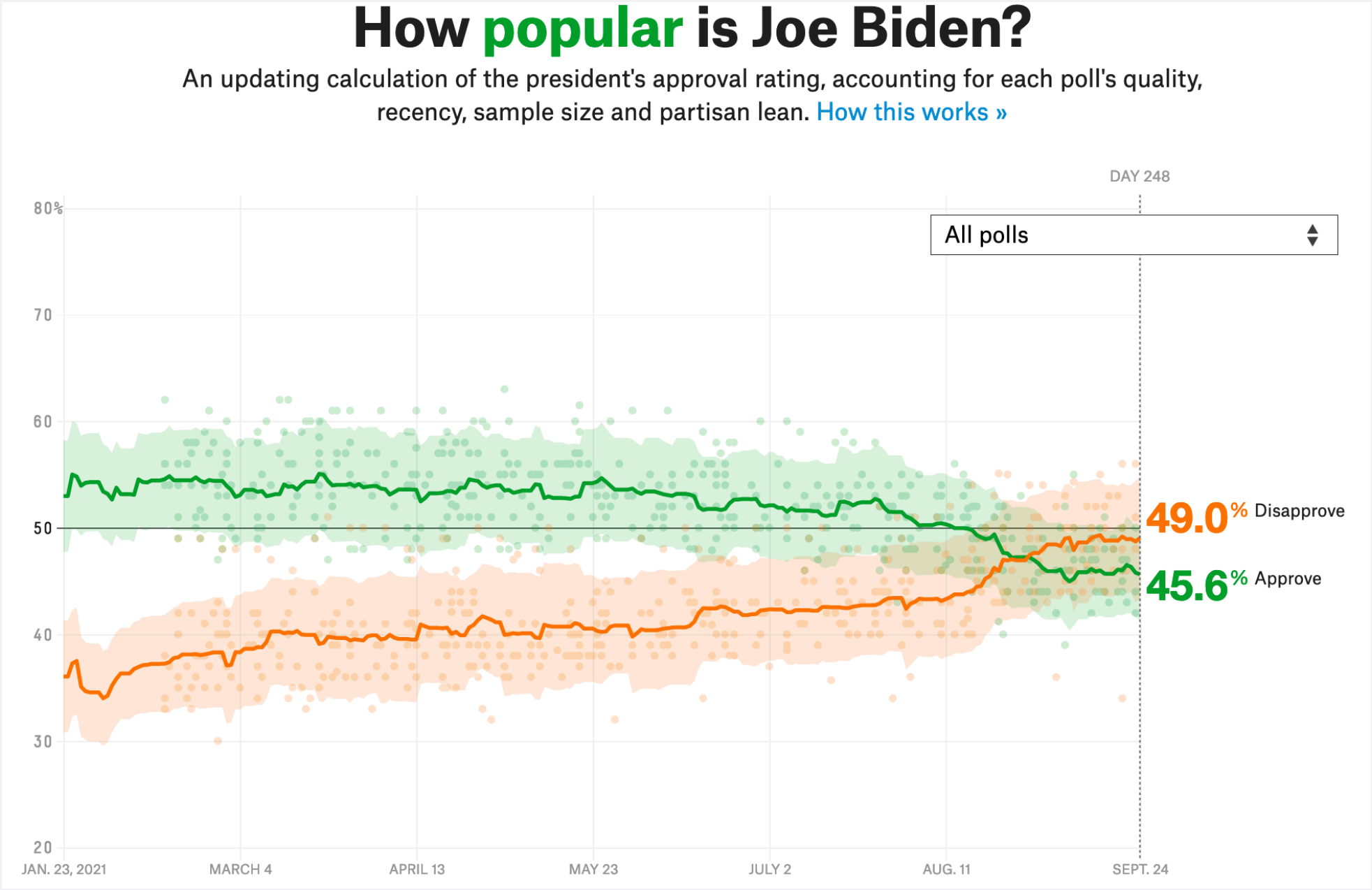
Triumph of the Nerds: Nate Silver Wins in 50 States

Share on Facebook Share on Twitter +

Numbers nerd Nate Silver's forecasts prove all right on election night

FiveThirtyEight blogger predicted the outcome in all 50 states, assuming Barack Obama's Florida victory is confirmed

THAT'S HOW THIS IS POSSIBLE



- https://projects.fivethirtyeight.com/biden-approval-rating

AFTER 10,000 RANDOM SAMPLES