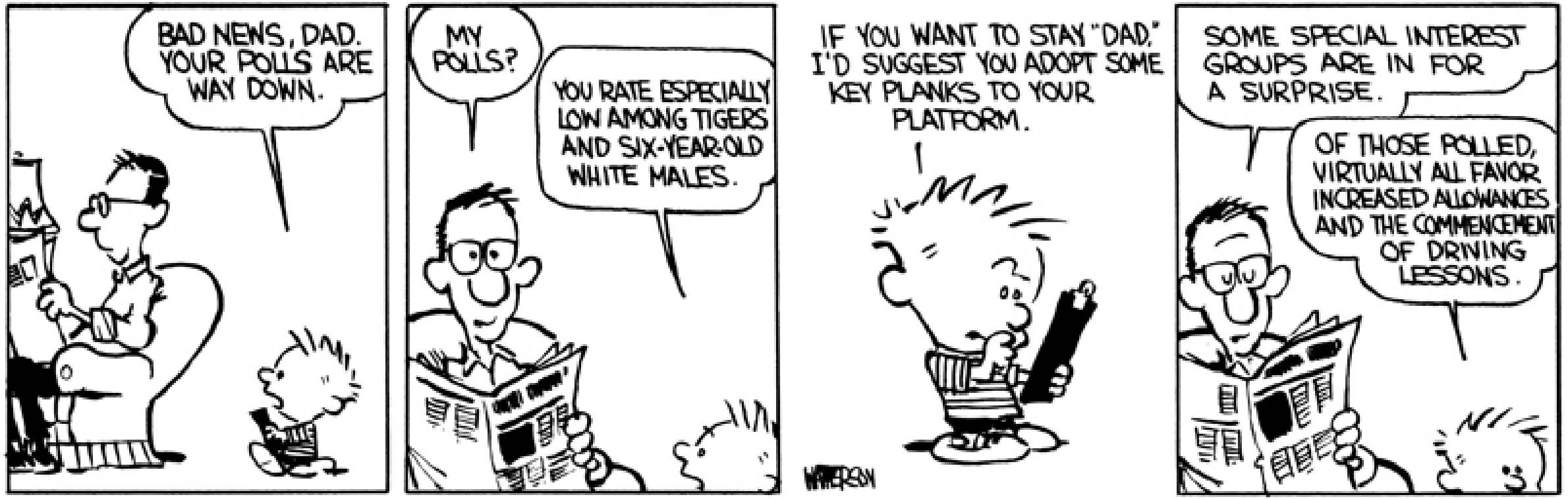
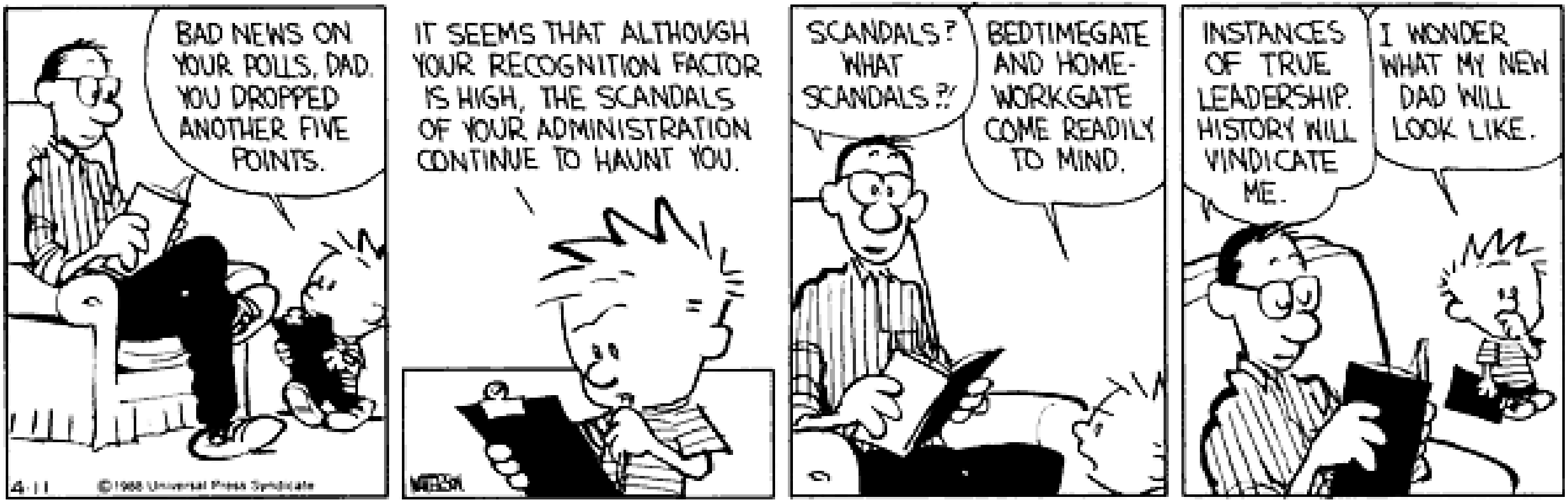
Random Sampling

Who to ask?


BAD NEWS, DAD. YOUR POLLS ARE WAY DOWN.
MY POLLS?
YOU RATE ESPECIALLY LOW AMONG TIGERS AND SIX-YEAR-OLD WHITE MALES.
IF YOU WANT TO STAY "DAD," I'D SUGGEST YOU ADOPT SOME KEY PLANKS TO YOUR PLATFORM.
SOME SPECIAL INTEREST GROUPS ARE IN FOR A SURPRISE.
OF THOSE POLLED, VIRTUALLY ALL FAVOR INCREASED ALLOWANCES AND THE COMMENCEMENT OF DRIVING LESSONS.

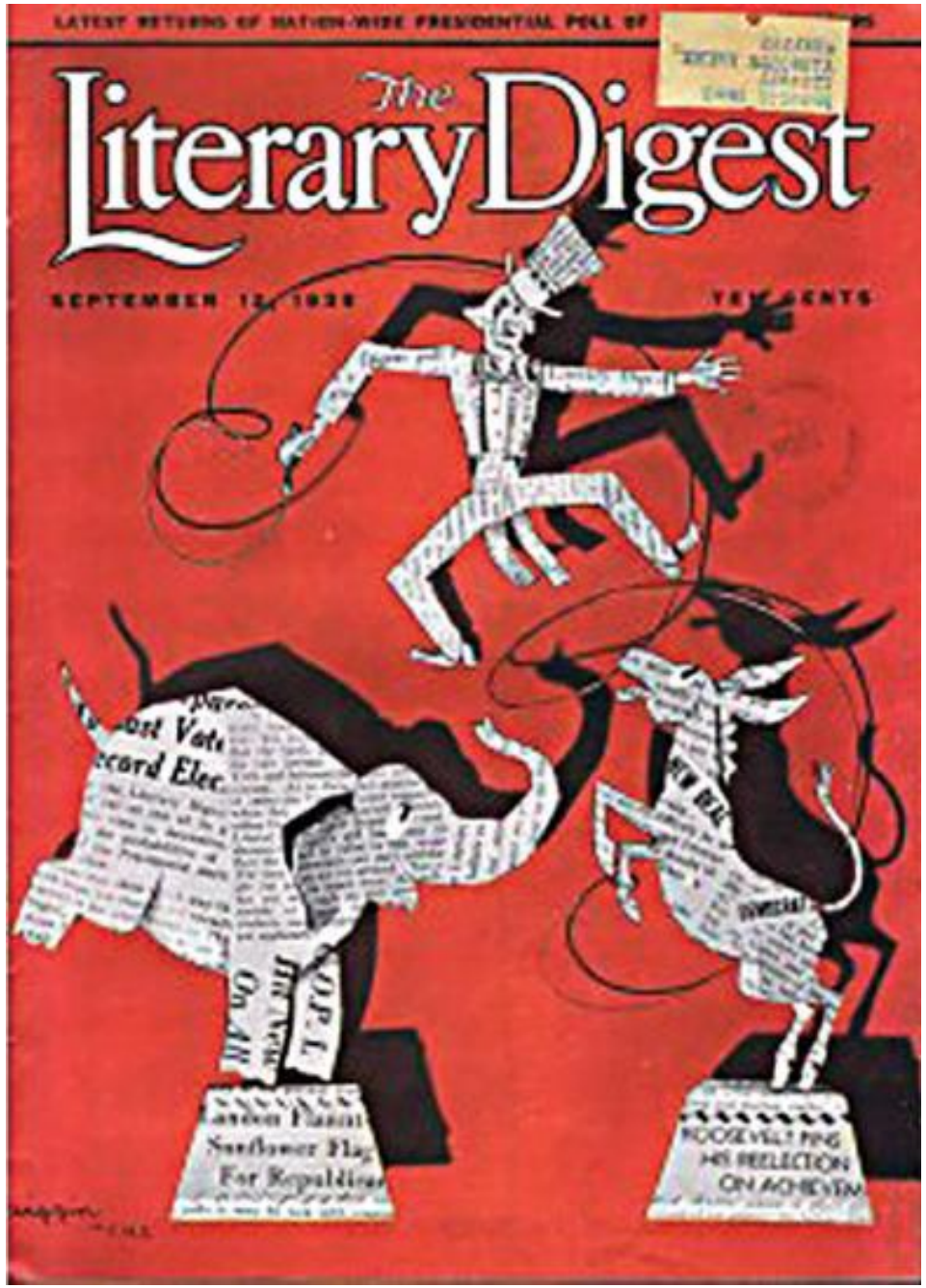

ELECTION DAY IS COMING UP, DAD. PEOPLE WANT TO KNOW WHERE YOU STAND ON THE ISSUES.
SUCH AS?
LATER BEDTIMES, EXPANDED TV PRIVILEGES, SHORTER SCHOOL WEEKS, AND LESS DISCIPLINE.
I'M AGAINST THEM ALL.
I SEE.
HOW'S YOUR IRA? PRETTY WELL FUNDED?
GO TO BED.
10-8


BAD NEWS ON YOUR POLLS, DAD. YOU DROPPED ANOTHER FIVE POINTS.
4-11
©1988 Universal Press Syndicate
IT SEEMS THAT ALTHOUGH YOUR RECOGNITION FACTOR IS HIGH, THE SCANDALS OF YOUR ADMINISTRATION CONTINUE TO HAUNT YOU.
WATTERSON
SCANDALS? WHAT SCANDALS?!
BEDTIMEGATE AND HOME-WORKGATE COME READILY TO MIND.
INSTANCES OF TRUE LEADERSHIP. HISTORY WILL VINDICATE ME.
I WONDER WHAT MY NEW DAD WILL LOOK LIKE.

Franklin D. Roosevelt

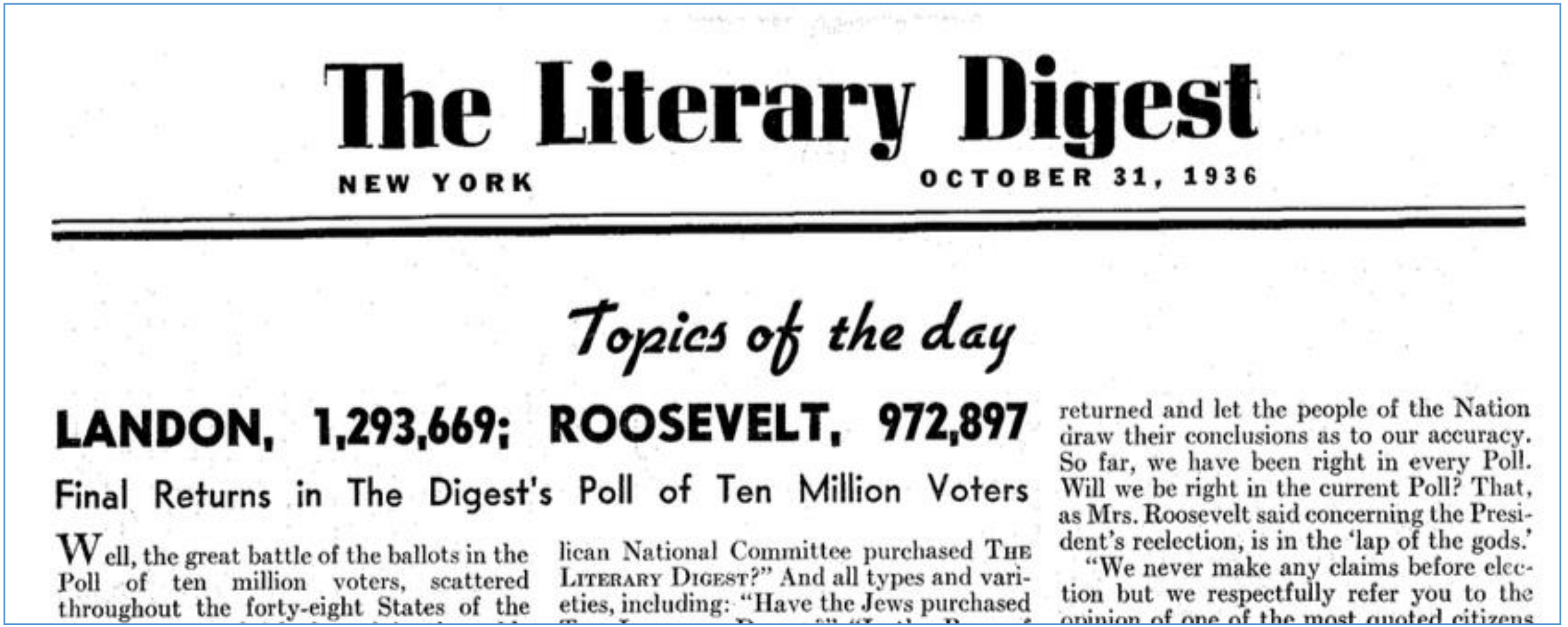
Alf Landon

1936



1936 Literary Digest poll

- largest and most expensive poll
- a sample size of 2.4 million people



The Literary Digest

NEW YORK — OCTOBER 31, 1936

Topics of the day

LANDON, 1,293,669; ROOSEVELT, 972,897

Final Returns in The Digest's Poll of Ten Million Voters

Well, the great battle of the ballots in the Poll of ten million voters, scattered throughout the forty-eight States of the

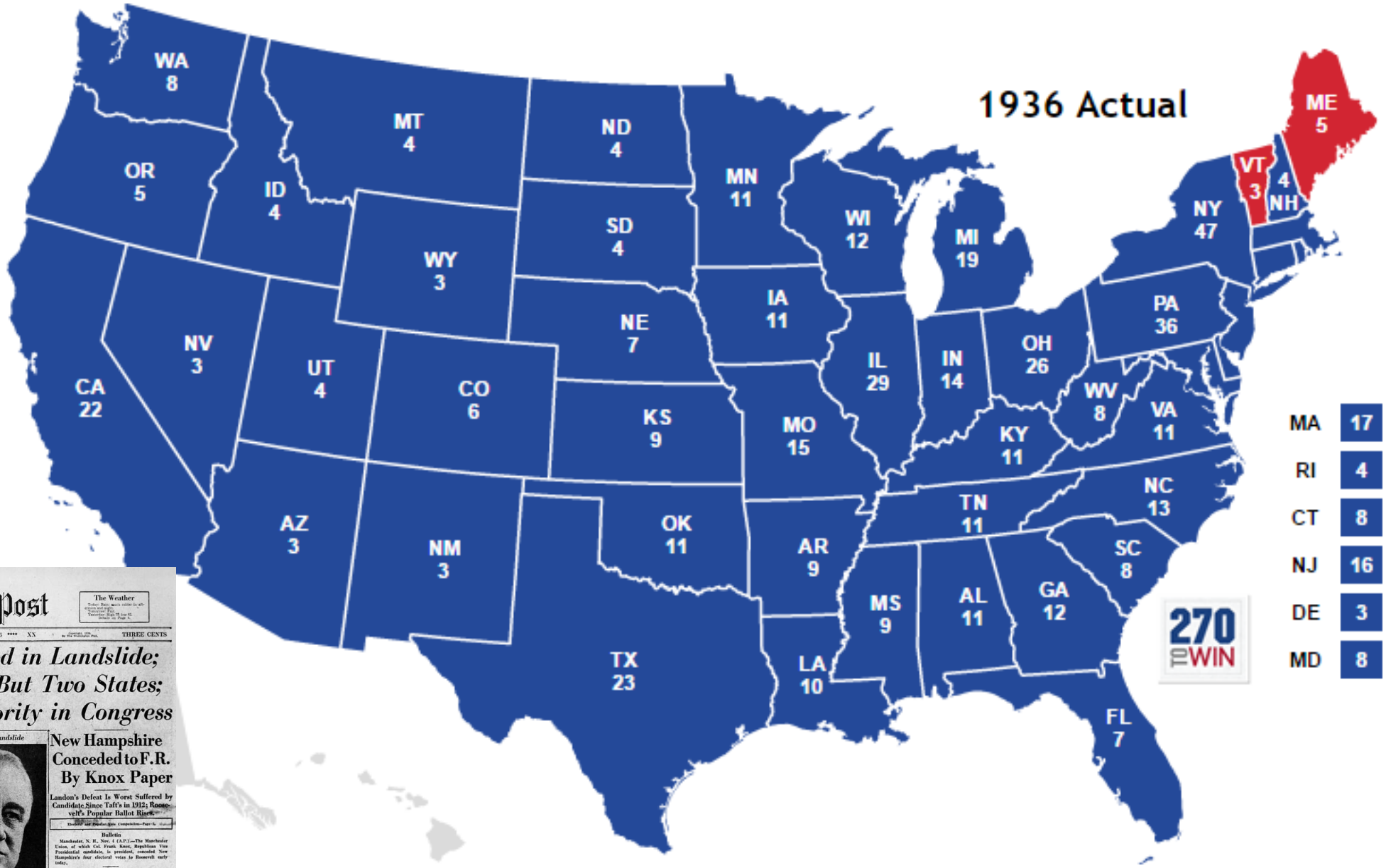
lican National Committee purchased THE LITERARY DIGEST?" And all types and varieties, including: "Have the Jews purchased

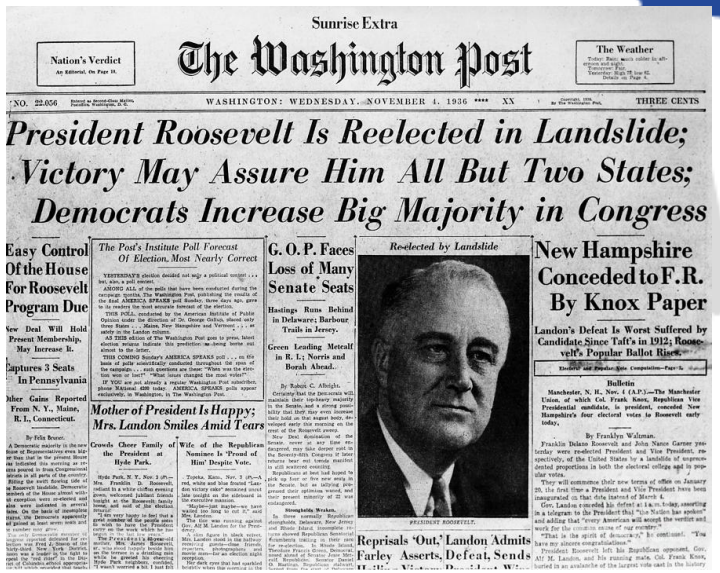
returned and let the people of the Nation draw their conclusions as to our accuracy. So far, we have been right in every Poll. Will we be right in the current Poll? That, as Mrs. Roosevelt said concerning the President's reelection, is in the 'lap of the gods.'

"We never make any claims before election but we respectfully refer you to the opinion of one of the most quoted citizens

- Landon would get 57% of the vote against Roosevelt's 43%
- Asked 1 out of 4 Americans.

Actual results were 62% for Roosevelt against 38% for Landon





Sunrise Extra

Nation's Verdict

The Washington Post

The Weather

WASHINGTON: WEDNESDAY, NOVEMBER 4, 1936 •••• XX

THREE CENTS

President Roosevelt Is Reelected in Landslide; Victory May Assure Him All But Two States; Democrats Increase Big Majority in Congress

Easy Control Of the House For Roosevelt Program Due

The Post's Institute Poll Forecast Of Election, Most Nearly Correct

G. O. P. Faces Loss of Many Senate Seats

Hastings Runs Behind in Delaware; Barbour Trails in Jersey.

Green Leading Metcalf in R. I; Norris and Borah Ahead.

Re-elected by Landslide

New Hampshire Conceded to F.R. By Knox Paper

Landon's Defeat Is Worst Suffered by Candidate Since Taft's in 1912; Roosevelt's Popular Ballot Rises.

Bulletin

Mother of President Is Happy; Mrs. Landon Smiles Amid Tears

Crowds Cheer Family of the President at Hyde Park.

Wife of the Republican Nominee Is 'Proud of Him' Despite Vote.

Reprisals 'Out,' Landon Admits

Farley Asserts, Defeat, Sends

WHAT WENT WRONG *with the*
PRESIDENTIAL POLLS?

IS OUR FACE RED?

Read the November 14th issue of
The Literary Digest
10¢

Not Random Sampling!

Based on every telephone directory in the United States, lists of magazine subscribers, rosters of clubs and associations, a mailing list of about 10 million names was created.

Not Random Sampling!

Based on every telephone directory in the United States, lists of magazine subscribers, rosters of clubs and associations, a mailing list of about 10 million names was created.

Non-response Bias

Only 2.4 million of 10 million returned their ballots.

Not Random Sampling!

Based on every telephone directory in the United States, lists of magazine subscribers, rosters of clubs and associations, a mailing list of about 10 million names was created.

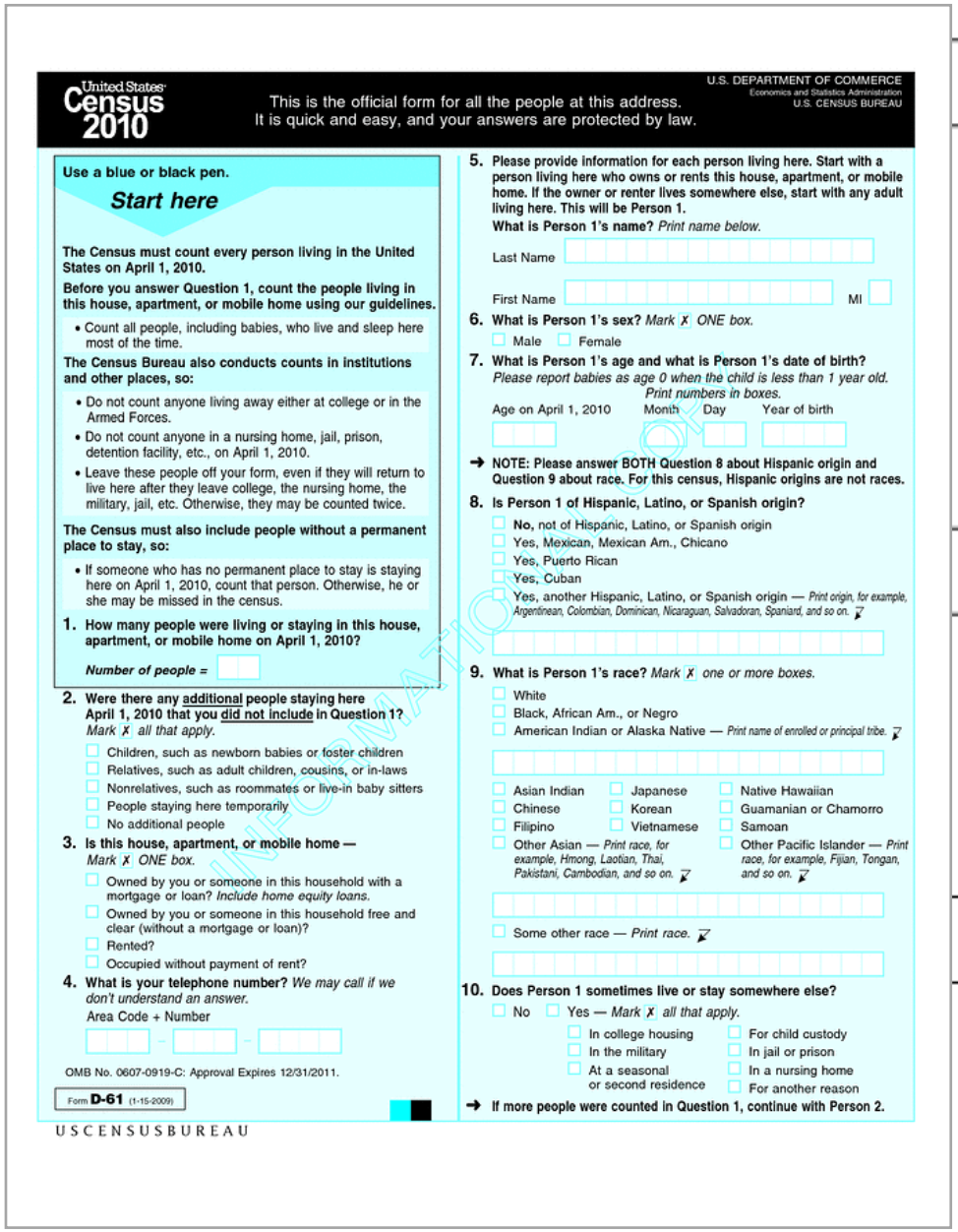
Non-response Bias

Only 2.4 million of 10 million returned their ballots.

Sampling Errors are not fixed by a large sample size.

Simple Random Sample.

- Use Excel to generate a series of random numbers.
- Poll those people.
- Benefit: members are likely to represent population well
- Problem: you need a list of all the people.



United States Census 2010

U.S. DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. CENSUS BUREAU

This is the official form for all the people at this address.
It is quick and easy, and your answers are protected by law.

Use a blue or black pen.

Start here

The Census must count every person living in the United States on April 1, 2010.

Before you answer Question 1, count the people living in this house, apartment, or mobile home using our guidelines.

- Count all people, including babies, who live and sleep here most of the time.

The Census Bureau also conducts counts in institutions and other places, so:

- Do not count anyone living away either at college or in the Armed Forces.
- Do not count anyone in a nursing home, jail, prison, detention facility, etc., on April 1, 2010.
- Leave these people off your form, even if they will return to live here after they leave college, the nursing home, the military, jail, etc. Otherwise, they may be counted twice.

The Census must also include people without a permanent place to stay, so:

- If someone who has no permanent place to stay is staying here on April 1, 2010, count that person. Otherwise, he or she may be missed in the census.

1. How many people were living or staying in this house, apartment, or mobile home on April 1, 2010?

Number of people =

2. Were there any additional people staying here April 1, 2010 that you did not include in Question 1? *Mark ✗ all that apply.*

- Children, such as newborn babies or foster children
- Relatives, such as adult children, cousins, or in-laws
- Nonrelatives, such as roommates or live-in baby sitters
- People staying here temporarily
- No additional people

3. Is this house, apartment, or mobile home — *Mark ✗ ONE box.*

- Owned by you or someone in this household with a mortgage or loan? *Include home equity loans.*
- Owned by you or someone in this household free and clear (without a mortgage or loan)?
- Rented?
- Occupied without payment of rent?

4. What is your telephone number? *We may call if we don't understand an answer.*

Area Code + Number

OMB No. 0607-0919-C: Approval Expires 12/31/2011.

Form **D-61** (1-15-2009)

5. Please provide information for each person living here. Start with a person living here who owns or rents this house, apartment, or mobile home. If the owner or renter lives somewhere else, start with any adult living here. This will be Person 1.

What is Person 1's name? *Print name below.*

Last Name

First Name MI

6. What is Person 1's sex? *Mark ✗ ONE box.*

Male Female

7. What is Person 1's age and what is Person 1's date of birth?

Please report babies as age 0 when the child is less than 1 year old.
Print numbers in boxes.

Age on April 1, 2010 Month Day Year of birth

➜ NOTE: Please answer BOTH Question 8 about Hispanic origin and Question 9 about race. For this census, Hispanic origins are not races.

8. Is Person 1 of Hispanic, Latino, or Spanish origin?

- **No,** not of Hispanic, Latino, or Spanish origin
- Yes, Mexican, Mexican Am., Chicano
- Yes, Puerto Rican
- Yes, Cuban
- Yes, another Hispanic, Latino, or Spanish origin — *Print origin, for example, Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.*

9. What is Person 1's race? *Mark ✗ one or more boxes.*

- White
- Black, African Am., or Negro
- American Indian or Alaska Native — *Print name of enrolled or principal tribe.*
- Asian Indian
- Japanese
- Native Hawaiian
- Chinese
- Korean
- Guamanian or Chamorro
- Filipino
- Vietnamese
- Samoan
- Other Asian — *Print race, for example, Hmong, Laotian, Thai, Pakistani, Cambodian, and so on.*
- Other Pacific Islander — *Print race, for example, Fijian, Tongan, and so on.*
- Some other race — *Print race.*

10. Does Person 1 sometimes live or stay somewhere else?

No Yes — *Mark ✗ all that apply.*

- In college housing
- In the military
- At a seasonal or second residence
- For child custody
- In jail or prison
- In a nursing home
- For another reason

➜ If more people were counted in Question 1, continue with Person 2.

USCENSUSBUREAU

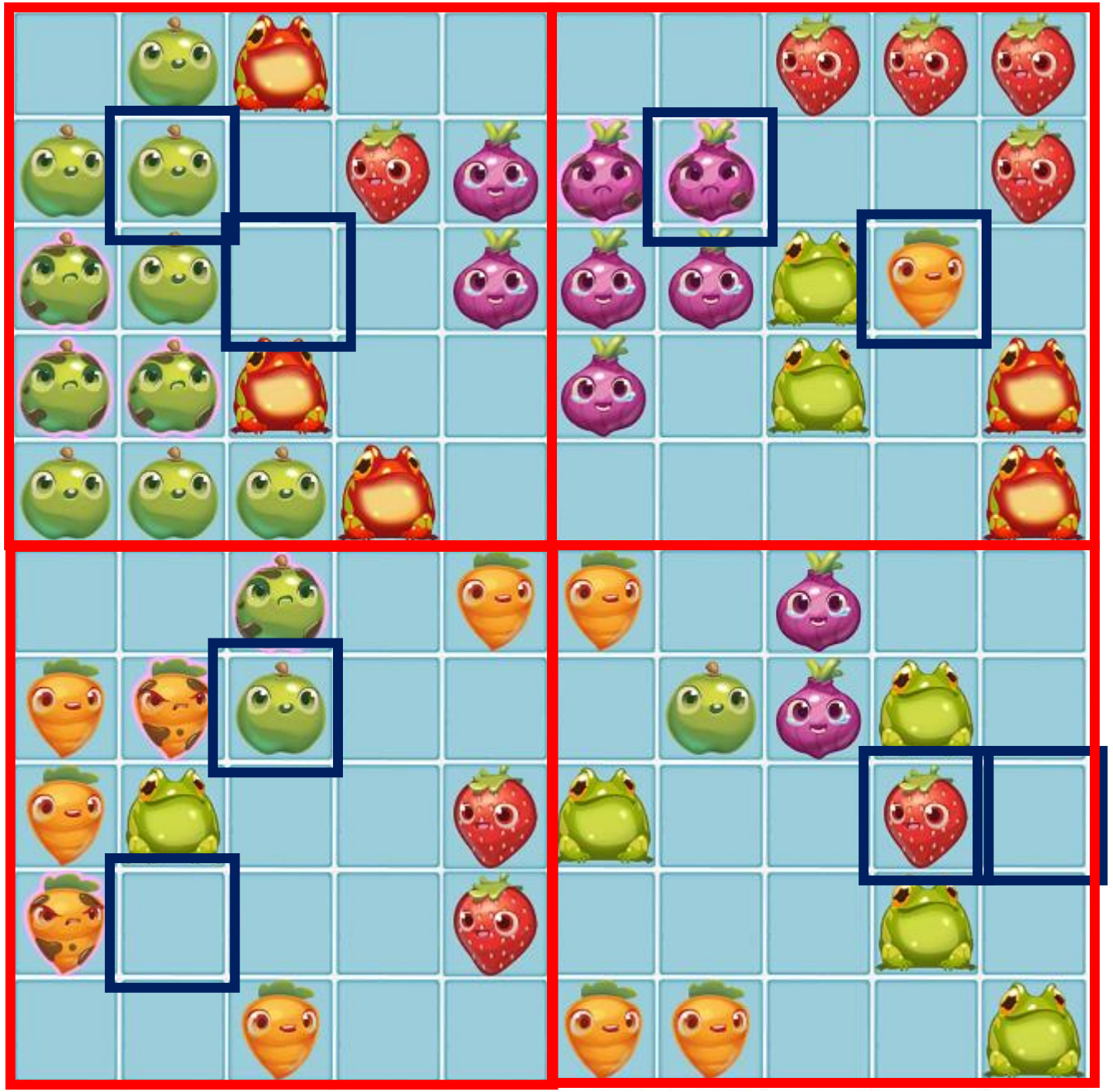
2006	2011	2016
CENSUS FORMAT		
SHORT FORM + LONG FORM (MANDATORY)	SHORT FORM + NATIONAL HOUSEHOLD SURVEY (VOLUNTARY)	SHORT FORM + LONG FORM (MANDATORY)
RESPONSE RATE		
LONG FORM **94%** (SENT TO 1 IN 5 HOUSEHOLDS)	NATIONAL HOUSEHOLD SURVEY **69%** (SENT TO 1 IN 3 HOUSEHOLDS)	LONG FORM **??%** (SENT TO 1 IN 4 HOUSEHOLDS)
PERCENTAGE OF CANADIANS FILLING OUT THE CENSUS ONLINE		
TARGET 20% (18.5% ACHIEVED)	TARGET 40% (54% ACHIEVED)	TARGET 65%

Systematic Sample

- Choose a random start point and then gather every X people after.
- Poll those people.
- Benefit: easier than simple random in a real world situation
- Problem: population may have groups, need large sample.

Stratified Sample

- Choose randomly from naturally occurring groups.
- Poll those people.
- Benefit: easier than simple random in a real world situation
- Problem: groups might not be representative



- Quota sampling introduced by George Gallup to successfully to predict the winner of the 1936, 1940 and 1944 elections.
- Quota sampling forces the sample to fit a certain national profile by using quotas: The sample should have so many women, so many men, so many under 40, so many over 40...

- Quota sampling introduced by George Gallup to successfully to predict the winner of the 1936, 1940 and 1944 elections.
- Quota sampling forces the sample to fit a certain national profile by using quotas: The sample should have so many women, so many men, so many under 40, so many over 40...

1948 Presidential Election:

- Gallup's predicted breakdown of the vote was 50% for Dewey, 44% for Truman, and 6% for third-party candidates Strom Thurmond and Henry Wallace.

- Quota sampling introduced by George Gallup to successfully to predict the winner of the 1936, 1940 and 1944 elections.
- Quota sampling forces the sample to fit a certain national profile by using quotas: The sample should have so many women, so many men, so many under 40, so many over 40...

1948 Presidential Election:

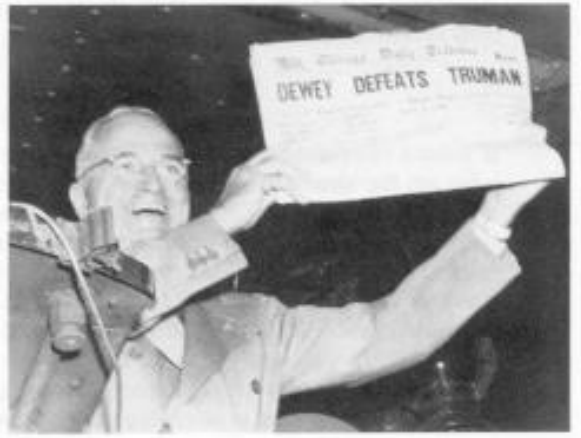
- Gallup's predicted breakdown of the vote was 50% for Dewey, 44% for Truman, and 6% for third-party candidates Strom Thurmond and Henry Wallace.
- The actual results of the election turned out to be almost exactly reversed: 50% for Truman, 45% for Dewey, and 5% for third-party candidates.

- Quota sampling introduced by George Gallup to successfully to predict the winner of the 1936, 1940 and 1944 elections.
- Quota sampling forces the sample to fit a certain national profile by using quotas: The sample should have so many women, so many men, so many under 40, so many over 40...

1948 Presidential Election:

- Gallup's predicted breakdown of the vote was 50% for Dewey, 44% for Truman, and 6% for third-party candidates Strom Thurmond and Henry Wallace.
- The actual results of the election turned out to be almost exactly reversed: 50% for Truman, 45% for Dewey, and 5% for third-party candidates.

Ain't the way I heard it



Chicago Tribune

Convenience Sample

- Choose people near to you.
- Poll those people.
- Benefit: really easy to implement.
- Problem: no accuracy what-so-over.

Actual results were 62% for Roosevelt against 38% for Landon

Remember the problems here! Convenience samples are even worse.

Sunrise Extra

Nation's Verdict

The Washington Post

The Weather

WASHINGTON: WEDNESDAY, NOVEMBER 4, 1936 •••• XX

THREE CENTS

President Roosevelt Is Reelected in Landslide; Victory May Assure Him All But Two State

Democrats Increase Big Majority in Con ess

Easy Control Of the House For Roosevelt Program Due

The Post's Institute Poll Forecast Of Election, Most Nearly Correct

G. O. P. Faces Loss of Many Senate Seats

Hastings Runs Behind in Delaware; Barbour Trails in Jersey.

Green Leading Metcalf in R. I.; Norris and Borah Ahead.

New Hampshire Conceded to F.R. By Knox Paper

Landon's Defeat Is Worst Suffered by Candidate Since Taft's in 1912; Roosevelt's Popular Ballot Rises.

Re-elected by Landslide

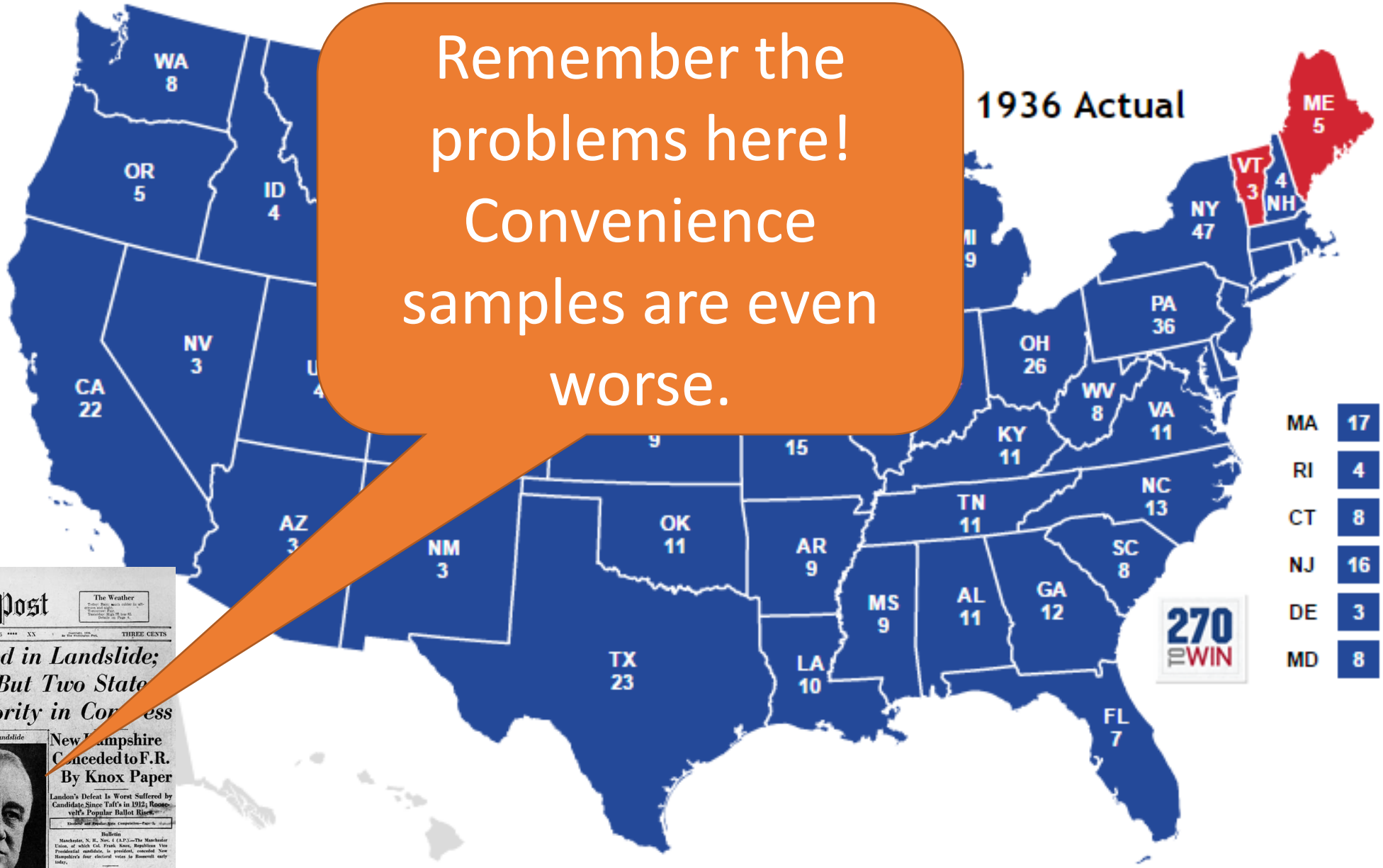
PRESIDENT ROOSEVELT.

Mother of President Is Happy; Mrs. Landon Smiles Amid Tears

Crowds Cheer Family of the President at Hyde Park.

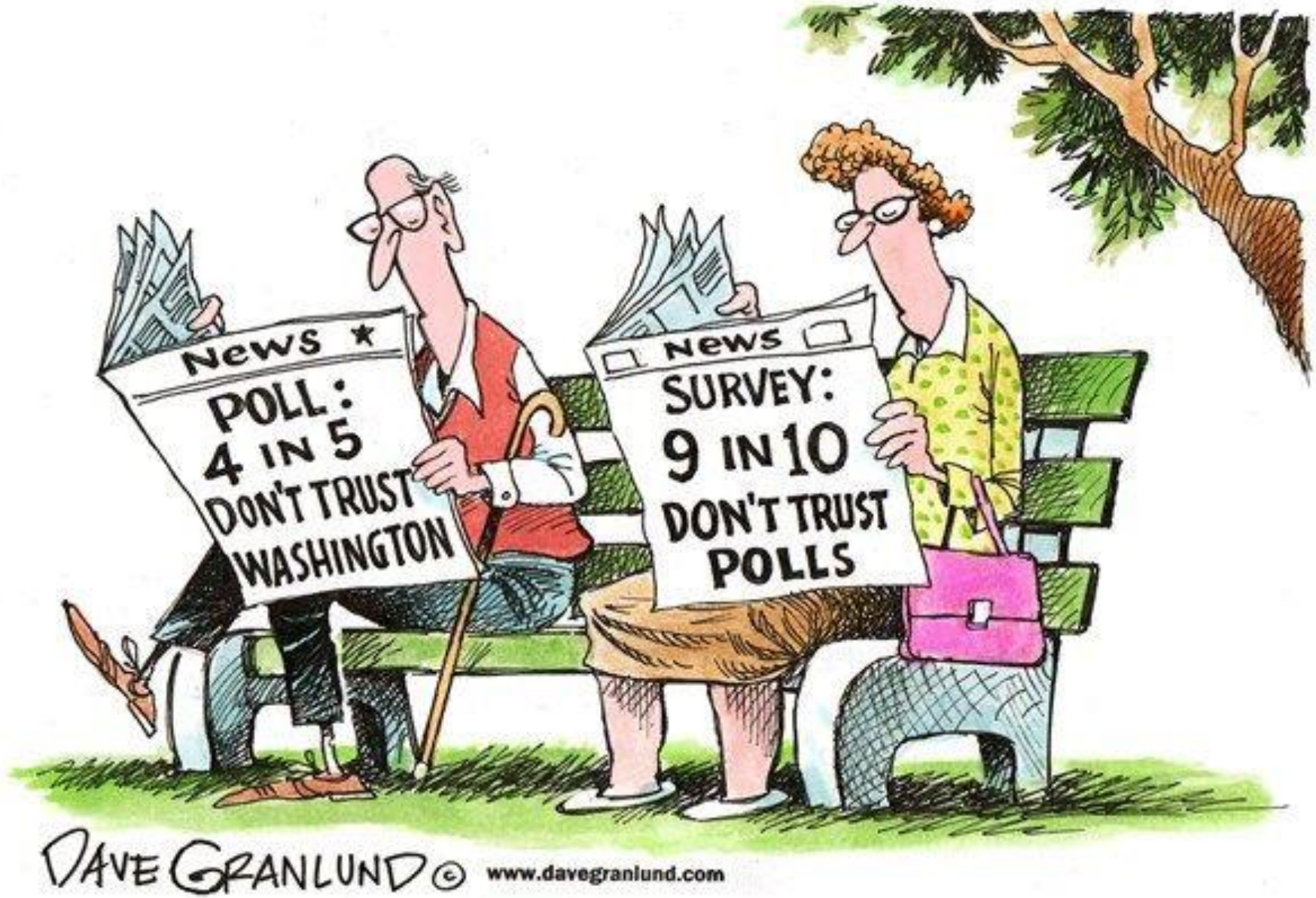
Wife of the Republican Nominee Is 'Proud of Him' Despite Vote.

Reprisals 'Out,' Landon Admits Farley Asserts, Defeat, Sends



Voluntary Response Sample

- Ask who would like to participate
- Poll them
- Benefit: really easy to implement.
- Problem: no accuracy what-so-over.


News ★
POLL:
4 IN 5
DON'T TRUST
WASHINGTON
News
SURVEY:
9 IN 10
DON'T TRUST
POLLS
DAVE GRANLUND ©
www.davegranlund.com

a)How much Replication?	• **How many people were in the study?** • **To be sufficient, there should be thousands.**

a) Sampling Technique?	**Random Sampling** • Subjects are selected from a group/list/phone book using random numbers from a computer or drawing from a hat. **Convenience Sampling** • No list/group/phone book. • Went to a location close by and asked whoever was there. **Voluntary Sampling** • Posted it in a public medium (internet, social media) • Whoever wants to can respond.

a) Identify the Problem Unit	• **Who you want your results to apply to** • **Very general, no specifics** • **Often: "A person"**
a) Identify the Plan Unit	• Who you actually tested • Specific: include who, when, where if possible. • Eg. A university student in the subject pool at the University of Iowa in 2017 (Problem Unit – A person) • Eg. A lab rat in University of Iowa in 2017 (Problem Unit – A person… hmm, that's some diversity bias for you)

a)What are the Diversity Limitations?	• **First think of the subjects in your research pool. Then, think who wasn't included in that group.** • **Generally, studies are conducted from university research pools.** • **The lens of "Power" is helpful here. Groups in power conduct studies on other people in their group. For example, medical studies often are done on university aged white men.**

Law of Large Numbers

A.1.4 determine, through investigation using class generated data and technology-based simulation models (e.g., using a random-number generator on a spreadsheet or on a graphing calculator; using dynamic statistical software to simulate repeated trials in an experiment), the tendency of experimental probability to approach theoretical probability as the number of trials in an experiment increases (e.g., “If I simulate tossing two coins 1000 times using technology, the experimental probability that I calculate for getting two tails on the two tosses is likely to be closer to the theoretical probability of than if I simulate tossing the coins only 10 times”)

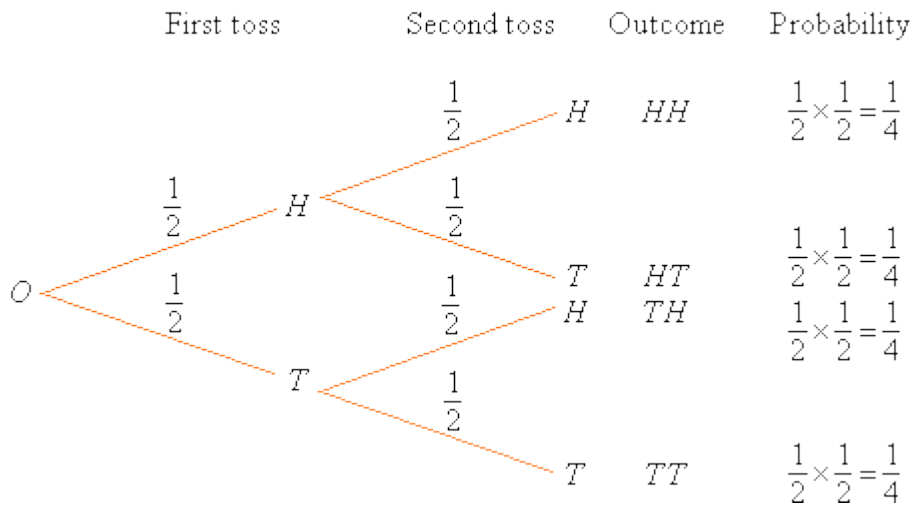
Two ways of getting probability:

1. Run an experiment

Flip a coin

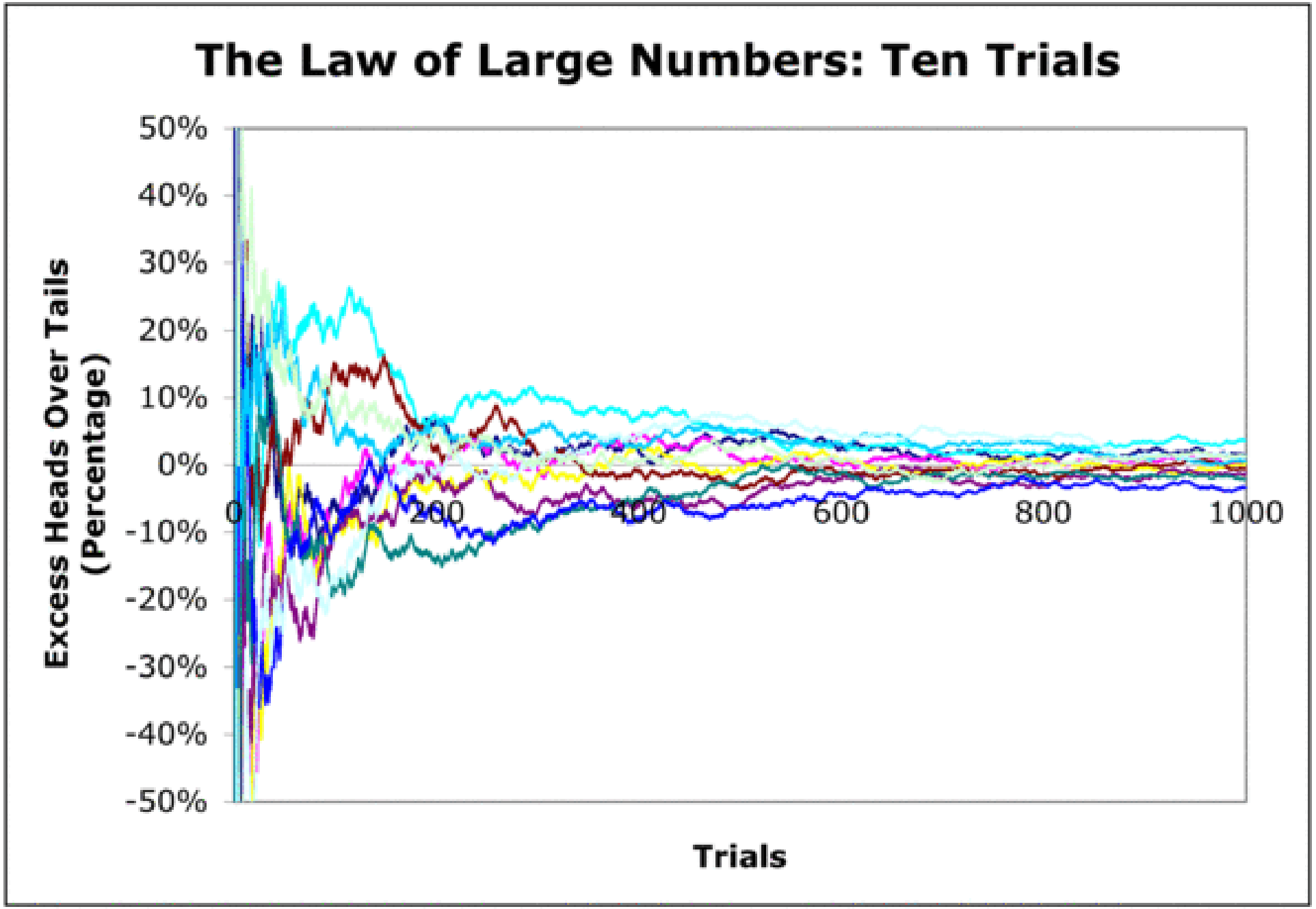
2. Calculate it

Draw the tree



If you run an experiment a really, really, really, large number of times, then your experimental probability will approach your calculated (theoretical) probability.

Number of Tosses	Number of Heads	Probability of Heads
4	1	25%
100	64	64%
1000	582	58.2%
10,000	4989	49.89%


The Law of Large Numbers: Ten Trials
Excess Heads Over Tails (Percentage)
50%
40%
30%
20%
10%
0%
-10%
-20%
-30%
-40%
-50%
0
200
400
600
800
1000
Trials

A Gambler is making a bet at a roulette table.
They are betting on RED or BLACK.
The last few rounds:

RED
RED
BLACK
BLACK
RED
RED
RED

They say: I'm going to bet on BLACK because that's got to come up soon.

Small samples often yield more extreme results that large ones.

Large ones will be close to the theoretical probability.

Small ones might be far off the theoretical probability.

Normally, Toronto has a murder every two weeks.

Last week there were three murders.

It’s time to increase the police force!!

In determining the accuracy of a statistic, the sample size is EXTREMEMLY important.

A common mathematical error is that people ignore the sample size.