PS0700
Survey Research and Sampling

Political Science Research Methods

Professor Steven Finkel
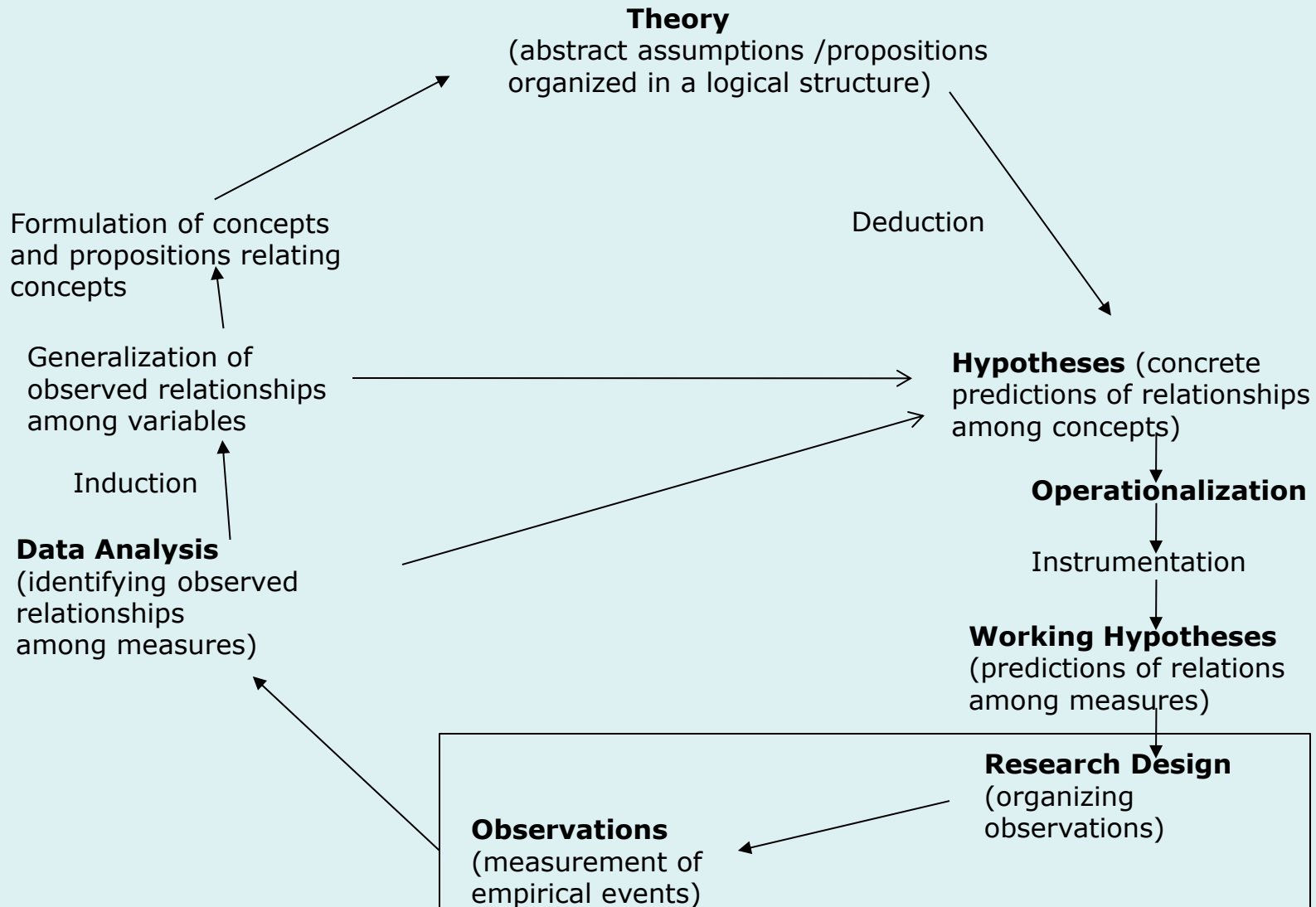
Fall Semester 2022

Week 8

Goals for the Sessions

- Discuss survey research as a means of data collection and as a general research strategy in political science

- Discuss strengths and weaknesses of different kinds of surveys (telephone, in-person, internet, etc.)

- Discuss issues related to sampling

- Discuss issues related to questionnaire design and measurement error in surveys

A Model of the Research Process

Theory
(abstract assumptions /propositions
organized in a logical structure)

Deduction

Formulation of concepts
and propositions relating
concepts

Generalization of
observed relationships
among variables

Hypotheses (concrete
predictions of relationships
among concepts)

Operationalization

Induction

Instrumentation

Data Analysis
(identifying observed
relationships
among measures)

Working Hypotheses
(predictions of relations
among measures)

Research Design
(organizing
observations)

Observations
(measurement of
empirical events)

What is Survey Research?

- Survey research is a **method for collecting data** by asking questions to a *sample* of individuals in a structured fashion and coding the responses

- Survey research is also a **research strategy** in its own right, like experiments and quasi-experiments, etc. We use surveys as one of the most prominent kinds of "passive observation designs" to:
  - obtain information on individual attitudes, perceptions, behaviors
  - generalize the information and the relationships we find in our sample to an overall population of interest
  - It is the workhorse methodology for the political science field of "Political Behavior"

- Some additional complexities:
  - Experiments (or quasi-experiments) often use surveys as the mode of data collection
  - Surveys often use experimental manipulations *within* the survey itself (in so-called "survey experiments")

Surveys are used to: *Describe, Predict, Explain, Explore*

- To *Describe* the characteristics, attitudes, perceptions, beliefs, behaviors of a larger population

  - How many people support Joe Biden's performance as President? https://fivethirtyeight.com; www.gallup.com

  - How many people were unemployed and looking for jobs in the past month? http://www.census.gov/cps/

  - What do people across the world think about democracy and the democratic performance of their political systems?
    www.worldvaluessurvey.org
    www.americasbarometer.org (Latin American political survey data)

- See, e.g., Pew Research Center (www.people-press.org) for excellent public opinion/attitudes/behavior data and reports

- To **Predict** future behavior

  - What is the likely outcome of upcoming Presidential primaries, general elections, congressional elections, or other contests? www.fivethirtyeight.com

  - What kinds of individuals are most likely to get divorced? To be victims of crime? To use illegal drugs or engage in binge drinking?

  - Which individuals are most likely to vote, join political groups, protest, participate in violent political actions?

- To *Explain* attitudes, perceptions, beliefs, behavior

  – What explains the levels and declines in happiness in the US?
  – What explains why some individuals vote and some do not?
  – What explains why individuals voted for Biden versus Trump in 2020? www.electionstudies.org
  – What factors explain individuals' support for democratic political systems?

- These ideas lead to hypotheses and testing of possible causal relationships between variables!

- We then use surveys to develop empirical models that explain individual opinions and behavior

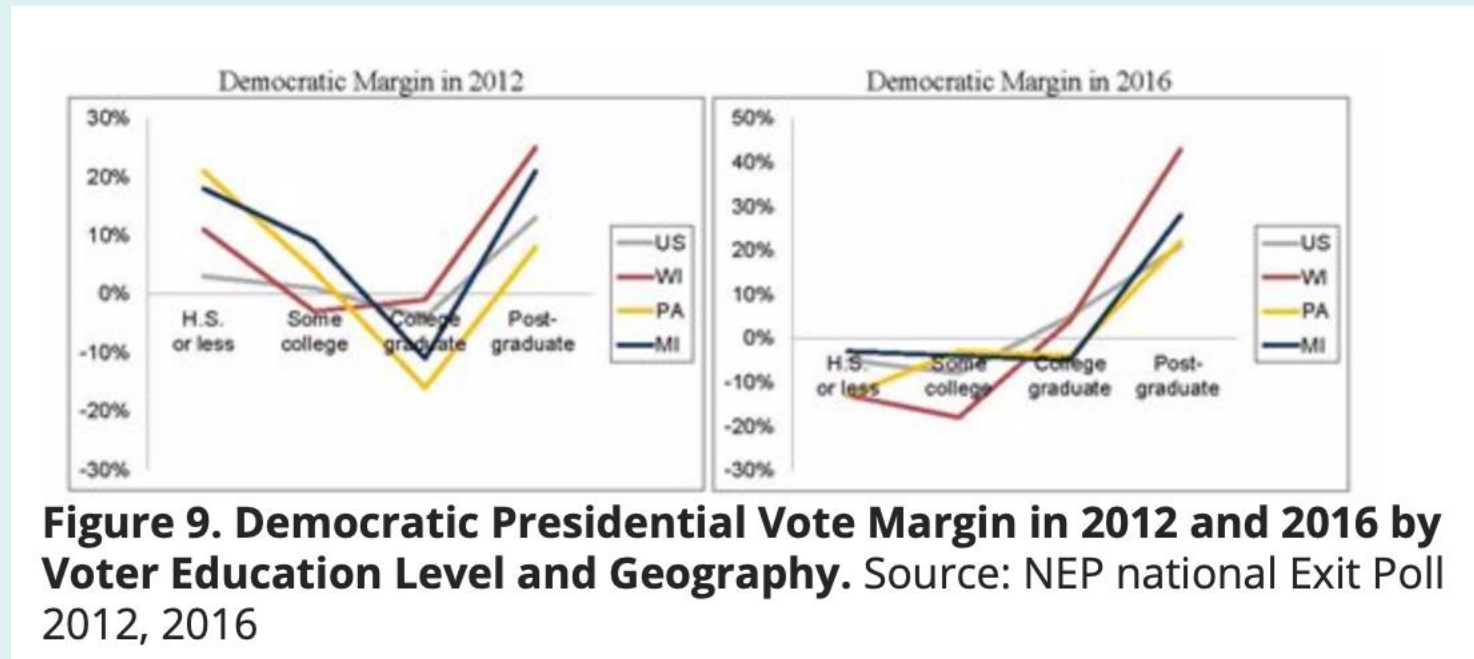# Survey-based Explanations of Voting Behavior in the 2012 and 2016 Elections

Figure 9. Democratic Presidential Vote Margin in 2012 and 2016 by Voter Education Level and Geography. Source: NEP national Exit Poll 2012, 2016

Survey-based Explanations of Voting Behavior in the 2012 and 2016 Elections
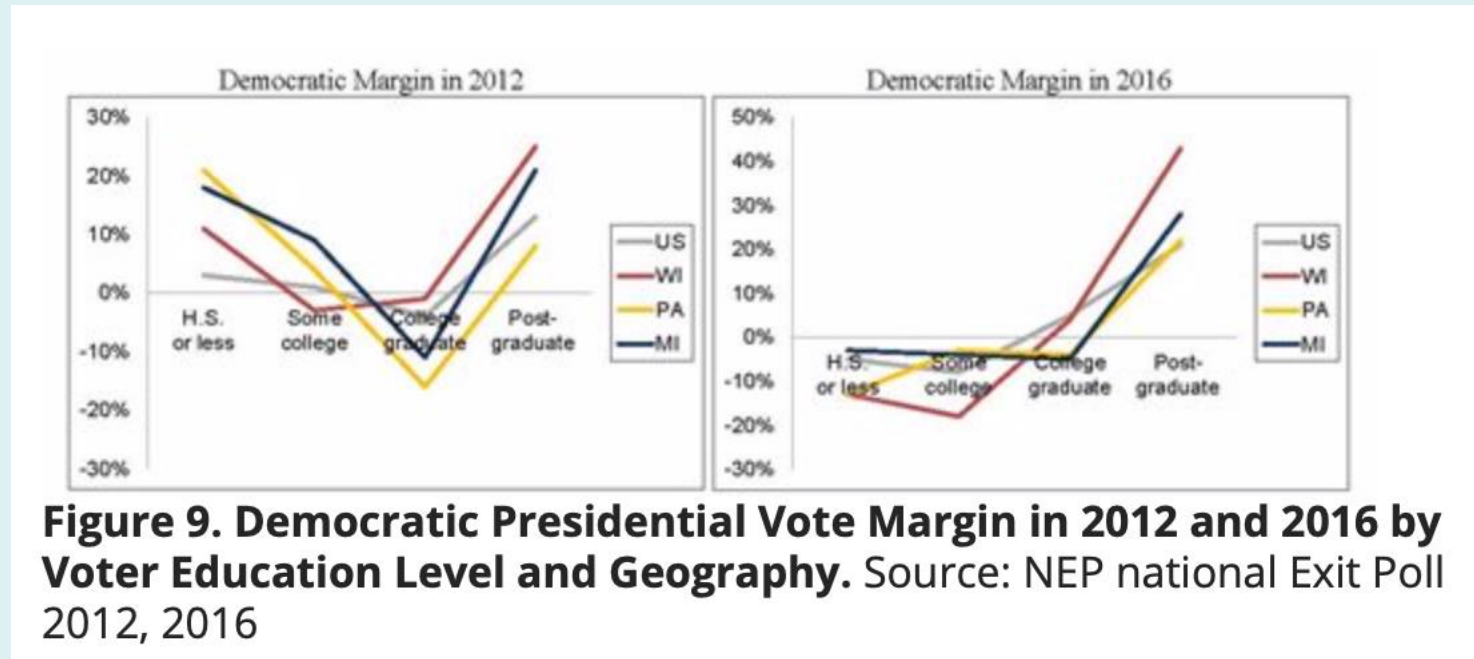

Figure 9. Democratic Presidential Vote Margin in 2012 and 2016 by Voter Education Level and Geography. Source: NEP national Exit Poll 2012, 2016

The effect of education changed dramatically from 2012-2016, especially at the lower ends of the education scale, and in key battleground states – survey research revealed this important factor in accounting for Trump victory. But why did the change occur? We need deeper analyses (surveys and non-surveys)

Mutz, "Status threat, not economic hardship, explains the 2016 presidential vote" *PNAS* (2018)

Table S4. Cross-sectional analysis of predictors of Trump support, 2016

Predictors	Trump thermometer advantage		Trump vote preference		Trump/Clinton vote	
	Coefficient	t Value	Coefficient	z Value	Coefficient	z Value
Party identification (Democratic)	−2.340	−25.010***	−1.107	−14.050***	−1.822	−13.880***
Education (not college graduate)	0.173	1.140	0.140	0.880	0.068	0.260
Race (white)	1.203	6.990***	0.591	3.080**	1.216	4.250***
Gender (female)	−0.548	−4.030***	−0.009	−0.060	−0.473	−2.070*
Age	−0.196	−4.380***	0.019	0.420	−0.151	−2.010*
Religiosity	0.029	1.130	0.033	1.290	0.063	1.450
Economic hardship/anxiety						
Income	0.017	0.960	0.048	2.600**	0.031	1.060
Looking for work	0.065	0.250	0.173	0.590	−0.035	−0.080
Concern about future expenses	0.042	0.430	−0.023	−0.230	0.016	0.100
Perceptions of family finances (better)	−0.001	−0.020	0.047	0.610	0.124	0.950
Support better safety net	−0.337	−4.180***	−0.154	−1.870	−0.350	−2.570*
Immediate economic context						
Median income	0.000	0.550	0.000	−1.210	0.000	−1.700
Unemployed, %	−3.107	−1.500	−2.832	−1.310	−6.116	−1.760
Manufacturing, %	0.686	0.630	−1.122	−1.090	−0.760	−0.420
Perceived status threat						
Perceive discrimination against high-status groups > low-status groups	0.565	8.060***	0.345	4.630***	0.572	4.600***
American way of life threatened	0.129	1.360	0.243	2.200*	0.330	1.930*
SDO	0.107	2.390*	0.077	1.720	0.144	1.940*
Domestic prejudice	0.098	1.580	0.124	1.960*	0.139	1.420
Support for isolationism	0.262	2.960**	−0.106	−1.200	0.266	1.750
China as opportunity	0.231	1.990*	0.080	0.680	0.354	1.900
Support for immigration	−0.776	−9.510***	−0.815	−10.020***	−1.050	−8.160***
Support for international trade	−0.302	−4.400***	−0.182	−2.650**	−0.315	−2.830**
National superiority	0.046	0.540	0.159	1.800	0.149	1.020
National economy (better)	−0.824	−10.970***	−0.376	−5.350***	−0.739	−6.210***
Terrorist threat	−0.135	−1.380	0.203	1.890	−0.079	−0.480
Constant	22.839	23.490***	2.640	2.610**	8.987	5.340***
R^2/pseudo-R^2	0.69		0.56		0.78	
Sample size	2,600		2,845		2,175	

Data were collected by Amerispeak/NORC, October 2016. All variables are described in detail in *Cross-Sectional Survey*. Trump thermometer rating is on a 20-point scale. Trump vote preference is dichotomous, indicating support for Trump (one) or anyone else (zero). Trump/Clinton vote is a dichotomous indicator of voting for Trump (one) or Clinton (zero), with third party voters eliminated. Trump thermometer advantage is analyz[ed using l]east squares regression. Trump vote preference and Trump/Clinton vote are analyzed using logit regression. *$P < 0.05$; **$P < 0.01$; ***$P < $

- To **Explore** potential changes in opinions, attitudes, beliefs that may be produced by changes in the information environment
  - Makes use of experimental methods *within* surveys
  - Will voters respond more positively to a political issue that is "framed" in one way as opposed to another?
  - Can people be "talked out" of their initial opinion and change their minds in response to particular arguments? Are people on one side of the issue more easily swayed than the other?
  - Should a candidate stress issue X or issue Y in her advertisements in order to maximize voter support?
- Examples at http://www.tessexperiments.org/introduction.html

General Strengths of the Survey Method

- With proper sampling methods and good questionnaire design, surveys can yield:
  - generally high *predictive accuracy* for many future behaviors
  - generally high *descriptive accuracy* of attitudes, perceptions, behaviors
  - Accurate assessments of the *covariation* between independent variables (X) and dependent variables (Y) in overall populations of interest
- Taken together, this means that relationships found in surveys have a good deal of **external validity**, and are usually **superior** to laboratory and field experiments in this regard
- Surveys can often be done across multiple contexts (countries, states, neighborhoods) simultaneously and/or over time, and this kind of "micro-macro" integration can be a very powerful combination.

- Survey researchers have access to standardized questions in many substantive areas of political science. This means that measurement error *to some extent* have been reduced through trial/error over the years

- Combining experiments *within* surveys is an *extremely* powerful way to explore and explain individual attitudes and behaviors, and to measure attitudes on sensitive topics

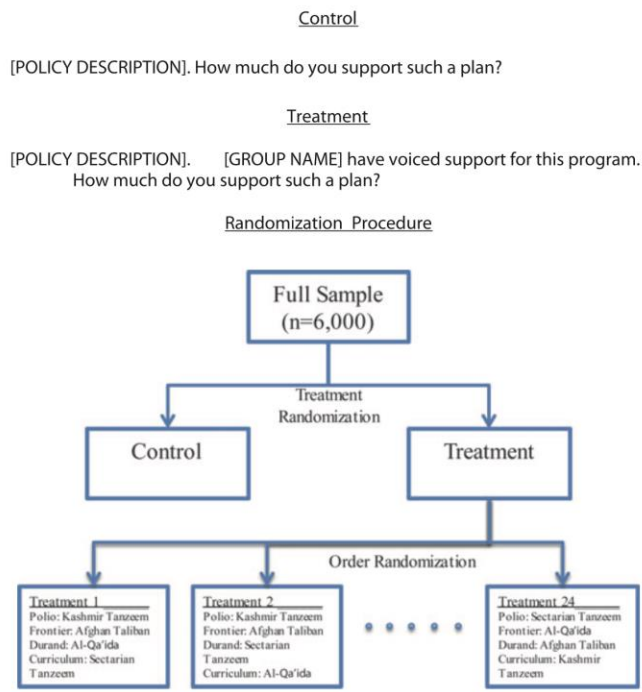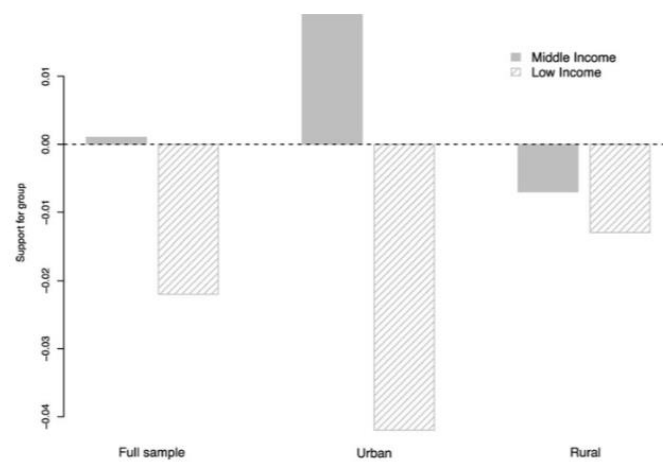FIGURE 1 Illustration of the Endorsement Experiment

Control

[POLICY DESCRIPTION]. How much do you support such a plan?

Treatment

[POLICY DESCRIPTION].   [GROUP NAME] have voiced support for this program. How much do you support such a plan?

Randomization Procedure

Full Sample (n=6,000)

Treatment Randomization

Control   Treatment

Order Randomization

Treatment 1
Polio: Kashmir Tanzeem
Frontier: Afghan Taliban
Durand: Al-Qa'ida
Curriculum: Sectarian Tanzeem

Treatment 2
Polio: Kashmir Tanzeem
Frontier: Afghan Taliban
Durand: Sectarian Tanzeem
Curriculum: Al-Qa'ida

• • • • •

Treatment 24
Polio: Sectarian Tanzeem
Frontier: Al-Qa'ida
Durand: Afghan Taliban
Curriculum: Kashmir Tanzeem

FIGURE 2 Treatment Effects by Income and Strata

Source: Blair *et al.,* 'Poverty and Support for Militant Politics: Evidence from Pakistan", *Am Journal of Political Science* (2013)

Sources of Error in Surveys

- However, there are many potential sources of error in the survey process, many biases that may distort the findings

- Some relate to the *mechanics* of survey research:
  - Sampling procedures and sampling error
  - Mode of data collection (telephone, in-person, internet)
  - Questionnaire construction

- Some are inherent to the method of survey research itself
  - Reliance on verbal reports
  - Passive observational data

- Some are specific to election polling and the use of surveys in predicting election outcomes

National Election Polls Predictive Accuracy, 1936-2020

Orange line represents average absolute error
Bars represent average signed error (**red bars** indicate overestimation of Republican vote margin; **blue bars** indicate overestimation of Democratic vote margin)
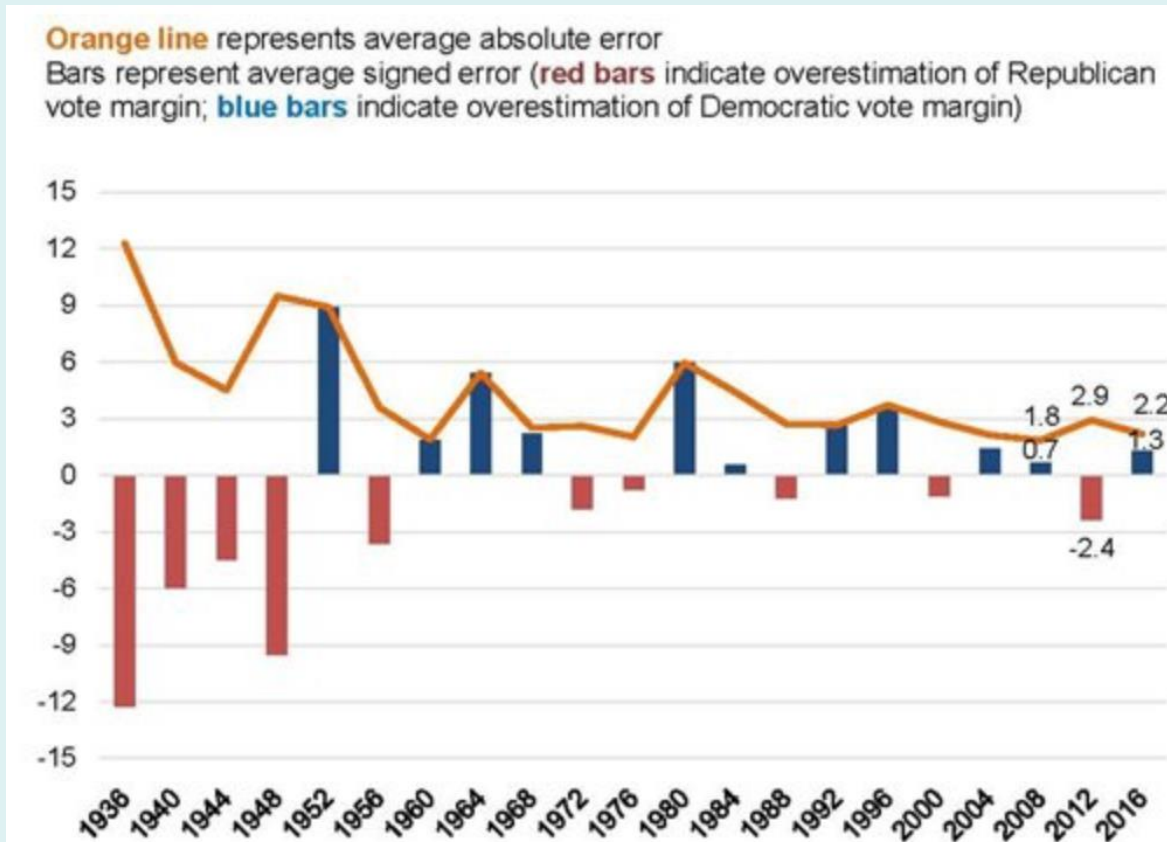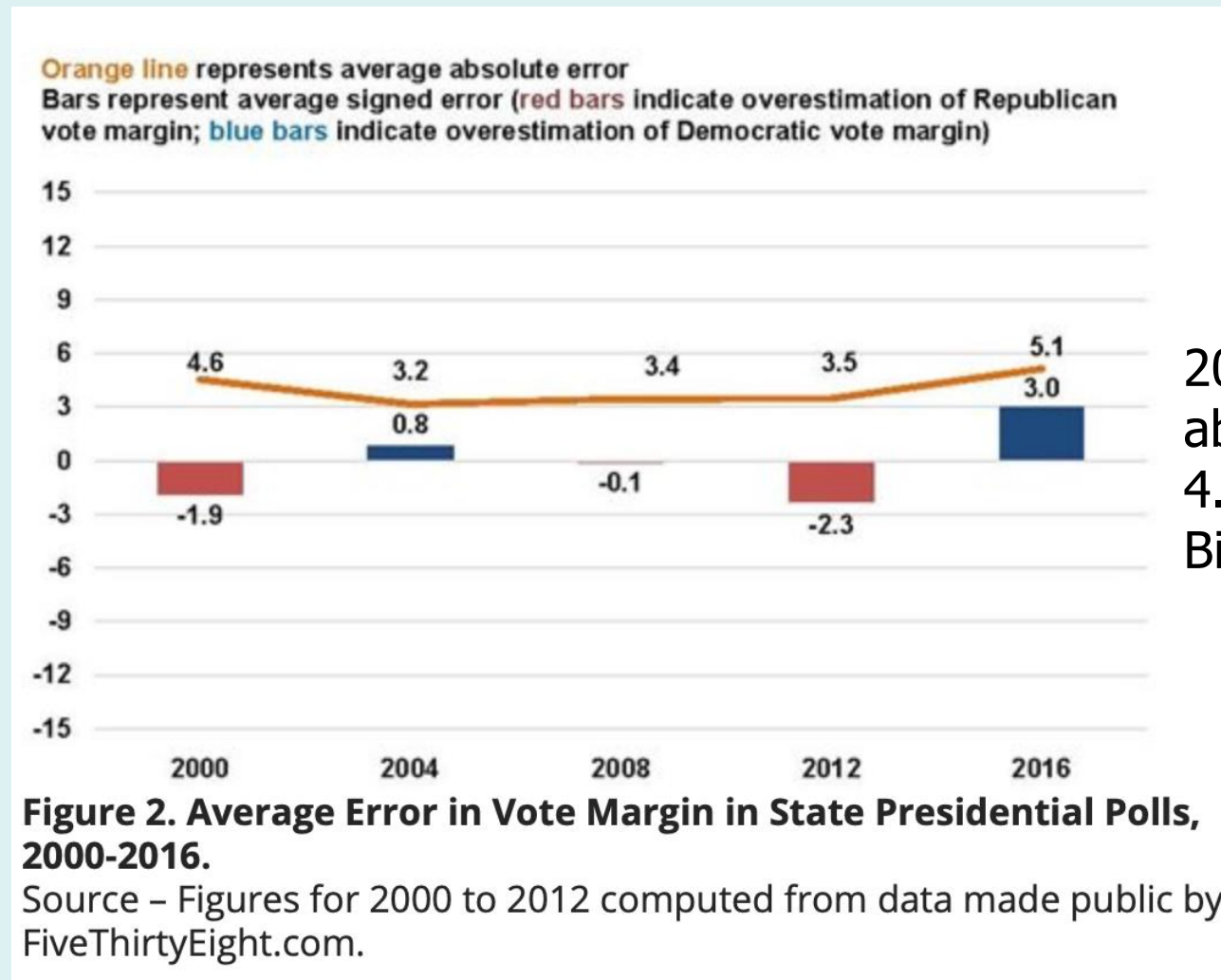
2020: 4.5% absolute error, 3.9% pro-Biden

Figure 1. Average Error in Vote Margin in National Presidential Polls, 1936-2016.

Note – The 2016 figures are based on polls completed within 13 days of the election. Figures for prior years are from the National Council for Public Polls analysis of final poll estimates, some occurring before the 13-day period. Figures for 1936 to 1960 are based only on Gallup.

State Election Polls Predictive Accuracy, 2000-2020

Orange line represents average absolute error
Bars represent average signed error (red bars indicate overestimation of Republican vote margin; blue bars indicate overestimation of Democratic vote margin)

Figure 2. Average Error in Vote Margin in State Presidential Polls, 2000-2016.
Source – Figures for 2000 to 2012 computed from data made public by FiveThirtyEight.com.

2020: 5.1% absolute error, 4.3% pro-Biden

Some Reasons Why Polls Predict Elections Incorrectly

- A Truism: Events intervene, voters change their mind
  - Polls farther away from election are more unreliable
  - Polls in contexts with large numbers of undecided voters more unreliable
- Sampling Error (i.e., "± 3%" error in surveys – we will discuss where this number comes from this in a few weeks)
- Pollsters use different methods for assessing how undecided voters are "leaning"
- In low-turnout contexts such as the U.S., polling firms must estimate who will be in the electorate, i.e., who is a "likely voter".
  - **This is an inexact science with different methods uses by different polling outfits**
  - **This is especially difficult in U.S. primary elections and caucuses**, where there are state-by-state and party-by-party differences in eligibility and election regulations, and where turnout is generally very low
  - This is one reason why exit (or entrance) polls, taken on the day of the election with voters exiting the polling place, are typically more accurate (though they have other problems, including sometimes very small sample sizes for subgroups in the voting population)

- In some election contexts pitting white candidates against African-American candidates, the so-called "Bradley effect" has led pollsters to *overestimate* the vote for the African-American candidate

- Recently it appears that Republican-leaning voters are less likely to agree to be interviewed, possibly due to low trust in government and in the institution of polling

- It may be that there is recently more social desirability bias *against* revealing support for Trump and MAGA-aligned candidates (i.e., the so-called "shy Trump voters")

- This is potentially overcome by **weighting the data to bring it in line with expected demographic and political distributions of the electorate, for example, education and party affiliation.** But this assumes that the high educated/low educated people who are in the sample are the same as the high educated/low educated people who declined to participate.

Recommendations for the Consumer of Election Polls

- Read and critically assess the methodology of the survey!

- Use multiple surveys taken in the same time period to get better sense of candidate support and trends in the campaign

- Trust polling aggregators such as www.fivethirtyeight.com over individual polling web sites

- Read respected pollster's methodology pages (e.g., Pew Research Center, AmericasBarometer)

Sampling Strategies in Survey Research

- Goal: Ensure that sample of individuals selected to be interviewed is representative of the population of interest
- Probability Samples: Each person in the population (of interest) has a ***known*** chance of being selected for inclusion in the sample
- Non-Probability Samples: no known chance of selection
  - Telephone call-in surveys
  - Surveys on web sites
  - Robocall surveys
  - Convenience samples
  - Snowball samples
  - Quota samples
- It is **only** with probability samples that we can accurately make generalizations from our sample to the overall population – with a known amount of "sampling error" (e.g., $\pm$ 3% for simple random samples of approximately 1000 individuals)

Types of Probability Samples

1. Simple Random Sample (SRS)

2. Systematic Sampling: Every *kth* member of population list

3. Stratified Sampling: Allows for *a priori* allocation of certain numbers of interviews to different sub-groups of the population

Most Large-Scale Surveys in Practice Use One of the Following:

4. Multi-Stage Cluster Sampling for Areas with Census Information

5. Random Route Sampling in Areas without Census Information

6. Random Digit Dialing (RDD) Sampling for Telephone Surveys (including banks relating to both cell phones and landlines)

7. Two stage surveys with RDD or "snail mail" initial contacts, followed by internet "opt-in" and tracking over time

Modes of Survey Data Collection

- Face to Face: Excellent for getting representative samples, good for generating rapport, poor for sensitive topics (that's why survey experiments can be valuable), most susceptible to interviewer and social desirability effects

- Telephone Surveys: Cheaper than face to face, good to excellent for getting representative samples -- if care is taken to reach respondents given high refusal rate (see NYT, October 13, 2022, "Who in the World is Still Answering Pollsters' Phone Calls?; less susceptible to interviewer effects but still vulnerable to social desirability effects

- Internet: Cheapest; least effective in getting representative samples, little control over who is completing the survey, least susceptible to social desirability effects and no interviewer effects whatsoever

- Mode Effects in survey responses exist, but are not *overwhelmingly* large for the most part. See, e.g. Pew's methods video on the topic.

Questionnaire Construction

- General Goal: Minimize measurement error induced by questions themselves, by the order or groupings of questions, and error induced through interviewer-respondent interaction

- Remember the concepts of indicator "validity" and "reliability"?
    - Does the question measure what it is supposed to measure (validity)?
    - Would asking the question 100 times yield consistent (reliable) responses, or would there be random error around the respondent's "true" score?

- General Guideline: The respondent must **understand** the question, must be **able** to respond to the question, and must be **willing** to respond to the question

Good Questions Should NOT:

- **Be vague:** *"Did you eat breakfast today?"*

- **Use Compound ("double-barrel") questions: "***Do you agree or disagree that legislators are honest and hard working?"*

- **Use biased wording:** *"Do you believe the Republican/Tea Party majority in Congress will heed President Obama's call for cooperation or escalate their attempts to obstruct his agenda"*? [From Democratic National Committee mail survey 2015]

- **Use abbreviations or scientific jargon:** *"Do you prefer the current D'hondt method of representation or the Single Transferable Vote?"*

- **Use objectionable or argumentative questions:** *"Radical minority religious leaders say that Government discriminates against them. Do you agree?"*

- **Use leading questions:** *"Most people think that government is too corrupt. Wouldn't you agree?"*

- Other issues for question design:
  - No overly emotional questions
  - Use a mixture of "closed" and "open-ended" questions to give respondent ways to express less structured opinions
  - Sensitive issues difficult, may require separate self-administered portion of the survey or other creative methods

    This is the logic of the "list experiment" used within the survey context by Kuklinski et al. (1997, discussed earlier in the course) and the "endorsement experiment" by Blair et al. (2012, slide 12 above)

- Minimize the problem of "non-attitudes"
  - Formal balance in questions (i.e., no reason from the wording of the question to answer one way or the other)
  - Vary what a "yes" or "strongly agree" response means in a series of questions to avoid "response set"
  - Use "filters" to give people an opt-out if they don't have an opinion (but probing an initial "don't know" is also okay)
  - Provide a neutral option within a question (but probing afterwards is also okay)

General Limitations of the Survey Approach

- Surveys generally provide only a "snapshot" of opinion at any given time
- Surveys are biased toward psychological explanations of behavior, as individuals are typically analyzed outside of their social context, networks, neighborhoods, workplaces
- Do individuals truly have access to the reasons for their behavior and their preferences? Many psychologists and political scientists think not
- As discussed, mechanical sources of survey errors such as non-response (refusals, unable to contact, hang-ups) impinge on external validity in ways that are sometimes difficult to estimate

- Limitations in surveys for causal inference: The *internal validity* problem
  - As in all passive observational designs, the survey researcher does not control X, but rather obtains information on X from the respondent or from information in the sampling design
  - In single-shot cross-sectional surveys, the causal direction between X and Y is not always clear
  - As we know, all observational studies are limited, as the researcher *cannot rule out completely* the possibility that some other variable(s) Z is responsible for the relationship between X and Y
  - These concerns mitigated for survey experiments, but these studies are only a small part of the survey research universe

- Survey solutions to the causality problem
  - Multivariate statistical analysis: bring all known Zs into the models as "control variables"
  - Panel (Longitudinal) designs for assessing causal order and possible reciprocal causality
  - With multi-wave panel data, also can control for some kinds of "unobserved" Z variables that confound the assessment of a causal relationship between X and Y
- These same issues apply to the use of other kinds of "passive observational" data in testing hypotheses and developing causal explanations