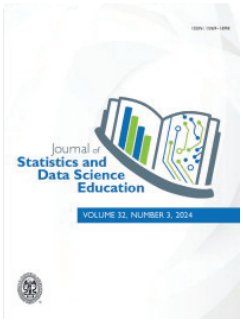


Journal of Statistics and Data Science Education

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ujse21

Taking the Next Step in Exploring the *Literary Digest* 1936 Poll

Beth Chance, Andrew Kerr & Jett Palmer

To cite this article: Beth Chance, Andrew Kerr & Jett Palmer (22 Aug 2024): Taking the Next Step in Exploring the *Literary Digest* 1936 Poll, Journal of Statistics and Data Science Education, DOI: 10.1080/26939169.2024.2395505

To link to this article: https://doi.org/10.1080/26939169.2024.2395505

Taking the Next Step in Exploring the *Literary Digest* 1936 Poll

Beth Chance*, Andrew Kerr, Jett Palmer

Department of Statistics, California Polytechnic State University, San Luis Obispo, CA

*bchance@calpoly.edu

ABSTRACT

While many instructors are aware of the *Literary Digest* 1936 poll as an example of biased sampling methods, this article details potential further explorations for the *Digest's* 1924-1936 quadrennial U.S. presidential election polls. Potential activities range from lessons in data acquisition, cleaning, and validation, to basic data literacy and visualization skills, to exploring one or more methods of adjustment to account for bias based on information collected at that time. Students can also compare how those methods would have performed. One option could be to give introductory students a first look at the idea of "sampling adjustment" and how this principle can be used to account for difficulties in modern polling, but the context is rich in other opportunities that can be discussed at various times in the course or in more advanced sampling courses.

KEY WORDS: Data cleaning, Data validation, Bias, Post-stratification

1. INTRODUCTION

We were recently reading Gelman and Vehtari's (2024) forthcoming text on teaching with stories which mentions Lohr and Brick's (2017) argument that the popular example of sampling bias in the 1936 *Literary Digest* poll has other useful lessons that have been overlooked. Mainly, how, rather than stopping at the bad news, adjusting polling results for known sources of biases could have improved the *Digest's* prediction of the election outcome. We initially started looking for the original data in order to replicate Lohr and Brick's analyses, but found several other interesting aspects for introductory statistics and data science students as well (i.e., we went down a rabbit hole). In this article, we present our data journey and findings. Our goal is to compile and simplify useful information together in one location and to provide sufficient resources (even images!) to allow instructors to design authentic experiences with the data for students at different levels. Our focus is not on exploring causes for the bad prediction (e.g., see Lusinchi, 2012; Squire, 1988), debating how to best evaluate the accuracy of straw polls (e.g., Robinson, 1932), or analyzing the effectiveness of different adjustment methods (e.g., Lohr & Brick, 2017), though we do present some of this detail for additional background for teachers. Instead, our goal is to present enough background, historical context, data access, and suggestions for instructors to utilize these data in several possible ways. These include applying data wrangling principles in an interesting real-world context, using technology for basic data manipulation, and introducing introductory students to the principle of post-stratification. Rather than telling students nothing can be done if sampling methods are biased or convincing them that simple random samples are practical and solve all of our problems, we can raise their awareness of possible alternatives and put

them in the role of data detective. Students may even learn a little history along the way. While the electoral college system will be of less interest outside the United States, learning from errors can be powerful for anyone. As one student commented, "Learning and visualizing the pitfalls of past studies gives me a sharper eye for identifying procedural and analytical weaknesses today."

2. BACKGROUND

The *Literary Digest* (*LD*) was a prominent political magazine in the early 1900s. Starting in 1916, they began conducting multi-state postal polls to predict the U.S. presidential election outcomes, surveying in Illinois, Indiana, New Jersey, New York, and Ohio. They expanded to six states in 1920 (adding California) and went nationwide in 1924 (as well as conducting polls on the Mellon tax and prohibition).

Each election year, *LD* sent out increasing record numbers (e.g., on the order of 18 million in 1928) of hand-addressed postcards, drawing names and addresses from club rosters, city directories, and (mostly) vehicle registration lists and telephone books. Their methods correctly predicted the presidential election outcome each time, until 1936. They, and the nation, were quite surprised when their poll incorrectly predicted a landslide victory for Republican Alf Landon over Democratic (and current president) Franklin Roosevelt. Ten million postcards had been sent throughout the country and almost 2.4 million were returned. Landon (then governor of Kansas) was predicted to receive 54 percent of the popular vote and 370 electoral votes. Instead, Landon won 40 percent of the popular vote and just 8 electoral votes. The magazine went out of business soon after.

This story is often used in introductory statistics courses to caution against sampling bias and nonresponse bias. Students are often quick to point out the problems with the sampling frame (e.g., owners of cars and telephones tended to be wealthier, "white collar," older, and more Republican) and the response rate (e.g., supporters of the challenger may be more likely to respond than supporters of the incumbent). These two arguments provide evidence that the magazine was overrepresenting the Republican voters (although the response rates of 15-20% in the *LD* polls are still larger than many modern polls). Yet, as Lohr and Brick (2017) point out, evidence of such bias had occurred in earlier years, which the *Digest* stubbornly refused to fully consider, citing their high sample sizes ("the most extensive straw ballot in the field", 10/31/1936) and past success, and wanting to allow the readers to "draw their own conclusions" from the published poll results.

What could the *Literary Digest* have done differently? Lohr and Brick (2017) compare several potential adjustments, examining the results that would have been obtained by *LD* in 1928, 1932, 1936, using the information available at that time. However, the article does not provide the raw data or code for replicating the analyses. Our goal in this paper is to provide this ability for instructors and students to play with the raw data. So we first needed to access the data. Along the way, we found several opportunities for classroom use.

3. DATA PROCESSING AND VALIDATION

Our first attempt to find the 1936 published magazine results led us to *History Matters*. This website provides the original article and a data table (Figure 3).

Step one was to confirm that the totals in the table matched the headline of the article. Step two was to copy and paste the data into a spreadsheet program and see whether the totals in the data matched the column totals. Alas, they did not.

TEACHING TIP (Data quality): Provide the website to students and ask them to spot check the data and see whether they can identify the data entry errors.

Visual inspection of the provided data table reveals a few issues:

- Some rows have identical counts to other states. This seems highly unlikely given the sample sizes (see Ark. vs. Calif and La. vs. Maine)
- Some states (e.g., Calif again) also show a very disproportionate Republican share
- The "State Unknown" row includes a count for electoral college votes

For the last item, some students are able to puzzle out the correct fix (a shift in the data values: moving the data values over by one position to yield 7158, 6545, perhaps a comma was misplaced).

Back to our search, we were able to acquire the *Literary Digest* issues from 1924–1936, with the published data tables, from our university library as pdf files (see Supplementary Materials). We then set out to extract the raw data from these pdf files.

TEACHING TIP (Data scraping. Data validation): Give students strategies for extracting the data from pdf files. Have students resolve inconsistencies with the data and the information provided in the original articles.

There are various optical character recognition options. One we feel is useful to show students is a newer feature in Excel (Data → From Picture → Picture From File). This method is far from perfect and

depends strongly on the quality of the image, but should still be faster than typing in the data. (See also pdftools in R among others.) Some common issues we encountered with the Excel conversions:

- Misreading commas as periods, rendering some values (thousand integers) as non-integer.
- Misreading 6 as 0.
- Combining two counts into a single cell.
- Misaligned columns.
- Incorrect handling of missing values and repetitious periods.

Reviewing and making these corrections by hand in a small dataset (e.g., using "Find and Replace" to locate all instances of periods and replacing them with commas) can be a gateway to helping students automate this process (e.g., using gsub in R) with larger datasets. In the files shared in the Supplementary Materials, we did minimal processing (e.g., corrected some text errors in the state names), but you can also ask students to think about/apply strategies for harmonizing the data across the years for example (e.g., capitalization of state names).

Once the initial fixes are made, you can consider teaching other techniques for exploring the data. We recommend starting with a spreadsheet program to allow students to quickly scan the data file for potential problems. Our students often aren't familiar with how to make a formula or to fill down. Spreadsheet packages also have tools like row sums, freeze panes, SumIf, AND (for composite conditionals), and Conditional formatting of cells to help students explore and clean the data. For example, calculating the row totals (by state) and comparing to the totals provided by the *Literary Digest* (*LD*) and then using conditional formatting to highlight nonzero rows helped us identify

discrepancies faster. Using SumIf was useful for confirming electoral count totals. Many of our students, including our majors, have appreciated this light exposure to spreadsheet skills.

There is one discrepancy we were not able to resolve. In the 1928 "Other Candidates" table, we identified 370 extra votes for prohibitionist candidate William Varney. In 1928, we do not have the state level row totals to compare to, so then we found the proportion of Varney votes in each state to identify any states that seemed outside the norm. This caused us to question Maryland. We obtained the previous issue of the magazine which had earlier returns, and Maryland was not unusual for those initial responses in the same way. We decided that Varney's vote count of 416 in Maryland should actually be 46. We can't verify this with certainty, and it is important to remind students to document such decisions.

One other note, in the Excel file in the Supplementary materials we also pasted in the Wikipedia data on the actual vote counts (e.g., https://en.wikipedia.org/wiki/1936_United_States_presidential_election). The Wikipedia data can be easily copied and pasted from the website (or you could use this as an opportunity to use R for web scraping), but Wikipedia does not include the District of Columbia. We recommend deciding whether you want to remove DC from the *Literary Digest* results (1924, 1928) or add a blank line to the Wikipedia results so the results for the same state are all in the same row. Furthermore, the Wikipedia data include lots of dashes for missing values which need to be handled (e.g., changing to zero) before applying formulas. The *Literary Digest* results also include a final "unknown state" row.

The data file we compiled for the four election years can be foundin the Supplementary Materials, along with an example (but not necessarily idealized) RMarkdown file and a separate Excel "solutions" file with more calculations for 1936. (These files are not necessarily intended to be given directly to students, but as a reference for instructors.) There are also many opportunities for learning

some data processing skills (e.g., reading data from different spreadsheet tabs, only reading in some columns, starting in the second row of the data, merging datasets, map visualization, harmonizing column names, harmonizing state names and captialization).

4. INITIAL EXPLORATIONS

Using the data provided, students can compare the predicted vote to the actual vote. We can start by focusing on the error in the prediction of the Republican vote share.

The percentages in Table 1 are with respect to all voters rather than only those voting for the two leading candidates. (If instead you compare only the Republican and Democratic votes in 1936 (see Figure 2), the predicted Republican vote share is 57.1%, so you may see this value reported in many classroom activities instead.) From Table 1, the results do seem to indicate reasonable accuracy by the *Literary Digest* until the 1936 poll, but also some consistency in overestimating the Republican share.

TEACHING TIP (Data exploration, Parameter vs. Statistics): Have students replicate these results, using the raw data. For class discussion: Do these results "justify all the praise the *Literary Digest* received for the reliability of its forecast" (Robinson, 1932, p. 62) in the earlier polls? This is also an opportunity to practice having students be careful with proportions vs. percentages.

In 1928, the Republican share prediction error was 63.3 – 58.2 = 5.1 percentage points, and the *Literary Digest* argued the poll was 100% – 5.0% = 95.0% accurate. Robinson (1932) considered whether this was an appropriate statistic (should the 5.0 value be compared to 100% or 58.2%?), and whether it is reasonable to focus only on a single candidate vs. the adding the predictive error across the first and second place candidates (Hoover was predicted to win over Smith by a "plurality" of 27.6 percentage

points, but was actually only by 17.4 percentage points, an error of 10.1 percentage points). Without a strong third candidate, plurality error is roughly double the prediction error. We don't necessarily recommend discussing plurality with your students, but you can ask them to suggest other ways of measuring/comparing poll accuracy.

Students can also explore how electoral votes are assigned for each state. (Remember to address the inconsistent formatting across the different data sources, and again whether you want to focus on all the returned postcards or focus on only the two-party candidates. The third-party candidate is most interesting in 1924 when LaFollete had a strong showing in WI in the actual election and in 11 of the state results for *LD.*)

TEACHING TIP (Data visualization): Have students determine which states and how many electoral votes were predicted for Landon in 1936.

Some visualizations you could incorporate here include color-coded maps for the state-level predictions, heat maps to see how close the states were to 0.5, or dotplots with a line at 0.5, also color-coded on each side. Results for Coolidge (R) vs. Davis (D) in 1924 are shown in Figure 4.

In 1924, the *Literary Digest* correctly predicted whether the majority would vote for Coolidge or Davis in all but two states, Kentucky and Oklahoma. When considering all the candidates, Wisconsin was (correctly) predicted to go to LaFollete. In 1928, all but four states were predicted correctly (MA, RI, AL, AK). In fact, when they published the 1928 poll results, the *Literary Digest* suggested AL and AK

("the two States where the vote is close") should be credited to Governor Alfred Smith, "on the ground that they are normally Democratic" (11/3/1928, p. 7).

TEACHING TIP (Measuring accuracy): Possible questions for class discussion: Is determining the correct winner in each state a reasonable way to judge the accuracy of the poll? Is expecting state-level accuracy rather than focusing on overall election results reasonable?

Critics (e.g., Franklin in *New York Times* editorials in 1928 cited by Robinson, 1932) suggested that a poll's accuracy was better determined by comparing the predicted proportion for each candidate to the actual proportion of votes received in the official election rather than only predicting the electoral outcome. As noted by Robinson (1932), "Only when the election is close will the electoral-college predictions of such a poll be upset and the error in forecasting the proportion of the total vote received by each candidate be revealed. In both the 1924 and the 1928 presidential campaigns, the Literary Digest overpredicted the popular vote of the winner, and the subsequent official returns proved to be one-sided; hence few states were mispredicted in the electoral college."

5. EXPLORING AND ADJUSTING FOR BIAS

A very simple explanation for the errors in 1936 is overestimation of the Republican vote share. Students can explore this claim by comparing the predicted Republican proportions to the actual results state by state.

$$\textit{Error per state} = \textit{predicted Republican \%} - \textit{actual Republican \%} \quad (1)$$

This can be done through simple differences, but again decide whether you want to look at all candidates or just Republican vs. Democrat (the R code in the Supplementary Materias illustrates both approaches), and watch for the extra DC row and for proportions vs. percentages.

TEACHING TIP (Sampling bias, treating states as separate samples): Help students understand the meaning of "bias" by showing graphs of results that are consistently "off center" in one direction.

The graphs in Figure 5 show the difference in the *Literary Digest* Republican share minus the actual Republican share for the 1924 –1936 election years for each state.

We notice that in 3 of the 4 years, most of the distribution is above 0 (means 0.049, 0.073, 0.003, 0.174), indicating that the Republican share predicted by *Literary Digest* was larger than the actual share in most states. In fact, in 1928 and 1936, every single state overestimated the Republican vote share, a nice illustration of consistent errors across multiple samples.

By the way: Robinson (1932), using plurality error, compared the *Literary Digest* results to those of other straw polls in 1924 and 1928 and found the *Digest* errors to be on par with other national polls ("not the best nor the worst"). Plurality error is computed (here for each state) by comparing the observed and actual difference between the winning candidate vote share to the second-place vote getter (which in the 1924 *Digest* poll was LaFollette in 11 states rather than Davis).

Plurality error = predicted (1st place – 2nd place vote share) - actual (1st place – 2nd place vote share) (2)

The median (absolute value) plurality error across the states was 12 percentage points in both 1924 and 1928 (compared to 12 and 5, respectively for *Hearst Newspapers*). Again, we only recommend discussing plurality error if you want to have students replicate Robinson's results.

TEACHING TIP (Adjusting for bias): Have students brainstorm a simple way to adjust for the overestimation of the Republican voters. In particular, have them think about what additional information was available to them from the postcards or previous elections.

Students may be quick to suggest that we just need to subtract off the bias! But of course, we don't usually know the bias, at least until after the election. In this case, we can use results from the previous election. Robinson (1932) suggested that *Literary Digest* adjust their 1928 poll results for the bias in the 1924 poll results, as the "tel-auto" population utilized (using telephone books and automobile ownership frame) can expect more Republican votes from election to election. For example, in each state, we could subtract the overestimate from 1924 from the 1928 poll prediction:

$$\textit{New LD prediction} = \textit{1928 prediction} - (\textit{1924 prediction} - \textit{1924 actual}) \quad (3)$$

The output in Figure 6 for this "additive method" shows the adjusted prediction errors (percentage points) after comparing the adjusted Republican vote shares (using the previous year's bias) to the actual vote shares for 1928–1936. (The results are similar using plurality error as well. The median plurality error drops to 6 points in 1928. Note: We did find some discrepancies with Robinson's Table 8 as noted in the R Markdown file in the Supplementary Materials.)

This method assumes the polling trends are similar from year to year. This is clearly true from 1924 to 1928 (e.g., correlation between 1924 and 1928 error = 0.670) and the errors are smaller with less bias in

1928 after the correction. However, this method overcorrects in 1932 (corr 1928, 1932 = 0.718), and is not sufficient for the overrepresentation of Republican votes in 1936 when the overall Republican vote was much smaller and the election much closer (corr 1932, 1936 = 0.303).

TEACHING TIP (Evaluating impact and effectiveness of the adjustment): Ask students to consider how the correlation in errors with the previous election increased between 1928 and 1932 but 1932 had more bias after the correction. Have students apply this simple correction to 1936 and see which states' electoral votes change. How does this impact the overall prediction?

To decide whether the state prediction changes (along with who gets the electoral votes), you can:

- See whether the predicted share switches which side of 50% it falls on. This ignores the fact that there were more than two candidates.
- See whether the predicted Republican share is still larger/smaller than the predicted Democrat share.
- Recompute the two-party predictions in terms of the split between only Republican and Democratic voters and see whether the predicted share switches which side of 50% it falls on.

For example, you can compute the 1936 predicted proportions for Republicans and Democrats and then apply the 1932 errors to each and then create a Boolean variable for whether the Republican > Democrat comparison is the same before and after adjusting (bullet 2):

New LD prediction = predicted Repub prop 1936 – (1932 predicted Repub – 1932 actual Repub)

(4)

We find 5 states change: Kentucky, Maryland, Nevada, New York, and North Dakota. But the first two are predicted to change to Republican and the last 3 to Democrat, resulting in only one more state and a net of 54 – 19 = 35 electoral votes, still predicting a Republican majority of electoral votes. These results are shown in Figure 7. (The other two methods, first and third bullets, largely agree, identifying KY, MD, NY, and ND, but not NV, see RMarkdown file in Supplementary Materials.)

6. OTHER ADJUSTMENTS

What other corrections could have been applied with the information available to the *Literary Digest*? If you show students Figure 1 again, they may notice that the magazine's postcards also collected information about how the respondent previously voted. This provides a window to whether Republican voters are overrepresented in the *Literary Digest* samples.

In the 1928 *LD* poll, 56.6% said they voted Republican in 1924. In the actual election, 54% voted Republican, showing slightly higher response from Republican voters in the *Literary Digest* poll.. Comparatively, in 1936, 46% of the Landon/Roosevelt/Lemke intended voters in the *LD* poll claimed to have voted Republican in 1932 (50.7% of those who claimed to have voted), compared to less than 40% in the general election voting Republican. So another adjustment strategy would be to "downscale" the Republican votes in the *Digest* poll to reflect this overrepresentation. As suggested by Lohr and Brick (2017), we can compute the ratio between the actual proportion and the claimed proportion. For example, overall:

Republicans: Actual vote share in 1932 / (claimed vote share in 1932) = 0.3965/0.5070 = 0.782.

Democrats: Actual vote share in 1932 / (claimed vote share in 1932) = 0.574 / 0.480 = 1.197

So when we tally the *LD* vote totals, we want to use roughly 78% of the Republican voters but 120% of the Democratic voters.

Similarly, Lohr and Brick find a ratio of 2.228 for the third party and "other" candidates; and also provide ratios of 0.87 and 1.13 for the Republicans and Democrats to apply for the non-voters and missing responses. We provide students these ratios so they can find that the overall predicted vote share becomes 0.504 for Landon and 0.500 for Roosevelt. State by state, the Republican ratio is less than 1 in only 5 states (all in the south), meaning Republicans were overrepresented in all the other states. Applying the ratios to the individual states, ten change from predicting Landon to have more votes to predicting Roosevelt to have more votes than Landon, including CA and NY. These changes correspond to 115 electoral votes. That is, Roosevelt is predicted to win 26 states and 276 electoral votes (52%). "Thus, this ratio adjustment predicts the correct winner of the election, although it still does a poor job of predicting the margin of the win in terms of the popular vote and the electoral college" (Lohr & Brick, 2017, p. 71).

Another approach explored by Lohr and Brick is to use a regression model, regressing the 1932 actual vote share on the *LD* predicted vote share, and then using that model to adjust the 1936 *LD* predictions. Applying this approach to the Republican proportion (for the two parties), we find

$$\widehat{1932\ actual\ Rep\ prop} = -0.23 + 0.97\ \textit{1932 LD pred Rep prop}\ (R^2 = 0.85) \quad (5)$$

If we use the 96 observations from 1928 and 1932, the fitted equation is

$$\widehat{actual\ Rep\ prop} = 1.40 + 0.88\ \textit{LD pred Rep Prop}\ (R^2 = 0.89) \quad (6)$$

The regression equation is then used to predict the 1936 vote share from the *LD pred prop Rep* 1936 poll results. With this approach, Roosevelt is predicted to win 22 states. Figure 8 shows the various

adjustments applied to predicting Roosevelt's vote share (vs. Landon rather than all vote getters), and Figure 9 shows the updated prediction errors.

TEACHING TIP (Measuring sampling bias, Evaluating corrections, Interpreting graphs): Ask students to verify the results of how many states switch from a Republican to a Democrat predicted majority. Ask students to use Figure 8 to decide which method was most effective in 1936 and whether they believe that method should have been used in 1940. Have students identify which states were more accurately predicted with the adjustments and whether there are any (geographic) patterns. While these ideas are probably more time than you want to spend on this topic in the introductory course, you could consider asking them to interpret the graphs and discuss what they notice.

Again, the adjustments help but not dramatically. The mean prediction error reduces from 18.6 percentage points to about 13 and 14. The states that most benefit are from the northeast (e.g., NH, MA, VT, ME), all lowering the predicted Republican vote. For example, *LD* overestimated the Republican vote by over 30 percentage points in RI and MA. These reduce to 24 percentage points with the regression adjustment.

7. SUMMARY

The 1936 *Literary Digest* poll has become a go-to example of polling gone wrong. Students can usually identify common explanations on their own (e.g., bad sampling frame, response bias) and instructors can also cite the rise of George Gallup at the time, who heralded the use of random sampling to improve polling accuracy, even with much smaller sample sizes. Gallup's methods were followed successfully by modifications such as random digit dialing, but recent polling errors (e.g., Clinton vs. Trump; see Cohn, 2017) have demonstrated that such methods are now subject to new types of biases as technology

and people's patterns of behavior change. Current survey sampling methods are much more complicated, but ideas of stratified sampling and sampling adjustments are understandable by introductory statistics students.

Examples of possible classroom use (either individually or as a case study with repeated applications throughout the course), mostly focusing on introductory courses:

- A homework problem early in the term when discussing data processing steps, scraping and validating web data. For example, we now have a low-stakes assignment asking students to be data detective with the *History Matters* website. You could show this webpage in class and have a "prize" for the first students to notice the issues.
- Expanding current discussion on sampling biases to discuss possible remedies for sampling bias. For example, show Figure 8 and have students compare the results and think about bias in terms of systematic tendency to err in one direction to help them internalize sampling bias as a property of a sampling method. Or even just a quick demonstration that while proper random sampling is difficult, we should not give up (e.g., Gelman and Vehtari, 2024). This discussion could focus on using the adjustments to see how the electoral counts change and which candidate would have been predicted the winner.
- Homework exercise asking students to perform the simple ratio adjustment based on the 1932 results as part your discussion of the *Literary Digest* poll to illustrate the simple downscaling of the Republican percentages that was possible. See Appendix B for a recent homework exercise along these lines.

- A lab assignment asking students to work with the data, produce the graphs, and carefully discuss differences in percentages. Similar to the homework exercise but asking students in pairs to work with directly in the software, focusing more on the computing skills.
- A lab assignment asking students to take the "basic" RMarkdown file in the Supplemenatry Materials and to improve the R coding, incorporate more tidyverse, add more documentation, etc.
- A case study in a data visualization class asking students to create maps/highlight states that change (e.g., 1936 with the ratio adjustment). Can also focus on the challenges of reading in the data.
- A lab assignment giving a practical application of regression method to adjust predictions.
- A lab assignment asking students to verify the results of the Lohr & Brick and/or Robinson papers as a way to develop their statistical programming skills and checking their work (and spotting inconsistencies).
- Ask students to delve deeper into one of the references and present their findings to other students.

One or two of these further explorations of the *Literary Digest* poll results offers students opportunities to work with original (and messy) data, explore data cleaning issues, visualize bias, and be introduced to principles of adjusting poll results to reduce bias in an authentic context. After the homework exercise (Appendix B), one student commented "It was super interesting to learn about adjusting the results using additional information to make a more accurate prediction!" In particular, students should realize that these methods depend on consistency in the voting behavior between elections and when that assumption is or is not reasonable. Considerations can include how adjustments like weights are used in

current polls, and reasons more recent polls still have issues with sampling bias and non-sampling errors, whether poll results can be adjusted, and when they should or should not be trusted. See Lohr and Brick for more discussion, including uncertainty measures and Bayesian modeling.

Instructors of a survey sampling or political methodology course may want to delve further into sampling weights, including other historical examples of polling errors, including the 1944 election, as well as more recent non-sampling errors including the 2106 election in the United States and Brexit polling (see additional references: Sturgis et al., 2016; AAPOR, 2017; 2021). But we feel all students can benefit from reminders to inspect data, explore data from different perspectives, learn from mistakes, and apply those lessons to future investigations.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Andrew Gelman, Sharon Lohr, Anelise Sabbag, and the reviewers for their helpful comments and suggestions in revising this paper or the homework exercise. This work was also partially supported by NSF grant DUE-# 2235355.

OTHER INTERESTING LINKS

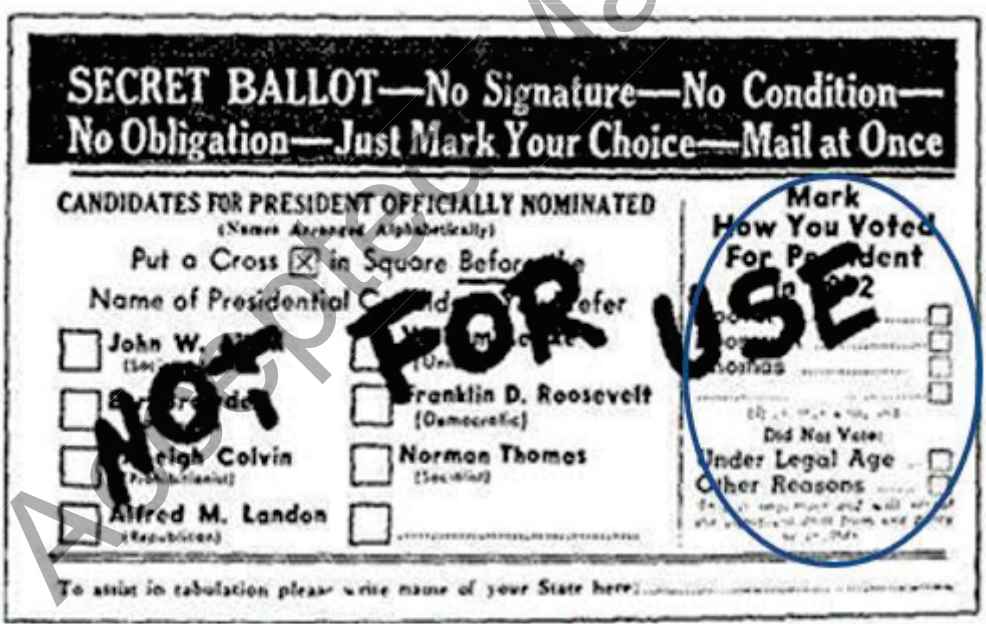
- https://www.historyofinformation.com/detail.php?entryid=1652
- https://potus-geeks.livejournal.com/tag/alf%20landon
- https://www.c-span.org/video/?c5024222/straw-polls-american-history
- https://studylib.net/doc/10302848/has-polling-enhanced-representation%3F-unearthing-evidence-...
- https://www.pewresearch.org/short-reads/2020/08/05/key-things-to-know-about-election-polling-in-the-united-states/
- PBS documentary, "The First Measured Century", which had a 10-minute segment on the 1936 poll (posted by Dennis Sun: https://youtu.be/KNar7iSdN0Y)
- Gelman, A. (2007). “Struggles with Survey Weighting and Regression Modeling,” http://www.stat.columbia.edu/~gelman/research/published/STS226.pdf

APPENDIX A: SUPPLEMENTARY MATERIAL

- LitDigestFull.xlsx: Has a tab with the raw data for each of the four years (see Data dictionary)
    - Data dictionary
- Literary Digest1936 Sols.xlsx: Contains columns showing "results" for the suggestions offered in the paper (e.g., computing proportion voting Republican, comparing actual to predicted results, bringing over the 1932 error, computing the additive adjustment, comparing the results focusing on all parties vs. only 2 of the parties, highlighting changes, exploring the regression adjustment)
- Copies of *Literary Digest* articles containing original 1924, 1928, 1932, 1936 results
- RMarkdown file illustrated many of the possible anlayses, creating figures for this paper

APPENDIX B: Example homework exercise

Recall the *Literary Digest* example (see Inv 1.16), where we blamed the poor estimate (41% voting for Roosevelt when actually 60.8% did) largely on an *incomplete sampling frame* (the wealthier Republicans were more likely to be sampled) and *voluntary response bias* (those who had more time/money to respond or who were more unhappy with the incumbent were more likely to respond). These seem like obvious explanations in hindsight, but should the *Digest* have realized this was happening? And could they have done anything about it? Normally, we don't know whether the size of the bias or even if there actually is bias until we know the parameter (we may never happen), but if we suspect a sampling method is biased, and if we have other information about the individuals in our sample, can we make adjustments in advance? For example, the *Digest* postcards sent out in 1936 also asked individuals to report whom they had voted for in 1932.



The goals of this exercise are to explore whether using this information would have been helpful to the *Digest* in predicting the 1936 election (related to the idea of post-stratification which you can learn more about in Stat 421), as well as to practice a few "spreadsheet skills" and think about data quality checks.

Open the LitDigest1936.xlsx file in Excel or Google Sheets. This contains the raw counts for the three main candidates: Landon (column C), Roosevelt (column K), and William Lemke, Union Party (column S), in each state and overall (row 51), as well as the overall total number of straw votes cast in *Digest* poll in each state (column AA).

(a) For the 3 "major" candidates, what percentage of the poll respondents said they would vote for Republican Landon in the 1936 election? [*Hint*: Set up a column formula in row 53. *Be sure to include the formula you used. You can type it out, or screen capture the formula bar, or in Excel, for example, you can got to Formulas > check Show Formulas. Make sure I can replicate what you have done.*]

Columns D-I is the breakdown of how all of the "Landon voters" in the 1936 *Digest* poll voted (or not) in 1932, for each individual state. For example, in Alabama, 3060 *Digest* respondents said they planned to vote for Landon. Of those, 1,218 said they voted for the Republican candidate in 1932.

Focus on **row 52** (state totals).

(b) Set up a formula for determining the number of respondents to the 1936 poll who said they voted in **1932** for either the Republican, Democratic, or Socialist or Other candidate. [*Hint*: Use columns D-G, L-O, and T-W.] What proportion of these voted for the Republican candidate. [Include your formulas.]

Total voters in 1932 for Republican, Democrat, Socialist, Other:

Proportion of 1932 R, D, S, O voters who voted Republican in 1932:

(c) Now examine the actual 1932 election results, what proportion of voters voted for the Republican (Hoover) candidate (among the three major candidates)? (Show your work.)

(d) Is there initial evidence that the *Literary Digest* sampling methods tend to overrepresent Republican voters? Cite your (numerical) evidence.

One way to adjust the 1936 poll results would be to "scale down" the number of Republican voters and "scale up" the number of Democratic voters. Consider the following ratios:

	Republican	Democrat	Socialist, Other	Non-voters, Missing
Ratio	0.782	1.197	2.228	Dem: 1.1275 Rep: 0.871

(e) Verify that the first value is the ratio of the actual Republican turn out in 1932 to the *Digest* claimed turn out in 1932.

(*Hint*: Use your results from (b)-(d).)

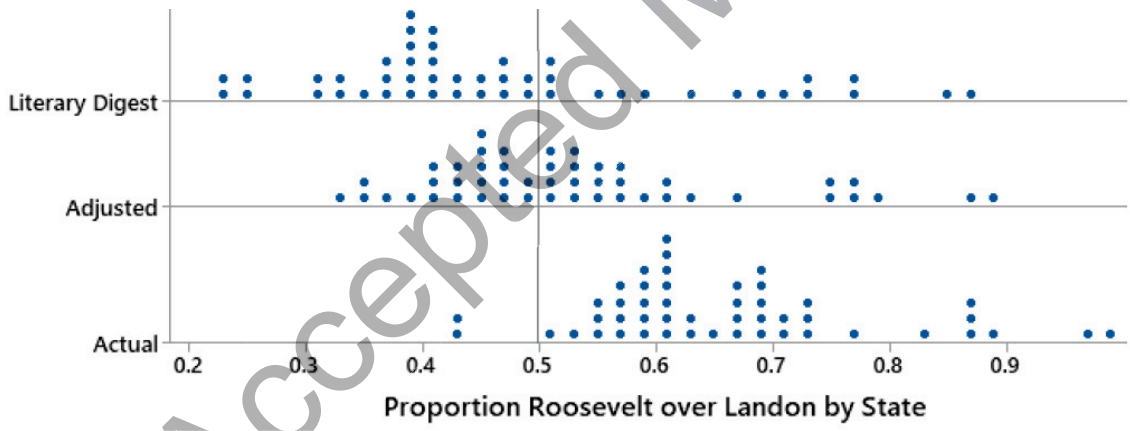
(f) Start with the 1,293,669 Landon "voters" in the 1936 poll, arising from folks who voted Republican, Democrat, Socialist etc. in the previous election (columns D-I). Create a formula that multiplies each of these counts by the corresponding ratio (e.g., 0.782 * 920225), using 0.871 for nonvoters and missing, and then sum these "adjusted counts" from each party. What is the adjusted number of Landon voters? *(Remember to copy and paste your formulas as well.)*

(g) Repeat (f) for Roosevelt (columns L-Q, 1.1275 for non-voters, missing). *(Remember to include your formulas/documentation.)*

(h) Let's focus on Roosevelt now, and on the 'two-party breakdown.'

- In the original *Digest* poll, of those saying there were going to vote for either Landon or Roosevelt, what proportion said they would vote for Roosevelt? (Show your work.)
- In the actual 1936 election, what proportion of those voting for Landon or Roosevelt actually voted for Roosevelt? (Show your work.)
- Using your results from (f) and (g), what is the adjusted percentage of voters for Roosevelt in 1936? Is this larger or smaller than the two-party breakdown without adjusting/closer or further from the actual vote in 1936?

(i) The graph below shows the results from this same process but applied to the individual states the *Digest* proportion planning to vote for Roosevelt (top graph), the actual proportion from the 1936 election results, and the adjusted proportions (middle graph).



- Does the *Digest's* original method appear to be biased? Explain how you are deciding.
- Does the adjustment appear to help? How are you deciding?
- In the U.S. election, what really matters is the electoral vote; that is, which candidates has the most votes in the state. Between the *Digest* poll and the adjusted proportions, how many states changed which candidate would receive their electoral votes?

(j) When I first went looking for the original *Digest* results, I first found the *History Matters* webpage, but soon realized there were some data errors on this page. Using only the data provided on that webpage:

- If you check the totals, they don't quite match up. Inspect the numbers in the table: Do any numerical values look suspicious to you? Do any states behave unusually?
- The State Unknown row also looks suspicious to me. Why is it suspicious? Based on the values given in that row, what do you think the counts for Landon and Roosevelt for individuals with unknown states actually were?

More recently

- 2016 election issues.
- 2020 election issues.
- Failure and success in political polling and election forecasting (Gelman, 2021)

Student feedback:

In a quick follow-up survey in two sections of Statistics and Economics majors (67 out of 71 responding anonymously online), most (67%) indicated that they found the context fairly to very interesting; only 10% chose "not so interesting," none selected "not at all interesting," and one selected "I have much less interest in historical data." The rest (21%) rating the context as "average" compared to the rest in the course. About one-fifth (18%) found the instructions on this first implementation difficult to follow, and future revisions will aim to focus on improving the instructions. Another option we are considering is breaking the initial spreadsheet into smaller chunks to be less overwhelming initially, and to explain the 1932 vote breakdown more clearly.

REFERENCES

Cohn, N. (2017), "A 2016 Review: Why Key State Polls Were Wrong About Trump," *New York Times*, May 31, 2017. https://www.nytimes.com/2017/05/31/upshot/a-2016-review-why-key-state-polls-were-wrong-about-trump.html

Gelman, A., and Vehtari, A. (2024). *Active Statistics: Stories, Games, Problems, and Hands-On Demonstrations for Applied Regression and Causal Inference*, Cambridge University Press.

Lohr, S., and Brick, J. M. (2017). "Roosevelt Predicted to Win: Revisiting the 1936 *Literary Digest* Poll," *Stat Polit Pol*, 8(1): 65-85.

Lusinchi, D. (2012). "'President' Landon and the 1936 *Literary Digest* Poll: Were Automobile and Telephone Owners to Blame?" *Social Science History*, 36:23–54.

Robinson, C. (1932). *Straw Votes.* New York, NY: Columbia University Press.

Squire, P. (1988). "Why the *Literary Digest* Poll Failed," *The Public Opinion Quarterly*, Spring, 1988, 52(1), 125-133. https://www.jstor.org/stable/2749114

ADDITIONAL REFERENCES

American Association for Public Opinion Research (AAPOR). *An Evaluation of 2016 Election Polls in the U.S. (2017)*. https://aapor.org/wp-content/uploads/2022/11/AAPOR-2016-Election-Polling-Report.pdf.

American Association for Public Opinion Research (AAPOR).*2020 Pre-Election Polling: An Evaluation of the 2020 General Election Polls (2021)*. https://aapor.org/wp-content/uploads/2022/11/AAPOR-Task-Force-on-2020-Pre-Election-Polling_Report-FNL.pdf

Sturgis P., Baker N., Callegaro M., Fisher S., Green J., Jennings W., Kuha J., Lauderdale B., & Smith P. (2016). *Report of the Inquiry into the 2015 British General Election Opinion Polls.* London: MRS/British Polling Council/NCRM. https://eprints.soton.ac.uk/390588/1/Report_final_revised.pdf

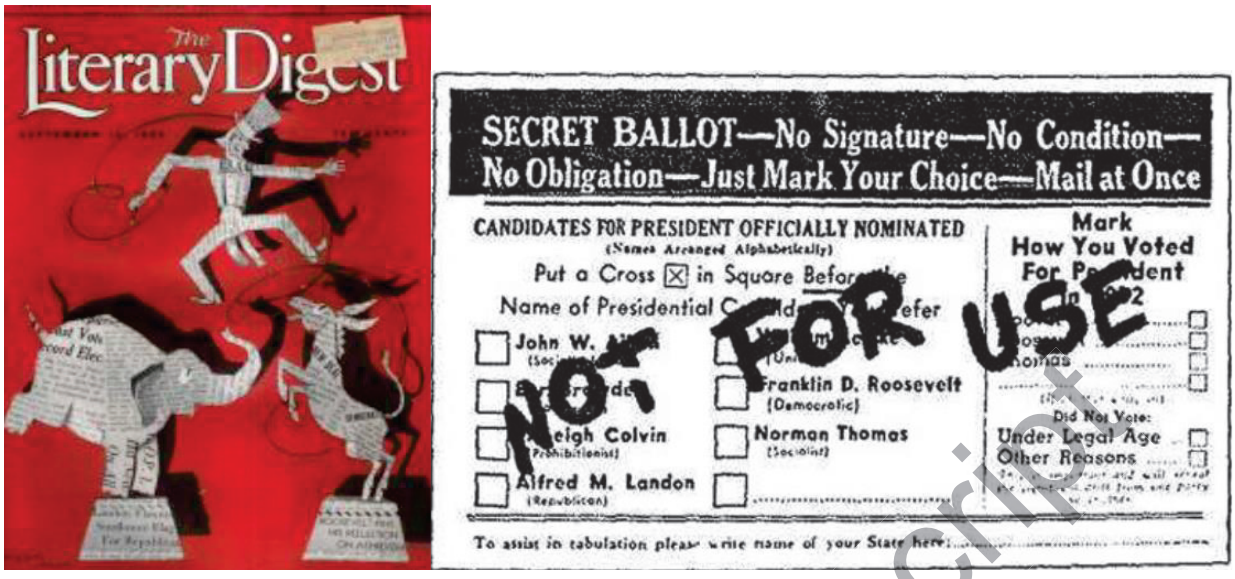


Figure 1: The cover of the September 1936 election issue (Source: HistoryofInformation.com) and "secret ballot" from 1936 (Source: *Literary Digest*, August 22, 1936, p. 3)

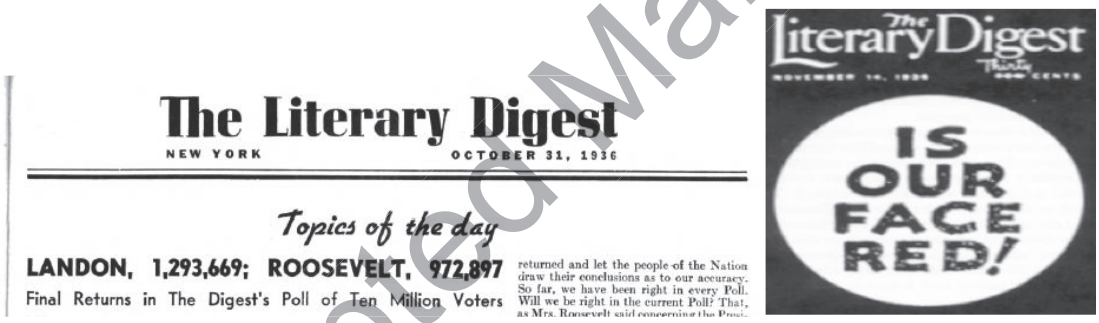


Figure 2: The original headline (Source: *Literary Digest,* Oct. 31, 1936 p. 5) and post-1936 election cover (Source: NextTV MBPT Spotlight)

State	Electoral Vote	Landon 1936 Total Vote For State	Roosevelt 1936 Total Vote For State	State	Electoral Vote	Landon 1936 Total Vote For State	Roosevelt 1936 Total Vote For State
Ala.	11	3,060	10,082	Nebr.	7	18,280	11,770
Ariz.	3	2,337	1,975	Nev.	3	1,003	955
Ark.	9	2,724	7,608	N.H.	16	9,207	2,737
Calif.	22	89,516	7,608	N.J.	16	58,677	27,631
Colo.	6	15,949	10,025	N.M.	3	1,625	1,662
Conn.	8	28,809	13,413	N.Y.	47	162,260	139,277
Del.	3	2,918	2,048	N.C.	13	6,113	16,324
Fla.	7	6,087	8,620	N. Dak.	4	4,250	3,666
Ga.	12	3,948	12,915	Ohio	26	77,896	50,778
Idaho	4	3,653	2,611	Okla.	11	14,442	15,075
Ill.	29	123,297	79,035	Ore.	5	11,747	10,951
Ind.	14	42,805	26,663	Pa.	36	119,086	81,114
Iowa	11	31,871	18,614	R.I.	4	10,401	3,489
Kans.	9	35,408	20,254	S.C.	8	1,247	7,105
Ky.	11	13,365	16,592	S.Dak.	4	8,483	4,507
La.	10	3,686	7,902	Tenn.	11	9,883	19,829
Maine	5	3,686	7,902	Texas	23	15,341	37,501
Md.	8	17,463	18,341	Utah	4	4,067	5,318
Mass.	17	87,449	25,965	Vt.	3	7,241	2,458
Mich.	19	51,478	25,686	Va.	11	10,223	16,783
Minn.	11	30,762	20,733	Wash.	8	21,370	15,300
Mis.	9	848	6,080	W.Va.	8	13,660	10,235
Mo.	15	50,022	8,267	Wis.	12	33,796	20,781
Mont.	4	4,490	3,562	Wyo.	3	2,526	1,533
State Unknown	7	1586	545				
Total	531	1,293,669	972,897				

Source: *Literary Digest,* 31 October 1936.

Figure 3: Copy of data table provided by *History Matters*

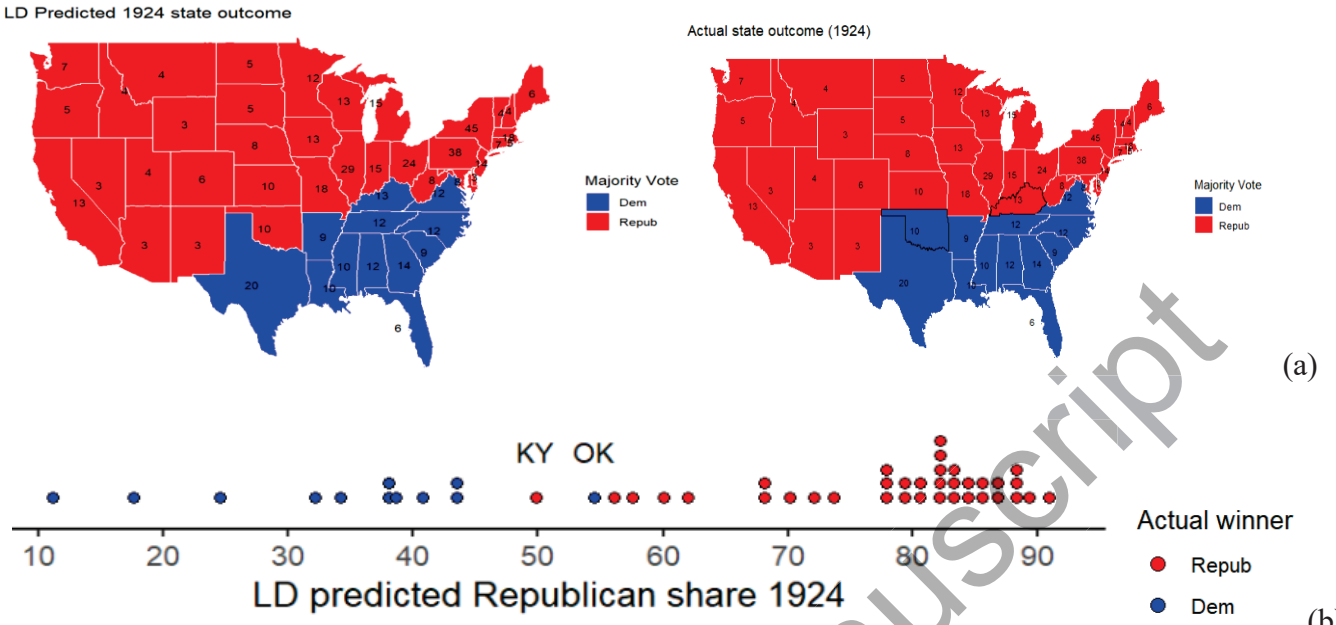


Figure 4: (a) Maps showing the predicted and actual state winners between Coolidge and Davis. (b) The predicted two-party proportion voting Republican (over Democrat) in 1924 in each state, color-coded by the actual winner (Red = Republican majority; Blue = Democrat majority).

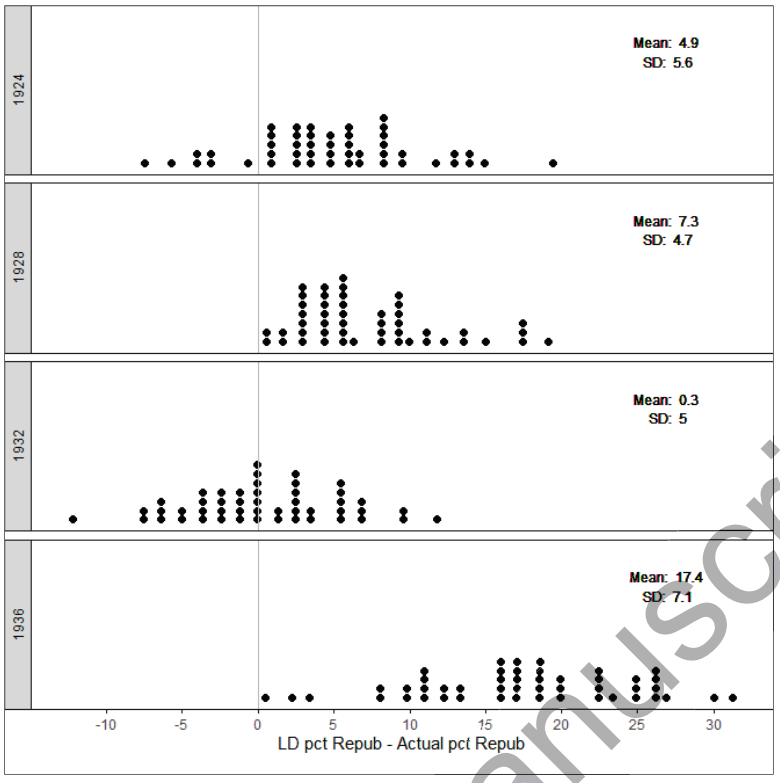


Figure 5: *Literary Digest* predicted Republican vote share minus actual for 1924–1936 in each state

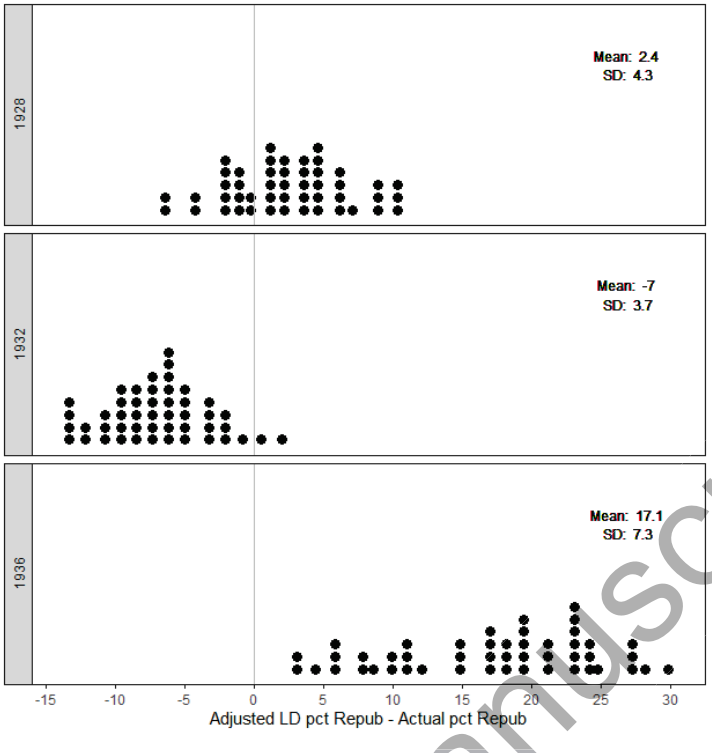


Figure 6: Errors in predicted Republican vote share in each state after adjusting for previous election poll results

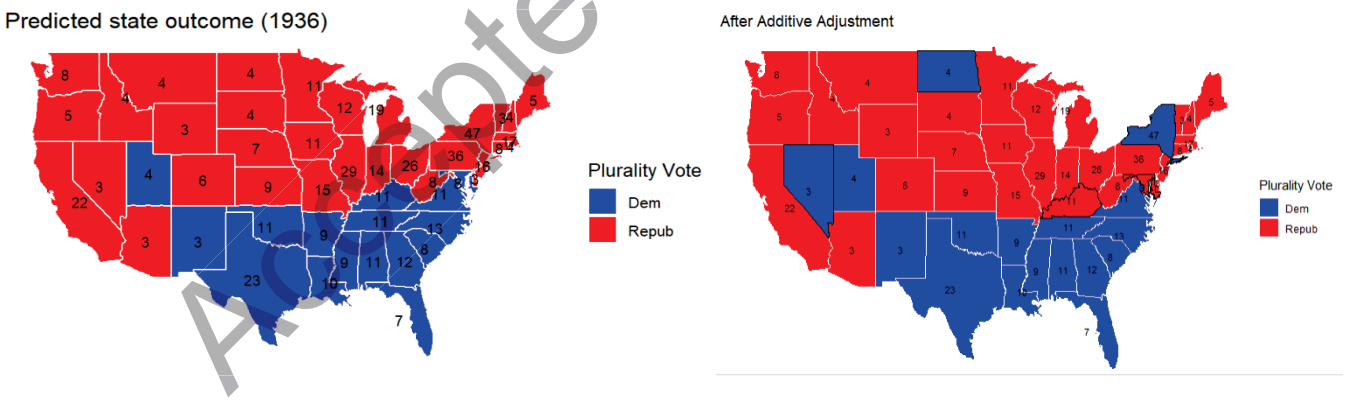
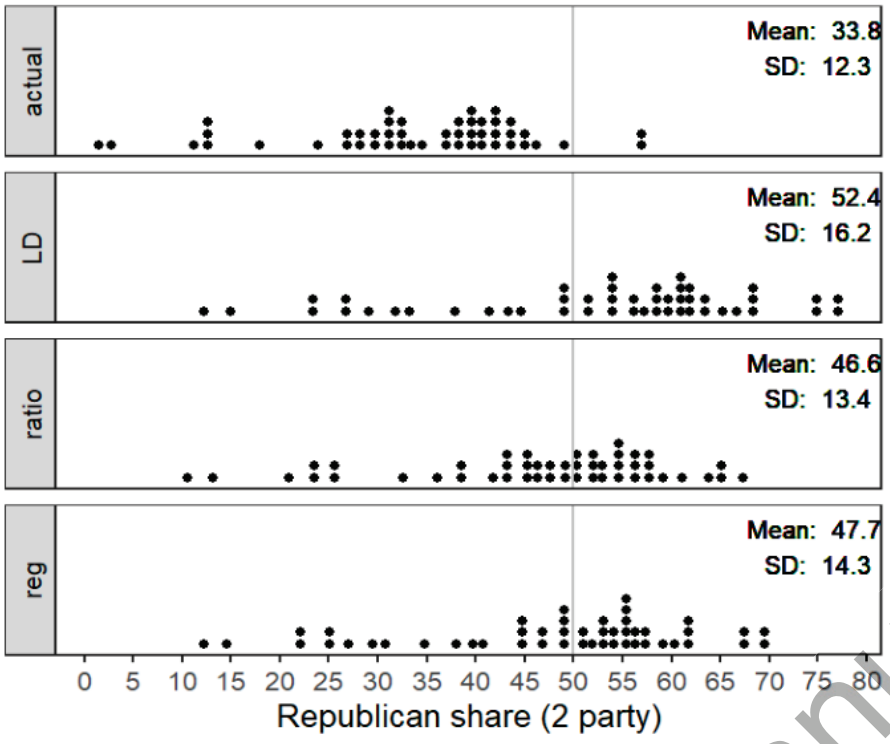


Figure 7: Electoral map with original 1936 predictions and after applying additive adjustment



Actual proportion Republican (vs. Democrat)

Original *Literary Digest* predictions (16 Dem states)

Adjusting the *LD* predictions with the relative ratios (26 Dem states)

Using the relation between predicted and actual votes in 1928 and 1932 (22 Dem states)

Figure 8: Comparison of adjustment methods to Republican vote share in 1936 *LD* poll

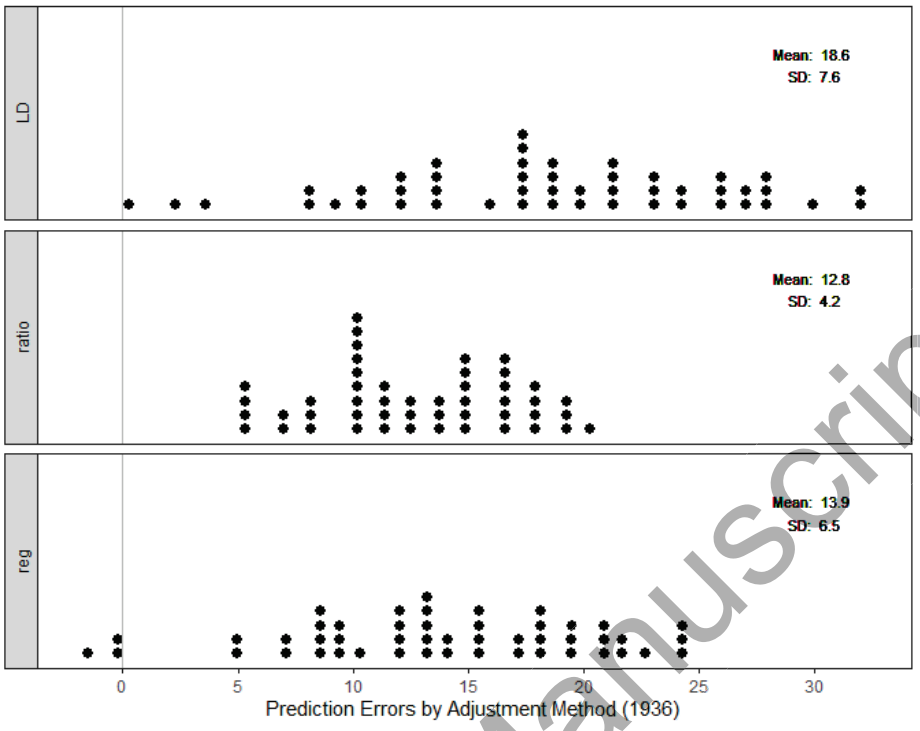


Figure 9: Prediction errors (predicted – actual) in Republication vote share in each state by method comparing Republican to Democrat proportions

Table 1: The *Digest* results and the actual election results for four election years

Election	Predicted Republican share	Actual Republican share	Difference
1924	56.6%	54.0%	2.6 perc. pts
1928	63.3%	58.2%	5.1 perc. pts
1932	37.5%	39.6%	-2.1 perc. pts
1936	54.4%	36.5%	17.9 perc. pts