

WHERE DID THE ELECTION POLL DATA GO WRONG?

FEELING LUCKY? WHERE DID THE ELECTION POLL DATA GO WRONG?



Let me start off by saying that we're not in the business of talking politics. We're not going to discuss the candidates or their views. I don't care who grabbed what, who deleted what, or who you voted for. But, when an election suffers from such a huge discrepancy between what the poll data predicted and the actual results, we just have to explore.

Political elections are one of the foremost uses of predictive data. For the last year, from the primaries up until election day, we've seen countless polls and predictions about who would win the Presidency.

Hillary Clinton had been enjoying a comfortable lead in most polls since the beginning of this campaign. Right up to election day, polling evangelist Nate Silver gave her a 72% chance of winning. By midnight, he had flipped his odds to 84% in favor of Donald Trump.

About two weeks prior to election day, a [Huffington Post](#) article foreshadowed this turn of events by saying, "If Trump wins the popular vote on Nov. 8, it would be a polling miss unlike we've ever seen before in a general presidential elec-

tion." However, the article went on to compare the situation to the "Dewey Defeats Truman" debacle of 1948, adding, "That would never happen now."

Well it could, and it did. They were right about one thing: Trump did not win the

popular vote. He did, however, earn more electoral college votes to secure the victory. The newspapers didn't run the wrong headline this time, but the poll data was wrong and Donald Trump is our next President. The vast majority of analysts and media got it wrong. So how did it happen?

Missing Information

Maybe the polls were flawed. Key demographics could have been carelessly excluded by the pollsters. Some believe that many Trump supporters did not participate in polls, assuming that those voters chose to remain silent in the face of harsh criticism for their political leaning. A "silent majority" if you will. This may have skewed the results to appear that more voters favored Hillary Clinton when, in fact, the polls were missing a large part of the population.

Geoff Garin, a veteran Democratic pollster, said many surveys had under-sampled non-college-educated whites, a group that Trump appealed to. He also argued there had been an over-emphasis on the belief that the country's rising demographic diversity would put Clinton over the top. (Politico.com)

Media Bias

We the public can only make assessments based on what we are shown. Depending on where you get your information, you may encounter some bias. Pro-Trump operatives argued that even when some polls indicated Trump's strength as a candidate, it was ignored or explained away by analysts and members of the media.

"Most of the press and folks in DC were science deniers when it came to this election," said veteran GOP operative Curt Anderson, an adviser to a pro-Trump super PAC. "Even in the face of polls that showed it very close, they all said that Trump had almost no chance. It was because they couldn't imagine it happening." (Politico.com)

Misinterpretation

Polling data is meant to be pretty cut-and-dry, but as shown by this election, the raw numbers may not tell the whole story. There is some interpretation and assumption to be made where the data isn't specific. As Garin mentioned, there were those that simply believed certain things about key demographics and how they would vote.

Arie Kapteyn is the director of the University of Southern California's (USC) Dornsife Center for Economic and Social Research, which jointly ran an LA Times/USC internet poll that accurately predicted Trump's victory in the final months of the campaign.

In a recent interview, Kapetyn said, "There's some suggestion that Clinton supporters are more likely to say they're a Clinton supporter than Trump supporters are to say they're a Trump supporter," Kapteyn said late Tuesday in an interview.

Kapteyn also suggested that many pollsters may have incorrectly ruled out the prospect that people who didn't vote in 2012 would nonetheless cast ballots in 2016.

"But the people who didn't vote last time are more likely to be Trump supporters," he said, noting that the LA Times/USC poll gave more weight to whether voters said they planned to vote. "If you eliminate people who didn't vote last time, you may have eliminated too many Trump supporters." (USA Today)

time, you may have eliminated too many Trump supporters. (USC Today)

Only 6 out of 61 polls showed Trump in the lead from the start of October. All six were ran by the LA Times/USC.

The Impact of Inaccurate Data

These polls were important. They were supposed to be a climate check of the United States, and tell us what Americans were thinking. For too long, the media and analysts were missing key information, using inaccurate poll data, and were unable to accurately predict the outcome of this election. Was this missing information even attainable? Had it been more accurate from the beginning, would it have influenced other voters? Would either of the candidates have altered their strategy? Would the outcome have changed?

These are rhetorical questions now, but they're interesting to ponder. Many Americans didn't realize how close it would be, or that their candidate was more/less favored than they thought. The majority of reports were making assumptions like those mentioned above, based on flawed or incomplete poll data. But there were also voters who believed it when a few outlets challenged the reports.

As data analysts, we think about how this can impact other scenarios. Predictive analytics in marketing and business are incredibly important to their respective organizations. Future plans, strategies, and spending are dependent on accurate data.

That's why we felt this was an important topic to discuss. If a retailer made decisions based on inaccurate data, it could cost them. I cannot stress enough how important clean, accurate data is to a business.

RELATED RESOURCES

Blog: Why Marketers Need Clean Data

Infographic: Finding Customer Data Blind Spots

Whitepaper: Spring Cleaning for Dirty Data

It starts with your data sources and ends with how well you can organize and segment that data to deliver actionable information. Consider this statistic from a Forrester research study: *A lack of analytics tools and repressive data silos lead companies to ignore 88% of their customer data.*

That is poor percentage on which to base major company decisions. Unite your teams, departments, and data sources into one database to enable more organized and actionable customer information. See the related resources listed above to learn more about cleaning and organizing your data for better business decisions.

