Three Modern Case Studies

Moneyball: A Sports Case Study

Setting the Scene

- The Athletics (A’s) are a baseball team that plays in Oakland, CA.
- In the early 2000s, the A’s were a low budget baseball team.
- The owners of the team refused to increase their budget.
- The A’s lost their 3 best players to teams that were willing to pay more.
- Since there is no salary cap in baseball, the A’s had to find a way to remain competitive with much wealthier teams.

Baseball Management meets Data Science

- Billy Beane was the general manager of the A’s at that time.
- Beane began to look for undervalued players who he thought could become stars in the league. He took the revolutionary step of using data to help decide how to formulate his roster.
- Before Beane, baseball scouts decided which players were destined for stardom after observing a few games, and their gut feelings.
    - Scouts would judge appearances of players, and even their spouses!
    - Using his data-driven approach, Beane often came up with conclusions that contradicted what his top scouts thought.
    - Beane was derided by professional scouts and the media.

The 2002 Athletics Season

- The A's began the season with 20 wins and 26 losses. Deriders of the "Moneyball" strategy gloated, contending a data-driven method could never beat the work of professional scouts.

- The A's, however, recovered by winning 16 out of 17 games in June. Between August and September, the A's won 20 games in a row, setting the American League record at that time (2016).

- Overall, the A's finished with 103 wins, tied with the Yankees for the most in the major league.
  - The A's had the 3rd smallest payroll in the league, at less than $40Mn
  - The Yankees had the largest payroll, at over $125Mn

How Did the A’s Do It?

At the core of the A’s success was Sabermetrics (SABR is the Society for American Baseball Research): i.e., using data to predict baseball-player performance.

Fielding:

- More athletic players have better range on the field. This enables these players to get to more baseballs.
- Being able to get to more balls (especially if you get to them in a difficult spot) leads to more errors.
- Players with a lot of errors were seen as poor fielders by scouts, even though they’d reach balls that others couldn’t reach.
- Beane innovated here, using advanced statistics such as “defensive runs saved”, to measure fielding ability.

Batting:

- Consider the RBI (runs batted in): this statistic is meant to capture how many runs a batter generates, but ignores the batter himself!
- For example, if a batter hits a triple with no one on base, the batter has still put the team in a good position to score.
- Beane created and used other novel statistics such as “on-base percentage” and “runs created” to more accurately capture a batter’s ability to generate runs.

Conclusion

- The success of the Oakland A’s provided the first really compelling evidence that a data-driven method could work in sports.
  - The Moneyball strategy has been adopted by other sports.
  - The Rockets of the NBA and the Blackhawks of the NHL provide evidence this approach can work throughout sports.
- Statistics cannot guarantee a playoffs win, but can get a team to the playoffs!
  - Moneyball requires large sample sizes; over the 162 game baseball season, a player’s statistics will balance out to reveal his true value.
  - However, playoff series are much shorter (either 1, 5, or 7 games), so invariably, other factors, including luck, come into play.

Tennis

- In tennis, an unforced error occurs when a player hits the ball either into the net or out of bounds.
- While it may appear that having a high number of unforced errors is bad, players make unforced errors when they go for tough shots.
- So players with a high rate of unforced errors could actually be going for more shots, so they could likewise be hitting more winners.
- A more useful, but less cited, statistic is the ratio of winners to unforced errors; indeed, the player with the higher ratio wins about 90% of the time.

Hockey / Basketball

- In hockey and basketball, turnovers (when the ball changes hands from one team to the other) are tracked.
- Common sense would dictate: giveaways are bad and takeaways are good.
- However, players can only record takeaways when their team does not have the ball, and can only record giveaways when their team does have the ball.
- Good teams have the ball more, giving the players more opportunities for giveaways rather than takeaways.
- In the 2015-16 NBA season, 7 of the 10 leaders in giveaways were recent all-stars; they had the ball more, and in turn, they gave it away more.
- Giveaways turns out to be a misleading statistic.

Conclusion

- In Moneyball, Beane created statistics that more accurately measured what traditional statistics were trying to capture.
- In the other sports, data analysts followed suit, discrediting statistics which sports commentators and fans had come to believe in.
- Because shortcomings of the “eye-test” (watching and judging games without statistics), data science, specifically predictive analytics, is gaining acceptance in sports.

FiveThirtyEight

A Politics Case Study

FiveThirtyEight

- Founded by Nate Silver, who first won acclaim for his work in Sabermetrics
- Correctly predicted the outcome in *every* state in the 2012 presidential election
- Correctly predicted *every* Senate election outcome in 2008

Triumph of the Nerds: Nate Silver Wins in 50 States

Image Source

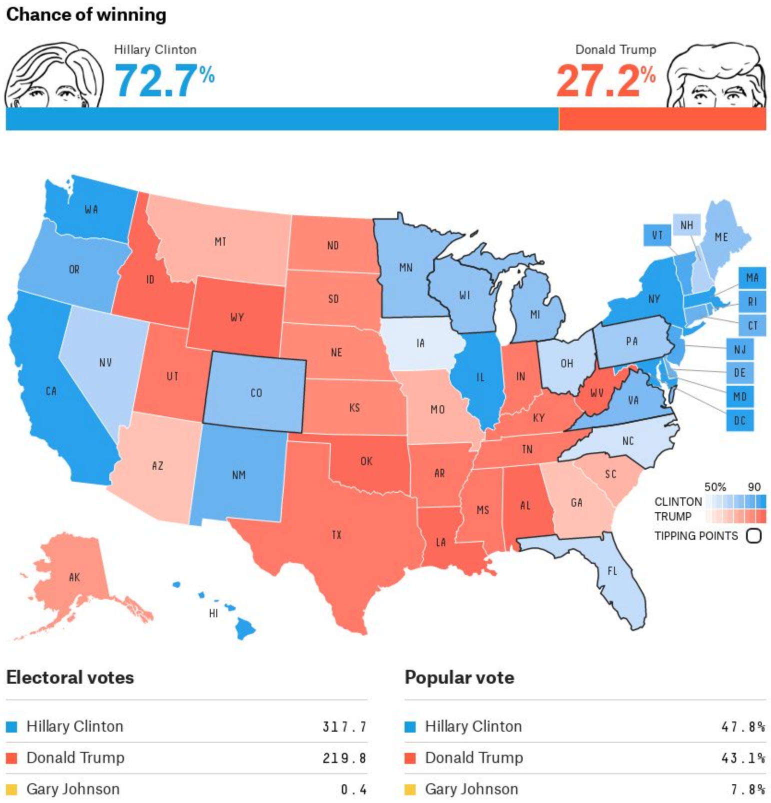


Image Source

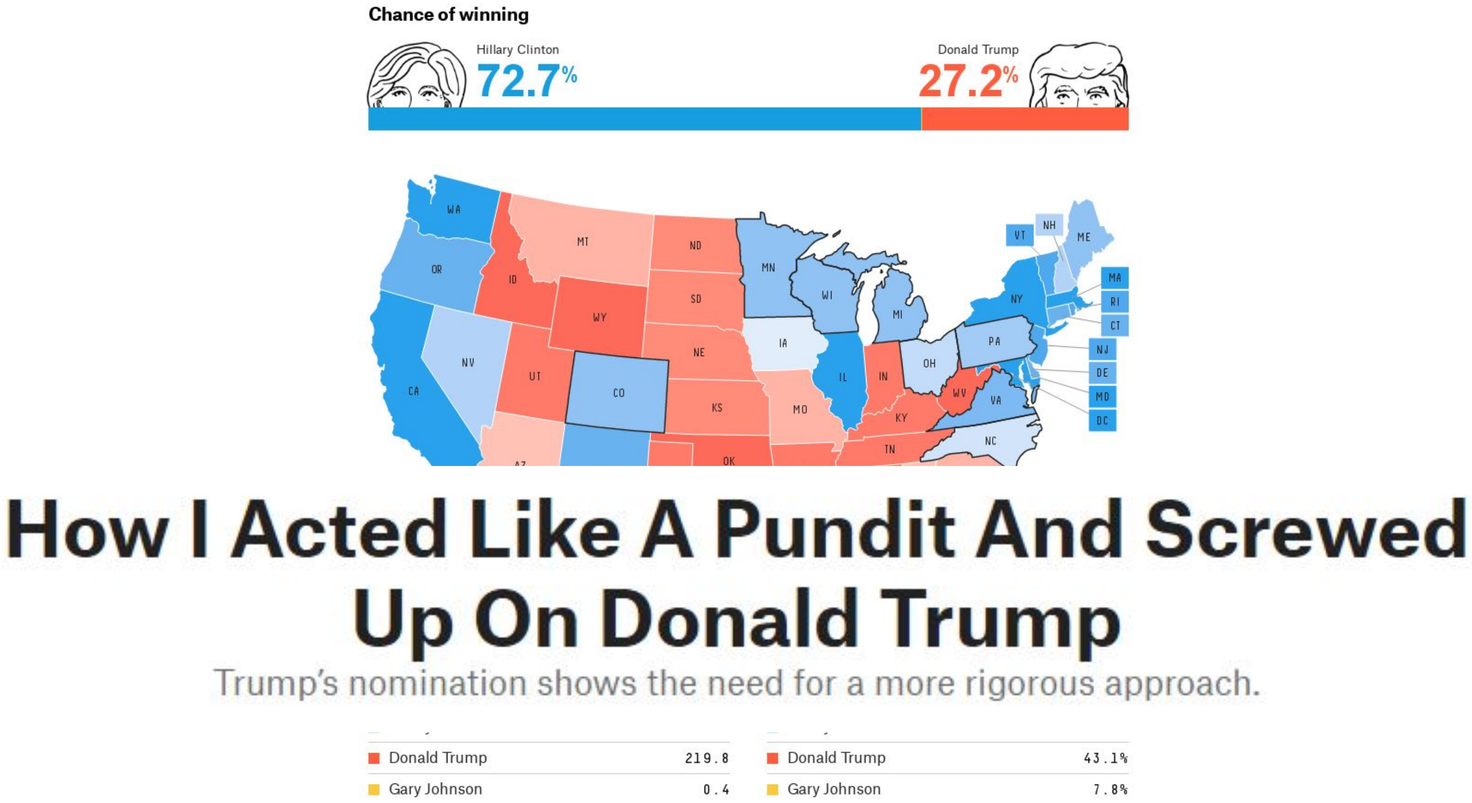


Image Source

Goal (in a presidential election)

Predict a probability that the Democrat wins, a probability that the Republican wins, and perhaps a probability that a third or fourth party candidate wins, as well.

Intermediate Goal

Build a probability distribution over all possible electoral vote outcomes (e.g., 332 vs. 206; 173 vs. 365; etc.). Then tally the probabilities associated with the various wins.

What if the election were today?

Inter-state poll aggregation:

1. Using the results of today's per-state polls, compute a probability distribution over all possible electoral vote (i.e., national) outcomes

2. Aggregate over electoral vote outcomes to compute each party's win probability

Who will win some state?

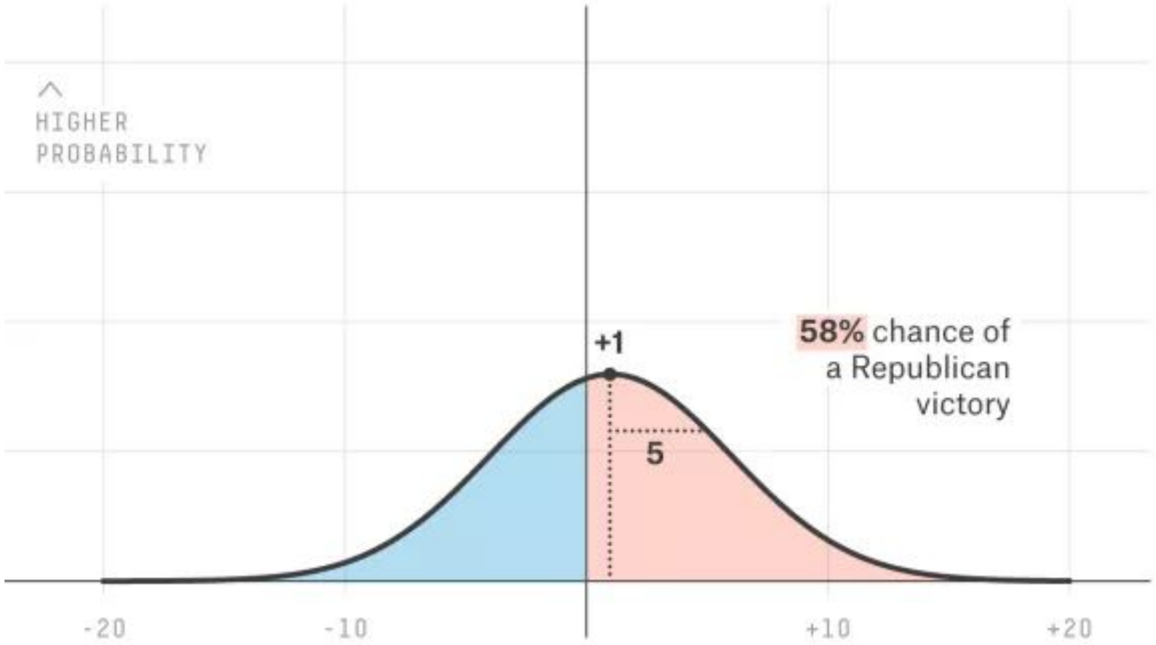


Image Source

Who will win some state?

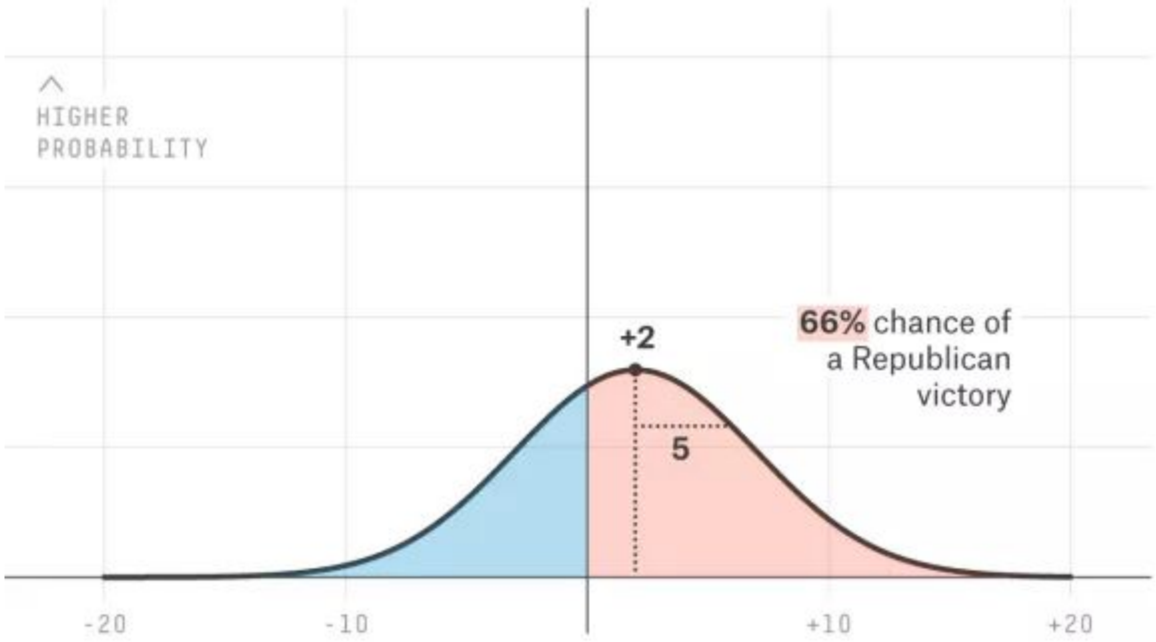


Image Source

Who will win some state?

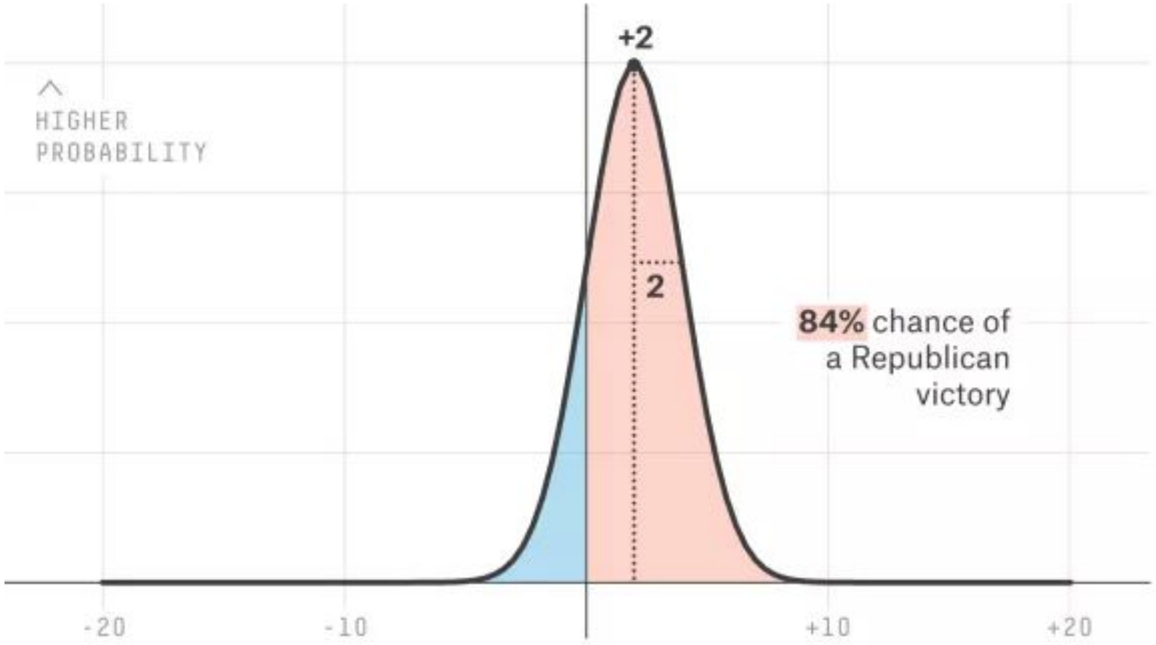


Image Source

Underlying assumption

Per-state polls taken on election day correctly predict national election outcomes

Success of Gallup Polls

Further Intermediate Goal

Use past and current poll data to predict the results of election-day polls

How can we predict future polls from past?

Answer

- (Statistical) machine learning
- Specifically, regression, to infer a trend line

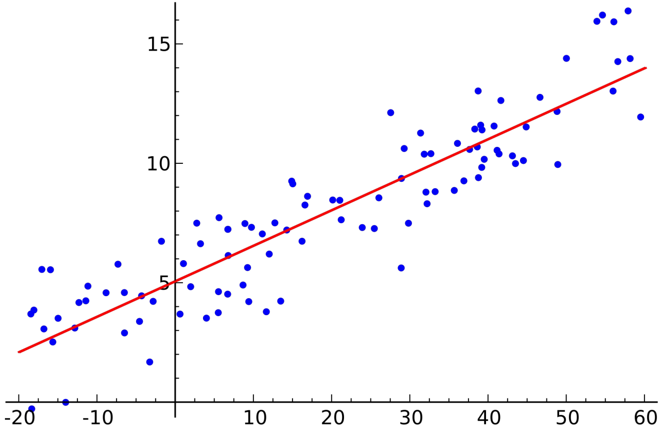


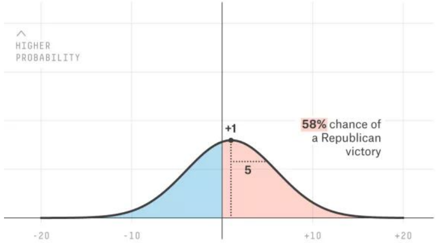
Image Source

Knowing that the election is not today...

Inter-state poll aggregation:

1. Predict future (election-day) polls from past polls
2. Using the predicted future per-state polls, compute a probability distribution over all possible electoral vote (i.e., national) outcomes
3. Aggregate over electoral vote outcomes to compute each party's win probability

Aggregating intra-state polls



Calculate the center of all the polls (mean, median, etc.)

Calculate the spread among the polls:

- Professor Wang’s (Princeton) approach:
  how much do the polls differ from one another?

- 538 approach: how accurate have these pollsters been in the past?

- Correct approach: how do the polls vary from how the population would have voted had the actual election been the day of the poll?

Competing approaches

- Professor Wang: assumes independence across state polls, and calculates the probability distribution in closed-form
- 538 calculates correlations between states (e.g., outcomes in OH and PA might be highly correlated), but is then forced to resort to simulation to estimate the probability distribution

Similar states usually have similar outcomes

Correlation matrix after 20,000 simulations, polls-only model, June 27, 2016

	Ala.	Calif.	Fla.	Minn.	N.C.	N.M.	R.I.	Wis.
Alabama		.60	.61	.53	.72	.54	.41	.55
California	.60		.73	.67	.69	.80	.61	.68
Florida	.61	.73		.67	.75	.70	.63	.76
Minnesota	.53	.67	.67		.68	.58	.64	$\boxed{.84}$
N. Carolina	.72	.69	.75	.68		.60	.53	.67
New Mexico	.54	.80	.70	.58	.60		.54	.64
Rhode Island	.41	.61	.63	.64	.53	.54		.69
Wisconsin	.55	.68	.76	$\boxed{.84}$	.67	.64	.69	

Image Source

538 Modeling Principles

Model Desiderata

1. Probabilistic, not deterministic
2. Empirically sound (empirically justifiable assumptions)
3. *Not* unstable: it should not oscillate wildly given additional inputs
4. *Do not* tweak if the prediction is undesirable

Extras

What are the best predictors?

- Fundamentals at the national, regional, or state level
  - Demographic makeup (race, religion, etc.)
  - Economic indices
- Per-state polls

Polls-only vs. polls-plus models

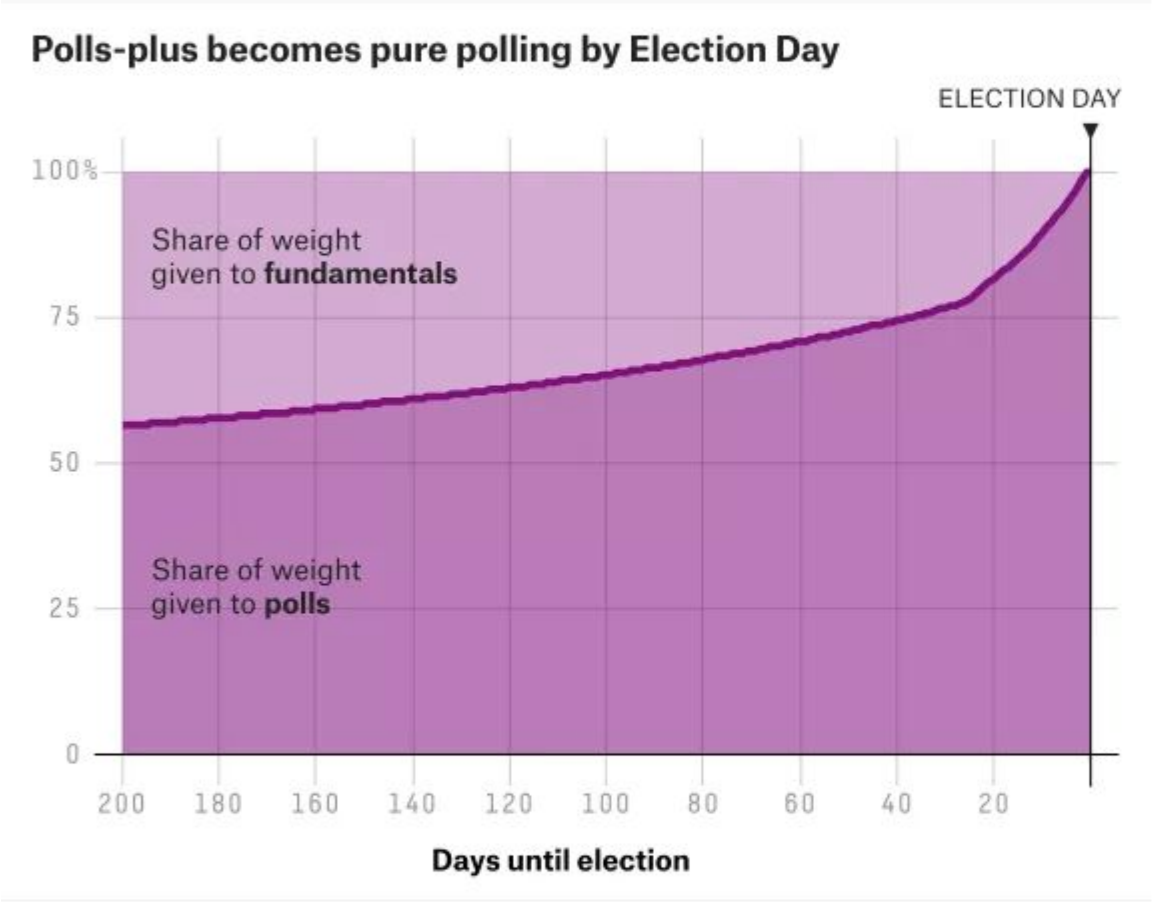


Image Source

How can we predict future polls from past?

Answer

- (Statistical) machine learning
- Specifically, regression, to infer a trend line

Technical detail:

- Predict per-poll, and then aggregate to produce per-state probability distributions

- Or simplify, by aggregating state polls first, and then predicting a single per-state probability distribution

Netflix: An Internet Case Study

Image Source

Netflix is data-driven

- A typical subscriber will only look at about 10 to 20 titles, and only a couple of rows of recommendations
- If the user does not find something of interest in the first 90 seconds, Netflix runs the risk of them abandoning the service
- Netflix has spent more than a decade and lots and lots of money refining its recommendation system
- 75 percent of viewer activity is driven by other users' recommendations
- Netflix wants its service to be so engaging that users pick it over other end-of-the-day activities, like reading a book or magazine, TV or Facebook

Collaborative filtering

- Method of making predictions (filtering) about the interest of one user based on the tastes and preferences of other users (collaborative)

Underlying assumption

- If A has the same opinion as B on an issue (e.g., liking or disliking a movie), then on some other issue, it is more likely that A will have the same opinion as B than the same opinion as some other person chosen at random

Image Source

Netflix Prize

- A competition for the best collaborative filtering algorithm
- Goal: to predict user ratings for films based on previous user ratings, without any additional information about the films or the users
- The data consisted of 100,480,507 ratings by 480,189 users of 17,770 movies
- On 21 September 2009, BellKor's Pragmatic Chaos team won the grand prize of US $1,000,000, besting Netflix’s own algorithm by 10.06%

Image Source

Exploration vs. Exploitation

- Netflix is careful not to over-personalize
- Chris Jaffe, Netflix's vice president of product innovation, loves dark TV dramas
- The algorithm knows this, so it usually points him to shows of that genre (exploitation), but every now and then it throws in something else (exploration), like a horror film, to gauge Jaffe's interest
- Netflix risks upsetting Jaffe by doing so, but learns more about Jaffe's interests

Image Source

Experimental design at Netflix

- Netflix A/B tests everything from images to the size of font on the screen

Image source:
http://static4.businessinsider.com/image/56d04b996e97c61a008b9e1e-3264-2448/netflix.jpg

Experimental design at Netflix

- Netflix A/B tests everything from images to the size of font on the screen

- Netflix tested whether images that appear at the top of the screen should be static or a carousel
  - A rotation of three images worked best

- Netflix tested whether detailed synopsis make users more likely to watch a title
  - It turned out that a couple of short and clear sentences is best

- Netflix uses a large set of users (roughly 300,000, worldwide) in its A/B tests to increase its confidence that its design changes will indeed increase engagement

Sources

1. Netflix lifted the lid on how the algorithm that recommends you titles to watch actually works
2. Netflix knows exactly how long it has before it loses you
3. The Science Behind the Netflix Algorithms that Decide What You'll Watch Next
4. Collaborative Filtering (Wikipedia)
5. Netflix Prize (Wikipedia)