

# How a small team of academics correctly predicted a Trump victory

 [blogs.lse.ac.uk/europpblog/2016/11/11/how-a-small-cambridge-based-startup-correctly-predicted-a-trump-victory/](https://blogs.lse.ac.uk/europpblog/2016/11/11/how-a-small-cambridge-based-startup-correctly-predicted-a-trump-victory/)

11/11/2016

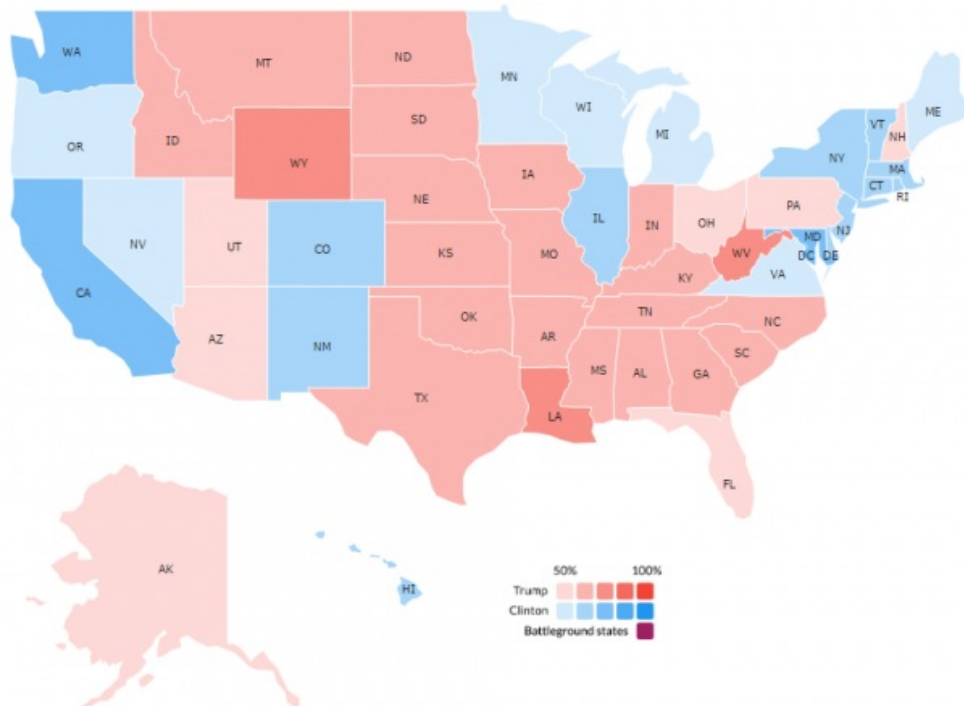
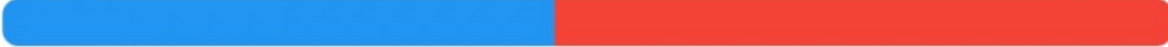
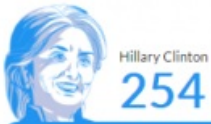
*Much like the UK's vote to leave the European Union, few polling experts predicted Donald Trump would win the US presidential election. But this was not the case for all polling companies. [Vuk Vukovic](#) outlines a prediction model he developed with a small team of colleagues that proved remarkably accurate in calling the final result of the election and assesses how the method adopted could be used to provide more accurate predictions of future elections.*



The US election result came as a shock to many, but it was the pollsters that took the biggest hit. All the major poll-based forecasts, most models, the prediction markets, and even the 'super-forecaster' crowd all got it wrong. They estimated high probabilities for a Clinton victory, even though [some were more careful](#) than others in claiming that the race would be very tight.

Our prediction survey, however, was [spot on](#). We, at [Oraclum Intelligence Systems](#), a Cambridge-based start-up, predicted a Trump victory, and we called all the major swing states in his favour: Pennsylvania (which no single pollster gave to him), Florida, North Carolina, and Ohio. We gave Virginia, Nevada, Colorado, and New Mexico to Clinton, along with the usual red states and blue states to each. We only missed three – New Hampshire, Michigan, and Wisconsin (although for Wisconsin we didn't have enough survey respondents to make our own prediction so we had to use the average of polls instead). Essentially the only genuine misses in our method were Michigan, where we gave Clinton a 0.5 point lead, and New Hampshire where we gave Trump a 1 point lead. The figure below shows how our prediction map looked prior to the election.

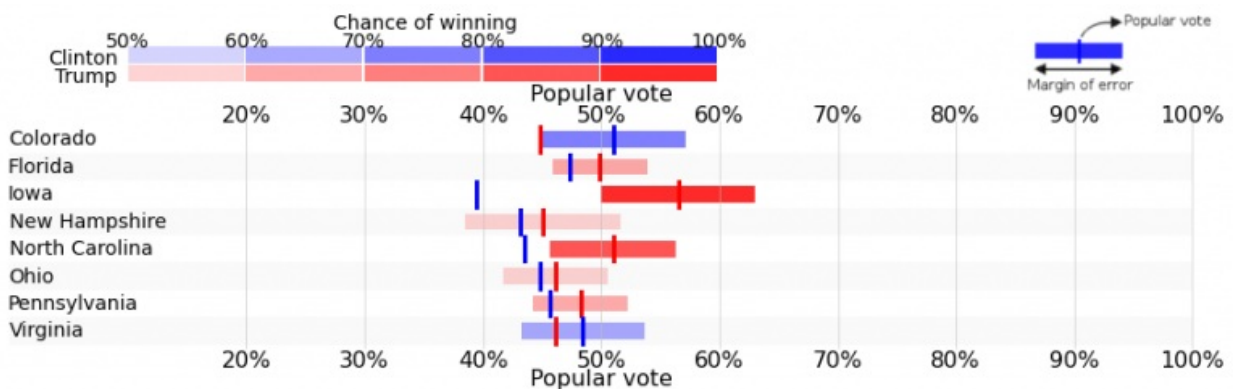
**Figure 1: US presidential election prediction map**



Source: [Oraclum](#)

Every other state, although close, we called right. For example, in Florida we estimated 49.9% to Trump vs. 47.3% to Clinton. In the end it was 49.1% to 47.7%. In Pennsylvania we gave 48.2% to Trump vs. 46.7 for Clinton (it was 48.8% to 47.6% in the end). In North Carolina our method said 51% to Trump vs. 43.5% for Clinton (Clinton got a bit more at 46.7% but Trump was spot on at 50.5%). Our model even gave Clinton a higher chance to win the overall vote share than the electoral vote, which also proved to be correct. Overall for each state, on average, we were right within a single percentage point margin – the full prediction can be read [here](#) and a snapshot of our results is reproduced in the table below.

Table: Prediction of state results and popular vote in the US presidential election



## Source: [Oracleum](#)

It was a big risk to 'swim against the current' with our prediction, particularly in the US where the major predictors and pollsters were always so good at making correct forecasts. But we were convinced that the method was correct even though it offered, at first glance, very surprising results.

### How did we predict the result?

We used a different type of survey to produce our results – a prediction survey. The established poll-based forecasters all usually pick up the 'low-hanging fruit' polling data and run it through an elaborate model. We, on the other hand, needed to get actual people to come to our site and take the time to make a prediction for their state. So instead of just picking up raw data and twisting it as much as we could, we compiled our own data. Given that we were doing this with limited resources, our sample size was rather small (445). This indicates that our method did not require a representative sample to make a good prediction, nor is it sensitive to typical problems of online polls such as self-selection.

However, even with a small and unrepresentative sample the method worked. Why? Our survey asked respondents not only who they intended to vote for, but also who they thought would win and by what margin, as well as their view on who other people thought would win. It is essentially a 'citizen forecaster' concept adjusted to take account of 'groupthink'. The idea is to incorporate wider influences, including peer groups that shape an individual's choice on voting day. This is why our method was perfect for being conducted via social networks.

Our respondents had to log into the survey via either Facebook or Twitter in order for us to gather a single piece of information from them – an average of how their friends (followers and followees) voted. So our survey did require some social network sharing and re-tweeting. All the data was anonymous so we cannot see who anyone voted for, but we do have enough data to perform a type of a network analysis. This helped us uncover whether our respondents were living in their own 'bubbles' where everyone around them simply thinks the same way. This would obviously make for a worse forecast so such biases were adjusted for.

Finally, we performed 100,000 simulations to get the most probable outcome. This ultimately ended up in the prediction that Donald Trump taking Florida, Pennsylvania, North Carolina, and Ohio, and winning the presidency by electoral college vote, was the most likely scenario. Believe it or not, we tested the same method on the Brexit referendum and it provided the same results. We had 6 models tested, three of which showed Leave and three of which showed Remain. We did not bother with being correct at the time, we just wanted to see which method was the best one. The one method that gave us a 51.3% for Leave is the same one that predicted the victory for Donald Trump. We intend to test it further.

### The logic behind the approach

Perhaps the most puzzling question is why would this approach work at all? When people make choices, for example in elections, they usually succumb to their standard ideological or otherwise embedded preferences. However, they also carry an internal signal which tells them how much chance their preferred choice has. In other words, they think about [how other people will vote](#). This is why people tend to vote strategically and do not always pick their first choice, but opt for the second or third choice, only to prevent their least preferred option from winning.

Each individual therefore holds some prior knowledge as to who he or she thinks will win. This knowledge can be based on current polls, or drawn from the information held by their friends or individuals they find more informed about politics. Based on this it is possible to draw upon the wisdom of crowds where one searches for informed individuals thus bypassing the necessity of having to compile a representative sample.

However, what if the crowd is systematically biased? For example, many in the UK believed that the 2015 election would [yield a hung parliament](#). In other words, information from the polls created a distorted perception of reality

which was returned back to the crowd and thereby biasing their internal perception. To overcome this, we need to see how much individuals within the crowd are diverging from opinion polls, but also from their internal networks of friends.

In conclusion, we realise that by leaving the domain of the representative sample we are entering uncharted territory. It is going to take quite a few of these social experiments to answer the question of whether polling is just a glorified 'art' or whether it is actually based in real science. This is therefore just the beginning for our prediction method.

*Please read our comments policy before commenting.*

*Note: This article gives the views of the author, and not the position of EUROPP – European Politics and Policy, nor of the London School of Economics. Featured image credit: Gage Skidmore (CC-BY-SA-2.0)*

Shortened URL for this post: <http://bit.ly/2fVv5WO>

---

### About the author

**Vuk Vukovic** – *University of Oxford*

Vuk Vukovic is a DPhil student of politics at the University of Oxford and the Director and Co-founder of Cambridge-based start-up Oraclum Intelligence Systems (together with Dr Dejan Vinkovic and Prof Dr Mile Sikic).

