

# THE DATA SCIENCE OF PREDICTING ELECTIONS

## HOW A MODEL PREDICTED THE RESULTS OF THE ONT. ELECTION

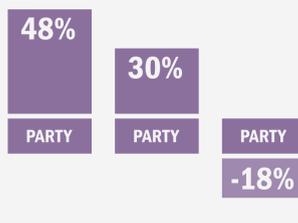
### PSEPHOLOGY

Psephology is the statistical analysis of elections and voting trends. Psephologists use polls, and demographic and historical data to create models that predict the outcomes of elections before they happen.



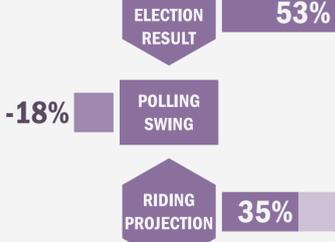
### THE UNIFORM SWING MODEL

In countries like Canada and the United Kingdom, election prediction models use the metric of swing to translate topline polling data into first-past-the-post riding predictions. Swing represents the amount of support a given party gains or loses from one election to the next. This USRI project built a swing model to predict the results of the 2022 Ontario Election.



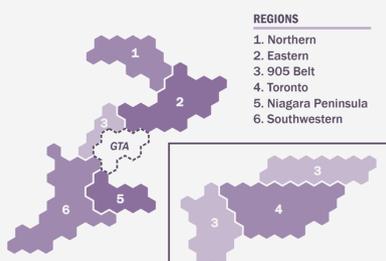
### HOW THE MODEL WORKS

The model first calculates swing by comparing each party's share of vote from the previous election to its standing in current opinion polls. It then adds the swing to that party's previous share of vote in every individual riding. For instance, the model assumes that if a party gains 12% in a given region, it will gain roughly 12% in every riding in that region.



### REGIONAL TRENDS AND POLLS

Ontario is a politically diverse province, with different parties gaining and losing support in different places. To reflect this, the model uses polling localized to distinct regions, such as Toronto or Southwestern Ontario, rather than province-wide data. CBC News compiles polls from different firms into a common set of six regional polling averages.



### PARTY SEAT RANGES

Professional psephologists run hundreds to thousands of simulations to determine the probability of victory for each party in every riding. This model uses a simpler system, categorizing a riding as safe, likely or a tossup depending on the margin of victory of the party projected to win. The number of ridings in each category is totalled, providing a range of possible seat outcomes for each party.

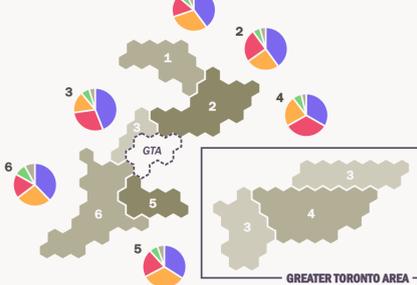


## 2022 ONTARIO UNIFORM REGIONAL SWING MODEL ELECTION PREDICTION

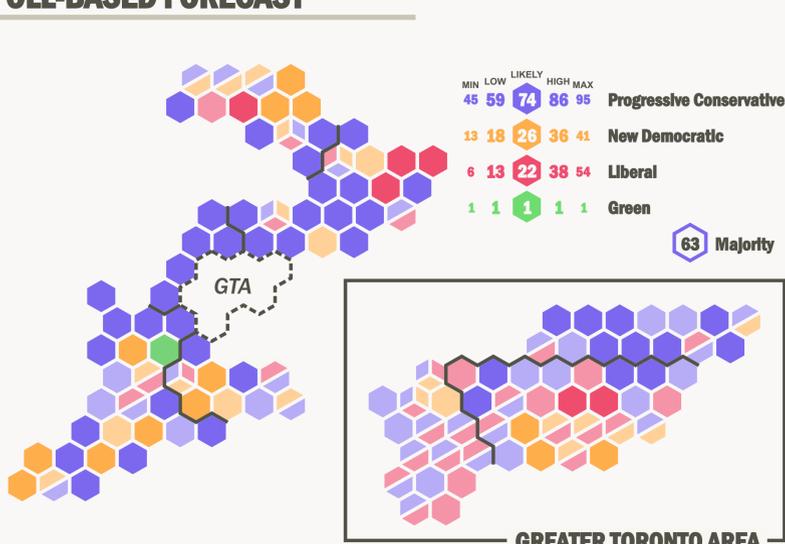
### AND HOW IT COMPARED TO THE RESULTS

#### FORECASTING

The model predicts the result of an election if it were held on the date of the poll or polls provided. The prediction below is based on polling averages released by the CBC Ontario Poll Tracker on June 2, the day of the election. Each hexagon represents a single riding, with dark lines separating each region.

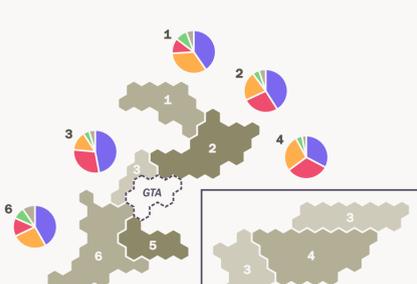


#### POLL-BASED FORECAST

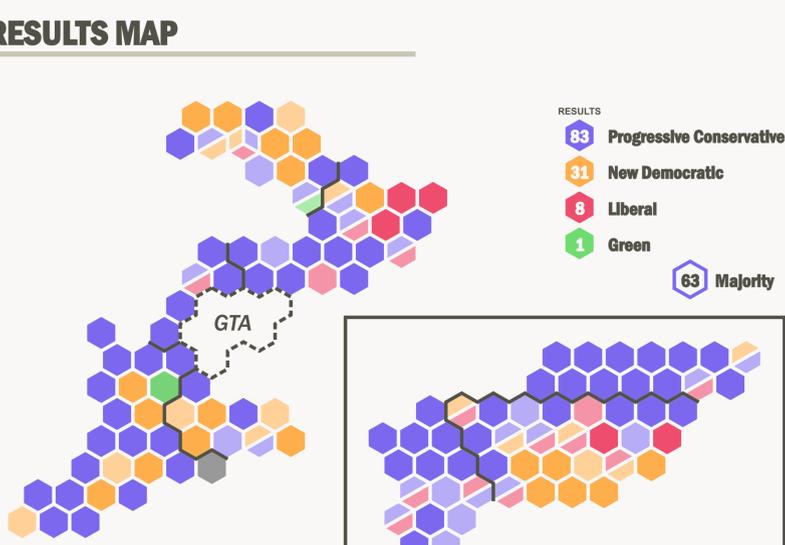


#### ELECTION RESULT

As ridings reported it became clear that polls had vastly overestimated Liberal performance. While the other three parties posted results between the model's low and high ranges, the Liberals just barely outperformed their minimum expected seat count.

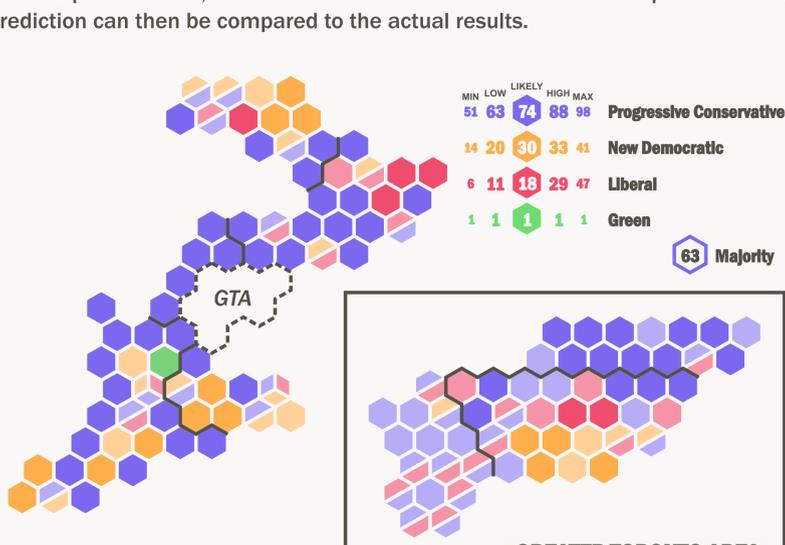


#### RESULTS MAP



#### RESULTS CONVERSION

The model can only perform as well as the polls it is provided. To better evaluate the model's performance, the actual election results can be used as inputs. The model's prediction can then be compared to the actual results.



#### MODEL PERFORMANCE

Though the model projected a lower Liberal total when using real results, it still overestimated the Liberal result. This Liberal underperformance despite the model's accuracy for the other parties suggests that the Liberals have unusually low vote efficiency. Despite the Liberals winning 5,000 more votes than the NDP, only 8 Liberals will go to Queen's Park compared to 31 New Democrats.

Despite the Liberal underperformance, the model performed well overall. In 115 of the province's 124 ridings, the winning party was correctly identified as either the victor, or a contender in a tossup riding.

A major disadvantage of this model is its overreliance on regional opinion polls. This results in low accuracy, particularly in diverse regions like Northern Ontario, where demographic differences from riding to riding are not taken into account.

Providing the model with historical voting data for each riding would allow it to make a more accurate and specific prediction, taking past party performance into account.