Welcome to Dataland, an imaginary country whose elections have frustrated many an aspiring political forecaster. Dataland is a young democracy, having thrown off the shackles of an autocratic one-party system in 1984. The authors of its constitution were so eager to hold free and fair elections that they scheduled a full slate of them every single year, making the 2024 cycle the country’s 41st contest.

Dataland consists of three geographic regions (all subsequent proper names are courtesy of ChatGPT): Metaflux Realm, with five provinces; Circuit Confederation, with four; and the Synapse Territories, with three. The country’s population is drawn from three ethnic groups: the Pythons, the Javarians and the Cobolites.

The electoral system for Dataland’s presidency is a near carbon copy of America’s. Its 12 provinces each hold a first-past-the-post election for the office—whichever candidate gets the most votes wins the province. Provinces are assigned electoral votes in proportion to their representation in the legislature, in which each province is guaranteed at least six seats regardless of population. All provinces allocate their electoral votes on a winner-take-all basis. The electoral college is also chosen via first-past-the-post: whichever candidate gets the most electoral votes becomes president, even if they do not obtain an absolute majority. In the event of a tie in the electoral college, the first tiebreaker is the number of provinces won, followed by the nationwide popular vote.

Much to the surprise of acolytes of Maurice Duverger, Dataland has failed to develop a two-party system despite its pure first-past-the-post electoral rules, thanks mainly to regional and ethnic cleavages. Instead, four different political parties have won electoral votes in its brief democratic history, and all of them continue to draw support today. The party with the highest average historical vote share is the Progressive Data Action League (PDAL): a big-tent, centre-left party with a solid base of support in every region and in every ethnicity, which held power during Dataland’s transition to democracy. Although the majority of voters usually support right-of-centre parties, they split their votes between two parties: the smaller Cobolite Coalition (CC), a Cobolite nationalist party that concentrates its support in Metaflux Realm, and the larger, business-friendly Data Growth Movement (DGM), which is more popular in the other regions. Finally, the Sovereign Synapse Party (SSP) is a separatist party advocating secession for its region. It only runs candidates in provinces in the Synapse Territories.

As an imaginary country—albeit one whose economic data up through mid-2023 (but not subsequently) we may have copied from the randomly selected real country of New Zealand—Dataland is mercifully free of a few factors that complicate real-world electoral analysis. Each party’s voter turnout rate is fixed over time, meaning that all changes in vote shares from year to year result from voters changing which party they support, rather than from deciding whether or not to vote at all. (As a result, we are providing data on vote shares and total population, but not on absolute vote totals.) The populations and mix of ethnicities in Datalandish provinces never change. Similarly, although parties’ vote shares in each province oscillate from year to year—possibly for quantifiable reasons like the strength of the economy, and possibly for unquantifiable ones like the comparative appeal of their candidates or of their proposed policies—the baseline popularity of each party in each province is constant all the way from 1984 to 2024, as are the characteristics of each pollster. There is no equivalent in Dataland of, say, West Virginia in the United States shifting over time from a reliably Democratic-voting state to a reliably Republican-voting one, because there are no long-run time trends. As a result, when predicting elections in 2024, you should treat historical results and polls from 1984 as being just as informative as those from 2023 are.

As befits the country’s name, data on Datalandish elections are relatively abundant. Forecasters seeking to predict the country’s 2024 presidential contest will have access to the following data sources:

* Historical presidential-election results: vote shares for every party in every province in every presidential election, as well as the identities of the party in power at the time, and the winners of each province, the presidency, and the national popular vote. Whichever party wins the presidency in a given election takes power immediately, and holds it for one year. These appear in the dataland\_election\_results\_1984\_2023.csv file.
* Provincial demographics: Region, population size, electoral votes, and share of the provincial population belonging to each ethnic group. These appear in the dataland\_demographics.csv file.
* Electoral calendar: A list of start and end dates in each election year for the biennial conferences for each political party, which tend to attract a flurry of media coverage; the dates of televised presidential debates; and the date of the presidential election. These appear in the dataland\_electoral\_calendar.csv. file.
* Economic context: National-level GDP growth, unemployment, inflation, and stock-market returns at quarterly intervals since 1980. Provincial-level economic data are not available. These appear in the dataland\_economic\_data\_1984\_2023.csv file for the 1984-2023 period, and in the dataland\_economic\_data\_2024\_scenarios.csv file for the five scenarios in 2024. The listed date corresponds to the end of the quarter in question, and data become publicly available as soon as each quarter ends.
* Polls: A complete dataset of presidential voting-intention polls for each election cycle. The entry for each survey lists the name of the pollster; the geographic area surveyed (nationwide, regional or provincial); the sample size; the mode (live-caller, interactive voice response (“robocalls”) or online); the date of field work; the date the poll was published; the population surveyed (all adults, registered voters or likely voters); the identity, if any, of the political party that paid for the survey; and the vote shares for each party and for undecided voters. These appear in the dataland\_polls\_1984\_2023.csv file for the 1984-2023 period, and in the dataland\_polls\_2024\_scenarios.csv file for the five scenarios in 2024.