The Malaysian Election Corpus (MECo): Federal and State-Level Election Results from 1955 to 2025

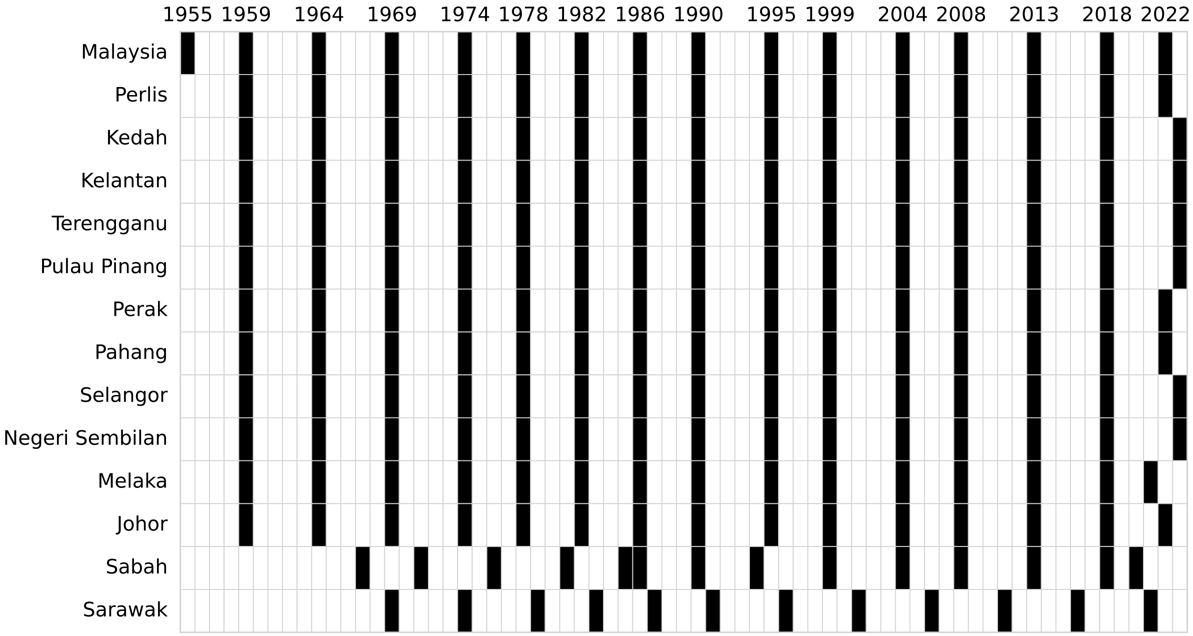
Thevesh Thevananthan

Empirical research and public knowledge on Malaysia's elections have long been constrained by a lack of high-quality open data, particularly in the absence of a Freedom of Information framework. We introduce the Malaysian Election Corpus (MECo; ElectionData.MY), an open-access panel database covering all federal and state general elections from 1955 to the present, as well as by-elections from 2008 onward. MECo includes candidate- and constituency-level results for nearly 10,000 contests across seven decades, standardised with unique identifiers for candidates, parties, and constituencies. The database also provides summary statistics on electorate size, voter turnout, rejected votes, and unreturned ballots. This is the most well-curated publicly available data on Malaysian elections, and will unlock new opportunities for research, data journalism, and civic engagement.

Background & Summary

Malaysia's electoral history is among the most dynamic in Southeast Asia, encompassing 2,770 federal general elections and 6,935 state general elections (Figure 1), as well as hundreds of off-cycle by-elections across a multiethnic, multi-party system. Furthermore, Malaysia offers significant scope for the study of democratisation, having experienced its first change of ruling party in 2018, and its first ever hung Parliament as recently as 2022. However, conducting empirical studies of Malaysian elections is challenging due to the lack of a comprehensive, standardised, and publicly available dataset that provides a single source of truth for scholars. The Election Commission (EC) does not publish machine-readable open data which abides by best practices for data sharing, preferring instead to limit citizens to searching up isolated results.[1] Slightly richer information is available via gazetted election results published as subsidiary legislation,[2] but these are published in PDF format only. The lack of a Freedom of Information Act further complicates efforts to acquire electoral returns.

Figure 1: Federal and state general election years



Amidst this paucity of data, global initiatives such as the Constituency-Level Elections Archive (CLEA)[3] provide immensely valuable cross-country coverage, including for Malaysian elections. However, they generally focus on federal contests, and within that scope, only on the number of votes won by each candidate (thus omitting information such as the electorate size, voter turnout,unreturned ballots, rejected votes, etc). Similarly, international turnout or election integrity datasets[4,5] capture only high-level national indicators. Locally, a number of news and civil society organisations[6–8] compile election data to varying degrees of completeness and quality, but these efforts – while laudable for their public service, and valuable as a stopgap measure – typically lack proper data hygiene and standardisation, and are not subjected to systematic validation and review, thus limiting their usefulness for rigorous empirical research and long-term preservation.

In this paper, we address this gap by providing the first comprehensive open database of Malaysian election results at the federal and state level – we intend for this to be a living resource which provides the go-to empirical foundation for research and journalism on Malaysian elections. The database covers *all* general elections since the pre-independence general election in 1955, and 115 by-elections held since 2008. In total, it records 25,552 candidates representing 99 political parties in 9,705 unique electoral contests from 1955 to 2025. The dataset comprises two main components:

- **Ballots**: The final results for each state legislative assembly constituency (DUN) and federal Parliament constituency (Parliament), with the number of votes received by each candidate.
- **Summaries**: The electorate size, ballots issued, unreturned ballots, and votes rejected in each constituency. For each constituency, we also derive the margin of victory (majority), voter turnout rate, vote rejection rate, unreturned ballot rate, and majority as a share of valid votes.

Furthermore, our database offers three key advantages built on the use of unique identifiers (UIDs). First, we encode a UID for each candidate, thus allowing the tracking of a single individual across time, even if they run under different political parties or change the name which they use in public life; this is especially important in Malaysia, because politicians are not required to use their official name on electoral ballots. Second, we encode a UID for each party, allowing for tracking of party evolution, such as the expansion of the Alliance Party (*Parti Perikatan*) into the National Front (*Barisan Nasional*). Third, we provide a lookup for each unique date-state-constituency combination, allowing for the creation of geospatial lineage and tracking of changes in constituency names. In general, the use of UIDs makes our database highly extensible – a deliberate choice to help other scholars seamlessly build on our work by creating new lookups, rather than altering core datasets.

To the best of our knowledge, no comparable database exists. As a living resource, this paper lays the foundation for future data curation, as well as research into areas like malapportionment, gerrymandering, local-level voting patterns, and the spatial dynamics of political competition. We also hope that MECo will serve as a catalyst for broader collaborations between academics, civil society, journalists, and election observers, supporting both scholarly inquiry and public accountability.

Finally, we note that while our work is the first of its kind for Malaysian elections, it follows a growing body of recent academic work focused on compiling and curating country-specific election data for reuse.[9–14] By situating MECo within this emerging tradition of high-quality electoral data curation, we contribute to the rapidly-improving global infrastructure of comparative political research. This work reflects a commitment to transparency, reproducibility, and the democratisation of access to electoral information.

Methods

There are three sources of data we used to construct our database:

1) Physical post-election reports.[15–45]

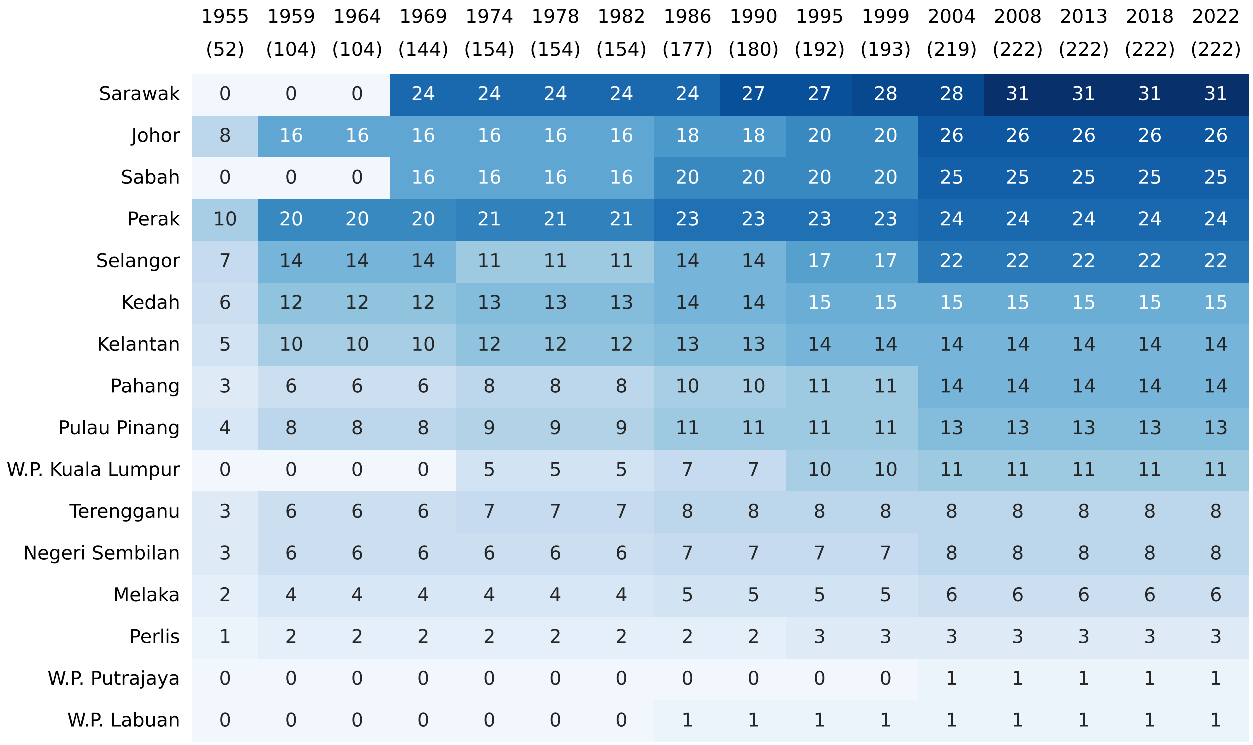
2) Gazetted election results published as subsidiary legislation.[2]

3) Digital results published officially by the EC via a dashboard.[1]

To enable checking and validation, we made images of the post-election reports and PDFs of the gazetted election results (sources 1 and 2) available via this anonymous Dropbox link. We constructed MECo beginning with federal general elections, then state general elections, and finally by-elections. This is because state legislative assembly constituencies (DUNs) must lie completely within the boundaries of a federal parliamentary constituency (Parliament), so it was sensible to begin with the superset. By-elections coming last is an intuitive choice, since a by-election must follow a general election by definition; validation of by-election data is therefore dependent on having complete federal and state-level results.

Federal General Elections

Figure 2: Federal election coverage (number of seats)



	1955 (52)	1959 (104)	1964 (104)	1969 (144)	1974 (154)	1978 (154)	1982 (154)	1986 (177)	1990 (180)	1995 (192)	1999 (193)	2004 (219)	2008 (222)	2013 (222)	2018 (222)	2022 (222)
Sarawak	0	0	0	24	24	24	24	24	27	27	28	28	31	31	31	31
Johor	8	16	16	16	16	16	16	18	18	20	20	26	26	26	26	26
Sabah	0	0	0	16	16	16	16	20	20	20	20	25	25	25	25	25
Perak	10	20	20	20	21	21	21	23	23	23	23	24	24	24	24	24
Selangor	7	14	14	14	11	11	11	14	14	17	17	22	22	22	22	22
Kedah	6	12	12	12	13	13	13	14	14	15	15	15	15	15	15	15
Kelantan	5	10	10	10	12	12	12	13	13	14	14	14	14	14	14	14
Pahang	3	6	6	6	8	8	8	10	10	11	11	14	14	14	14	14
Pulau Pinang	4	8	8	8	9	9	9	11	11	11	11	13	13	13	13	13
W.P. Kuala Lumpur	0	0	0	0	5	5	5	7	7	10	10	11	11	11	11	11
Terengganu	3	6	6	6	7	7	7	8	8	8	8	8	8	8	8	8
Negeri Sembilan	3	6	6	6	6	6	6	7	7	7	7	8	8	8	8	8
Melaka	2	4	4	4	4	4	4	5	5	5	5	6	6	6	6	6
Perlis	1	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3
W.P. Putrajaya	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
W.P. Labuan	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1

For the 3 elections after 8th March 2008 (Figure 2), official results are available digitally via the aforementioned dashboard, in a manner that enabled us to directly scrape the results, albeit with the trial and error generally required for frontend web scraping. Where unavailable (all 13 elections prior to and including the 2008 federal general election), we manually (i.e. by hand) digitised the data from the aforementioned post-election reports and gazetted results. We made a deliberate methodological choice to avoid the use of optical character recognition (OCR) and PDF parsing tools after initial experimentation revealed an average error rate of approximately 10%, primarily due to frequent changes in formatting and layout, even within the same document. We considered this to be unacceptably high for a resource intended to serve as a single source of truth. Moreover, the downstream process of error detection and correction proved to be more time-consuming and error-prone than simply transcribing the data by hand, especially given the relatively manageable size of this data (25,552 ballots across 9,705 electoral contests).

As a result, all records acquired from physical books or PDFs were transcribed by hand. The Technical Validation section further explains why the way in which Malaysia reports election results made it possible for us to do this with near-total confidence in the accuracy of the final product.

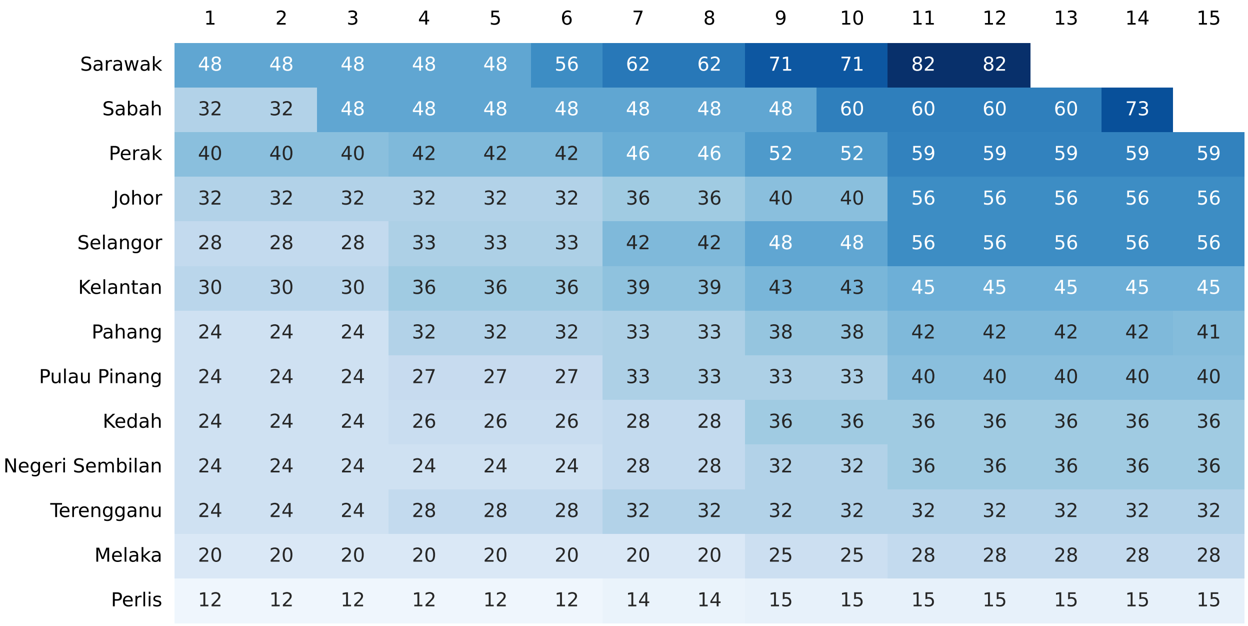
Records for seats in Peninsular Malaysia begin in 1955, while records for seats in Sabah and Sarawak begin in 1969. Although there was a federal general election in 1964, one year after Sabah and Sarawak (and Singapore) joined then-Malaya to form the Federation of Malaysia in 1963, seats in Sabah and Sarawak were not contested since the transition agreement allowed their respective state legislatures to appoint (and not elect) their representatives to the federal Parliament.[46] There are no records for Singapore, which was not contested in the 1964 general election for the same reason as Sabah and Sarawak, and which exited the Federation prior to the next federal general election in 1969.

State General Elections

For state-level general elections, we compiled data using the same three primary sources: post-election reports published by the EC, gazetted election results, and digital results where available. The same combination of web scraping and manual digitisation used for federal general elections was applied here.

For states in Peninsular Malaysia, records begin in 1959, when general elections for all 13 state legislative assemblies were held concurrently with the federal general election. For Sabah and Sarawak, records begin in 1967 and 1969 respectively, the years of the first state general elections held after the formation of the Federation of Malaysia. In all, MECo contains records for 15 elections for all states in Peninsular Malaysia, 14 elections for Sabah, and 12 for Sarawak (Figure 3). The reason for the discrepancy between the number of observations for Sabah and Sarawak is that there were two instances in Sabah's electoral history where state general elections were held in relatively quick succession. The first was in 1986, when Sabah went to the polls just one year after the previous state general election due to increasing civil and political instability.[47] The second was in 2020, when Sabah held a state general election two years after the watershed election of 2018 due to a collapse of the state government.[48]

Figure 3: State election coverage (number of seats)

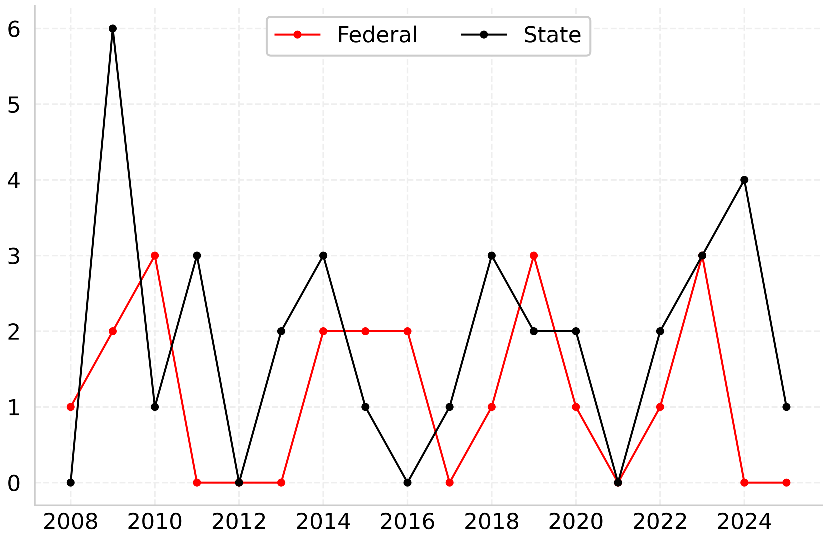


	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Sarawak	48	48	48	48	48	56	62	62	71	71	82	82			
Sabah	32	32	48	48	48	48	48	48	48	60	60	60	60	73	
Perak	40	40	40	42	42	42	46	46	52	52	59	59	59	59	59
Johor	32	32	32	32	32	32	36	36	40	40	56	56	56	56	56
Selangor	28	28	28	33	33	33	42	42	48	48	56	56	56	56	56
Kelantan	30	30	30	36	36	36	39	39	43	43	45	45	45	45	45
Pahang	24	24	24	32	32	32	33	33	38	38	42	42	42	42	41
Pulau Pinang	24	24	24	27	27	27	33	33	33	33	40	40	40	40	40
Kedah	24	24	24	26	26	26	28	28	36	36	36	36	36	36	36
Negeri Sembilan	24	24	24	24	24	24	28	28	32	32	36	36	36	36	36
Terengganu	24	24	24	28	28	28	32	32	32	32	32	32	32	32	32
Melaka	20	20	20	20	20	20	20	20	25	25	28	28	28	28	28
Perlis	12	12	12	12	12	12	14	14	15	15	15	15	15	15	15

By-Elections

The by-elections dataset covers all Parliament and DUN by-elections held since the 12th federal general election in 2008, totaling 115 contests as of end-April 2025 (Figure 4). Unlike for general elections, we could not locate any systematically compiled post-election reports or gazetted results for older by-elections. As a result, we relied solely on digital results made available via the EC's official website, which only covers by-elections since 2008. These digital records were scraped and harmonised with the same schema used for general elections.

Figure 4: Byelection coverage (number of elections)



We plan to expand the dataset to include an exhaustive record of previous by-elections once we are able to acquire a reliable source, likely via a combination of EC reports and Hansards from Parliament and State Legislative Assemblies. However, we made the decision not to delay the dissemination of the primary MECo datasets any further. Given that the most recent three election cycles are already fully covered, the exclusion of older by-elections is likely negligible for almost all prospective users.

Unique Identifier (UID) Generation

The database incorporates unique identifiers (UIDs) for parties and candidates to enable consistent tracking across elections and to facilitate longitudinal analysis. Additionally, constituency harmonisation is handled through a pragmatic approach that avoids imposing subjective judgments about boundary changes.

Parties. Party UIDs were generated manually, given that the total number of distinct parties in the dataset is relatively small (approximately 100). No historical harmonisation was conducted for parties that evolved or rebranded over time. For example, *Parti Perikatan* and *Barisan Nasional* are assigned separate UIDs, even though the latter is widely regarded as the successor to the former. This approach maintains full historical fidelity and allows users to apply their own ex-post harmonisation logic.

Candidates. Candidate UIDs were generated through a combination of fuzzy matching and manual correction, enabling the tracking of individual candidates across multiple elections even when their names changed in spelling, formatting, or title usage over time. This is particularly important in the Malaysian context, where candidates frequently acquire new honorifics, adopt different formatting conventions, or have minor spelling variations across election cycles. Two illustrative examples highlight the complexity of this task. 8-time Parliamentarian **Rafidah Aziz** appeared on ballots across 8 elections in forms including:

Rafidah Aziz
Rafidah Bt. Ab. Aziz
Rafidah Bt. Abdul Aziz
Datuk Seri Rafidah Aziz

Another 8-time Parliamentarian **Samy Vellu A/L Sangalimuthu** appeared on ballots across 9 elections with variations such as:

S. Samy Vellu
Datuk Seri Samy Vellu
S. Samy Vellu A/L Sangalimuthu
Dato' S. Samy Vellu A/L Sangalimuthu

Through careful matching and correction, we assigned a single UID to all instances of a candidate to support longitudinal analysis.

Constituencies. We deliberately avoided assigning UIDs to constituencies. Instead, we rely on the intrinsic uniqueness of the `date-state-constituency` combination, which guarantees that each contest is uniquely identifiable within its context (since no two constituencies within the same state can have the same name in the same election). A left join on these three fields suffices to generate a lookup table via which additional data can be merged. To aid users in harmonising constituencies across elections, we provide a field called `area_name`, which standardises name changes and spelling variations across time. Examples include:

Temerluh → Temerloh
Telok Anson → Teluk Intan
Bagan Datoh → Bagan Datok → Bagan Datuk

We do not currently implement harmonisation for changes involving boundary adjustments or the creation of new constituencies. For example, although the seat of Bayan Baru in Penang was carved out of Balik Pulau, we do not merge or tag these together because Balik Pulau remains an active seat name today, and such harmonisation would require subjective judgment. Our aim is to provide a flexible foundation that allows users to apply their own harmonisation strategies according to their specific research needs.

Data Records

The data records supporting this paper are available through multiple platforms to facilitate access, transparency, and long-term preservation:

- **Harvard Dataverse**
  All primary datasets (Table 1) are published and versioned on Harvard Dataverse at https://doi.org/10.7910/DVN/O4CRXK. This serves as the canonical archive for replication and reuse.

- **GitHub**
  The full source code used to process, compile, and validate the data is publicly available under a CC0 license at https://github.com/Thevesh/paper-malaysian-election-corpus. This repository also contains all files made available on Harvard Dataverse in both CSV and the highly efficient Parquet format, as well as raw source files split by state and election type.

- **Companion Website**
  The data can be interactively explored via ElectionData.MY. In addition to making the data accessible to non-technical users, the site is intended to serve as the primary channel for continuous updates and future enhancements to MECo.

Table 1: Description of primary datasets

Filename	**Description**
consol_ballots	Candidate-level results for all federal and state elections
federal_ballots	Subset of consol_ballots for federal general elections
state_ballots	Subset of consol_ballots for state general elections
byeelection_ballots	Subset of consol_ballots for by-elections
consol_summary	Summary statistics for all federal and state elections
federal_summary	Subset of consol_summary for federal general elections
state_summary	Subset of consol_summary for state general elections
byeelection_summary	Subset of consol_summary for by-elections
lookup_candidates	Standardised candidate names
lookup_parties	Standardised party names and abbreviations
lookup_dates	Election dates, by state and election type
lookup_seats	Standardised constituency names
logs/corrections	Log of manual corrections applied to ballot issued counts to ensure consistency as described in the Technical Validation section

Table 2 describes the structure of *_ballots files, while Table 3 describes the structure of *_summary files.

Table 2: Structure of all `*_ballots` files

Variable	**Description**
`date`	Date of the election (YYYY-MM-DD)
`election`	Election name (e.g., GE-14, SE-10, BY-ELECTION)
`state`	State in which the constituency is located
`seat`	Code and full name of the seat (e.g., P.052 Bayan Baru)
`ballot_order`	Order in which the candidate appeared on the ballot
`name`	Candidate name as it appeared on the ballot paper
`candidate_uid`	Unique identifier for the candidate
`party`	Party name as it appeared on the ballot paper
`party_uid`	Unique identifier for the party
`votes`	Number of valid votes received by the candidate
`votes_perc`	Share of valid votes received by the candidate (%)
`result`	Outcome (won, won uncontested, lost, lost deposit)

Technical Validation

There are four components of the database which require validation: Numerical data, candidate names, party names, and constituency names.

For numerical data, the way in which Malaysia reports election results makes a unique form of validation possible. First, we define the following variables:

$$\begin{aligned} I &= \text{Ballots Issued} \\ U &= \text{Unreturned Ballots} \\ R &= \text{Votes Rejected} \\ V_i &= \text{Valid Votes for Candidate } i \end{aligned}$$

Table 3: Structure of all `*_summary` files

Variable	Description
`date`	Date of the election (YYYY-MM-DD)
`election`	Type of election (e.g., GE-14, SE-10, BY-ELECTION)
`state`	State in which the constituency is located
`seat`	Full name of the seat (e.g., P.049 Tanjong)
`voters_total`	Total number of registered voters
`ballots_issued`	Number of ballots issued
`ballots_not_returned`	Number of ballots not returned (often postal votes)
`votes_rejected`	Number of rejected (spoiled) votes
`votes_valid`	Number of valid votes
`majority`	Margin of victory between the top two candidates
`voter_turnout`	Ballots issued as a share of registered voters (%)
`majority_perc`	Majority as a share of valid votes (%)
`votes_rejected_perc`	Rejected votes as a share of ballots returned (%)
`ballots_not_returned_perc`	Unreturned ballots as a share of ballots issued (%)

For a given contest involving N candidates, the following relationship must hold:

$$I - U - R = \sum_{i=1}^{N} V_i \tag{1}$$

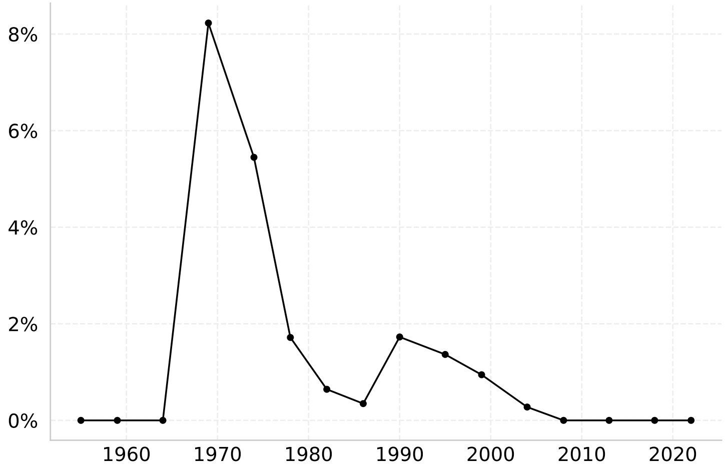
In plain language, the implied number of valid votes must be equal to the sum of votes received by all candidates. Because I, U, R, and V_i are (and have been) reported separately in all historical election reports and gazetted results, we can leverage this relationship as an almost-foolproof way to validate the accuracy of our manual data entry process. For transparency, we detected and corrected errors for 82 out of 9,705 contests, implying an error rate of 0.84%, which is significantly lower than the 10% error rate we encountered in our initial OCR experiments. Importantly, these errors are detectable and fixable, and come without the additional overhead of correcting errors in text data; our familiarity with local geography, politicians, and culture enabled us to transcribe text data rapidly and accurately.

This validation procedure also revealed 115 contests (excluding the 2006 Sarawak state general election, which did not report I, U, or R) for which the data was digitised accurately based on our source material, but which nevertheless failed to satisfy Equation (1). These errors presumably occurred due to mistakes in data entry which were not caught and corrected during the original publication process. In order to ensure a clean dataset, we applied a standard correction – for all 115 contests failing validation, we adjust the number of ballots issued (as documented in `corrections.csv`) such that Equation (1) holds. We choose `ballots_issued` as the variable to adjust for two reasons. First, until the 1989 amendment of the Elections (Conduct of Elections) Regulations 1981 (implemented at the 1989 federal general election), the number of unreturned ballots was not reported. Second, because the number of rejected votes is very small relative to the number of ballots issued, using rejected votes as the adjusted variable would have resulted in a much larger correction in percentage terms – specifically, the correction would have required a 20% change to the number of rejected votes on average (with several exceeding 50%), relative to a 0.6% change to the number of ballots issued (with only one above 5%, and none above 20%).

An interesting trend emerges when we plot the error rate arising from these 115 contests (Figure 5). The error rate was 0% up to 1964, but spiked to nearly 8% in 1969, when elections were severely disrupted by the Emergency. By 1978, error rates stabilised below 2%, and have been constant at 0% since 2008. We posit that this reflects general improvements in data management technology over the decades; we do not have a clear explanation for why no errors were made up to 1964.

Wrapping up our checks on numerical values, we note that the validation procedure can fail if two errors exactly offset each other. Furthermore, it cannot detect errors in the number of registered voters, which does not enter Equation (1). To address these limitations, we plotted histograms of 4 derived variables (Figure 6): voter turnout rate, vote rejection rate, unreturned ballot rate, and majority as a share of

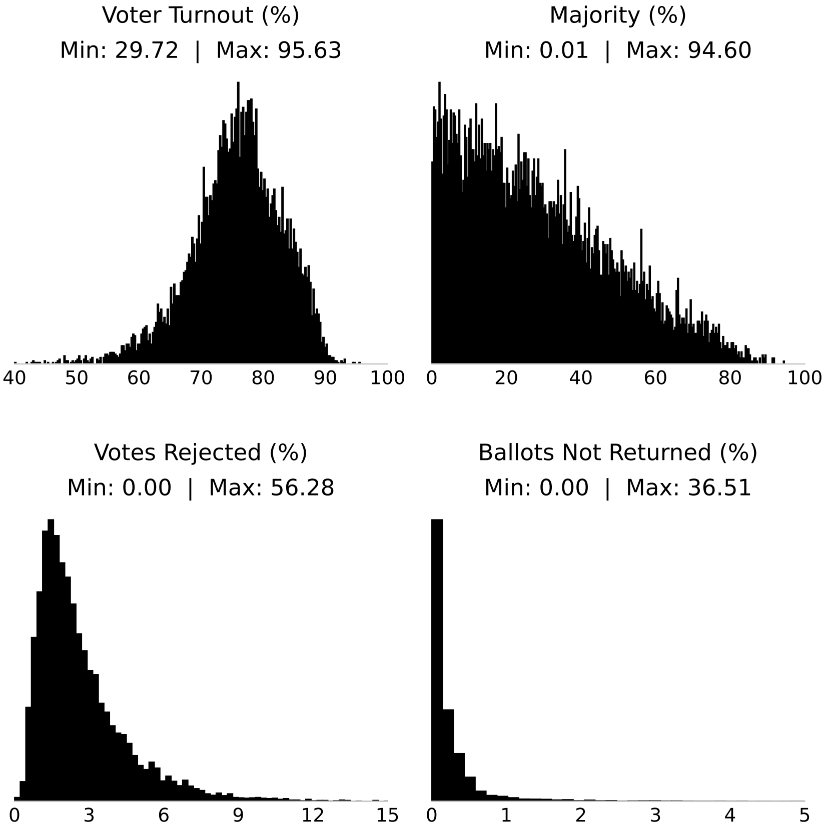
Figure 5: Error rate in general election years



valid votes. All four variables pass the check of being bounded between 0 and 100%, and display smooth distributions as would be expected if data entry was accurate. All outlier values were double-checked; in particular, we verified that extreme instances of high rejected votes or unreturned ballots were not due to mistakes in data entry. For example, the DUN of Pangkor in Perak had two instances where over 30% of ballots were not returned; this was due to historical inefficiencies in the implementation of postal voting for navy personnel.[49]

For non-numerical data, we conducted 3 types of checks. For parties, the full list of parties was only 100 long, so we manually checked each against publicly available information. For constituency names, we generated a list of unique constituency names (1,437 in total), which we manually checked for errors in spelling or syntax. The most challenging component was candidate names. The scale of the dataset – 25,552 candidates across 9,705 contests – made full manual validation infeasible. While we applied rigorous formatting standards during data entry and conducted targeted manual checks for prolific or high-profile candidates, we acknowledge that minor inconsistencies may persist. These challenges are amplified by the fact that candidates often change the spelling of their names across elections, acquire new titles and honorifics over time (e.g., Dato', Tan Sri, Haji, Ustaz, academic degrees), and format their names differently in different years (for example, omitting or reordering

Figure 6: Histogram of derived variables



titles). Such variations make consistent longitudinal tracking inherently complex. As with all living datasets of this scale, we anticipate incremental improvements over time as users engage with the data and identify potential refinements.

Finally, to ensure accuracy of the database as a whole, we derive seats and votes by party for all elections in the dataset, and ensure that these match against the record of parties in Parliament and State Legislative Assemblies.

Usage Notes

The Malaysian Election Corpus (MECo) is designed to support a wide range of use cases, ranging from rigorous empirical research to rapid data journalism, and even casual civic technology projects. In the academic realm, MECo – as the first database of its kind for Malaysia – serves as a foundational resource for research in electoral studies, political science, and public policy. Researchers can employ the data to answer important questions about the evolution of Malaysia's electoral system, employing both cross-sectional and longitudinal analysis.

We anticipate that many users will want to extend or adapt the dataset. The standardised schema and use of lookup tables enable seamless enrichment and flexible integration with other datasets. Some valuable examples include:

- `lookup_candidates` can be enriched with data on candidates' demographic characteristics, such as age, gender, ethnic group, and education level.
- `lookup_parties` can be enriched with data on party affiliations, or even component parties of coalitions. For instance, by adding a `coalition` field to the lookup, it is possible to assign different UIDs to two candidates representing the same coalition, such that they can be traced to different component parties.
- `lookup_seats` can be enriched with a `lineage` field, which might track the geospatial lineage of a constituency, including changes in boundaries over time. However, this will require the user to define how constituency identity is determined.

As a practical reference for users, the `dashboards.py` script illustrates how to merge the core datasets with lookup tables to create panel data suitable for interactive dashboards and advanced analyses.

Finally, while rigorous validation has been applied (see Technical Validation), this remains a living database and is intended as such. Minor inaccuracies, particularly in the candidate name field, may persist as discussed above. We encourage users to report issues or submit improvements via the GitHub repository, which provides full transparency of changes between releases.

Code Availability

All data processing, validation, and compilation was conducted in Python. The full source code is publicly available via GitHub under a CC0 license. Three key scripts which users should find particularly useful are:

- `compile.py`, which generates and validates all tabular datasets published on Harvard Dataverse, with the exception of the `lookup*.csv` files, which were manually curated.
- `dataviz.py`, which generates all visualisations used in this paper.
- `dashboards.py`, which generates the source files for the interactive visualisations available at ElectionData.MY.

References

[1] Election Commission. MySPR Semak, 2025. Accessed: 2025-05-05.

[2] Attorney General's Chambers of Malaysia. Federal Legislation: P.U.(B), 2025. Accessed: 2025-05-05.

[3] Ken Kollman, Allen Hicken, Daniele Caramani, David Backer, and David Lublin. Constituency-level elections archive, 2024. Accessed: 2025-05-05.

[4] Ferran Martínez i Coma and Diego Leiva Van De Maele. The global dataset on turnout (GD-Turnout). *Electoral Studies*, 86:102681, 2023.

[5] Pippa Norris, Richard W Frank, and Ferran Martínez i Coma. Measuring electoral integrity around the world: A new dataset. *PS: Political Science & Politics*, 47(4):789–798, 2014.

[6] MalaysiaKini. undi.info, 2025. Accessed: 2025-05-05.

[7] Sinar Project. GE15 Open Data, 2025. Accessed: 2025-05-05.

[8] Tindak Malaysia. Historical election results. `https://citethis.link/meco1-tindak`, 2025. Accessed: 2025-05-05.

[9] Virgilio Pérez, Cristina Aybar, and Jose M Pavía. Spanish electoral archive. SEA database. *Scientific Data*, 8(1):193, 2021.

[10] Samuel Baltz, Alexander Agadjanian, Declan Chin, John Curiel, Kevin DeLuca, James Dunham, Jennifer Miranda, Connor Halloran Phillips, Annabel Uhlman, Cameron Wimpy, et al. American election results at the precinct level. *Scientific Data*, 9(1):651, 2022.

[11] Justin de Benedictis-Kessner, Diana Da In Lee, Yamil R Velez, and Christopher Warshaw. American local government elections database. *Scientific Data*, 10(1):912, 2023.

[12] Bruno Calderón-Hernández, Horacio Larreguy, John Marshall, and José Luis Pérez-Castellanos. Electoral precinct-level database for Mexican municipal elections. *Scientific Data*, 12(1):582, 2025.

[13] Vincent Heddesheimer, Hanno Hilbig, Florian Sichart, and Andreas Wiedemann. GERDA: The German election database. *Scientific Data*, 12(1):618, 2025.

[14] Francesca R Jensenius and Gilles Verniers. Studying Indian politics with large-scale data: Indian election data 1961–today. *Studies in Indian Politics*, 5(2):269–275, 2017.

[15] T. E. Smith. *Report on the First Election of Members to the Legislative Council of the Federation of Malaya.* Supervisor of Elections, 1955.

[16] Election Commission. *Report on the Parliamentary and State Elections 1959.* Election Commission of the Federation of Malaya, 1959.

[17] Election Commission. *Report on the Parliamentary (Dewan Ra'ayat) and State Legislative Assembly General Elections, 1964 of the States of Malaya.* Election Commission of the Federation of Malaya, 1964.

[18] Election Commission. *Report on the Parliamentary (Dewan Ra'ayat) and State Legislative Assembly General Elections 1969 of the States of Malaya, Sabah and Sarawak.* Election Commission of Malaysia, 1971.

[19] Election Commission. *Report on the Parliamentary (Dewan Rakyat) and State Legislative Assembly General Elections 1974 of the States of Malaya and Sarawak.* Election Commission of Malaysia, 1974.

[20] Election Commission. *Report on the General Elections to the House of Representatives and the State Legislative Assemblies Other Than the State Legislative Assemblies of Kelantan, Sabah and Sarawak 1978.* Election Commission of Malaysia, 1978.

[21] Election Commission. *Report on the Malaysian General Elections 1982.* Election Commission of Malaysia, 1982.

[22] Election Commission. *Report on the Malaysian General Elections 1986.* Election Commission of Malaysia, 1986.

[23] Election Commission. *Report on the Malaysian General Elections 1990.* Election Commission of Malaysia, 1990.

[24] Election Commission. *Report of the General Election Malaysia 1995.* Election Commission of Malaysia, 1995.

[25] Election Commission. *Report of the General Election Malaysia 1999.* Election Commission of Malaysia, 1999.

[26] Election Commission. *Report of the General Election Malaysia 2004.* Election Commission of Malaysia, 2004.

[27] Election Commission. *Report of the General Election Malaysia 2008*. Election Commission of Malaysia, 2008.

[28] Election Commission. *Report of the General Election Malaysia 2013*. Election Commission of Malaysia, 2013.

[29] Election Commission. *Report on the State Legislative Assembly General Election 1967 of the State of Sabah*. Election Commission of Malaysia, 1967.

[30] Election Commission. *Report on the State Legislative Assembly General Election 1971 of the State of Sabah*. Election Commission of Malaysia, 1971.

[31] Election Commission. *Report on the State Legislative Assembly General Election 1976 of the State of Sabah*. Election Commission of Malaysia, 1976.

[32] Election Commission. *Report on the State Legislative Assembly General Election 1981 of the State of Sabah*. Election Commission of Malaysia, 1981.

[33] Election Commission. *Report on the State Legislative Assembly General Election 1985 of the State of Sabah*. Election Commission of Malaysia, 1985.

[34] Election Commission. *Report on the State Legislative Assembly General Election 1986 of the State of Sabah*. Election Commission of Malaysia, 1986.

[35] Election Commission. *Report on the State Legislative Assembly General Election 1990 of the State of Sabah*. Election Commission of Malaysia, 1990.

[36] Election Commission. *Report on the State Legislative Assembly General Election 1994 of the State of Sabah*. Election Commission of Malaysia, 1994.

[37] Election Commission. *Report on the State Legislative Assembly General Election 1999 of the State of Sabah*. Election Commission of Malaysia, 1999.

[38] Election Commission. *Report on the State Legislative Assembly General Election 1979 of the State of Sarawak*. Election Commission of Malaysia, 1979.

[39] Election Commission. *Report on the State Legislative Assembly General Election 1983 of the State of Sarawak*. Election Commission of Malaysia, 1983.

[40] Election Commission. *Report on the State Legislative Assembly General Election 1987 of the State of Sarawak*. Election Commission of Malaysia, 1987.

[41] Election Commission. *Report on the State Legislative Assembly General Election 1991 of the State of Sarawak*. Election Commission of Malaysia, 1991.

[42] Election Commission. *Report on the State Legislative Assembly General Election 1996 of the State of Sarawak*. Election Commission of Malaysia, 1996.

[43] Election Commission. *Report on the State Legislative Assembly General Election 2001 of the State of Sarawak*. Election Commission of Malaysia, 2001.

[44] Election Commission. *Report on the State Legislative Assembly General Election 2006 of the State of Sarawak*. Election Commission of Malaysia, 2006.

[45] Election Commission. *Report on the State Legislative Assembly General Election 1978 of the State of Kelantan*. Election Commission of Malaysia, 1978.

[46] Gordon P. Means. *Malaysia: The Second Generation*. Oxford University Press, Singapore, 1991.

[47] Mohammad Agus Yusoff. Sabah politics under Pairin. *JATI - Journal of Southeast Asian Studies*, 6:29–48, 2001.

[48] P.O. Lee. The collapse of the state government in Sabah: Back to the drawing board. *ISEAS - Yusof Ishak Institute Commentaries*, 2020. Available at: https://citethis.link/meco1-iseas.

[49] Iskandar Dzulkarnain Ahmad Junid. The battle for Perak. In *Malaysia's 14th General Election and UMNO's Fall*, pages 223–239. Routledge, 2019.

Author Contributions

TT conceived, designed, and carried out all aspects of this study, including data collection, data curation, analysis, and manuscript preparation.

Competing Interests

The author declares no competing interests. The author is not affiliated with any political party, or any organization which actively participates in or aims to influence the outcome of Malaysian elections.

Acknowledgements

This work did not receive any specific funding. We thank Rosmadi Fauzi and Zulkanain Abdul Rahman for their helpful feedback and suggestions on the dataset and manuscript. We also gratefully acknowledge Nimesha Thevananthan for their assistance with data collection, as well as Danesh Prakash Chacko of Tindak Malaysia for providing historical insights and filling data gaps, particularly for the Sabah state election in 1971. Finally, we thank the EC for granting access to physical records of historical election results via *Pusat Sumber SPR*.