Indian Election Dataset (1951 to 1962) 1.0
<link to dataset>
Last Updated: 8th May, 2023

Contents

1. Introduction

The Indian Election Dataset (TCPD_IED_1951-62) is a data repository of all Indian election results between 1951 and 1952. The data comes from the Statistical Reports published by the Election Commission of India (ECI). The data extracted was cleaned and structured to fit into a tabular format. Additional variables, mostly derived from ECI data, have been added.

The Indian Election Dataset has two disjoint versions, the current version ranging from 1951 to 1962, and the other one that can be found here ranging from 1962 till present day. These exist as separate entities due to irreconcilable differences in the underlying structure. For instance, before 1962, it was possible for one constituency to have multiple winners (2-3). As a result, it became impossible to generate a ranking for these elections by just considering the votes polled by a certain candidate.

This codebook provides information about the variables included in the Indian Election Dataset (1951-1962). The codebook will be updated as and when new information is added to the dataset.

We hope that you will find this dataset useful and request that you cite the data as indicated in the codebook, whenever you use the data in a publication or public presentation.

With thanks,

TCPD Team

2. Terms and conditions

The users are free to download, display or include the data in other products for non-commercial purposes at no cost subject to the following limitations:

Data Citation: "TCPD Indian Elections dataset (TCPD-IED), 1951-1962". Trivedi Centre for Political Data, Ashoka University.

Codebook Citation: Navya Asopa, Shivam Gangwani, Aishwarya Sunaad, Gilles Verniers. 2023. "TCPD Indian Elections dataset (TCPD-IED), 1951-1962 Codebook 1.0", Trivedi Centre for Political Data, Ashoka University.

No Endorsement: The user must not claim or imply that the Trivedi Centre for Political Data endorses the user's use of the data or use of the Centre's logo(s) or trademarks(s) in conjunction with the same.

No Warranty: The Centre makes no warranties with respect to the accuracy of the data. While we do our best to ensure that the numbers are accurate, there is room for error. And as such the user must agree that the Centre shall not be held responsible or liable to the user for any errors, omissions, misstatements and/or misrepresentations of the data though the user is encouraged to report the same to us ([send us a message](#) or write us at tcpd-contact@ashoka.edu.in).

Personal data: The Centre may record visits to the TCPD website without collecting the personal information of the users. The records shall be used for statistical reports only.

Acknowledgement: The user must agree that the use of the Data presented within the application can be seen as the acknowledgement of unconditionally accepting the Terms of Use presented by the Centre.

3. Table 1: Variable Names, Type, and Description

#	Variable Name	Variable Type	Variable Description

1	Election_Type	string; categorical	The Type of Election:
			   1. AE (Assembly Election [Vidhan Sabha])
			   2. GE (National General Election [Lok Sabha])

2	State_Name	string	Name of the State
3	Assembly_No	numeric	The Assembly Number as per the ECI reports. Assembly number #1 corresponds to the first election. For AE, this is a unique indexing for the elections of a state. For GE, we ignore the State_Name and only consider Assembly_No to look at a distinct election.
4	Constituency_No	numeric	This is the constituency number assigned by the ECI to each constituency. This number does not necessarily stay the same across each election
5	Candidate	string	Name of the Candidate as per ECI
6	Gender	string; categorical	Gender of the Candidate as per ECI.
7	Party	string	Party (abbreviation) of the Candidate as per ECI.
8	Votes	numeric	Number of Votes won by the Candidate.
9	Winner	logical	Binary variable to indicate whether the candidate won the election (True) or not (False).
10	Constituency_Name	string	Name of the Constituency
11	Idx	numeric	The index number assigned to the candidate under the Detailed Results section of the ECI statistical report.

12	Party_Type	string; categorical	Party category assigned by ECI in the statistical report.
			
			The Candidate's party may be categorised in the following buckets:
			
			   1. STATE PARTIES
			   2. NATIONAL PARTIES
			   3. INDEPENDENTS
			   4. REGISTERED (Unrecognised) PARTIES
			   5. OTHER STATE PARTIES

13	Party_Expanded	string	The full expanded name of the Candidate's party, specified by ECI in the statistical report.

14	NumberOfSeats	numeric	The number of seats available in the constituency, as per ECI.
15	ElectorsTotal	numeric	Total number of registered electors in the constituency as per ECI
16	ElectorsWhoVoted	numeric	Total number of electors who voted in the given constituency as per ECI.
17	VotesValid	numeric	Sum of all valid votes to all candidates in a particular election for a particular constituency.
18	PollingDate	date	The date as per the ECI report on which the election was held.
19	Year	date	Year in which the election was held. This variable is derived using PollingDate.
20	Runner up.PARTY	string	The party of the runner-up candidate in the given constituency under the concerned election.
21	Runner up.CANDIDATE	string	The name of the runner-up candidate in the given constituency under the concerned election.
22	Runner up.VOTES	numeric	The votes of the runner-up candidate in the given constituency under the concerned election.
23	Winner X.PARTY	string	The party of Winner X (X=1,2,3) candidate in the given constituency under the concerned election.
24	Winner X.CANDIDATE	string	The name of Winner X (X=1,2,3) candidate in the given constituency under the concerned election.
25	Winner X.VOTES	numeric	The votes of Winner X (X=1,2,3) candidate in the given constituency under the concerned election.

4. Additional Notes about the Data

1. This dataset does not contain any variable to indicate the 'Position' of a candidate in a given election. This is due to the presence of double members and triple members constituencies. Before 1962, some seats were given additional representatives, elected on distinct SC and ST lists. This means that these seats had more than one winner, elected on different lists. The ECI reports, unfortunately, do not indicate which candidates contested, and therefore won, on which list.

   As a result, we have the 'Winner' logical variable to denote the winners within a constituency, and no 'Position' variable can be computed.

2. The 'Idx' variable of a candidate appears to be a dummy variable assigned by the ECI. It has been retained for possible updates to the data that may be made in the future.

3. The 'Winner' logical variable has been calculated by comparing (Party, Candidate, Votes) to (Winner X Party, Winner X Candidate, Winner X Votes) - if one such match is found, then 'Winner' is set to True.

   Expanding on that, we found that the following variables can be used to form a 'primary key' of this data:
   a. Election_Type
   b. State_Name
   c. Party
   d. Candidate
   e. Votes

4. The 'PollingDate' variable has been extracted from the constituency-wise results within the ECI report. This often clashes with the date specified on the ECI website (For instance, Madras Assembly #1 year has been listed as 1951 on the website but the polling date specified is 27-03-1952 inside the report itself).

5. 'VotesValid' and 'ElectorsWhoVoted' variables are identical in all places, except for a certain election in Kerala. This discrepancy emerges within the source of this dataset (the ECI statistical report).

6. The 'Gender' of the candidates are available for all AE elections. For GE elections, these are missing from the ECI report and consequently from our dataset as well.