

Political Parties of India Codebook 1.0

<https://github.com/tcpd/ppi>

Last Updated: 17 August, 2021

Contents

Title	Page
1. Introduction	2
2. Terms and Conditions	3
3. Table 1: Variable names, type, and description	4-5
4. Data Sources	6
5. Documentation Process	7
6. Exceptions	8



1. Introduction

This dataset is a repository of all parties who contested national and state elections in India since 1962.

Over the years, many parties have changed their names or the ways their name is abbreviated. This can lead to discrepancies in the analysis of party performance over time. In this dataset, we have used our trademark tool Surf, to assign a unique identifier, or numerical party ID (for example: Congress, INC, Indian National Congress were all assigned the ID number 3482). With these IDs, it becomes possible to study the trajectory of parties and also collate aggregate figures on party performance over time. Furthermore, it also offers more credibility to the study of turncoats who can now be mapped through their party ID rather than simply through their names.

In time, we will add more variables to this dataset, such as the year of registration and the name of founding presidents (or equivalent). We hope you 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.

Team TCPD

2. Terms and conditions

The Political Parties of India dataset is a list of all political parties that have contested state and national elections in India from 1962 to 2021. 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 Political Parties of India Dataset (TCPD-PPI), 1962-2021". Trivedi Centre for Political Data, Ashoka University.

Codebook Citation: Ananay Agarwal, Neelesh Agrawal, Saloni Bhogale, Sudheendra Hangal, Francesca Refsum Jensenius, Mohit Kumar, Chinmay Narayan, Basim U Nissa, Priyamvada Trivedi, and Gilles Verniers. "TCPD Political Parties of India Dataset (TCPD-PPI), 1962-2021 Codebook". 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 the Statistical Reports published by the ECI comprise the primary data source, we also rely on multiple other sources in cases of discrepancies in ECI's 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 Lok Dhaba 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, types and labels

#	Variable Name	Variable Type	Variable Description
1	Assembly	string	Name of the Assembly (Lok Sabha or Vidhan Sabha)
2	State_Name	string	Name of the state. "All_States" is for all India consolidated values for Lok Sabha elections.
3	Party_Name	string	Expanded name of the party
4	Party_Type	string	Classification of the party : National, state-based, local, independent or NOTA
5	PartyID	numeric	Unique numeric identifier for each party
6	Frequent_Abbreviation	string	Most common abbreviation used by ECI
7	Last_Abbreviation	string	Most recent abbreviation used by ECI
8	Abbreviations	list	List of all unique abbreviations by ECI
9	Start_Year	numeric	First year of contest of party in Assembly and State
10	Last_Year	numeric	Last year of contest of party in Assembly and State
11	No_Assemblies_Contested	numeric	Total number of assemblies contested by party in Assembly and State
12	Assemblies_Contested	list	List of all unique assemblies contested by party in Assembly and State
13	Candidates_Contested	numeric	Total number of candidates contested by the party in all assemblies.
14	Candidates_Represented	numeric	Total number of candidates represented from the party in all assemblies.
15	Females_Contested	numeric	Total number of female candidates contested by the party in all assemblies.
16	Females_Represented	numeric	Total number of female candidates represented from the party in all assemblies.
17	SC_Seats_Contested	numeric	Total number of Scheduled Caste(SC) seats contested by the party in all assemblies.
18	SC_Seats_Represented	numeric	Total number of Scheduled Caste(SC) seats

			represented from the party in all assemblies.
19	ST_Seats_Contested	numeric	Total number of Scheduled Tribe(ST) seats contested by the party in all assemblies.
20	ST_Seats_Represented	numeric	Total number of Scheduled Tribe(ST) seats represented from the party in all assemblies.
21	BiPoll_Contested	numeric	Total number of bye elections contested by the party in all assemblies.

4. Data Sources

- Statistical reports to general elections from Election Commission of India (<https://eci.gov.in/statistical-report/statistical-reports/>)

5. Documentation Process

The raw data to generate this dataset was obtained through the statistical reports released by ECI. For each assembly election, the respective statistical report contains a list of participating political parties. This was collated across assemblies and elections, and included information on the party name, abbreviation, and category. This collated dataset was then processed through Surf, and unique IDs were assigned to parties across assemblies and elections. For example, CPM, CPI(M), CPI (M), were all assigned the same ID.

While Surf accounted for most of the parties, there were some parties with major name changes that could not be accounted for using Surf. For example, AITC and TMC are both abbreviations of All India Trinamool Congress. Such parties were assigned the same ID through manual inspection.

These IDs were then assigned to the [TCPD-IED](#), and variables #6-#21 were accordingly aggregated from that dataset. For more granular information at the candidate level based on the party IDs, one can directly use the [TCPD-IED](#).

6. Exceptions

For some elections, the party category is not defined in the statistical reports. Those values are blank in this dataset.

For some byelections, the expanded party names have not been provided in the statistical reports. The following values (as provided in the statistical reports) are present for the Party_Name variable in such cases:

- NA's
- XXX
- Expanded Party name not released by the ECI