Trivedi Centre
for Political Data

Lok Dhaba Codebook 1.0

http://lokdhaba.ashoka.edu.in

Last updated: November 12, 2019

Contents

Introduction

Welcome to Lok Dhaba!

Lok Dhaba is a repository of Indian election results beginning 1962. The data comes from the Statistical Reports published by the Election Commission of India (ECI from here on). Once the data is extracted from the statistical reports, we clean the data for it to fit a tabular format. This codebook provides information about the variables in Lok Dhaba such as the labels, types, brief description, and a few summary statistics. The codebook will be updated as and when new data is added to Lok Dhaba.

Thank you.

TCPD Team

Trivedi Centre
for Political Data

Terms and Conditions

Lok Dhaba is an online web interface provided by the Trivedi Centre for Political Data. In these terms of use of the data provided by the Centre, 'Data' includes all visualizations, texts, graphics and compilations of data and other material presented within the application. 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: Saloni Bhogale, Sudheendra Hangal, Francesca Refsum Jensenius, Mohit Kumar, Chinmay Narayan, Basim U Nissa, and Gilles Verniers. 2019. "TCPD Indian Elections Data v1", Trivedi Centre for Political Data, Ashoka University.

Codebook Citation: Saloni Bhogale, Sudheendra Hangal, Francesca Refsum Jensenius, Mohit Kumar, Chinmay Narayan, Basim U Nissa, and Gilles Verniers. 2019. "TCPD Indian Election Data Codebook v1", 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.

Table 1: Variable names, types and labels

#	Variable Name	Variable Type	Variable Label
1	State_Name	string	Name of the state
2	Assembly_No	numeric	Assembly/election number
3	Constituency_No	numeric	Assembly constituency number
4	Year	numeric	Year in which election was held
5	month	numeric	Month in which election results were announced
6	Poll_No	numeric	Poll number for the constituency
7	DelimID	numeric	Delimitation number
8	Position	numeric	Rank of the contestant
9	Candidate	string	Name of the candidate
10	Sex	string	Gender of the candidate
11	Party	string	Party of the candidate
12	Votes	numeric	Number of voters for the candidate
13	Candidate_Type	string	Category of the candidate: General (GEN) or reserved for Scheduled Castes (SC) or Scheduled Tribes (ST)
14	Valid_Votes	numeric	Number of valid votes in the constituency
15	Electors	numeric	Total number of electors in the constituency
16	Constituency_Name	string	Name of the constituency
17	Constituency_Type	string	Type of the constituency: General (GEN) or reserved for Scheduled Castes (SC) or Scheduled Tribes (ST)
18	Sub_Region	string	Subregion of the constituency
19	N_Cand	numeric	Number of candidates contesting the constituency
20	Turnout_Percentage	numeric	% turnout in the constituency
21	Vote_Share_Percentage	numeric	% vote share received by the candidate in a given constituency
22	Deposit_Lost	logical	Did the candidate lose their deposit?
23	Margin	numeric	Difference in votes between a candidate and

			the next ordered candidate (ordered by position)
24	Margin_Percentage	numeric	Percentage margin of a candidate in compared to the next position candidate
25	ENOP	numeric	Effective number of parties (derived from vote shares of each candidate in a constituency)
26	pid	numeric	Unique identifier for a candidate
27	Party_ID	numeric	Unique identifier for a political party
28	Max_Poll_No	numeric	The last poll number for the constituency
29	last_poll	logical	Whether this poll was the last?
30	Contested	numeric	The number of times a candidate has contested so far
31	Last_Party	numeric	The name of the last party the candidate contested from
32	Last_Party_ID	numeric	The unique party ID of the last party the candidate contested from
33	Last_Constituency_Name	string	Name of constituency this candidate contested from the last time
34	Same_Constituency	logical	Is the candidate contesting from the same constituency?
35	Same_Party	logical	Is the candidate contesting from the same party?
36	Mandate	numeric	If this person wins then this is 1, else 0
37	No_Mandates	numeric	Number of contests won by the candidate
38	Turncoat	logical	Is this candidate a turncoat?
39	Incumbent	logical	Is this candidate an incumbent?
40	Recontest	logical	If this person contested in the last assembly or not?

Table 2: Detailed variable description

#	Variable Name	Variable description
1	State_Name	Name of the state as per the ECI. A tabulation of the values for this variable will show the following values: "Andaman_&_Nicobar_Islands","Andhra_Pradesh","Arunachal_Pradesh","Assam","Bihar","Chandigarh","Chhattisgarh","Dadra_Nagar_&_Haveli","Daman_&_Diu","Delhi","Goa","Gujarat","Haryana", "Himachal_Pradesh","Jammu_&_Kashmir","Jharkhand","Karnataka", "Kerala","Lakshadweep","Madhya_Pradesh","Madras","Maharashtra","Manipur","Meghalaya","Mizoram","Mysore","Nagaland","Odisha", "Puducherry","Punjab","Rajasthan","Sikkim","Tamil_Nadu", "Telangana","Tripura","Uttar_Pradesh","Uttarakhand","West_Bengal". We have data on erstwhile states such as 'Mysore' and 'Madras' which are now a part of Karnataka and Tamil Nadu respectively. In a similar way, states such as Chhattisgarh, Jharkhand and Uttarakhand were carved out of Madhya Pradesh, Bihar and Uttar Pradesh in 2000.
2	Assembly_No	Our data begins in 1962 when the third Lok Sabha elections were held, and the third assembly was constituted. The "Assembly_No" for the 1962 election is thus 3 as these were the *third* Lok Sabha elections. The following shows the mapping of each assembly number to the corresponding year in which a Lok Sabha election was held: 4 → 1967, 5 → 1971, 6 → 1977, 7 → 1980, 8 → 1984, 9 → 1989, 10 → 1991, 11 → 1996, 12 → 1998, 13 → 1999, 14 → 2004, 15 → 2009, 16 → 2014. This works in a similar way in the case of the state/Vidhan Sabha elections too. For instance, Uttarakhand which was created in 2000, has had 4 assembles (1 → 2002, 2 → 2007, 3 → 2014, 4 → 2017).
3	Constituency_No	This is the constituency number assigned by the ECI to each constituency. This number does not necessarily stay the same across each election year or across delimitations (see variable #7, "DelimID").
4	Year	The year in which the election took place.
5	month	The month in which the election results are announced where 1 → January, 2 → February, 3 → March, 4 → April, 5 → May, 6 → June, 7 → July, 8 → August, 9 → September, 10 → October, 11 → November, 12 → December
6	Poll_No	This is the poll number for the constituency which can have only positive values with 0 being the minimum. This variable indicates the number of times a constituency is up for election within a given assembly. For instance, for an election held in 2009, the "Poll_No" will be 0 for all the

		constituencies. If however for any given constituency, elections are held again (in case the incumbent resigns or dies), then this number changes. In essence, it tracks each election in a constituency per assembly.
7	DelimID	This refers to the delimitation done by the ECI. The purpose of the exercise is to redraw boundaries of constituencies in order to account for changing demographics. There have been 4 delimitations in India post-Independence - 1952, 1962, 1976 and 2008. In the data, this variable has the following values [1, 2, 3, 4] which correspond to the following years [(1962-1963), (1964-1972), (1973-2007), (2008-current)]
8	Position	At any time there can be several candidates contesting in any constituency. The "Position" variable shows the rank of the particular candidate for the constituency per assembly. It is possible for the votes of two people to be the same. However this has not been true for a winner.
9	Candidate	Name of the candidate as per the ECI results. What has been observed is that for a particular candidate, if the person contests several elections, there is no guarantee that the spelling will be same. We have proposed a solution to this problem. Check out our project "Incumbency Mapping" here. Note: NOTA/None of The Above is also an option that this variable can take.
10	Sex	Gender of the candidate as per the ECI. This variable can take on three values - "F" for Female, "M" for Male and "O" for Other. The "Other" category was introduced in 2013. See here for details.
11	Party	Political party of the candidate as per the ECI.
12	Votes	Number of votes that the candidate received as per the ECI.
13	Candidate_Type	The candidate can be any of the three types - General (GEN) or reserved for Scheduled Castes (SC) or Scheduled Tribes (ST)
14	Valid_Votes	Number of valid votes that the candidate received in the constituency as per the ECI.
15	Electors	This is the total number of people who are registered to vote in a given constituency as per the ECI.
16	Constituency_Name	Name of the constituency as per the ECI.
17	Constituency_Type	The constituency can either be non-reserved (General (GEN)) or reserved (Scheduled Castes (SC) or Scheduled Tribes (ST)).
18	Sub_Region	The subregion within which the constituency lies. This subregion has been assigned to constituencies by TCPD in consultation with state experts. For example, Andhra Pradesh

		is divided into two subregions - Coastal Andhra and Rayalaseema or Jammu & Kashmir is divided into three regions - Jammu, Kashmir and Ladakh.
19	N_Cand	The number of candidates contesting a constituency. This number can range from the single digits to a few hundred. It can also be 1 for an uncontested constituency.
20	Turnout_Percentage	This is the percentage of eligible voters who turn out to vote in a constituency. This is calculated by dividing the number of valid votes ("Valid_Votes") by the number of electors.
21	Vote_Share_Percentage	The vote share percentage is calculated for every contestant by the dividing the number of votes received by the candidate by the total number of valid votes that have been cast in a given constituency.
22	Deposit_Lost	When contesting elections, each candidate needs to deposit a certain amount of money. If the candidate receives less than @ vote share, the deposit is refunded else they lose the deposit.
23	Margin	Difference in votes between a candidate and the next ordered candidate (ordered by position/rank). The margin for the candidate who comes last will be 0.
24	Margin_Percentage	Percentage margin of a candidate and the next ordered candidate (ordered by position/rank). The margin percentage for the candidate who comes last will be 0.
25	ENOP	Effective number of parties (derived from vote shares of each candidate in a constituency). This is calculated by using a formula (1/sum(pi^)), where pi is the proportion of the total votes for of each candidate in a constituency (Laakso and Taagepera 1979).
26	pid	This is a unique politician identifier for a candidate assigned by TCPD. If 2 rows in the dataset have the same identifier, this means the same person contested both elections. Please note: there is no intrinsic meaning to this identifier. It should be treated as an opaque string. This identifier could very well be different between different versions of the TCPD dataset. It is only consistent within the a single version of the dataset.
27	Party_ID	This is a unique identifier for a political party assigned by TCPD.
28	Max_Poll_No	The last poll number for the constituency. By default, this number is 0. However it will be non-zero in case of a bye-poll. 0 → no bye-poll, 1 → one bye-poll, 2 → 2 bye-polls and so on.

29	last_poll	This is a dichotomous variable which indicates whether the poll number is the last election that was held for this constituency. For the calculation of various incumbency related metrics we need to know the last election (regardless of whether it was a bye-poll or not).
30	Contested	The number of times the candidate has contested so far (in any constituency).
31	Last_Party	The name of the last party on whose ticket the candidate contested.
32	Last_Party_ID	The unique party ID of the last party on whose ticket the candidate contested.
33	Last_Constituency_Name	The name of the last constituency the candidate contested from.
34	Same_Constituency	This is a dichotomous variable which indicates whether the candidate is contesting from the same constituency?
35	Same_Party	This is a dichotomous variable which indicates whether the candidate is contesting from the same party?
36	Mandate	This is a dichotomous variable which is 1 if the candidate is winner and 0 otherwise.
37	No_Mandates	Number of contests ever won by the candidate.
38	Turncoat	This is a dichotomous variable which indicates whether the candidate is a turncoat.
39	Incumbent	This is a dichotomous variable which indicates whether the candidate is an incumbent.
40	Recontest	This is a dichotomous variable which indicates whether this person contested in the last assembly or not.

Appendix: Distribution of each variable and patterns of missingness[1]

The following statistics have only been generated for numeric variables.

Variable name	N	# of blanks	# of NAs	Minimum value	Maximum value
State_Name	90879	NA	0	NA	NA
Assembly_No	90879	0	0	3	17
Constituency_No	90879	0	0	1	86
Year	90879	0	0	1962	2019
month	90879	0	2576	1	12
Poll_No	90879	88303	0	0	2
DelimID	90879	0	0	1	4
Position	90879	0	0	-185	480
Candidate	90879	NA	0	NA	NA
Sex	90879	NA	0	NA	NA
Party	90879	NA	0	NA	NA
Votes	90879	16	8052	0	922416
Candidate_Type	90879	NA	0	NA	NA
Valid_Votes	90879	16	8052	0	1620444
Electors	90879	0	197	1	4587998
Constituency_Name	90879	NA	0	NA	NA
Constituency_Type	90879	NA	0	NA	NA
Sub_Region	90879	NA	0	NA	NA
N_Cand	90879	0	0	1	480
Turnout_Percentage	90879	16	8072	0	15300.52
Vote_Share_Percentage	90879	74	8068	0	97.69
Deposit_Lost	90879	NA	0	NA	NA
Margin	90879	16598	0	0	696321
Margin_Percentage	90879	19371	8068	0	97.19

[1] This will be updated as soon as the 2019 data comes in.

ENOP	90879	0	8068	1.05	10
pid	90879	NA	0	NA	NA
Party_ID	90879	0	0	1	18944
Max_Poll_No	90879	85566	0	0	2
last_poll	90879	NA	0	NA	NA
Contested	90879	0	0	1	543
Last_Party	90879	NA	0	NA	NA
Last_Party_ID	90879	0	72481	9	18721
Last_Constituency_Name	90879	NA	90879	NA	NA
Same_Constituency	90879	NA	90879	NA	NA
Same_Party	90879	NA	72481	NA	NA
Mandate	90879	83145	0	0	11
No_Mandates	90879	79115	0	0	11
Turncoat	90879	NA	0	NA	NA
Incumbent	90879	NA	0	NA	NA
Recontest	90879	NA	0	NA	NA