TRIVEDI CENTRE
FOR POLITICAL DATA

Lok Dhaba Codebook 2.1

http://lokdhaba.ashoka.edu.in

Last updated: February 12, 2022

Contents

TRIVEDI CENTRE FOR POLITICAL DATA

Introduction

Welcome to Lok Dhaba!

Lok Dhaba is a visual and data repository of the results of all Indian election results held since 1962. The data comes from the Statistical Reports published by the Election Commission of India (ECI from here on). The data extracted was cleaned and structured to fit into a tabular format. Additional variables, mostly derived from ECI data, have been added. We have also used information, specifically, the education and profession of winners, from the Association for Democratic Reforms. However the original information has been modified according to schema made available in Appendix 1.

This codebook provides information about the variables included 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.

In Version 2.0, we have added more information on political parties (variables #28 and #29), and sociological variables extracted from candidate affidavits (variables #43 - #47). These new variables are described in Table 1 and Table 2 below.

We hope that you will find this repository useful and request that you cite it whenever you use it in publications or public presentations.

With thanks,

TCPD Team

TRIVEDI CENTRE FOR POLITICAL DATA

Terms and Conditions

Lok Dhaba is an online web interface designed and hosted 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: Ananay Agarwal, Neelesh Agrawal, Saloni Bhogale, Sudheendra Hangal, Francesca Refsum Jensenius, Mohit Kumar, Chinmay Narayan, Basim U Nissa, Priyamvada Trivedi, and Gilles Verniers. 2021. "TCPD Indian Elections Data v2.0", 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. 2021. "TCPD Indian Election Data Codebook v2.0", 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 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	Election_Type	string	*Type of Election:* General Election (GE) or Assembly Election (AE)
2	State_Name	string	Name of the State
3	Assembly_No	numeric	Assembly Number as per ECI
4	Constituency_No	numeric	Assembly Constituency Number as per ECI
5	Year	numeric	Year in which the election was held
6	month	numeric	Month in which election results were announced
7	Poll_No	numeric	This is an indicator of whether the election was a bye-election or regularly scheduled election. The variable is 0 in case it was a regularly scheduled election, 1 for the first bye-poll for that assembly and constituency, 2 for the second bye-poll, and so on.
8	DelimID	numeric	Delimitation Number
9	Position	numeric	Rank of the contestant based on Vote Share
10	Candidate	string	Name of the Candidate as per ECI
11	Sex	string	Gender of the Candidate as per ECI
12	Party	string	Party of the Candidate as per ECI
13	Votes	numeric	Number of Votes for the Candidate as per ECI
14	Candidate_Type	string	*Category of the Candidate:* General (GEN) or reserved for Scheduled Castes (SC) or Scheduled Tribes (ST)
15	Valid_Votes	numeric	Sum of all votes to all candidates in a particular election for a particular constituency
16	Electors	numeric	Total number of registered electors in the constituency as per ECI

17	Constituency_Name	string	Name of the Constituency
18	Constituency_Type	string	*Type of the Constituency:* General (GEN) or reserved for Scheduled Castes (SC) or Scheduled Tribes (ST)
19	Sub_Region	string	Subregion of the constituency
20	N_Cand	numeric	Number of candidates contesting the constituency
21	Turnout_Percentage	numeric	Percentage turnout in the constituency
22	Vote_Share_Percentage	numeric	Percentage Vote Share received by the candidate in a given constituency
23	Deposit_Lost	logical	Binary variable that indicates if the candidate lost their deposit (yes) or not (no)
24	Margin	numeric	Difference in votes between a candidate and the next ordered candidate (ordered by position)
25	Margin_Percentage	numeric	Percentage margin of a candidate in compared to the next position candidate
26	ENOP	numeric	Effective Number of Parties (derived from vote shares of each candidate in a constituency)
27	pid	numeric	Unique identifier for a candidate
28	Party_Type_TCPD	string	Classification of political parties done by TCPD
29	Party_ID	numeric	Unique identifier for a political party
30	last_poll	logical	Binary variable that indicates if this was the last election held in this constituency (TRUE) or not (FALSE)
31	Contested	numeric	The number of times a candidate has contested so far, including this election
32	Last_Party	numeric	The name of the party the candidate contested from in their previous contest
33	Last_Party_ID	numeric	The unique party ID of the party the candidate contested from in their previous contest
34	Last_Constituency_Na	string	Name of constituency this candidate

	me		contested from in their previous contest
35	Same_Constituency	logical	Binary variable that indicates if the candidate contested from the same constituency in their previous contest.
36	Same_Party	logical	Binary variable that indicates if the candidate contested from the same party in the previous contest.
37	No_Terms	numeric	Number of contests won by the candidate, including this election
38	Turncoat	logical	Binary variable that indicates if the candidate has switched their party from the immediately preceding election (TRUE) or not (FALSE)
39	Incumbent	logical	Binary variable that indicates if the candidate is a sitting member of the house at the time of this election (TRUE) or not (FALSE).
40	Recontest	logical	Binary variable that indicates if the candidate contested in the immediately preceding election (TRUE) or not (FALSE).
41	Age	numeric	Age of the candidate as per ECI
42	District_Name	string	Name of district of constituency as per 2001 Census
43	Education Level	string	Indicates the education level of the candidate as filled by them in their nomination affidavit.
44	Primary Profession Category	string	This is the primary profession category based on the profession filled by the candidate in the nomination affidavit.
45	Primary Profession Type	string	This is the primary profession type based on the profession filled by the candidate in the nomination affidavit.
46	Secondary Profession Category	string	This is the secondary profession category based on the profession filled by the candidate in the nomination affidavit.
47	Secondary Profession Type	string	This is the secondary profession type based on the profession filled by the candidate in the nomination affidavit.

48	PC_Name	string	Name of the parliamentary constituency within which the AC segment lies. Note: This applies when "Show AC segment wise results" has been selected.
49	PC_No	numeric	Number of the parliamentary constituency within which the AC segment lies. Note: This applies when "Show AC segment wise results" has been selected.

Table 2: Detailed variable description

#	Variable Name	Variable description
1	Election_Type	This variable is used to denote the *Type of Election:* General Election (GE) conducted for the Lok Sabha, or Assembly Election (AE)conducted for Vidhan Sabhas
2	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.
3	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, 17 → 2019. This works in a similar way in the case of the State or Vidhan Sabha elections too. For instance, Uttarakhand which was created in 2000, has had 4 assemblies (1 → 2002, 2 → 2007, 3 → 2014, 4 → 2017).
4	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").

5	Year	The year in which the election took place.
6	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
7	Poll_No	This variable indicates the number of times a constituency has been up for election within a given assembly. For instance, for an election held in 2009 for the Lok Sabha, 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 accordingly. In essence, it is used to keep track of bye-polls, with 0 indicating an election held when the previous assembly was dissolved, 1 indicating the 1st bye-poll, 2 indicating the second bye-poll, and so on.
8	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)]
9	Position	At any time there can be several candidates contesting in any constituency. This variable shows the rank of the particular candidate in the constituency, calculated based on Vote Share. It is possible for the votes of two people to be the same, however, this has not been true for a winner.
10	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 the 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.
11	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.
12	Party	The Political Party on whose ticket the candidate contested from as per the ECI.
13	Votes	Number of votes that the candidate received as per the ECI.
14	Candidate_Type	The candidate can be any of the three types - General (GEN) or reserved for Scheduled Castes (SC) or Scheduled Tribes (ST)
15	Valid_Votes	Sum of all votes to all candidates in a particular election for a particular constituency
16	Electors	This is the total number of people who are registered to vote in a given 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 ("Electors"). = Valid_Votes / Electors
21	Vote_Share_Percentage	This variable is calculated for every contestant by dividing the Number of votes received by the candidate ("Votes"0) by the total number of valid votes that have been cast in a given constituency ("Valid_Votes"). = Votes / Valid_Votes
22	Deposit_Lost	When contesting elections, each candidate needs to deposit a certain amount of money. If the candidate receives less than 1/6th of 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). The margin for the candidate who comes last will be 0.
24	Margin_Percentage	Difference in Vote_Share_Percentage between a candidate and the next ordered candidate (ordered by Position). The margin 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 each candidate in a constituency (Laakso and Taagepera 1979).
26	pid	This is a unique alpha-numeric identifier for each 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 a single version of the dataset. For more information, check out our project "Surf".
27	Party_Type_TCPD	Political parties and candidates have been classified into four categories: National Parties, State-based Parties, Local Parties and Independent candidates. This classification does not follow the ECI's legal definition of national and state-based parties, but reflects TCPD's assessment of the national, regional or local character of parties. Thus, parties contesting in several states but being principally associated with one state are classified as state-based parties, even though they may meet the ECI definition of national party (for ex.: BSP). State-based parties typically contest most seats across a state's sub-regions, although exceptions are possible (for ex.: JD(S) in Karnataka)
28	Party_ID	This is a unique numerical identifier for a political party assigned by TCPD. It is analogous to pid, but for political parties. 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 a single version of the dataset. For more information, check out our project "Surf".
29	last_poll	This is a dichotomous variable which indicates if this election was the last election held in this constituency before the Assembly was dissolved. 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, including this election
31	Last_Party	The name of the party the candidate contested from in their previous contest
32	Last_Party_ID	The unique Party_ID of the party the candidate contested from in their previous contest
33	Last_Constituency_Name	The name of the constituency the candidate contested from in their previous contest (TRUE) or not (FALSE)
34 35	Same_Constituency	Binary variable that indicates if the candidate contested from the same constituency in their previous contest (TRUE) or not (FALSE)
36	Same_Party	Binary variable that indicates if the candidate contested from the same party in their previous contest (TRUE) or not (FALSE)
37	No_Terms	The total number of unique assemblies in which a candidate has won. If a candidate has won more than one election within the same Assembly (either by winning in multiple constituencies or multiple bye-polls), the counter does not increase. Eg - Candidate Kalidas Nilkanth Kolambkar from the 13th Assembly of Maharashtra has won a total of 8 times (7a,8a,9a,10a,10a,11a,12a,13a). But he has won in a total of 7 unique assemblies (7a,8a,9a,10a,11a,12a,13a). So the No_Terms displays 7, and not 8.
38	Turncoat	Currently a candidate is marked as 'Turncoat' if - The candidate contested from a party with a particular ID

		as given in the TCPD-PPI dataset, and then stood in a subsequent election from a party with a different ID assigned from the TCPD-PPI dataset. Currently the TCPD-PPI dataset **does not** yet take into account mergers, factions, offshoots or successor parties – • For example, the SP was created in 1991 and emerged from the Janata Dal. Currently, a candidate who contested from Janata Dal in 1991, and subsequently in 1993 from SP will be marked as a turncoat • Similarly, the Lok Dal in UP split into multiple factions in the 1980s like Lok Dal (A), Lok Dal (B) etc. Many of these factions then merged with other parties and so on. Currently, each of these factions has a separate unique ID, thus a candidate contesting from these factions (despite Lok Dal also contesting) will be marked as turncoat as well. The data also **does not** take into account individuals who may have held official but non-elected positions in a party, or an elected position at a level lower than assembly (panchayat, zilla, municipality) and contested subsequently on a different party affiliation. We are also yet to code turncoats across national and state elections (Lok Sabha candidates who were MLAs from other parties, or vice versa). The TCPD-PPI **does** however account for party name changes (ex.: TMC and AITC for All India Trinamool Congress)
39	Incumbent	Binary variable that indicates if the candidate is a sitting member of the House at the time of this election (TRUE) or not (FALSE)
40	Recontest	Binary variable that indicates if the candidate contested in the immediately preceding election (TRUE) or not (FALSE).
41	Age	Age of the candidate as per ECI. This is only available for Assembly Election data
42	District_Name	Name of district of constituency as per 2001 Census. This is only available for Assembly Election data.

43	Education Level	Indicates the education level of the candidate as filled by them in their nomination affidavit. The information has been sourced from the Association for Democratic Reforms. The scale is as follows: 1. Illiterate 2. Literate 3. 5th Pass 4. 8th Pass 5. 10th Pass 6. 12th Pass 7. Graduate 8. Graduate Professional 9. Post Graduate 10. Doctorate
44	Primary Profession Category	This is the primary profession category based on the profession filled by the candidate in the nomination affidavit. The information has been sourced from the Association for Democratic Reforms. This is a categorisation done by TCPD based on the various professions filled by multiple candidates in their affidavits across elections. The 18 categories used are as follows: 1. Agricultural Labour 2. Agriculture 3. Business 4. Education 5. Former Government 6. Government 7. Labourer or Daily Wage 8. Liberal Profession or Professional 9. Politics 10. Religious Occupation 11. Retired or Pension 12. Salaried Work or Employed 13. Small Business or Self-employed 14. Social Work 15. Student 16. Traditional Occupation 17. Unemployed 18. Other or Unspecified

		The primary profession category is the one that is likely to be the primary source of income for the candidate.
45	Primary Profession Type	This is the primary profession type based on the profession filled by the candidate in the nomination affidavit. The information has been sourced from the Association for Democratic Reforms. This is the type within the primary profession category. The list of types under each category are given in Appendix 2.
46	Secondary Profession Category	This is the secondary profession category based on the profession filled by the candidate in the nomination affidavit. The information has been sourced from the Association for Democratic Reforms. This is a categorisation done by TCPD based on the various professions filled by multiple candidates in their affidavits across elections. The 18 categories used are as follows: 1. Agricultural Labour 2. Agriculture 3. Business 4. Education 5. Former Government 6. Government 7. Labourer or Daily Wage 8. Liberal Profession or Professional 9. Politics 10. Religious Occupation 11. Retired or Pension 12. Salaried Work or Employed 13. Small Business or Self-employed 14. Social Work 15. Student 16. Traditional Occupation 17. Unemployed 18. Other or Unspecified The secondary profession category is the one that is less likely to be the primary source of income for the candidate than the primary profession category. This is populated only when the candidate has filled in more than

		one profession in their affidavit.
47	Secondary Profession Type	This is the secondary profession type based on the profession filled by the candidate in the nomination affidavit. The information has been sourced from the Association for Democratic Reforms. This is the type within the secondary profession category. The list of types under each category are given in Appendix 2.
48	PC_Name	Name of the parliamentary constituency within which the AC segment lies. Note: This applies when "Show AC segment wise results" has been selected.
49	PC_No	Number of the parliamentary constituency within which the AC segment lies. Note: This applies when "Show AC segment wise results" has been selected.

Appendix 1: Profession Categories

Category	**Type**
Agricultural Labour	Coolie
Agriculture	
Business	Real Estate or Builder or Developer or Construction, Contractor, Industrialist, Trader, Transport, Petrol Pumps, Security Services, Film Producer, Distributors, Agricultural Business, Rent, Private Company, Hospitality, Unspecified, Manufacturer, Textile
Education	School Teacher, Principal or Administrator, College Professor or Lecturer, Home Teacher or Tuition
Former Government	Ex Military, Ex Civil Service, Ex Police, Ex Judiciary, Ex Govt Employee
Government	Aanganwadi, Government Employee, Civil Service, Judiciary
Labourer or Daily Wage	Construction, Taxi Driver, Auto-rickshaw Driver, Coolie, Mechanic, Service, Artisan, Truck Driver, Unspecified, Carpenter, Painter

Liberal Profession or Professional	Lawyer, Magistrate, Doctor, Journalist or Editor, Actor, Musician or Artist, Sports, Architect, Film Director, Author, Consultant, Designer
Politics	MP, MLA, Ex MP, Ex MLA, Party Worker, Panchayat Member, Municipality Member, Minister, Ex Minister, Panchayat Head, District Council, Ex Municipality Member
Religious Occupation	Sadhu, Pastor, Imam, Priest, Religious Teacher, Preacher
Retired or Pension	Retired or Pension, Retired Teacher
Salaried Work or Employed	Medical Professional, Engineer, Accountant, Insurance Agent, Managerial Position, Legal Professional, Qualified Professional, Computer Scientist
Small Business or Self-employed	Shop Keeper, Workshop or Small Manufacturer, Self-employed
Social Work	NGO, Activist, Advocacy, Trade Union
Student	
Traditional Occupation	Yoga or Meditation Teacher, Astrologer, Traditional Medicine, Traditional Education
Unemployed	
Other or Unspecified	Household

Appendix 2: User Interface

Figure 1: The *Data Visualization* tab allows the user to select whether they want to look at Lok Sabha or Vidhan Sabha results. There is also an option for users to explore Lok Sabha results at the assembly constituency level.

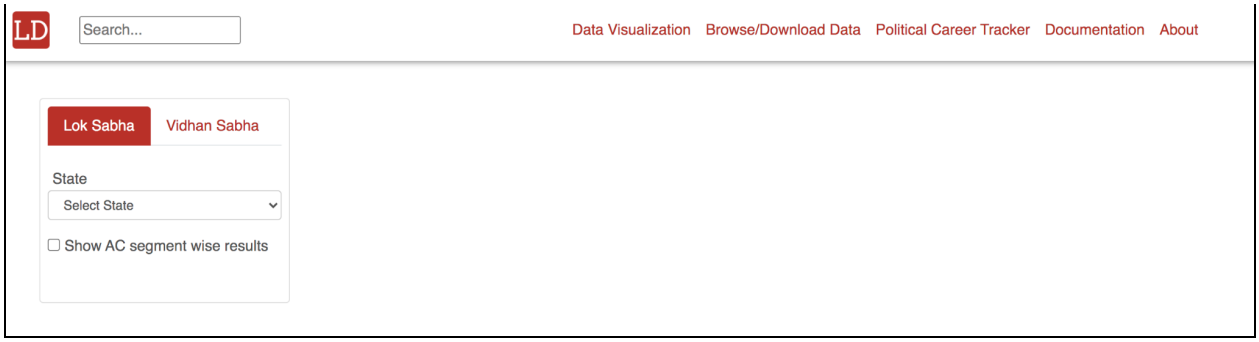


Figure 2: Once the following selections are made - Lok Sabha or Vidhan Sabha, and then the state/or all states - the user can explore different variables which are grouped under (1) voter turnout, (2) party performance, (3) candidate performance, (4) constituency information and (5) affidavit information.

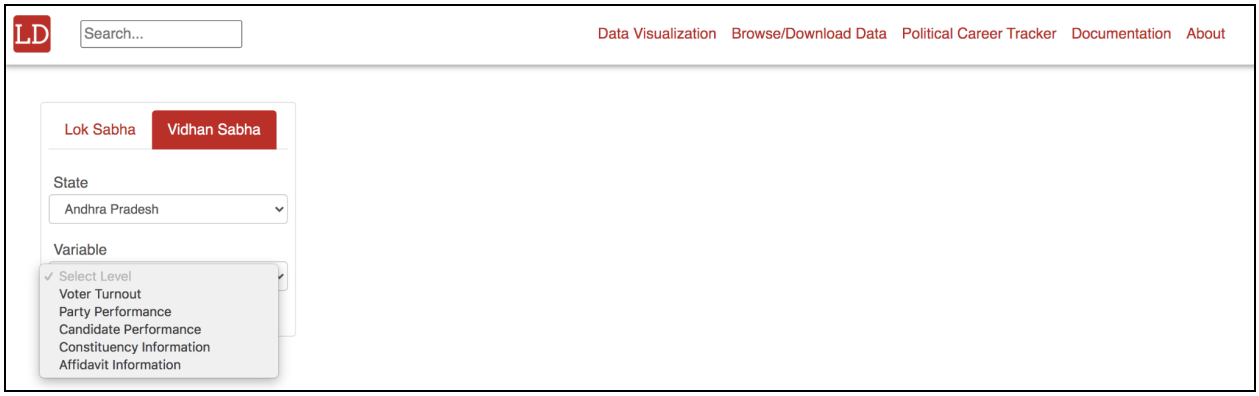
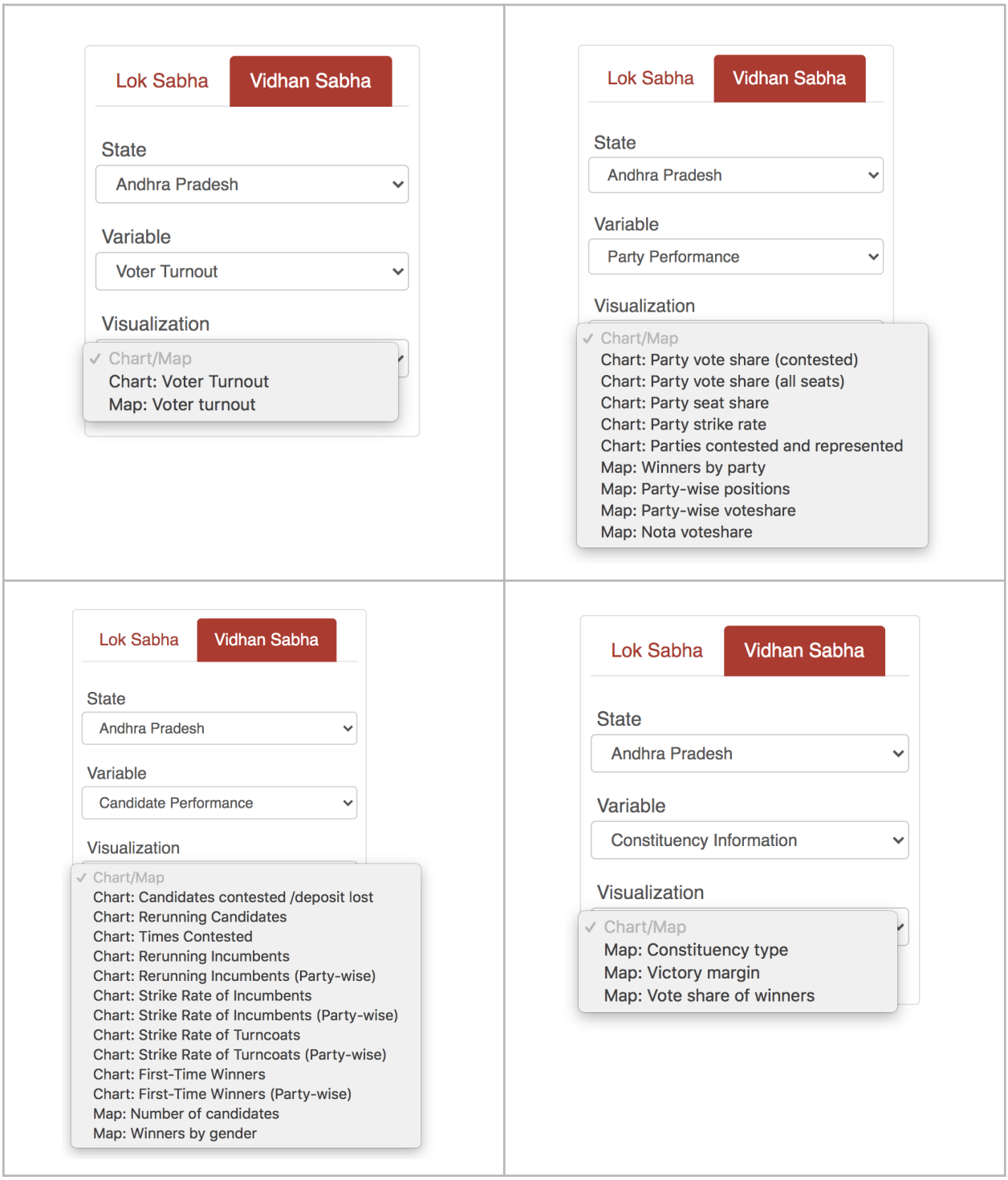
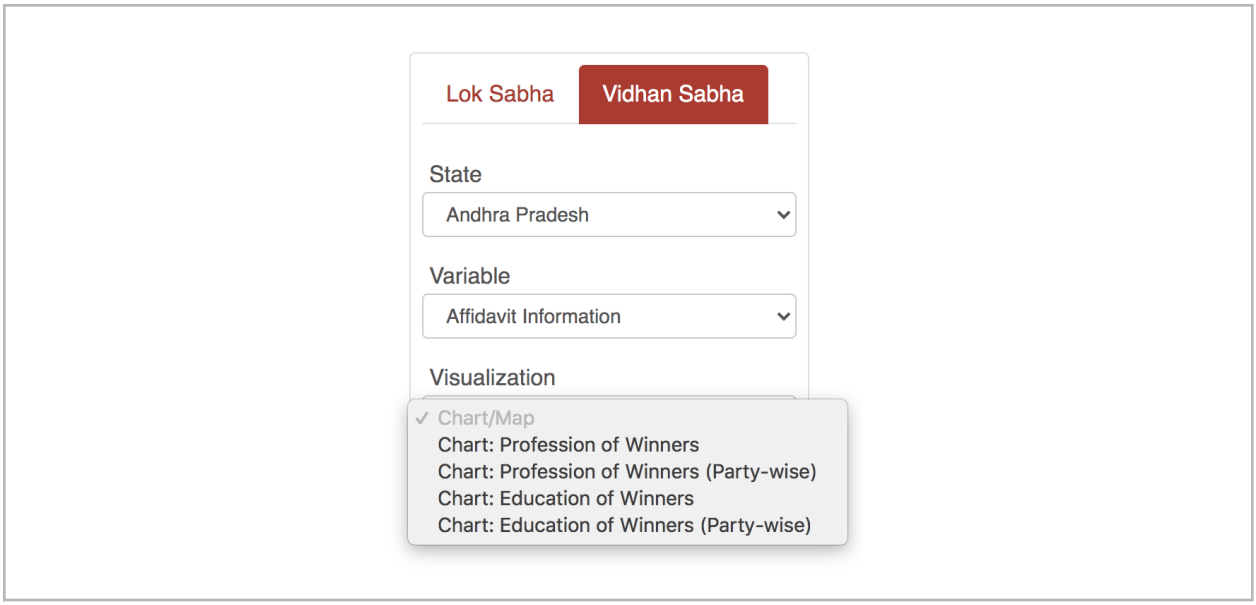


Figure 3: Each of the levels of aggregation presents the user with a menu of charts and maps that can be explored.




Lok Sabha
Vidhan Sabha
State
Andhra Pradesh
Variable
Affidavit Information
Visualization
Chart/Map
Chart: Profession of Winners
Chart: Profession of Winners (Party-wise)
Chart: Education of Winners
Chart: Education of Winners (Party-wise)