



## Data Article

# Dataset of Historical Elections in Catalonia (DHEC): Micro-level Historical Results (1890–1923)



Pau Vall-Prat

*Department of Political Science, Universitat de Barcelona, Av. Diagonal 684, 08034, Barcelona, Spain*

## ARTICLE INFO

*Article history:*

Received 8 March 2021

Accepted 23 March 2021

Available online 29 March 2021

*Keywords:*

Dataset

Historical Political Economy

Elections

Election fraud


Spanish Restoration

Catalonia

## ABSTRACT

The Dataset of Historical Elections in Catalonia covers the most detailed information available for elections that took place in Catalonia during the period 1890–1923. It describes all the specificities of a unique dataset encompassing census tract electoral returns, together with candidate-level data and GIS information. This note presents the context of the electoral data by describing the electoral system and political characteristics of Catalonia and Spain for the period of interest. This dataset expands information regarding electoral data in Catalonia for the period, which until now was only available at the electoral district level. Micro-level electoral data has been collected from multiple different sources ranging from official gazettes, original electoral tallies preserved in a wide range of archives, and secondary sources. This note descriptively presents the data to show its reuse potential for a large array of social scientists interested in historical, political, economic, or demographic processes that took place at the early 20th century. This new dataset has a large potential for future social science research on electoral behavior or party system formation and can be employed to better understand how social processes influenced politics and vice versa.

© 2021 The Author. Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)*E-mail address:* [pau.vall@ub.edu](mailto:pau.vall@ub.edu)*Social media:* <https://doi.org/10.1016/j.dib.2021.107017>2352-3409/© 2021 The Author. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

---

**Specifications Table**


---

Subject	Social Sciences-Political Science
Specific subject area	Electoral Data for Historical Political Science. It focuses on the electoral returns and all the variables generally associated with it: registered voters, turnout, or support for more than one thousand candidates.
Type of data	Table Shapefile
How data were acquired	Archival work, primary sources (official gazettes) and secondary sources (mainly digitized local newspapers).
Data format	Raw. Data provided in .csv/.dta format for the tables and .shp for the shapefiles.
Parameters for data collection	All lower chamber and provincial elections that took place in Catalonia between 1890 and 1923 were included in the dataset. Electoral data was recorded at the census-tract level.
Description of data collection	Data was hand coded from original and official sources when possible (mostly official gazettes, electoral tallies, and electoral censuses) and supplemented with secondary sources. First, electoral data in official gazettes were coded, then tallies and censuses in archival sources were consulted, finally-if necessary-secondary sources were employed to complete the information if official sources were not available. Provincial and electoral district capital cities archives were consulted. In order to reconstruct the map of municipalities, I departed from current day municipality borders and reconstructed the municipalities that disappeared over the years through original old maps displaying municipality-level borders employing QGIS software.
Data source location	Institution: Universitat de Barcelona City/Town/Region: Barcelona, Catalonia Country: Spain  Primary data sources: <ul style="list-style-type: none"> <li>- Official Gazettes <ul style="list-style-type: none"> <li>• Boletín Oficial de la Provincia de Barcelona (<a href="https://www.diba.cat/web/arxiu/boph">https://www.diba.cat/web/arxiu/boph</a>)</li> <li>• Boletín Oficial de la Provincia de Gerona (<a href="http://www.ddgi.cat/bopH/">http://www.ddgi.cat/bopH/</a>)</li> <li>• Boletín Oficial de la Provincia de Lleida (<a href="https://www.diputaciolleida.cat/publicacions-i-bases-de-dades/butlleti-oficial-de-la-provincia-bop/bop-historic/">https://www.diputaciolleida.cat/publicacions-i-bases-de-dades/butlleti-oficial-de-la-provincia-bop/bop-historic/</a>)</li> <li>• Boletín Oficial de la Provincia de Tarragona (<a href="https://prensahistorica.mcu.es/es/consulta/registro.do?id=1604">https://prensahistorica.mcu.es/es/consulta/registro.do?id=1604</a>)</li> </ul> </li> <li>- Archives <ul style="list-style-type: none"> <li>• Arxiu General de la Diputació de Barcelona</li> <li>• Arxiu General de la Diputació de Girona</li> <li>• Arxiu General de la Diputació de Tarragona</li> <li>• Arxiu General de la Diputació de Lleida</li> <li>• Arxiu Comarcal de l'Alt Camp</li> <li>• Arxiu Comarcal de l'Alt Penedès</li> <li>• Arxiu Comarcal del Bages</li> <li>• Arxiu Comarcal del Baix Camp</li> <li>• Arxiu Comarcal del Baix Llobregat</li> <li>• Arxiu Comarcal del Baix Penedès</li> <li>• Arxiu Comarcal del Berguedà</li> <li>• Arxiu Comarcal de la Conca de Barberà</li> <li>• Arxiu Comarcal del Garraf</li> <li>• Arxiu Comarcal de la Segarra</li> <li>• Arxiu Comarcal del Vallès Occidental</li> <li>• Arxiu Comarcal del Vallès Oriental</li> <li>• Arxiu Històric Fidel Fita</li> <li>• Arxiu Històric Municipal de Tarragona</li> <li>• Arxiu Històric de Sabadell</li> <li>• Arxiu Municipal de Badalona</li> <li>• Arxiu Municipal de Borredà</li> <li>• Arxiu Municipal de Granollers</li> <li>• Arxiu Municipal de Vic</li> <li>• Arxiu Municipal del Districte d'Horta-Guinardó</li> </ul> </li> </ul>

- 
- Local Newspapers Repositories
    - Pandora (<https://xacpremsa.cultura.gencat.cat/pandora/>)
    - Trencadís (<https://trencadis.diba.cat/>)
    - Arxiu Administratiu de Girona (<https://www.girona.cat/sgdap/cat/premsa.php>)
    - Biblioteca Hemeroteca Municipal de Tarragona (<https://www.tarragona.cat/patrimoni/fons-documentals/biblioteca-hemeroteca/hemeroteca-1/premsa-digitalizada-1>)

Data accessibility

Repository name: Mendeley Data

Data identification number: <http://dx.doi.org/10.17632/88vvxk4ydp.1>

Data, codebooks, and GIS files are all available at the Mendeley Data site.

## Value of the Data

- The data are useful to analyze important political and social processes in a context of partial democratization: from democratization of the political system, emergence of new political parties. The availability of disaggregated micro-level Historical electoral returns can be employed to better understand the micro-level mechanisms behind these processes.
- Historical and micro-level data in Spain have not received enough attention, although elections in Spain at that time were similar to other European countries. The political specificities of Spanish elections can contribute to better understand dynamics of political domination and manipulation of the electorate and to compare it to other similar European cases.
- Historical electoral data can be useful to a wide range of social science scholars: political scientists, historians, economists, or demographers. The largely detailed data and its long time-span can be useful to all social scientists who employ history as their laboratory and that are interested in understanding better social changes that took place in the early 20th century.
- Interdisciplinarity is a growing value in social sciences and Historical Political Economy is a growing field in between these three intertwined social science fields. Most scholars working in this field have focused on canonical cases for which rich information is publicly available. The availability of data for new cases will help to provide results that are more accurate.
- These data are also helpful for individuals interested in the study of the causes and preconditions that led to the rise of new political parties in contexts of partial democratization, as well as their associated political and social consequences.
- These data can be used to gain further insight into the micro-conditions for electoral fraud. They also permit a more detailed look into the interactions between social, economic, and political characteristics on fraud.

## 1. Data Description

Detailed electoral results are essential data in many social science fields. Political scientists, economic historians, demographers, or sociologists are likely to employ electoral data in their research because poll figures are good indicators of power struggles and reflect multiple characteristics of the society under study.

In fact, there has been a recent increase of Historically oriented works in comparative politics and political economy [1,2]. Micro-level Historical data are nowadays extensively used, which shows that Historical micro-level electoral data can be fundamental to answer relevant social science puzzles.

**Table 1**

Regular election date.

Date	Type of Elections	Date	Type of Elections
07/12/1890	Provincial	21/04/1907	Lower Chamber
01/02/1891	Lower Chamber	24/10/1909	Provincial
11/09/1892	Provincial	08/05/1910	Lower Chamber
05/03/1893	Lower Chamber	12/03/1911	Provincial
09/09/1894	Provincial	09/03/1913	Provincial
12/04/1896	Lower Chamber	08/03/1914	Lower Chamber
06/09/1896	Provincial	14/03/1915	Provincial
27/03/1898	Lower Chamber	09/04/1916	Lower Chamber
12/09/1898	Provincial	11/03/1917	Provincial
16/04/1899	Lower Chamber	24/02/1918	Lower Chamber
10/03/1901	Provincial	01/06/1919	Lower Chamber
19/05/1901	Lower Chamber	06/07/1919	Provincial
08/03/1903	Provincial	13/12/1920	Lower Chamber
26/04/1903	Lower Chamber	12/06/1921	Provincial
12/03/1905	Provincial	29/04/1923	Lower Chamber
10/09/1905	Lower Chamber	10/06/1923	Provincial
10/03/1907	Provincial		

In Spain, micro-level Historical electoral data have been scarce and mostly focused on the Second Republic period (1931–1939),<sup>1</sup> the first time Spain was considered a full democracy [3]. However, Spain had a long tradition of elections before the 1930s [4]. Some of these elections were rigged, coercion was recurrent, and fraud was widespread [5], like in many other European countries (e.g. see [6] for the German case).

In those years, Catalonia was a rapidly industrializing region, where processes of social change accelerated, and new modern parties emerged and consolidated. Hence, micro-level electoral data covering this period can contribute to disentangle the logic behind political and partisan dynamics during these intense years.

In this note, I present the Dataset of Historical Elections in Catalonia (DHEC), a unique and newly collected dataset that sheds light on micro-level Historical electoral records in Catalonia, a northeastern region of Spain. This Data in Brief article includes 3 Tables and 3 Figures.

The DHEC contains electoral data at the census-tract level for all elections comprised between the enactment of universal male suffrage (1890) until a military coup put an end to regular elections (1923); see Table 1 for a detailed list of all regular elections included in the DHEC dataset.<sup>2</sup> The data presented here can contribute to understand, for instance, electoral fraud, turnout in multilevel settings under partial democratization, or the heterogeneous political consequences of industrialization. The DHEC dataset is available at the [Mendeley Data repository](#).

The DHEC comprises not only electoral returns data, but it is also associated with candidate-level data. Furthermore, geographical data are provided to link each observation in the dataset with other geographical levels of aggregation. The DHEC dataset is presented at two different levels of geographic aggregation (census-tract and municipality) and in two different formats: one observation per geographical unit and electoral contest (wide) or one row per geographical unit, electoral contest, and candidate (long). The DHEC files also include several shapefiles for geographical representation purposes; in those shapefiles, municipality borders for different periods are represented. The [Mendeley Data repository](#) stores all the files.

Table 2 displays all the variables included in the DHEC dataset. It briefly describes each variable, the type of data, whether the variable is available at the census-tract (CT) and/or municipality (M) level, and whether it is available in the long (L) and/or wide (W) dataset format. Finally, it also specifies the source from which the data was obtained.

<sup>1</sup> See, for instance, the work by Vilanova [18].

<sup>2</sup> By-elections were also included in the DHEC but not listed in Table 1.

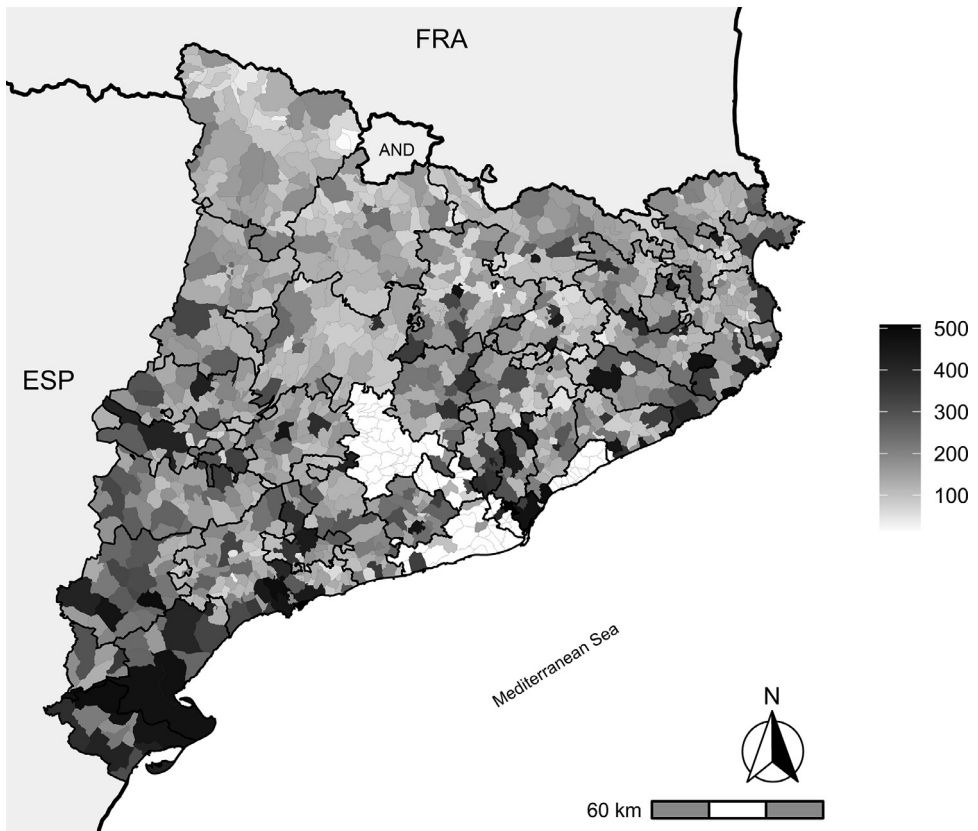
**Table 2**  
DHEC description variables.

Variable	Type	Dataset	Source
<b>Election data</b>			
Type of Elections	Categorical	CT/M; L/W	Balcells et al. [7] and primary sources.
By-election	Categorical	CT/M; L/W	Balcells et al. [7] and primary sources.
Election Date	Date	CT/M; L/W	Balcells et al. [7] and primary sources.
Article 29	Categorical	CT/M; L/W	Balcells et al. [7] and primary sources.
Census Year	Ordinal	CT/M; L/W	Archival.
Number of Registered Voters	Numeric	CT/M; L/W	Primary sources: census books, gazettes, or tallies.
Number of Voters	Numeric	CT/M; L/W	Primary sources, or calculated from votes casted (if Magnitude equals one).
Turnout	Numeric	CT/M; L/W	Calculated: $\text{Voters} / \text{Census} \times 100$
Votes to Candidate <sub>i</sub>	Numeric	CT/M; W	Primary sources.
Votes to Candidates	Numeric	CT/M; L/W	Calculated: sum of all votes to candidates.
Number of Candidates Voted	Numeric	CT/M; L/W	Calculated: number of candidates receiving 1 vote or more.
Votes to Other Individuals	Numeric	CT/M; L/W	Primary sources.
Blank Votes	Numeric	CT/M; L/W	Primary sources.
Null Votes	Numeric	CT/M; L/W	Primary sources.
Blank and Null Votes	Numeric	CT/M; L/W	Calculated: blank votes + null votes.
Votes casted	Numeric	CT/M; L/W	Calculated: votes to candidates + votes to other individuals + blank and null votes.
Calculated Number of Voters	Numeric	CT/M; L/W	Calculated: $(\text{votes to candidates} + \text{votes to other individuals}) / \text{maximum number of votes allowed per individual}$ .
Calculated Turnout	Numeric	CT/M; L/W	Calculated: $\text{calculated number of voters} / \text{census} \times 100$ .
<b>Candidate-level data</b>			
Candidate Name	Text	CT/M; L	Primary sources.
Candidate Id	Categorical	CT/M; L	Own.
Number of Votes Received by the Candidate	Numeric	CT/M; L	Primary sources.
Votes to candidate over votes to all candidates (%)	Numeric	CT/M; L	Calculated: $\text{Votes to specific candidate} / \text{Votes to all candidates} \times 100$ .
Elected Candidate	Categorical	CT/M; L	Balcells et al. [7] and other secondary sources.
General Political Affiliation	Categorical	CT/M; L	Own.
Further Political Affiliation Details	Categorical	CT/M; L	Own.
Government Party Candidate	Categorical	CT/M; L	Own.
<b>District-level data</b>			
Number of Candidates Standing	Numeric	CT/M; L/W	Balcells et al. [7] and other secondary sources.
District Magnitude	Numeric	CT/M; L/W	Balcells et al. [7] and other secondary sources.
Maximum Number of Votes Allowed per Voter	Numeric	CT/M; L/W	1890 Electoral Law and 1907 Electoral Law.
Number of Registered Voters in the Electoral District	Numeric	CT/M; L/W	Balcells et al. [7] and other secondary sources.
Number of Voters in the Electoral District	Numeric	CT/M; L/W	Balcells et al. [7] and other secondary sources.
Turnout at the District Level	Numeric	CT/M; L/W	Calculated: $\text{District Voters} / \text{District Census} \times 100$ .
<b>Geographic data</b>			
Electoral District	Text	CT/M; L/W	Own.
Municipality	Text	CT/M; L/W	Own.
Municipality Id	Categorical	CT/M; L/W	CED (2002).
Local District	Ordinal	CT; L/W	Own.

(continued on next page)

**Table 2** (continued)

Variable	Type	Dataset	Source
Local District Tract	Ordinal	CT; L/W	Own.
Locality Census Tract Id	Text	CT; L/W	Own.
Province	Text	CT/M; L/W	Own.
Province Id	Categorical	CT/M; L/W	Own.
Lower Chamber District	Text	CT/M; L/W	Own.
Lower Chamber District Id	Categorical	CT/M; L/W	Own.
Provincial District	Text	CT/M; L/W	Own.
Provincial District Id	Categorical	CT/M; L/W	Own.
Judiciary Party	Text	CT/M; L/W	INE Census.
Judiciary Party Id	Categorical	CT/M; L/W	INE Census.
County ( <i>Comarca</i> )	Text	CT/M; L/W	IDESCAT [8].
County ( <i>Comarca</i> ) Id	Categorical	CT/M; L/W	IDESCAT [8].

**Fig. 1.** Mean number of registered voters per census tract in each municipality (1898 electoral census).

Note: Blank municipalities correspond to missing data. Dark borders within Catalonia correspond to lower chamber electoral districts.

Table 2 includes four different types of data. First, election data describes all relevant electoral data describing the type of election (lower chamber or provincial), whether the election was regular or a by-election, the date, the number of registered voters (see Fig. 1 for a mapped example), the number of voters, how many votes obtained each candidate, the number of blank or null votes, etc.

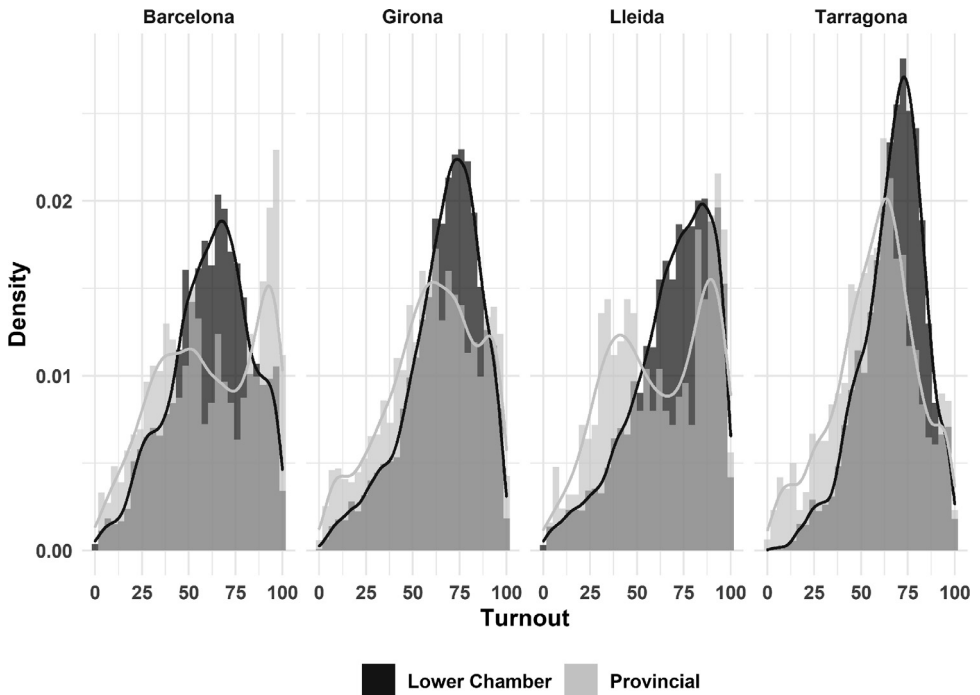


Fig. 2. Turnout by province and type of elections from 1890 to 1923.

Election data also include some easily calculated variables such as turnout (see Fig. 2) or total number of votes to candidates. It is important to note that some electoral variables were not found in any primary source but, nonetheless, can be approximated. For instance, in those elections in which the magnitude of the district, i.e., the number of seats to be elected was larger than 2, each voter could cast votes for more than one candidate. Thus, the number of voters cannot be simply calculated from all votes casted, but it can be approximated, and results are also included in the dataset.

The extent to which the different electoral-level variables are comprehensive is displayed in Table 3. This table describes, for each main type of electoral variables and different type of elections, to what extent census-tract data is available in each Catalan province.<sup>3</sup> Completeness is almost perfect in the province of Girona. In Lleida, the only large gap regards the number of voters in provincial elections. In Barcelona and Tarragona, completeness is lower due to partial availability of Census Books and tallies.<sup>4</sup>

The second type of data included in the dataset relates to candidate-level data. These variables are only available in the long-format datasets and include information regarding the number of votes received by each individual and personal characteristics of each candidate (name, unique ID, partisanship, etc.).

Third, the DHEC dataset includes variables covering district-level data for the elections. These variables describe the magnitude of the district, the number of candidates standing, the overall census and turnout, among others. Some of these variables were coded from secondary sources such as Balcells et al. [7], Gustems Torrent [9], or Mir [10]. When secondary sources were not

<sup>3</sup> Provinces are subnational administrative divisions in Spain. In Catalonia there are 4 provinces: Barcelona, Girona, Lleida, and Tarragona.

<sup>4</sup> This is because tallies and census books were missing or incomplete in provincial archives.

**Table 3**  
Completeness of the DHEC.

Province	Variable	Lower Chamber		Provincial	
		Observations	Completeness %	Observations	Completeness %
<i>Barcelona</i>	Census	3671	27.65	2004	25.53
	Voters	8448	63.62	1443	18.38
	Candidates Votes	12,061	90.83	6542	83.34
	Blank/Null	12,061	90.83	6542	83.34
	<b>N</b>	<b>13,279</b>		<b>7850</b>	
<i>Girona</i>	Census	5960	99.83	3465	100.00
	Voters	5863	98.21	2790	80.52
	Candidates Votes	5863	98.21	3236	93.39
	Blank/Null	5863	98.21	3236	93.39
	<b>N</b>	<b>5970</b>		<b>3465</b>	
<i>Lleida</i>	Census	6065	100.00	3513	99.97
	Voters	5346	88.15	309	8.79
	Candidates Votes	5346	88.15	2733	77.77
	Blank/Null	5346	88.15	2733	77.77
	<b>N</b>	<b>6065</b>		<b>3514</b>	
<i>Tarragona</i>	Census	3253	58.73	1915	59.60
	Voters	4837	87.33	1190	37.04
	Candidates Votes	5338	96.37	3003	93.46
	Blank/Null	5338	96.37	3003	93.46
	<b>N</b>	<b>5539</b>		<b>3213</b>	

Note: Completeness below 50% is signaled in gray.

available, district-level data are obtained from the aggregation of electoral data at the census-tract.

Finally, geographic variables are also included to link all electoral information reported previously to specific geographical units. These variables include both text information and categorical IDs to classify each census tract into a municipality, county, electoral district, judiciary party, and province. According to these geographic variables, the DHEC also provides municipality-level shapefiles that display all localities in Catalonia. Geographical boundaries allow mapping all electoral results at the municipality-level as exemplified in Fig. 1.

The DHEC provides 13 different shapefiles each one covering a different range of years. There are only minor differences between shapefiles. Mostly the changes correspond to mergers or segregations of municipalities, but also to changes in the adscription of localities to different electoral districts. Unfortunately, there were no maps describing the geographic boundaries for census tract divisions.

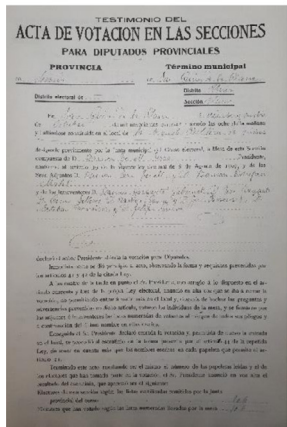
Finally, it is important to note that in most cases, DHEC municipality-level datasets have been generated from the aggregation of census-tract information, but in some (few) instances electoral data was only found at the municipality-level. Therefore, the DHEC dataset is provided at both geographical levels.

## 2. Experimental Design, Materials and Methods

The DHEC dataset encompasses the most extensive and detailed accounts of elections in Catalonia for the period 1890–1923. Micro-level electoral data presented here are mostly novel and have been compiled through original archival work. However, the DHEC database departs from existing lower chamber district-level electoral data [7] and province-specific works providing micro-level electoral data (e.g., [9,10]).



Provincial Elections Tally  
Sant Guim de la Plana



October 24, 1909

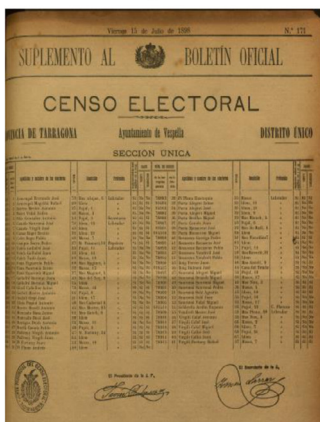
Official Gazette  
Girona Province

**DISTRITO electoral de Sta. Coloma de Farnés**  
*Relación de las votaciones hechas en dicho Distrito, según certificaciones recibidas á esta Junta por las Juntas de las respectivas Secciones.*

PUEBLOS	Secciones	Número de electores según censo	REG. VOT.	BLANCS	PROVINCIALES
Artenayas	1.ª	485	85	306	
	2.ª	453	21	243	
	1.ª	454	70	207	
	2.ª	286	83	197	
	3.ª	450	120	141	
Hostalrich	1.ª	164	52	67	
	2.ª	167	86	64	
Lloret de Mar	1.ª	354	46	300	
	2.ª	415	62	211	
Massanes	Un.	210	51	137	
Massanes de la Selva	1.ª	356	56	70	
	2.ª	168	37	50	
Riudarenas	1.ª	160	7	96	
	2.ª	154	61	61	
S. Felix de Baxallén	1.ª	159	5	105	
	2.ª	208	50	90	
S. Hilario Sacalm	1.ª	304	2	296	
	2.ª	269	10	277	
S. Miguel de Cladells	Un.	70	13	51	
S. Salvador de Breda	1.ª	200	68	92	
	2.ª	135	29	69	
S. Vicente de Ripolmayor	Un.	127	6	102	
	1.ª	423	80	189	
	2.ª	149	17	64	
	3.ª	425	20	202	
Sta. Coloma de Farnés	4.ª	155	6	85	
	1.ª	141	10	56	
	2.ª	151	58	34	
Toses	1.ª	254	4	214	
	2.ª	196	5	127	
Viabrea	Un.	124	9	81	
Vitres	1.ª	293	111	96	
	2.ª	240	111	55	
	1.ª	125	9	96	
	2.ª	129	10	100	
Viladrau					
		8811	1410	4381	

April 26, 1907. p. 2

Census Book  
Vespella



1898

Fig. 3. Example of different sources for electoral data.

2.1. Sources of information

Departing from the previous works, the goal of the DHEC dataset was to compile all available electoral results at the lowest level of aggregation possible. This required an effort to unearth pieces of information that had been hidden in the archives for a long time-period. Not only that, but electoral data was dispersed across multiple archives whose files' reach and comprehensiveness varied widely. This made the research project even more challenging.

The four different types of variables included in the DHEC dataset were compiled from different sources. To build the DHEC dataset I employed four main different data sources: electoral tallies, official gazettes, census books, and secondary sources. The following epigraphs will detail these sources and how the database was finally compiled.

2.1.1. Electoral data

Micro-level electoral data was constructed mainly from official gazettes (most of them digitized), electoral tallies and census books preserved in archives, and secondary sources such as local newspapers. The prioritization of each source followed the order in which have been described. Fig. 3 displays examples of each different data source.

Official gazettes, which were issued by public administrations (and thus should be conventionally trusted), generally displayed most of the relevant variables and concentrated large volumes of micro-level data in a few pages. However, when the information was not complete, then it was time to move to the archives to double check and to obtain the missing pieces of information.

Tallies contain all the relevant pieces of information: registered voters, number of voters, votes for each candidate, and blank (and/or null) votes. When tallies were not available, census books provided information for the number of registered voters in each census tract. Finally, if none of the above was available, secondary sources such as local newspapers were employed to complete the DHEC dataset, although information might be inaccurate or politically biased.

The availability of each data source was not homogeneous across all municipalities. Electoral results were always published in the Provincial Official Gazette a few days after the vote. However, it was usual that some municipalities' data were missing because tallies were sent with delay from remote localities or because publishers decided it was not worthy to publish the

electoral results of a scarce number of municipalities a few days later. Nevertheless, it is important to note that Official Gazettes are the most regularly found data sources for all elections across Catalonia.

Regarding archival data sources, availability is uneven across Catalonia. According to the 1890 and 1907 electoral laws, one copy of both electoral tallies and census books should be filed in each provincial archive. However, not all provincial archives preserved all electoral documentation. The Girona provincial archive is the most complete and almost all files are available; in Tarragona there are tallies (after 1914) and census books (only for certain years); in Lleida, all census books are available (and digitized<sup>5</sup>) but no tallies are available; in Barcelona, neither tallies nor census books are filed. Despite the absence of some of this documentation in provincial archives, electoral data remains in local archives and I have been able to complement those missing documents in provincial archives. The complete list of archives where information was consulted is detailed at the Specifications Table.

Finally, secondary sources from local newspapers have been consulted through multiple online platforms where Historical newspapers are digitized and publicly available. In the days following the election it was usual to find detailed summaries of electoral results in these local newspapers.

### 2.1.2. District-Level and candidate data

As previously detailed, district-level data was widely available from secondary sources. Balcells et al. [7] coded all lower chamber elections in Catalonia between 1901 and 1923. In their work, they included data regarding the number of registered voters, voters, votes to candidates, blank/null votes, and party affiliation for each candidate at the electoral district level. All this information has been included in the DHEC dataset and it has also been useful to double check the micro-level electoral results. Balcells et al. [7] obtained their results from official district-level tallies and certified electoral results at the lower chamber archives.

It is important to note that in many instances the aggregation of micro-level results at the district level do not perfectly match the results presented by Balcells et al. [7], although the DHEC results do not change the rank of candidates votes. This is most likely because some census tracts results were invalidated or because of human errors when transcribing electoral results. Most inconsistencies between both datasets cannot be attributed to human error hand coding in the DHEC, all census-tract electoral results have been thoroughly revised and reflect the results in official or archival sources.

The DHEC dataset does not only include information on lower chamber elections but also on provincial elections. Aggregate level electoral results for provincial elections were not compiled in a single secondary source. However, partial information was available in the works of de Riquer [11] for Barcelona, Gustems Torrent [9] for Girona, and Mir [10] for Lleida. Even in the case of Girona, micro-level data were provided between 1901 and 1923. In all these sources, similar data were provided at the district-level (census, voters, etc.) as well as the party affiliation for each candidate.

District-level information such as the district magnitude or the maximum number of votes to candidates allowed per voter were defined in the electoral laws of 1890 and 1907, and contrasted in secondary sources.

Finally, candidate-level data has been one of the most difficult pieces of information to fill. Political parties at that time were weak-the two predominant parties were notable parties-and party affiliation was not straightforward to determine. District-level secondary sources were very useful to code winning candidates but were less specific on losing candidates. In order to fill the gaps regarding candidate-level affiliation other secondary sources (for specific time-periods [12] or electoral districts (e.g., [13])) have been employed to complete the DHEC dataset. Despite not all candidates have been assigned to a political party, the number of missing data for can-

<sup>5</sup> Available in this link: <https://www.diputaciolleida.cat/publicacions-i-bases-de-dades/arxiu/censos-electorals/>.

didates is low: more than 98% of legislative candidates and 93% of provincial candidates were assigned to a party.

### 2.1.3. Geographic data

Geographic Information Systems (GIS) information complete the DHEC dataset by linking electoral returns to geographically delimited units. Historical GIS are now available for France [14] or Germany [15] but a Historical GIS for Catalonia had not been created thus far.

The DHEC thus includes yearly locality-level boundaries for all municipalities in Catalonia between 1890 and 1923.<sup>6</sup> To create the map, I started with a current day municipality-level GIS. Then, I followed the description of the Centre d'Estudis Demogràfics (CED) [16] regarding localities aggregation and segregation trends over time. In order to delimit the exact location of municipality borders, I superposed old digitized maps from a digital map library [17] to current municipality shapefiles. Municipality-level maps also allowed me to reconstruct constituency-level borders.

## 2.2. Methodology

The methodology employed to construct the DHEC dataset has simply consisted to hand-code the electoral results displayed in each data source into an Excel spreadsheet where observations corresponded to each census tract and columns defined the multiple variables of interest: registered voters, voters, and blank/null/candidate votes. In some (unfortunately few) instances, electoral data was OCR'd in pdf files, then, copy-pasting the tables into the spreadsheet was easy; all these tables were oversight to make sure the totals in the pdf matched the totals in our data.

Once all electoral data at the census-tract and/or locality level was coded, candidate-level data was assigned. Each individual running for office was tagged a unique Id number. Given that candidate names might have some misspelling mistakes, I ordered all candidates in all districts and elections in alphabetic order, thoroughly revised the list of names, and ensured the same individual was not assigned to two different Ids. Then, and based on primary data or secondary sources, each candidate was assigned a partisan allegiance. This implies that the same candidate Id might be associated with different party labels in different election years.

Despite initially I followed party categorizations by primary and secondary sources, parties have been simplified to 9 categories and a few subcategories within each category when necessary. This helped to simplify the large myriad of small factions within parties. Both candidate and party Ids can be checked in the DHEC codebook.

Once micro-level and district-level electoral returns, and candidate-level data had been coded, I aggregated all spreadsheets in which the electoral data had been coded employing Stata16. Using this software, I was able to combine all the data and present it both at the census-tract level or municipality level, and also reshape the data format from wide (candidates' votes as variables) to long (candidate-municipality units of observation).

Also employing the software, I have been able to generate new variables not found in the original primary data sources. For instance, I calculated turnout following the simple formula:

$$\text{Turnout} = \frac{\text{Voters}}{\text{Registered Voters}} \times 100$$

Other variables that have been calculated include the total number of candidates receiving votes in census tract in each election, the sum of votes to all candidates, the percentage of support to each individual candidate, or the number of voters in multi-member districts-when the number of voters was not available in primary data.

In sum, the DHEC dataset has been compiled and mostly hand-coded from multiple original data sources. Then the results were compared to existing district-level secondary sources

<sup>6</sup> Unfortunately, census tract map boundaries were not mapped.

covering electoral and candidate-level data. Finally, once all data was compiled, processed, and restructured it is now available for publication and sharing.

## CRedit Author Statement

**Pau Vall-Prat:** Conceptualization, Methodology, Software, Validation, Investigation, Resources, Data curation, Writing - Original draft preparation, Writing-Reviewing and Editing, Visualization, Supervision, Project Administration, Funding Acquisition.

## Declaration of Competing Interest

I declare no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

I want to express my gratitude to Raimon Soler-Becerro, who helped coding some election years, and to the multiple historians who have generously shared their wise counsels with me. I would also like to recognize the support I received from all archivists who have facilitated my field work collection. Last but not least, I want to thank my dissertation supervisors: Francesc Amat, Jordi Muñoz, and Pep Vallbé, for their constant encouragement in this enterprise.

This research was conducted as a part of my dissertation project at the Universitat de Barcelona, which is currently funded by the Spanish Ministry of Education through a competitive scholarship (FPU17/01243). The collection and publication of the dataset was also supported by a competitive grant funding from the Spanish Ministry of Science, Innovation, and Universities: Ministerio de Ciencia, Innovacion y Universidades, Programa Estatal I+D+i "Retos de Investigación" (RTI2018-101064-A-I00, 2019–2021, Principal Investigator: Francesc Amat).

## References

- [1] G. Capoccia, D. Ziblatt, The Historical turn in democratization studies: a new research agenda for Europe and beyond, *Comp. Polit. Stud.* 43 (8–9) (2010) 931–968.
- [2] E. Finkel, A. K. Lawrence, A. Mertha, Comparative politics and history, *APSA-CP Newsletter* 29 (2) (2019) 2–4.
- [3] S.I. Lindberg, M. Coppedge, J. Gerring, J. Teorell, V-dem: a new way to measure democracy, *J. Democr.* 25 (3) (2014) 159–169.
- [4] R. Villa García, España en las urnas: una historia electoral (1810–2015), Libros de la Catarata, Madrid, 2016.
- [5] J. Varela Ortega, Los amigos políticos: partidos, elecciones y caciquismo en la Restauración (1875–1900), Alianza, 1977.
- [6] I. Mares, *From Open Secrets to Secret Voting*, Cambridge University Press, Cambridge, 2015.
- [7] A. Balcells, J.B. Culla, C. Mir, Les eleccions generals a Catalunya de 1901 a 1923: relació dels resultats electorals de 1869 a 1899, Fundació Jaume Bofill, Barcelona, 1982.
- [8] IDESCATCodis Territorials i d'entitats, 2020 <https://www.idescat.cat/codis/?id=50&n=9> (2020).
- [9] A. Gustems Torrent, Les eleccions provincials a Girona 1901-1923, Universitat Autònoma de Barcelona, 1985.
- [10] C. Mir, Lleida (1890–1936): caciquisme polític i lluita electoral, Publicacions de l'Abadia de Montserrat, 1985.
- [11] B. de Riquer, Història de la Diputació de Barcelona, Diputació de Barcelona, Barcelona, 1987.
- [12] E. Pujol, J. Santesmases, Diccionari biogràfic dels diputats de la Mancomunitat de Catalunya (1914–1925), 1st ed., Diputació de Barcelona, Barcelona, 2019.
- [13] G. Rubí, Els catalans i la política en temps del caciquisme: Manresa, 1875–1923, Eumo Editorial, Vic, 2006.
- [14] V. Gay, Mapping the third republic, *A Geogr. Info. Sys. France* (2020) 1870–1940 fffal-02951461f.
- [15] D. Ziblatt, J.C. Blossom, Electoral District Boundaries, Germany, 1890–1912, Harvard Geospatial Library, 2011.
- [16] Centre d'Estudis Demogràfics [CED], 2002. "Evolució i continuïtat dels municipis catalans 1497–2002." <https://ced.uab.cat/infraestructures/banc-de-dades-de-catalunya/evolucio-i-continuitat-dels-municipis-catalans-1497-2002/> (2020).
- [17] ICGC. 2020. "Cartoteca Digital." <https://cartotecadigital.icgc.cat/digital/custom/catalunya> (2020).
- [18] M. Vilanova, Atlas electoral de Catalunya durant la Segona República. Orientació del vot, participació i abstenció, Fundació Jaume Bofill, Barcelona, 1986.