LONDON ELECTORAL HISTORY – STEPS TOWARDS DEMOCRACY

7.1 PRINCIPLES OF CLASSIFICATION IN THE LED

Classification has been a foundation stone of science since ancient times. Indeed, in politics the classification of constitutional systems has yet to escape from the influence of Aristotle. Scientific method consists of testing generalities (propositions, hypotheses, and theories) in the light of the empirical evidence of individual cases. Generalisation by classification is a necessary way of making sense of the universe, for without classification it would be impossible to comprehend the vast range of observed phenomena.

Classification necessarily involves both relation and simplification. Two distinct cases sharing common characteristics (relation) may be placed in the same class. There are thus fewer classes than observations (simplification). The principle of the Identity of Indiscernibles is used to allocate indistinguishable observations to a common category.[1] Put crudely, if observations are indistinguishable they are held to be the same.

All classification involves a trade-off between precision and comprehensibility. Taxonomies should be exhaustive and exclusive, that is, each observation should be classified, and no two different observations should share a common top-level classification. In addition, a good taxonomy should group similar observations while distinguishing between those that are different.

Almost all of the information found in the LED consists of nominal variables: a vote for this or that candidate, a parish of residence, an occupation, or the livery company to which a voter belonged. While a vote is expressed by a number, in a binary notation, this identifier remains the nominal variable of a ‘tick in a box’. Only the rack rental valuations contained in rate books constitute truly numerical interval data. Nominal data do not just facilitate classification by aggregating similar cases: in many cases they require classification both to simplify

and to make comprehensible an immense range of observations.

In developing the LED, much use has been made of nested or cascading taxonomies. These are grouped like Russian dolls from the most general to the most specific: a general category, the *summum genus*, is broken down into progressively more specific sub-categories down to the *infirma species*. Table 77 shows a familiar example of a nested classificatory system, of the species *Homo sapiens*.

Table 77
Example of a nested classification

Domain	Eucarya
Kingdom	Animalia
Phylum	Chordata
Subphylum	Vertebrata
Class	Mammalia
Order	Primates
Family	Hominidae
Genus	Homo
Species	Sapiens

Source: B. Bryson, *A short history of nearly everything* (2004), p. 436.

Another example also clarifies the point. Many early county poll books contain information about the location of the freehold property that qualified its owner to poll. The immobility of real property gives classification by place of freehold or place of residence a strength that is lacking in other taxonomies. Middlesex county voters polled in respect of the freeholds which they owned. These freehold estates all lay within Middlesex, which is thus the most general spatial category available. From ancient times Middlesex had been divided into six Hundreds. So the location of the freehold might be expressed more specifically as lying within a certain Hundred. Meanwhile, the urbanised Ossulston Hundred comprised five Divisions. Each Hundred, and each Division of Ossulston Hundred, contained many parishes, and freehold property was often identified by the parish within which it lay.

Moreover, within the urbanised part of the county, many freeholds could be distinguished by a street within a parish. So here the historian can construct a nested classification in which streets were subsumed by a parish, parishes were subsumed by a Division, Divisions were subsumed by a Hundred, and all Hundreds lay within Middlesex. The principle is

the same as the familiar seven-character postcode. Theoretically it should be possible to address a letter with just the house number and full postcode, since the seven-character postcode usually identifies a group of houses. The four-character postcode identifies a region within a postal sorting area, while the two-character postcode identifies the location of the postal sorting office.

Table 78
Example of a nested spatial classification of Middlesex voters

County	Middlesex
Hundred	Ossulston
Division	Westminster
Parish	St James Westminster
Street	Marlborough Street
Alley	Great Marlborough Street

Source: LED.

However, spatial classification may break down at both the highest level and the lowest level. A landed estate did not have to be wholly contained within Middlesex, and a freeholder who was qualified to poll in more than one county could exercise his franchise in more than one. Likewise, a Middlesex freeholder who was also a Westminster householder and a London liveryman might poll in all three constituencies.

At the lowest level, furthermore, a street might cross a parish boundary or even a constituency boundary. Oxford Street formed the boundary between the post-reform constituencies of Westminster and Marylebone. Electors living on the southern side of Oxford Street might poll in Westminster, where the street extended from the St George's parish through St James to St Anne's parish. Those living on the northern side of Oxford Street might poll in Marylebone, where the parish of St Marylebone extended from Tyburn almost to Tottenham Court Road. And of course duly qualified electors on both sides of Oxford Street were eligible to poll in Middlesex elections. In these cases, classificatory decisions have to be taken, so that different classifications apply simultaneously for different purposes.

Hence the analogy of Russian dolls is imperfect in application to the LED. In such a system the largest doll contains the whole of the next

largest, which contains the whole of the next largest, and so on. The set is complete, without overlapping or contradictions. By contrast, classification of the social world takes place on a number of different intersecting planes. People may be classified by sex, by age, by place of birth, by occupation, and so on. No single category contains the whole of the other categories, though it will contain the whole of its own sub-categories. Voters varied widely in age, occupation, income, and political outlook. Hence different taxonomies are needed to describe and to explore each of these intersecting planes.[2]

It is a principle of good coding that it should be collectively exhaustive; that is, in any coding schema every observation should receive a code. The corollary of this is that coding should be mutually exclusive; that is, no two distinct categories should share a common code at the most specific level.

Many of the classifications within the LED are inherent in the data and users are unlikely to want to change them, although they may wish to aggregate them for analytical purposes. For example, different candidates may be held to have contested an election as a 'slate'. Other classifications have been imputed to the data in the editorial process which created the LED. In such cases, the preservation of the original poll book data strings allows users of the LED to adopt their own alternative classifications, if desired.

The distinction between categories that are inherent in the data and those that are imputed to the data is a fundamental one. Many basic categories are inherent in the data. The most obvious of these is the temporal classification: data from the London election of 1713 is distinguished from that of the London election of 1784. And many other spatial and temporal categories are inherent in the data or the process of data collection: thus data from the London parliamentary election of 1784 is distinguished from the Middlesex and Westminster elections in the same year.

The individuation of candidates is in itself robust information. However, the phonetic coding of surnames, while rule-driven, remains potentially problematic. Meanwhile occupational and social classification, discussed in sections 7.12 and 7.13, is more subjective. Those categories that are inherent in the data are necessarily pre-coded, but the imputed categories were post-coded; that is, the coding was implemented after data entry. However, retention of the original surname and occupation character strings, as an oft-repeated point of principle, allows

users to formulate their own solutions to phonetic and occupational coding.

Other key variables for social analysis are either redundant or unavailable. All of these candidates and voters were male, so the analytical variable of sex is redundant. Nothing is known of the voters' own perceptions of their place in society. Virtually nothing is known about voters' religious affiliations, although it may be guessed from their names that a few of the voters were Jews, and some Quakers may be identified by their affirmations. While nothing is known about voters' ages, a London livery list of 1734 is annotated with the date of each liveryman's admission to freedom, which might serve as a proxy for this variable.[3] Meanwhile, the ambitious might conceivably link the Westminster poll data of 1852 to the census data of the previous year.[4]

Despite these deficiencies, imaginative use of the available taxonomies will facilitate much data analysis. The livery companies of all of the London liverymen polling in parliamentary elections and in Common Hall are known. The parish of residence of all voters in Westminster and Marylebone is known. The occupations of almost all the pre-reform Westminster voters are similarly matters of record. Linking poll book and rate book records enables users to discover more about where a substantial proportion of Westminster and Marylebone voters lived. Such linkages allow historians to identify not only voters' place of residence within a parish, but also the relative valuations of the property which they occupied, so providing information that might serve as a proxy for their income. Moreover, linking poll book records from successive elections may reveal whether a voter polled in a previous election and, if so, how.

Collectively, the nested classificatory systems deployed in the LED lend themselves to the complementary processes of historical description of abstraction and extraction. Abstraction reveals the 'big picture', an overview of metropolitan society. Meanwhile extraction reveals the individual cases that illuminate the bare bones of abstraction.

In that respect, abstraction and extraction are akin to the two traditions of empirical research: the general survey and the case study. General surveys may lack colour and human interest, while case studies may lack evidence or even a discussion of the question of typicality. Hence a happy medium encourages historians to move between the abstract and the extract, identifying both the big picture and the significant detail.[5]

Notes

[1] For the principle of the 'identity of indiscernibles', see I. Winchester, 'On referring to ordinary historical persons', in E.A. Wrigley (ed.), *Identifying people in the past* (1973), esp. pp. 24-7.

[2] Colon classification facilitates classification by many characteristics in a single string in which the separate elements are divided by colons.

[3] The London livery list of 1734, which lacks a printed title-page, is to be found at BL 1303 d.12.

[4] The 1851 census reports appear in *BPP* (1852/3), LXXXV-LXXXVIII, pts 1-2.

[5] For example, this approach was adopted by Namier, *Structure of politics*. See too J.E. Miller, *The Chicago guide to writing about numbers* (Chicago, 2004), which advocates a threefold model of Generalisation, Example, Exception. This format is not unlike that adopted by K.V. Thomas, *Religion and the decline of magic: studies in popular beliefs in sixteenth- and seventeenth-century England* (1971); and idem, *Man and the natural world: changing attitudes in England, 1500-1800* (1983).