DATA PAPER

Vagrant Lives: 14,789 Vagrants Processed by the County of Middlesex, 1777–1786

Adam Crymble[1], Louise Falcini[2] and Tim Hitchcock[3]

[1] #1 conceptualisation; #2 methodology; #8 data curation; #14 funding acquisition, University of Hertfordshire, GB
adam.crymble@gmail.com

[2] #8 data curation; #4 validation, University of Reading, GB
louise@falcini.info

[3] #1 conceptualisation; #12 supervision; #14 funding acquisition, University of Sussex, GB
T.Hitchcock@sussex.ac.uk

This dataset makes accessible the uniquely comprehensive records of vagrant removal from, through, and back to Middlesex, encompassing the details of some 14,789 removals (either forcibly or voluntarily) of people as vagrants between 1777 and 1786. It includes people ejected from London as vagrants, and those sent back to London from counties beyond. Each record has been georeferenced (where possible). Significant background material is available on the 'London Lives' website, which provides additional context for these records. The authors also recommend the following article [1].

(1) Overview
Context
These data were produced as part of the 'Vagrant Lives' project, supported by a small grant from the British Academy / Leverhulme Trust (2012). Each record includes details of the name of the vagrant, his or her parish of legal settlement, where they were picked up by the vagrant contractor, where they were dropped off, as well as the name of the magistrate who had processed them as a vagrant. Each entry is georeferenced (using the WGS 84 standard), to make it possible to follow the journeys of thousands of failed migrants and temporary Londoners back to their place of origin in the late eighteenth century [2].

Each entry has 29 columns of data, all of which are described in the READ ME file available at the DOI address given above.

The original records were created by Henry Adams, the vagrant contractor for the county of Middlesex who had – as had his father before him – conveyed vagrants from Middlesex gaols to the edge of the county where they would be sent onwards towards their parish of legal settlement. His role also involved picking up vagrants on their way back to Middlesex, expelled from elsewhere, as well as those being shepherded through to counties beyond, as part of the national network of removal. Eight times per year at each Middlesex Session of the Peace, Adams submitted lists of vagrants conveyed as proof of his having transported these individuals, after which he would be paid for his services. Given the Middlesex Sessions

were traditionally convened eight times per year, between December 1777 (the first known record) and February 1786 (the last surviving record), the Middlesex Sessions would have met 67 times. As the vagrancy system ran year round, we would expect a list to be submitted at each of those eight sessions. However, some of these lists have not survived and we have 42 of a possible 67.

The gaps in the surviving records are unfortunately not evenly spaced throughout the year. Due to the happenstance of survival, we know more, for example, about removal in October than in May. The entries immediately following London's Gordon Riots of 1780 are amongst the missing ones. Likewise, the records for the whole of 1779 are absent from the archive. No complete twelve-month period survives in tact, however a list from each of the cyclical eight sessions survives in at least one of the years.

Spellings have been interpreted and standardised when possible (as described in the 'Methods' below). Location data (hereafter, georeferences) have been added when they could be identified.

This dataset was created for 21st century historians, and should not be construed as a true transcription of the original sources. Instead the goal was to use a limited vocabulary and to interpret the entries rather than recreate them verbatim. While this is undesirable for anyone interested in spelling variations of names and place names in the eighteenth century, it is the authors' hope that these interpretations will make it easier to conduct quantitative analysis and studies in historical geography.

The dataset includes the details of 14,789 removals. The individuals involved were transported as part of 11,017 family groups, including 3,099 children travelling with a parent and 653 were wives of the named vagrant. Only the lead vagrant in a family group was listed by name. Single parents with children were more likely to be travelling with their mother than their father. Nearly six in ten vagrants were travelling alone (8,852), of whom 54 per cent were male. Most vagrants were from England (71 per cent), followed by Ireland (21 per cent), Scotland (5 per cent), and Wales (2 per cent) [1].

(2) Methods

This dataset was produced through a series of steps that involved downloading and then structuring already extant digitised materials.

Steps

The original records had already been digitised and transcribed by the 'London Lives' (2010) project (http://www.londonlives.org/) of which Tim Hitchcock was the Principal Investigator [3]. His knowledge of the 'London Lives' website aided in the discovery of relevant records. The records of vagrant removal in the 'London Lives' collection were first microfilmed by the London Metropolitan Archives (the repository for the original manuscripts). The microfilm was then scanned and converted into 400 dpi JPEG image files. This image capture process was managed by the 'Higher Education Digitisation Service' at the University of Hertfordshire (2010).

The scans were then transcribed through a process known as 'double rekeying', whereby two typists transcribe the text independently. The resultant transcriptions are then compared by a computer to highlight discrepancies, which are then resolved manually. This resulted in a highly accurate set of transcriptions. The imprecise nature of transcription from handwritten documents means that some errors undoubtedly entered the corpus at this stage.

The records of interest to this project were tabular in nature. Unfortunately the transcription process did not preserve the tabular layout of the vagrancy removal material in the HTML-based website (www.londonlives.org), making the dataset difficult to use without extensive cleaning and reformatting. The records appeared in their original locations, spread across larger series of very mixed documents. The relevant sub-set of these documents that contributed to this dataset were brought together by Adam Crymble and Tim Hitchcock through a combination of keyword searching and targeted browsing of likely document sets. As we expected a single list for each of the eight Middlesex Sessions per year, targeted browsing was used to supplement the initial keyword searching. The records of 'London Lives' are arranged as in a physical archive (by collection and date), meaning it was possible to limit browsing to each session, minimising the likelihood of missed entries. Louise Falcini and Adam Crymble then worked to structure the data back into a tabular format using a Comma Separated Values (CSV) file, ensuring each entry lined up with its corresponding details.

Louise Falcini and Adam Crymble manually standardised the spelling of names (Wm to William, for example).

Louise Falcini then added geo-coordinates for each entry for each column of data containing geographic data. This included the place of settlement for each vagrant as listed in the original source, as well as the place they were picked up by Henry Adams (one of a limited set of depots in London and around the borders of the county of Middlesex), and where Adams dropped each person off. This provided a path for each vagrant either out of or into Middlesex.

For places of settlement, the granularity of the recorded point of origin differed greatly across entries. The place names mentioned in the original historical sources did not adhere to a limited vocabulary and were undoubtedly gathered orally from the vagrant at the time of his or her processing. This means that the data include everything from parish names, to hamlets, to towns and cities, to counties, and even to countries. To manage this differing granularity, each entry was categorized into four geographic units of increasing scale:

1) Micro-level (parish level or smaller)
2) Area-level (city or town with more than one Anglican parish)
3) County-level
4) Country-level

Not all entries contained all levels of granularity. Area-level was only used for vagrants from an urban area (eg, Bristol), for which it was plausible that a more detailed description could have been given that would further narrow down the parish of settlement. For users of this dataset, the smallest of micro- or area-level geo-references should be used when seeking to generate point coordinates for a vagrant.

Before they were geo-referenced, the spellings of all place names were modernised and standardised when an equivalent could be identified. Spellings in the original records were often archaic or phonetic, and frequently reflected the regional accent of the vagrant. While most entries were obvious, many places of settlement required a great deal of interpretation to convert, for example 'Bluebury, Berkshire' to 'Blewbury, Berkshire' before it could be georeferenced. This standardisation of spelling was done manually on a case-by-case basis and the spellings updated accordingly in the database. Place names should therefore be construed as modern places on the map, rather than a true transcription of the original source, whenever a geo-reference value is provided.

In most cases a county was listed for each entry (particularly English entries), which made it easier to identify the most likely geographic entity. Places appearing near the border, or in regions affected by later changes in the county boundaries, could be particularly difficult to identify.

A handful of entries, particularly in London, were ambiguous – the parish of 'St Luke', for example, could be either in Chelsea or Old Street. Other sufficiently vague entries such as 'St. Mary, Hertfordshire' also proved challenging. These entries were interpreted on a case-by-case basis, and left as ['unknown'] when two or more plausible entries were possible.

Finally, a significant subset listed what appear to be parish-level details, but no parish could be identified. When this was the case, the entry was georeferenced as '[unknown]' and the next largest identifiable area (usually area or county) was referenced instead. Entries within England typically contained much more specific settlement locations (micro- or area-level), while Irish vagrants were typically only listed as from 'Ireland' (see **Table 1**).

Once an appropriate modern location was identified, geo-coordinates were obtained using a range of online gazetteers and atlases, including but not limited to GeoNames (Geonames.org) [4], Wikipedia [5], GeoHack (https://tools.wmflabs.org/geohack/ [6]), Google Maps [7], and 'Vision of Britain' [8]. All coordinates for a given location were crosschecked against modern maps to ensure they pointed to the correct location. Some locations proved to be hamlets associated with a mother parish, in this instance the mother-parish geo-reference was used.

Cities and places with multiple parishes outside of the metropolis were aggregated and a single mid-point for that city or place was used. Parishes and extra-parochial places in the metropolis (the built-up area of London covering several jurisdictions including south of the river Thames) were given a separate geo-reference. For parishes, the location of the parish church was used as a reference point and for other smaller non-parochial areas, a mid-point reference was used.

The finished georeferenced dataset includes the number of vagrants at each level of geographic granularity as seen in **Table 1**.

Square brackets in a cell represent one of three types of interpreted values:

1) A cell for which no value is given in the original source. For example, '[unknown]' or ['n/a'] is used in a number of columns when no value is known or is not applicable. This is intended to make it clear that the entry is not in need of further interpretation, but has been given a value by the project team.
2) A cell that contains data that is considerably different than expected in that column. For example, '[runner]' is used in the 'Georeference (conveyed to)' column to make it clear that the person was not conveyed to a parish named 'ran away', but instead had run away.
3) A column of entirely interpreted data that did not appear directly in the original data, such as 'Relationship to Lead Vagrant'.

Square brackets were chosen to visually set these interpreted values apart and to make it easier to sort the dataset to isolate them either to the top or the bottom of the set as desired (useful when seeking to identify all entries that contain an area-level georeference, for example). More detail on each of the columns and the interpreted values can be found in the README file stored with the dataset.

Sampling strategy
We believe that this dataset contains all surviving lists of vagrants from the archival record. The dataset contains all 42 surviving lists out of a possible 67.

Quality Control
Louise Falcini checked roughly half of the place names against the images in the original documents to verify the accuracy of the transcription before modernising the spelling. Accuracy was high in the transcription of the documents on the 'London Lives' website, but was least successful with regards to personal names and places. Falcini checked for consistency of spelling of places/townships/counties in the finished dataset. 60 and 70 per cent of magistrates' names were checked to verify spelling and identify unique individuals.

(3) Dataset description
Object name
MiddlesexVagrants1777–1786.csv.
Readme.txt.

Format names and versions
– CSV, txt.

Creation dates
Start date: 2012-04; end date: 2013-07.

Dataset Creators
Adam Crymble, University of Hertfordshire.
Louise Falcini, University of Reading.
Tim Hitchcock, University of Sussex.

Language
English.

License
CC-BY 4.0.

Level	Entries georeferenced	Detail not provided	Unable to georeference	Not applicable
Micro	8,834	4,131	1,824	0
Area	4,190	0	0	10,599*
County	14,034	725**	9	22
Country	14,771	18	0	0

Table 1: The number of vagrants for each category of geographic granularity, showing the number of successful, failed and non-applicable entries. * only applicable to vagrants from known urban areas. ** only 13 of which are in England. The majority are in Scotland or Ireland.

(4) Reuse potential

This dataset contains the names, genders, and points of origins of 14,789 vagrants – men, women, and children expelled from eighteenth-century London. These data have reuse potential within studies of plebeian migration, demographic history, and the social history of London. The data could be combined with other similar sets of data, or could be mashed up with datasets that draw related variables together – cost of living at point of origin, or travel time to London in hours, for example.

The data also represent a clean set of records that can easily be geo-referenced and would be valuable as a teaching and learning tool for students trying digital mapping.

Anyone using the dataset for scholarly purposes is advised to consult the README file carefully, which explains the contents and limits of each column of data. In particular, the spelling of first names has been standardised (Wm = William, for example) and is not a direct transcription from the original historical source. All entries contain a URL where the original scanned primary source can be found for anyone needing to validate specific entries.

References
1. **Hitchcock, T, Crymble, A** and **Falcini, L** 2014 Loose, Idle and Disorderly: Vagrant Removal in Late Eighteenth-Century Middlesex. *Social History, 39*(4). DOI: http://dx.doi.org/10.1080/03071022.2014.975943
2. **Crymble, A** 2015 A Comparative Approach to Identifying the Irish in Long Eighteenth-Century London. *Historical Methods, 48*(3). DOI: http://dx.doi.org/10.1080/01615440.2015.1007194
3. **Hitchcock, T, Shoemaker, R, Howard, S, McLaughlin, J** et al. *London Lives, 1690–1800* (version 1.1). http://www.londonlives.org (Accessed 2012–2013).
4. **Wick, M** *GeoNames*. Available from http://www.geonames.org (Accessed 2012–2013).
5. *Wikipedia*. Available from http://www.wikipedia.org (Accessed 2012–2013).
6. *GeoHack*. Available from https://tools.wmflabs.org/geohack/ (Accessed 2012–2013).
7. *Google Maps*. Available from https://www.google.co.uk/maps (Accessed 2012–2013).
8. **Southall, H** et al. *Vision of Britain*. Available from http://www.visionofbritain.org.uk/ (Accessed 2012–2013).