

The Challenges and Rewards of Analysing Handwritten Records

Antriksh Dhand

School of Medical Sciences, The University of Sydney

SCDL3991: Science Dalyell Individual Research Project

Dr. Rebekah A. Jenkin and Professor Kevin Keay

15th June, 2024

Abstract

This research project investigates the challenges and rewards of analysing historical handwritten records through an analysis of Anatomy Registers at The University of Sydney. This paper discusses data verification, validation, and transformation techniques applied to improve data quality and prepare the dataset for meaningful analysis. It highlights the methods used for data standardisation and categorisation, and provide insights into societal trends over time through a geographical analysis of the data.

The Challenges and Rewards of Analysing Handwritten Records

The intersection of computational techniques with humanities research has historically encountered both skepticism and gradual acceptance (Jockers, 2013; Pieters, 2013). As early as the 1980s, the legitimacy of computer-based work in humanities was under debate, often seen as too “mechanical” for the exploration of human culture and history (Hockey, 2004; Olsen, 1993). It wasn’t until the turn of the century when advanced processing capabilities, simpler programming interfaces, and greater data availability came into fruition that the potential of “digital humanities” as a field of academia began to be recognised more broadly (Nyhan & Flinn, 2016).

This project is uniquely positioned at such an intersection, specifically intertwining humanities, anatomy, and data science. The broad aim of our project is to attain a better understanding of changes in anatomy body procurement practices during the formation of modern Australia. Central to this are the Anatomy Registers from The University of Sydney School of Medical Science, which chronicle over a century of body acquisition for anatomical study. This extensive dataset not only shows the evolution of medical and scientific practices in the nineteenth and twentieth centuries, but also opens a window to the sociocultural and ethical dimensions of the era. By exploring these registers, we are exploring how societal values, medical ethics, and the law all interact with the scientific pursuit of knowledge.

Background

In the digital era, it might seem that all necessary data for addressing contemporary challenges is readily available in a structured, digital format. In reality, researchers are repeatedly falling back onto the troves of information locked away in historical records to make ground on their inquiries – information which often remains underutilised due to its non-digital or unstructured format (Pieters, 2013).

Sea level records. One domain which has found good use of digitising archival records is the domain of geological sciences, specifically, the field of oceanography. A rise in global sea levels is expected to be one of the most devastating and costly effects of climate change in the next century (Mimura, 2013), and thus, having accurate data on sea levels across the world is essential for scientists to correctly understand the mechanisms involved behind these trends (Marcos et al., 2011). However, historical sea level data – crucial for long-term analyses – often only exists in analogue form, as digital recording only became

prominent in the late 20th century. Consequently, current datasets might be inadequate for centennial-scale forecasting (Talke et al., 2020). Recognising this gap, extensive digitisation efforts have emerged globally, including in Spain (Marcos et al., 2011, 2013, 2021), North America (Talke et al., 2014, 2018, 2020), France (Wöppelmann et al., 2014), and the United Kingdom (Inayatillah et al., 2022). These projects have significantly contributed to databases such as *Système d’Observation du Niveau des Eaux Littorales* (SONEL), which support ongoing oceanographic research (Łyszkowicz & Bernatowicz, 2019).

Ship logbooks. Researchers including Kostas Petrakis have aimed to uncover truths about the transition from sail to steam navigation in the Mediterranean through digitising and analysing ship logbooks from the nineteenth and twentieth centuries. Despite the challenges posed by outdated navigational references and archaic language unfamiliar to contemporary scholars, the records have helped trace out the expansion of maritime routes due to the advent of steam ships, and also the psychological impact on those experiencing these shifts firsthand: *“In the margin of the logbook [the captain] wrote twice about his disappointment and frustration... ‘The screw broke and I don’t know what to do, one difficulty after the other. God help me because I am going to lose my mind!’”* (Petrakis et al., 2021). It would be rare for such emotional insights to be captured through traditional research involving shipping records or calendars.

The above examples highlight the value of the research described in this project: transcribing and analysing historical data helps researchers attain a greater understanding of the physical and sociocultural changes occurring throughout history.

Project Scope

Originally the project aspirations were quite broad. For example, one of the initial aims of the project was to map advances in medical knowledge and diagnosis granularity through analysing changes in cause of death terminology. This was to be conducted through a technique called “dynamic topic modelling” (Blei & Lafferty, 2006), which utilises unsupervised machine learning techniques to extract the n most prevalent topics from a corpus of text. However, the limitations of using hand-transcribed records soon became apparent. For example:

1. No formal data cleaning procedure had been conducted on the dataset;

2. Historical records are usually transcribed with the aim of extracting an accurate account of events or people and not necessarily for the purposes of analysis. Due to this, no *deliberate* alterations were made from the source material during transcription e.g. fixing misspellings;
3. The process of digitising records written in traditional cursive scripts is incredibly laborious and error-prone.

Furthermore, though analysis on the dataset had been conducted before, none focused on geographical location or cause of death and as such these attributes were still in their raw form. The project aim thus shifted to (1) using programming techniques to transform the transcriptions into useable data; and (2) completing a geographical mapping of the place of death of all bodies received.

Methodology

This section will describe the methods used to improve the data quality of the Anatomy Registers dataset. All analysis work done was conducted using the Python programming language, specifically, the data processing library `pandas`. The appendices contain the complete analysis files.

Introduction to the dataset

The Anatomy Registers of The University of Sydney Medical School are a significant data repository that record the sourcing of human bodies for anatomical study from 1883 to present. Detailed records pertaining to the 7609 bodies received between 1883 to 1983 were transcribed into a digital dataset (Jenkin, 2023) and were deidentified for use in this project in accordance with ethics (approval number 2017/898).

The data, when contextualised with historical events, legal changes, and shifts in medical practices, allows for a comprehensive understanding of how societal values and norms influenced medical science. For example, in part due to a “lack of public awareness of the value and importance of donating bodies for dissection” (Jenkin, 2023), there was an extreme shortage of bodies for anatomical dissection during the 1940s and 1950s (Figure 1). Additionally, the dataset reveals a predominant male to female ratio, reflecting the gender disparities in residents of the asylums of the time (Figure 2).

Data verification and validation

Data quality can broadly be determined through three factors: (a) data completeness; (b) data consistency; and (c) data validity.

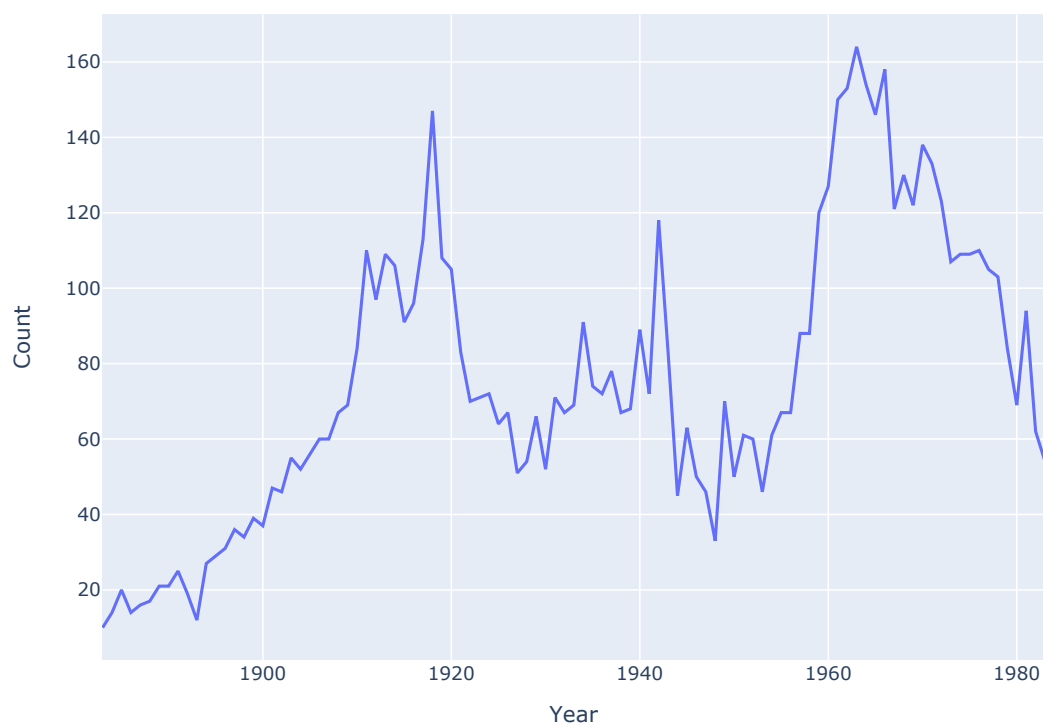
Data completeness

Data completeness refers to the extent that all expected records and attributes are present in a dataset. Initially, several of our attributes contained missing values (see Table 1).

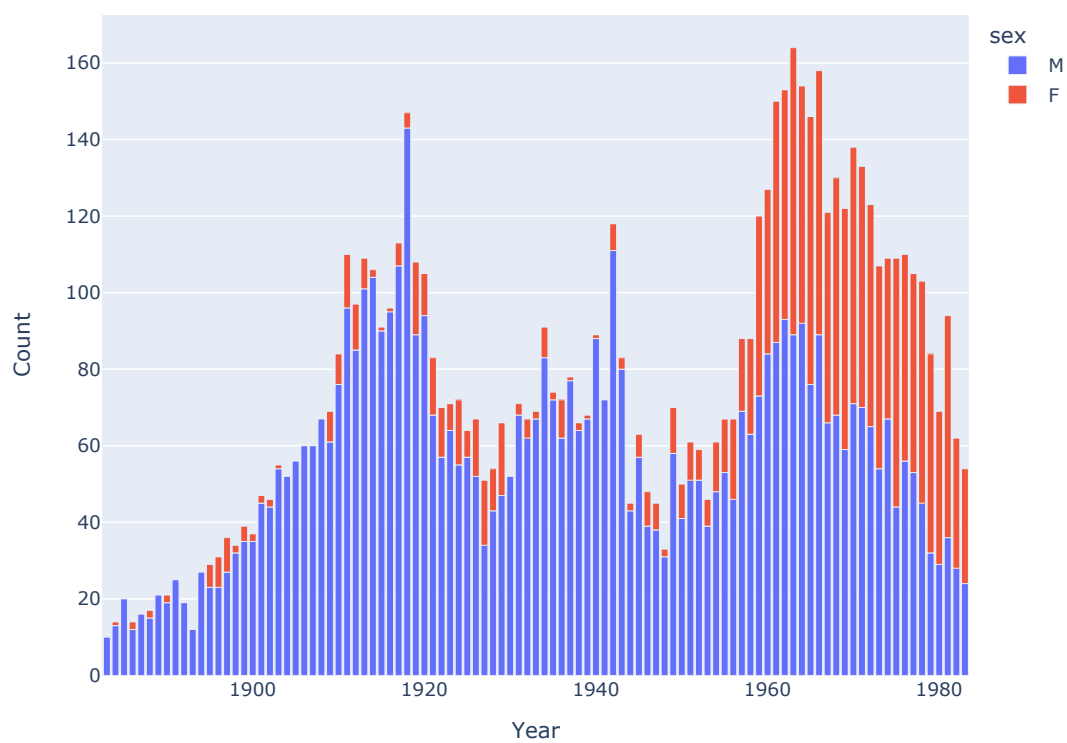
Imputing missing values. Over 2403 entries in the dataset were missing a date of death. However, there were zero missing entries for the date of reception. The date of reception was usually on, or close to (two to three days after) the patient’s death. Therefore, the date of reception values were used to impute the missing values of the date of death attribute.

Figure 1

Number of entries in USYD Anatomy Registers by year

**Figure 2**

Sex of bodies received over time



Removing entries with missing values. 15 cases had missing values for age, sex or place of death. As these characteristics could not be imputed from other information, these entries were removed from the dataset.

Data consistency

Data consistency refers to the dataset adhering to a set of rules and formats. More practically, it involves ensuring each attribute in a dataset fits the most precise datatype it can.

Checking dates. A regular expression (regex) was utilised to check if dates were in the standard DD/MM/YYYY format.

```
^(0?[0-9]|[12]\d|3[01])(0?[0-9]|1[0-2])(000\d|00\d{2}|0\d{3}|1\d{3}|2000)$ (1)
```

Some of the erroneous entries discovered through this method can be seen in Table 2. In total there were 21 such entries, ranging from a missing forward slash (“/”) in the date to the word “missing” or “[blank]” being used to indicate a missing value. These entries were then individually corrected by hand.

Data validity

Data validity refers to the data accurately representing the real-world values it is supposed to depict – i.e. the values in the dataset should make sense. Some of the checks conducted are listed below:

- Checking the range of `death_date` and `reception_date` (should be between 1883 and 1983)
- Checking that `death_date` and `reception_date` were no longer than 10 days apart
- Checking that `death_date` is always less than or equal to `reception_date`

56 invalid entries were found due to the above checks, mostly due to typos when entering dates. Most of these were corrected by double-checking with the original sources; a few were inferred through looking at surrounding entries for the correct values for day, month, or year.

Data transformation

It is difficult to apply traditional verification techniques such as those discussed in the previous section to the cause of death and place of death columns:

1. Cause of death and place of death are unstructured text attributes (unlike year, age, and dates).
2. Legislation mandated the maintenance of cause of death and place of death records but did not specify the level of detail required (see the Anatomy Acts of 1881, 1901, and 1977).
3. Medical advancements over the 20th century led to increasingly precise causes of death, with later entries often including detailed tertiary diagnoses.

These attributes are therefore difficult to validate or analyse in their raw form. Hence, *data transformation* was used in order to make programmatic analysis more feasible. Transforming the place of death attribute was our focus for this semester-long project.

Categorising place of death

Regular expressions and keyword analysis techniques were used to extract ~7000 entries into 9 categories based on their place of death (see Table 3). For example, the following set of rules were developed to extract private residences:

1. The location starts with “flat”, “unit”, “lot”, or contains “at home” e.g. “Flat 73, Darley St, Darlinghurst”
2. The location starts with a digit, and also contains an address suffix e.g. “5 York Rd, Bondi Junction”

These rules were converted into regular expressions, and any entries which matched the expressions were categorised as “private residence.” A similar procedure was then repeated for each other category, such as “private hospital” or “aged care”.

Extracting suburbs from place of death

In order to pursue a geographical analysis of the Anatomy Registers dataset, it was important to extract some form of location data from the place of death attribute. The most straightforward approach was to extract the suburb name, as it commonly appeared in the unstructured entries of both private residences and other categories of institutions.

A secondary dataset was first sourced which contained the names and geographical boundaries of all 4615 localities in NSW (see Tables 4 and 5). Any suburb name present in each place of death entry was then extracted from the records using regular expressions.

Interestingly, 950 entries from the dataset contained more than one suburb; examples of these can be seen in Table 6. This was mostly for the following reasons:

1. An institution contains a suburb in its name but is physically addressed in a different suburb e.g. “Berrima District Hospital, Bowral”;
2. A street shares its name with a suburb e.g. “Clovelly Rd”;
3. A combination of the above, often leading to three extracted suburbs e.g. “Ryde Hospital, Denistone Rd, Eastwood”.

These entries were cleaned programmatically by (a) checking if all classifications of an entry were identical and selecting any one of these classifications, and (b) if the place of death was an address, selecting the suburb which was extracted from the record *last*, as the suburb always comes after the street name and institution name in an address. These steps correctly assigned over 800 of the 950 multiply assigned entries in the dataset.

The remaining entries which could not be condensed through the above procedure were entries with typos, such as “2 balmoral *avemue* croydon park”; conjunct entries, such as “hornsby hospital *from* oatlands nursing home dundas”; or institutions which had multiple suburbs in their names, such as “*ryde* district soldiers memorial hospital *eastwood*”. These last entries were classified manually using a data entry UI created using Jupyter Notebook widgets (see Figure 3).

Finally, there were 478 entries in the dataset that did not contain any suburbs at all. A majority of these were hospital names, such as “callan park mental hospital” or “dalcross private hospital”, and many contained misspelled suburbs which were unable to be picked up through the regex checker, such as “wallandra private hospital *marrickville*” or “outside 1018 francis avenue *brightonlesands*”. These were also manually classified using a modified version of the data entry UI from Figure 3.

After this process, there were only 86 entries in the dataset (1.33%) which did not receive a suburb classification. As seen in Figure 4, about two-thirds of the remaining entries can be attributed to deaths at unspecified morgues, and a large majority of the remaining one-third entries were neonatal deaths which did not routinely include a place of death in the record associated with their receipt by the Medical School. These 86 entries were subsequently dropped from the dataset in order to facilitate an accurate merge with the

Figure 3
Data entry UI made to classify entries manually

Suburb:

Corrected entry:

| | |
|-----------------------|----------------------------------|
| year | 1945 |
| sex | F |
| age | 0 |
| death_date | 1945-04-11 00:00:00 |
| reception_date | 1945-04-11 00:00:00 |
| return_date_sepulture | NaT |
| burial_date | 1945-12-11 00:00:00 |
| death_place | graffton private hospital auburn |
| place_code | 70 |
| death_place_category | aged 2 or under |
| suburb_1 | auburn |
| suburb_2 | auburn |
| suburb_3 | NaN |

Name: 3819, dtype: object

Next to correct: booker bay aberdeen
 Pre-identified as 1. booker bay
 Pre-identified as 2. aberdeen

geographical dataset.

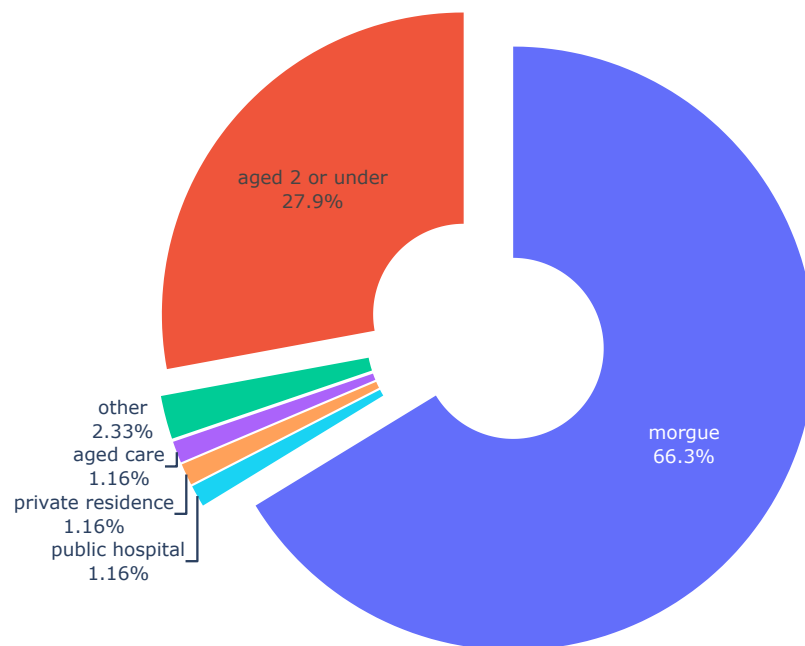
Merging with NSW geographical dataset. When first exploring the geographical dataset, it was noted that there were 202 localities across New South Wales with the same name. This would present a significant issue when matching a suburb from the anatomy records to a suburb in the table. In order to deal with this, the Anatomy Registers dataset was first merged with the NSW Localities dataset using a `LEFT JOIN` in order to retain any duplicates. The dataset was then queried to extract any duplicated suburbs, outputting:

1. Summer Hill (Inner West · Hunter Region)
2. Punchbowl (Canterbury-Bankstown · Northern Rivers Region)
3. Enmore (Inner West · New England Region)
4. Darlington (Inner West · Hunter Region)
5. Budgewoi

After plotting the duplicates on a map, it was found that Summer Hill, Punchbowl, Enmore, and Darlington were all inner city suburbs having rural villages sharing their name. It was confirmed through corroborating their place of death entries that these patients were not sourced from their rural counterparts. Therefore, the duplicate entries closest to Sydney city were chosen to remain in the dataset.

Lastly, Budgewoi is a coastal town which is naturally divided into two by the Budgewoi

Figure 4
Categories of entries with missing suburb classifications



Lake causing the duplicate suburb name. One Budgewoi geometry was chosen at random to remain in the dataset.

Final dataset

Ultimately, the work conducted in this section led to a cleaned, verified, and useable dataset containing 7510 entries (98.7% of the original dataset) with nine original and three new attributes (see Table 7). The quality and utility of the Anatomy Registers dataset was significantly enhanced through employing data verification, validation, and transformation techniques. Furthermore, each step – from ensuring data completeness to extracting geographical information – was not only important in preparing the dataset for detailed analysis, but also served as a valuable case study in techniques that can be employed when handling archival data from other sources.

Results and Discussion

The initial phase of our project was dedicated to cleaning and validating the Anatomy Registers dataset. Given the large volume and arcane lettering of the handwritten records, the dataset was prone to numerous transcription errors. Our systematic approach, involving techniques such as regular expressions, was critical for enhancing the data's reliability and accuracy, making it suitable for further analysis.

Following data validation, we focused on data transformation and feature extraction, particularly targeting the "place of death" attribute. This involved (a) categorising each entry by the type of institution the patient died in (e.g. private residence, public hospital etc.), and (b) extracting the suburb where the patient died in. This involved using text parsing and geographic information system (GIS) techniques, which facilitated the transition from basic descriptive statistics to a deeper geospatial analysis. The refinement of this data allowed us to generate comprehensive choropleth maps, illustrating the distribution and evolution of donor sources across NSW.

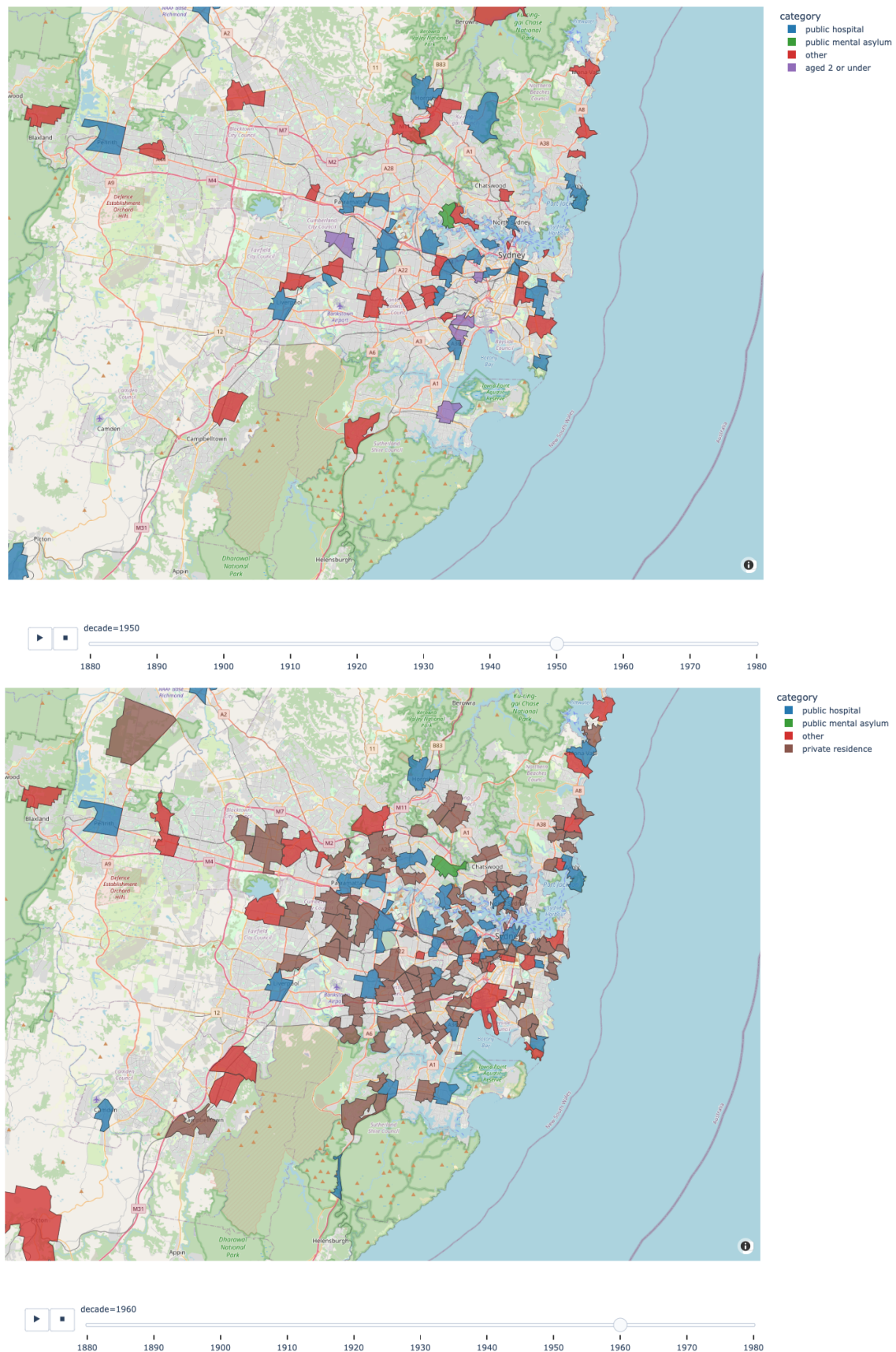
The animated choropleth map (Figure 5), showing the most common category of death place per suburb through each decade from the 1880s to 1980s, revealed a notable shift in the sources of bodies for anatomical study. Prior to 1950, a majority of bodies originated from public and mental asylums, indicating a period when body donation was more stigmatised and the Medical School was heavily reliant on unclaimed bodies from such institutions for anatomical dissection. However, post-1960s, there was a significant increase in donors from private residences, indicating a shift in public perceptions towards body donation as a more accepted and personal choice.

This shift in attitude was coincident with the transition to a body donor program. Voluntary body donation steadily increased from the year 1916, when the first documented body donation was recorded, to the year 1960, where 118 out of 127 bodies admitted were registered or private donors, until "by the late 1960s, a waiting list for would be donors had to be introduced, with restrictions on the registration of new donors sometimes lasting beyond a year" (Jenkin, 2023).

It appears that this trend of the public becoming more conscious over what happens to their body after they pass was not limited to just Australia. Indeed, Canada, the United Kingdom, and the United States also saw the upsurge of voluntary donation in the mid-20th

Figure 5

A choropleth map showing the stark difference between the largely institution-reliant donor program in the 1950s and the influx of private donors in the 1960s.



century. In a similar study to the one conducted at the University of Sydney, anatomists at McGill University in Canada saw the first bequeathments (or “gifts” as they were originally called) of bodies in the 1950s and also witnessed the rate of donations increasing rapidly after the year 1965 (Noël et al., 2022). In the United Kingdom, it was in the late ‘40s that the newly-initiated National Health Service saw a “dramatic” change in public perceptions and saw the beginnings of bodies “provided in a spirit of trust and generosity, by voluntary public donation” (Jones & Fennell, 1991; Richardson, 2006). The United States also shared this trend despite limited ties to the UK and its legislature (Garment et al., 2007).¹

We cannot be certain for the reasons behind this uptrend in body donation during the 1950s and 1960s. This is because the way a society deals with death reflects the way its citizens live, and, as a result, any trends to do with bequeathment programs are a culmination of changes in society itself.

A population boom, expansion of government, new legislation, changes in the population demographics, developments in science, and proliferation of mass media, all affected body acquisition. (Garment et al., 2007)

However, there are a few probable causes towards this trend.

1. **The effects of World War II.** The widespread establishment of blood transfusion services during World War II had a profound influence on public willingness to contribute to medical science, not just through blood donation but also through body donation for dissection. As Richardson notes, the collective effort in wartime medical services likely fostered a sense of communal responsibility towards advancing medical knowledge and healthcare, which persisted into the post-war era (Richardson, 2006). This shift contributed to a greater acceptance of body donation as another form of contribution to the medical field.
2. **Shifts in religious beliefs.** The mid-20th century witnessed significant shifts in societal and religious attitudes towards the human body post-mortem. As traditional views on the spiritual significance of the corpse began to wane, alongside an increase in the popularity of cremation, public attitudes towards scientific medicine and body donation

¹ Canada and Australia's Anatomy Acts both aligned themselves closely to the original British Anatomy Act of 1832.

started to change. Jones and Fennell touch on how these evolving beliefs, coupled with an enhanced understanding of the medical benefits of body donation, led to broader acceptance and support of anatomical donations (Jones & Fennell, 1991).

3. **The cost of dying.** The American funeral industry boomed due to the popularisation of embalming in the mid-19th century, and alongside this, the price of funerals skyrocketed. According to Garment et al., three pieces of journalism were instrumental in bringing to light the excessive costs paid to die in the country: “The High Cost of Dying” by Bill Davidson, “Can You Afford to Die” by Roul Tunley, and “The American Way of Death” by Jessica Mitford, all published in the ‘50s and ‘60s. Mitford’s book contained the first publicly available list of schools that would accept bodies for donation, provided as an alternative to succumbing to exorbitant funeral costs. It is likely that Davidson, Tunley, and Mitford’s works helped increase the public’s awareness of alternatives, a key one being body donor programs.

These insights into historical and societal trends are crucial as they highlight the impact of digitising and analysing historical datasets, echoing the broader themes discussed in our literature review. Just as the analysis of ship logbooks (Petrakis et al., 2021) provided a window into the psychological landscape of historical maritime transitions, our project shows the evolution of societal attitudes towards a once-stigmatised aspect in medical science.

Limitations

Despite our achievements, the project faced certain limitations, particularly in the categorisation of the ‘place of death’ entries. A significant number of entries remained classified under “Other”, including those with misspelled addresses and ambiguous locations, indicating the complexity of interpreting such historical data accurately. This limitation underscores the need for more advanced natural language processing techniques and machine learning models to handle the intricacies of data entry. Additionally, manual review and enhancement of the categorisation process would likely improve data accuracy and the quality of insights derived from our analysis.

Conclusion

The exploration of the Anatomy Registers from The University of Sydney Medical School has proven both challenging and enlightening. This project illuminated changes in historical practices and societal attitudes towards body donation, while reinforcing the value of digitising and analyzing historical datasets.

Working with manually transcribed data introduces significant complexities, primarily due to the extensive data cleaning and validation required to ensure analysis reliability. Transforming the "place of death" data into a structured format enabled us to identify geographical patterns and temporal changes in body donation sources. Our geospatial analyses, particularly the animated choropleth map, visually narrated the evolution of societal norms from a reliance on unclaimed bodies from public institutions to an increase in donations from private residences, highlighting a significant shift in public perception towards body donation.

Nevertheless, the project encountered some limitations, especially in data categorisation. The large number of entries labeled as "Other" highlights the potential for enhancements in methodology through employing advanced technologies like machine learning and natural language processing.

Future endeavors will likely focus on further refining data standardization processes and exploring our initial research questions, such as using the cause of death attribute to trace the emergence of new medical fields like psychiatry through dynamic topic modeling. This could potentially reveal historical trends in medical terminology and diagnostics.

In conclusion, while this project has provided valuable insights and advanced the field of digital humanities in relation to anatomical history, it also underscores the continuous need for innovation and reflection in the methodologies we employ to examine our past so that we may learn from it.

Table 1*Missing values for attributes in the Anatomy Registers dataset before and after data cleaning*

| Attribute | NULL entries before cleaning | NULL entries after cleaning |
|-----------------------|------------------------------|-----------------------------|
| year | 0 | 0 |
| id | 0 | 0 |
| sex | 5 | 0 |
| age | 10 | 0 |
| death_date | 2436 | 0 |
| reception_date | 0 | 0 |
| return_date_sepulture | 1680 | 1673 |
| burial_date | 4408 | 4394 |
| death_place | 6 | 0 |
| place_code | 0 | 0 |

Table 2*A few erroneous date entries from each date column*

| id | death_date | reception_date | return_date_sepulture | burial_date |
|------|------------|----------------|-----------------------|-------------|
| 1898 | 25/1/2020 | ... | ... | ... |
| 5014 | 26/87/1961 | ... | ... | ... |
| 5044 | 20/11/9161 | ... | ... | ... |
| 7116 | 5/101977 | ... | ... | ... |
| 1727 | ... | missing | ... | ... |
| 2408 | ... | 23 /4/1925 | ... | ... |
| 2644 | ... | ... | crossed out 27/9/1929 | ... |
| 3184 | ... | ... | [blank] | ... |
| 3185 | ... | ... | [blank] | ... |
| 3203 | ... | ... | [blank] | ... |
| 6726 | ... | ... | ... | 04/04//1975 |

Table 3*Categories extracted from the place of death attribute in the Anatomy Registers*

| Category | Classification criteria |
|----------------------|--|
| Private residence | A private address (commencing with a street, flat, or unit number) |
| Aged care | Locations containing "Nursing", "Convalescent", "Aged Care", ... |
| Hospice | Locations containing "Hospice" + some known hospices |
| Private hospital | Locations containing "Private Hospital" or "surgery" |
| Public hospital | All hospitals without "private" in their name |
| Public asylum | Macquarie St, Liverpool, George St, Rockwood |
| Public mental asylum | Newington + locations containing "mental asylum" or "psychiatric" |
| Aged 2 and under | Currently coded 70 |
| Morgues | Locations containing "morgue" |
| Other | Other |

Table 4*Attribution details of the NSW Localities dataset*

| Attribute | Details |
|--------------|---|
| Dataset name | NSW Local Government Areas - Geoscape Administrative Boundaries |
| Source | Data originally found on data.gov.au. Link here: https://shorturl.at/DMbhb |
| Attribution | Licensed by the Commonwealth of Australia under Creative Commons Attribution 4.0 International licence (CC BY 4.0). |
| Last updated | 27/02/2024 |

Table 5*First few rows of the NSW Localities dataset*

| suburb | geometry |
|-------------------|---------------|
| Aarons Pass | POLYGON (...) |
| Abbotsbury | POLYGON (...) |
| Abbotsford | POLYGON (...) |
| Abercrombie | POLYGON (...) |
| Abercrombie River | POLYGON (...) |
| ... | ... |

Table 6*A select few examples of place of death entries containing multiple suburbs*

| id | death_place | suburb_1 | suburb_2 | suburb_3 |
|------|---|-----------|--------------|----------|
| 5122 | Corner Clovelly Rd & Glebe Rd Randwick | Clovelly | Glebe | Randwick |
| 5533 | Berrima District Hospital, Bowral | Berrima | Bowral | - |
| 7567 | Ryde Hospital, Denistone Rd, Eastwood | Ryde | Denistone | Eastwood |
| 7596 | St George Nursing Home, 3 Verdun St, Bexley | St George | Bexley | - |
| 7598 | Kilbride Nursing Home, Appin Rd, Campbelltown | Appin | Campbelltown | - |

Table 7*Schema of the Anatomy Registers dataset (* indicates a column not present in the original dataset)*

| Attribute Name | Data Type | Missing | Description |
|-----------------------|-----------|---------|--------------------------------------|
| id | Integer | 0 | Unique identifier for each entry |
| year | Integer | 0 | Year the body was received |
| age | Integer | 0 | Age at the time of death |
| death_date | Date | 0 | Date of the donor's death |
| reception_date | Date | 1 | Date body was received at the school |
| return_date_sepulture | Date | 1645 | Date body was returned for burial |
| burial_date | Date | 4353 | Date of the body's burial |
| death_place | String | 0 | Location of death |
| place_code | String | 0 | Coded place of death |
| death_place_category* | String | 0 | Categorised place of death |
| suburb* | String | 0 | Suburb extracted from death place |
| geometry* | Geometry | 0 | Geospatial data for location |

References

- Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. *Proceedings of the 23rd International Conference on Machine Learning*, 113–120.
- Garment, A., Lederer, S., Rogers, N., & Boulton, L. (2007). Let the dead teach the living: the rise of body bequeathal in 20th-century America. *Academic medicine*, 82(10), 1000–1005.
- Hockey, S. (2004). The History of Humanities Computing. In *A companion to digital humanities*. John Wiley & Sons, Incorporated.
- Inayatillah, A., Haigh, I. D., Brand, J. H., Francis, K., Mortley, A., Durrant, M., Fantuzzi, L., Palmer, E., Miller, C., & Hogarth, P. (2022). Digitising historical sea level records in the Thames Estuary, UK. *Scientific Data*, 9(1).
- Jenkin, R. (2023). *Altruism in Death. Historic and contemporary use of mortal remains in anatomical examination for education and research in Australia and New Zealand* [Doctoral dissertation, The University of Sydney].
- Jockers, M. L. (2013). Part I Foundation. In *Macroanalysis : Digital Methods and Literary History* (pp. 1–32). University of Illinois Press.
- Jones, D., & Fennell, S. (1991). Bequests, cadavers and dissections: sketches from New Zealand history. *The New Zealand medical journal*, 104, 210–2.
- Łyszkowicz, A., & Bernatowicz, A. (2019). Geocentric Baltic Sea level changes along the southern coastline. *Advances in Space Research*, 64(9), 1807–1815.
- Marcos, M., Puyol, B., Amores, A., Gómez, B. P., Fraile, M. Á., & Talke, S. A. (2021). Historical tide gauge sea-level observations in Alicante and Santander (Spain) since the 19th century. *Geoscience Data Journal*, 8(2), 144–153.
- Marcos, M., Puyol, B., Calafat, F. M., & Wöppelmann, G. (2013). Sea level changes at Tenerife Island (NE Tropical Atlantic) since 1927. *Journal of Geophysical Research: Oceans*, 118(10), 4899–4910.
- Marcos, M., Puyol, B., Wöppelmann, G., Herrero, C., & García-Fernández, M. J. (2011). The long sea level record at Cadiz (southern Spain) from 1880 to 2009. *Journal of Geophysical Research*, 116(C12).
- Mimura, N. (2013). Sea-level rise caused by climate change and its implications for society. *Proceedings of the Japan Academy, Series B, Physical and biological sciences*, 89(7), 281–301.

- New South Wales Government. (1881). *Anatomy Act of 1881*.
- New South Wales Government. (1901). *Anatomy Act, 1901*.
- New South Wales Government. (1977). *Anatomy Act, 1977*.
- Noël, G. P., Heron, J., & Naismith, C. (2022). From 1883 to 2019; Variables Influencing Body Procurement at McGill University. *Anatomical Sciences Education*, 15, 599–608.
- Nyhan, J., & Flinn, A. (2016). Introduction. In *Computation and The Humanities: Towards An oral History of Digital Humanities* (pp. 1–16). Springer Open.
- Olsen, M. (1993). Signs, Symbols, and Discourses: A New Direction for Computer-Aided Literature Studies. *Computers and the Humanities*, 27(5/6), 309–314.
- Petrakis, K., Samaritakis, G., Kalesios, T., Domingo, E. G., Delis, A., Tzitzikas, Y., & Martin Doerr, P. F. (2021). Digitizing, Curating and Visualizing Archival Sources of Maritime History: the case of ship logbooks of the nineteenth and twentieth centuries. *Drassana*, 28.
- Pieters, T. (2013). Big Data for Global History: The Transformative Promise of Digital Humanities. *BMGN*, 128, 55–77.
- Richardson, R. (2006). Human dissection and organ donation: A historical and social background. *Mortality*, 11(2), 151–165.
- Talke, S. A., Kemp, A. C., & Woodruff, J. (2018). Relative sea level, tides, and extreme water levels in Boston harbour from 1825 to 2018. *Journal of Geophysical Research: Oceans*, 123(6), 3895–3914.
- Talke, S. A., Mahedy, A., Jay, D. A., Lau, P., Hilley, C., & Hudson, A. (2020). Sea level, tidal and river flow trends in the Lower Columbia River Estuary, 1853-present. *Journal of Geophysical Research-Oceans*, 125(3).
- Talke, S. A., Orton, P. M., & Jay, D. A. (2014). Increasing storm tides in New York Harbor, 1844–2013. *Geophysical Research Letters*, 41(9).
- Wöppelmann, G., Marcos, M., Coulomb, A., Míguez, B. M., Bonnetain, P., Boucher, C., Gravelle, M., Simon, B., & Tiphaneau, P. (2014). Rescue of the historical sea level record of Marseille (France) from 1885 to 1988 and its extension back to 1849–1851. *Journal of Geodesy*, 88(9), 869–885.

Appendix

All the code used to clean and analyse the Anatomy Registers dataset can be found at <https://github.com/antrikshdhand/SCDL3991-research>.