Report: Statistical Analysis and the Leeds General Cemetery Dataset

Imogen Gerard and Kelsie Root

Living with Dying: Everyday Cultures of Dying within Family Life in Britain, 1900-50s is a University of Leeds research project, supported by The Arts and Humanities Research Council and led by Dr Laura King.

One of the project's partnerships is a collaboration with Special Collections in the Brotherton Library, University of Leeds. Student interns Imogen Gerard and Kelsie Root are spending the summer of 2017 researching the Leeds General Cemetery collection, held by Special Collections. They are working closely with the cemetery's digitised burial records and producing resources for researchers.

Imogen graduated from University of Leeds in 2017 having studied English Literature. Kelsie also graduated from University of Leeds in 2017 with a joint honours degree in History of Art and History and Philosophy of Science. Kelsie is studying for a Masters in History of Health Medicine and Society at the University of Leeds for the academic year 2017-18.

In July 2017 Kelsie and Imogen attended a one week training course in statistics run by Q Step, a programme designed to improve quantitative data analysis in the social sciences. They applied what they learnt in the training to analyse the dataset which underpins the Leeds General Cemetery Burial Registers Index. This data consists of the transcribed and digitised burial records of the people interred at the Leeds General Cemetery. The following is a report on the research potential of using statistics to analyse this data.

Contents

Objectives

The purpose of this report is to investigate how far statistical analysis can aid interpretation of the Leeds General Cemetery (LGC). Having completed a week's training in statistical analysis, we will apply what we have learnt and use it to analyse the LGC dataset. We are interested in what statistical analysis can, and cannot, show us about this data.

The Data

- The Leeds General Cemetery Company was set up in 1833 and led to the establishment of the Leeds General Cemetery, which was open for burials between 1835 and 1969. The company's papers and burial records make up the Leeds General Cemetery Company Ltd Archive, which is held by Special Collections in the Leeds University Library.
- The 25 burial registers in the collections were transcribed and digitised in 2016. Detailed records of the people interred at the LGC can be searched and accessed online through the Leeds General Cemetery Burial Registers Index.
- Special Collections state on their website that researchers may request access the whole dataset by contacting Special Collections.
- Special Collections have provided us with a copy of the dataset in a Microsoft Excel spreadsheet.
- There are 97,122 observations. This is the number of entries the dataset contains about people who have been buried or cremated at LGC.
- This data has 18 variables. These are the different fields containing information about the deceased (i.e. name, age at death, cause of death etc.)
- We have written a Data Dictionary or guide, to supplement the dataset which provides explanatory descriptions of the field titles of the different variables used in the dataset as it exists in spreadsheet. This will be included in the report's appendix.
- We will use the software RStudio to handle the data and perform statistical analysis. Where appropriate, we will include the scripts we used in R in the footnotes of this report to aid any future researchers who may wish to replicate our actions.

Research Questions

In this report we will discuss and investigate trends which occur in LGC dataset under the following research questions:

- How might missing or unknown values affect our analysis?
- How can we measure, infer or observe the social class of people buried at the cemetery?
- How does a person's class, occupation or income affect the location of their grave within the cemetery?
- What are the most common causes of death for people of certain occupations?
- How can the data be used to evidence changes in medical terminology?
- How does age at death vary over time?

Points of Consideration

Limitations

Our data is largely categorical and the entries consist of words rather than numerical data. This limits the functions we can perform and the appropriate outputs for displaying the data. In particular, categorical data tends to be visually displayed using tables, bar charts and histograms.

Existing Statistical Charts

The online LGC Data Index already provides a range of charts to display the data's statistics. We will make use of these charts within this report, using them in conjunction with charts generated using R.

How might missing or unknown values affect our analysis?

There are many missing or unrecorded values within the dataset which have been marked as '[unknown]'. These indicate either information that was unavailable or entries that were too incomprehensible to transcribe.

After around 1940 a large proportion of the variables are marked as unknown. In these later entries information about cause of death, occupation and the deceased's parents is unrecorded. For this reason we chose to focus our analysis on the earlier part of the cemetery's life.

There are also a number of infant deaths in the records which provide no information about the families. We speculate that these instances may be the burials of infants who died at the Leeds Maternity Hospital. Perhaps records kept about infants who came to the cemetery from the hospital were not required to be detailed. It was also the case that stillborn deaths were not required to be officially registered until 1926,[1] and the burial and naming of a stillborn child could be a controversial subject, particularly in terms of religious observance. Thus, it may be the case that women preferred not to share these details with hospital staff and instead kept their mourning a private matter.

Infants who passed away at this hospital tended to be buried at the LGC.[2] This has an impact on the dataset. These records will skew the overall pattern in age at death and cause of death by increasing the representation of stillbirths and infant deaths. However, infant mortality is understood to have been common at this time so representing these deaths in the data is also important.

The software R can be used to summarise how much of the data is missing.[3] When we carried out this command R returned the following information:

```
> percentmissing
LeeAltReferenceCode_tab   EADLocationOfOriginals           EADUnitTitle        CreDateCreated           EADUnitDate
              0.6764757                0.7619361              0.0473636             8.3957126             0.2429959
    EADBiographyOrHistory        EADDescriptiveRules      EADPhysicalTechnical    EADScopeAndContent            Occupation
              0.9884577                0.9812502             14.3243995            17.3989148            16.7522987
        Parent Occupation          EADOtherName_tab            EADName_tab
             32.5315843               22.4369601             19.3562669
```

This tells us the percentages of missing or unknown values in each field. The same information is displayed in the following table:

---

[1] See: 'Stillbirths', Scottish Way of Birth and Death, University of Glasgow [http://www.gla.ac.uk/schools/socialpolitical/research/economicsocialhistory/historymedicine/scottishwayof birthanddeath/death/stillbirths/, accessed 11 Aug 2017].

[2] Causes of Death, Leeds Maternity Hospital, Special Collections, [https://library.leeds.ac.uk/special-collections/collection/730 , accessed 11 Mar 2017].

[3] To do this using RStudio, we had to load the following packages: library(mice), library(BaylorEdPsych), library(mvnmle), library(VIM). We recoded all the entries with '[unknown]' values as 'na'. We then used the following script. The first line creates a function to calculate the percentage of missing values. The second line executes the function for the dataset which is named 'data':
pMiss <- function(x){sum(is.na(x))/length(x)*100}
apply(data,2,pMiss)

| Dataset Variable	Field	Percent of missing / unknown values (%)
LeeAltReferenceCode_tab	Burial Number	0.68
EADLocationOfOriginals	Plot Number	0.76
EADUnitTitle	Name	0.05
CreDateCreated	Date of death	8.40
EADUnitDate	Date of Burial	0.24
EADBiographyOrHistory	Age	0.99
EADDescriptiveRules	Decimal Age	0.98
EADPhysicalTechnical	Gender	14.32
EADSScopeAndContent	Cause of Death	17.40
Occupation	Occupation	16.75
Parent Occupation	Parent Occupation	32.53
EADOtherName_tab	Father's name	22.44
EADName_tab	Mother's name	19.36

This table shows that information that was essential in keeping the burials records of LGC - such as burial number, plot number, date of burial and name of the deceased - is consistently recorded with less than 1% of those values missing. Additional information, particularly concerning the deceased's parents, has larger proportions of missing data.

How can we measure, infer or observe the social class of people buried at the cemetery?

We are interested in inferring the social class of the people in the registers in order to investigate how this might affect their deaths in terms of circumstance, cause and burial. The most useful variable within this dataset for indicating class is occupation. The original burial registers also record the deceased's address. This could also be used as a potential indicator of wealth or class but, as address has not been transcribed or digitised in the dataset, we will be concentrating on occupation.

This chart from the LGC Burial Registers Index displays the most frequently appearing occupations or roles.

Top 10: Occupation or role View details
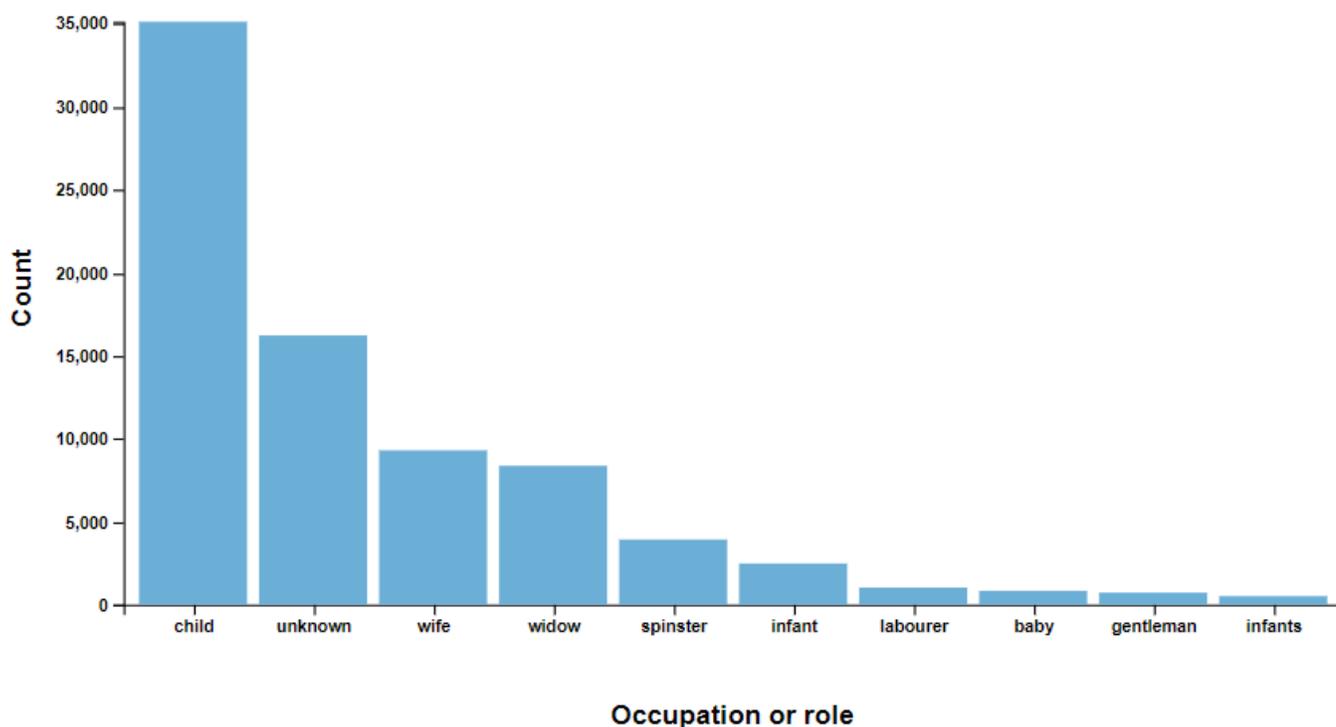
Some women in the burial registers had occupations recorded in the 'Rank, Trade, or Profession Field'; for example, servant, barmaid, actress, teacher. However, for the most part women were simply recorded as either wife, widow or spinster. Those in the register under 16 are generally recorded as either infants or children. Because of this, these generalised descriptions of the roles of children and women dominate the most frequently appearing occupations or roles in the registers. This is reflected in the Top 10 occupations or roles.

As the vast majority of women in the burial registers are recorded as either wife, widow or spinster, it is difficult to directly observe an indicator of class for women. For this reason our focus on occupation will lead to a focus on men.

As the chart shows, 'Labourer' and 'Gentleman' are two of the occupations that appear in the top 10 Occupations or roles. These occupations are helpful indicators of whether a given individual is working class or upper class. For this reason, we have focused on

these two groups – the gentlemen and the labourers buried in the cemetery – and will use these groups to make suggestions about the working class and the upper class.

How does a person's class, occupation or income affect the location of their grave within the cemetery?

Traditionally certain areas of a cemetery are more expensive than others. For this reason upper-class people and lower-class people tend to be buried in different places. We are interested in identifying which locations within the LGC were occupied by the upper class, and which areas were occupied by the lower class. This information will indicate which areas of the cemetery had more expensive burial plots and which areas cost less.

The LGC archival collection suggests that it was agreed that certain parts of the cemetery would be less expensive and were therefore designated for poorer people. The cemetery's chaplain, Reverend James Rawson, wrote to the committee in 1842 requesting the southeast corner of the cemetery be set aside solely for low cost burials of poorer residents.[4] By carrying out this analysis, we can investigate whether the intentions of the original LGC committee were honoured by subsequent committees.

Using R, we created a series of histograms showing the plot numbers of gentlemen and labourers buried at LGC. We did this in order to see whether or not the charts would corroborate the archive's suggestion that certain areas were set aside for the burials of the poorer residents of Leeds.

This histogram shows the plot numbers of the gentlemen buried in the cemetery during the first half of its life, before 1900.

---

[4] Leeds General Cemetery Collection, Special Collections MS421/121: https://library.leeds.ac.uk/special-collections-explore/482807/corporate_administrative_and_shareholding.

Plot Numbers of Gentlemen Buried from 1835-1899
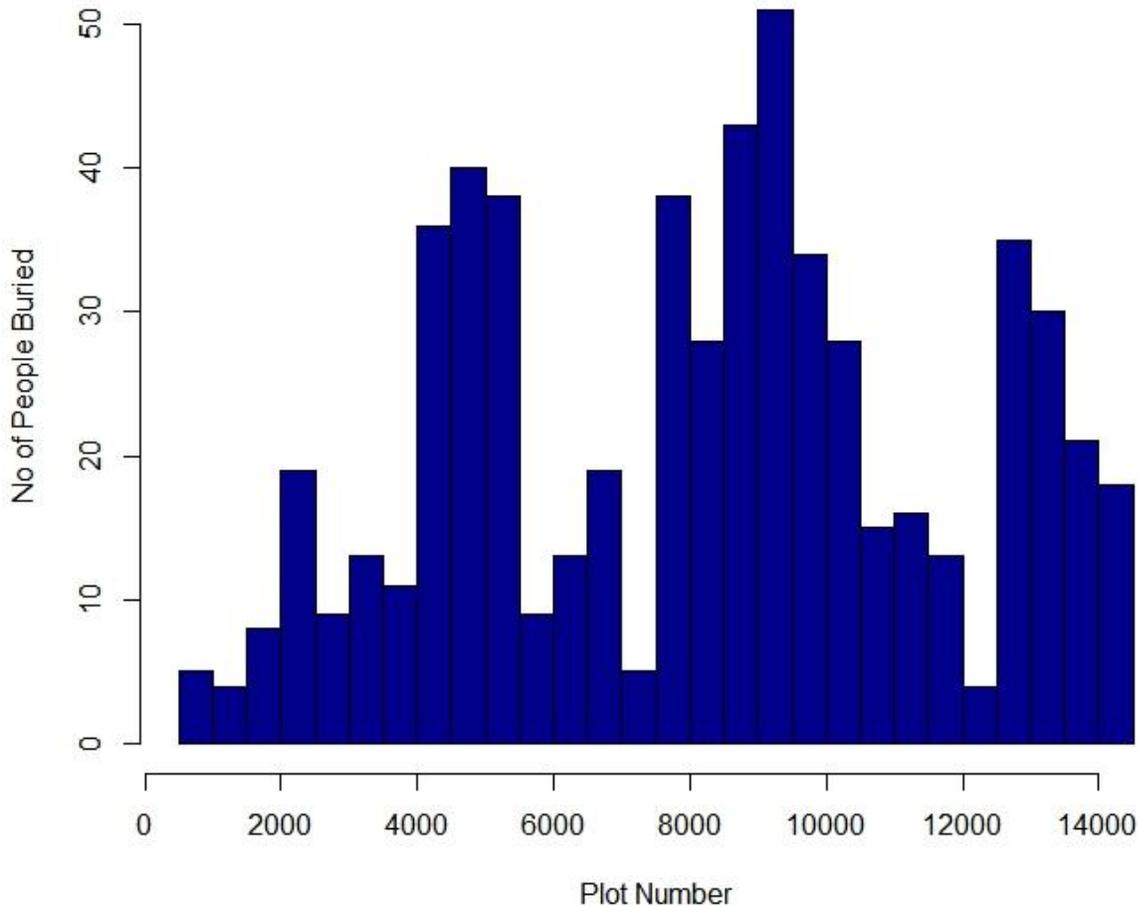
The histogram shows that many gentlemen were buried in certain areas of the cemetery, while small numbers of gentlemen occupied other areas. This pattern, with high peaks and low peaks, suggests that it was common for gentlemen to be grouped together, most likely in the more expensive locations within the cemetery. A particularly high peak in the histogram is around the plot numbers 8500-9500, showing a large cluster of gentlemen buried in proximity to each other. This means that the area of the cemetery covered by plot numbers 8500-9500 was highly concentrated with the graves of gentlemen.

The following histogram shows the plot numbers of labourers buried in cemetery before 1900.

Plot Numbers of Labourers Buried from 1835-1899
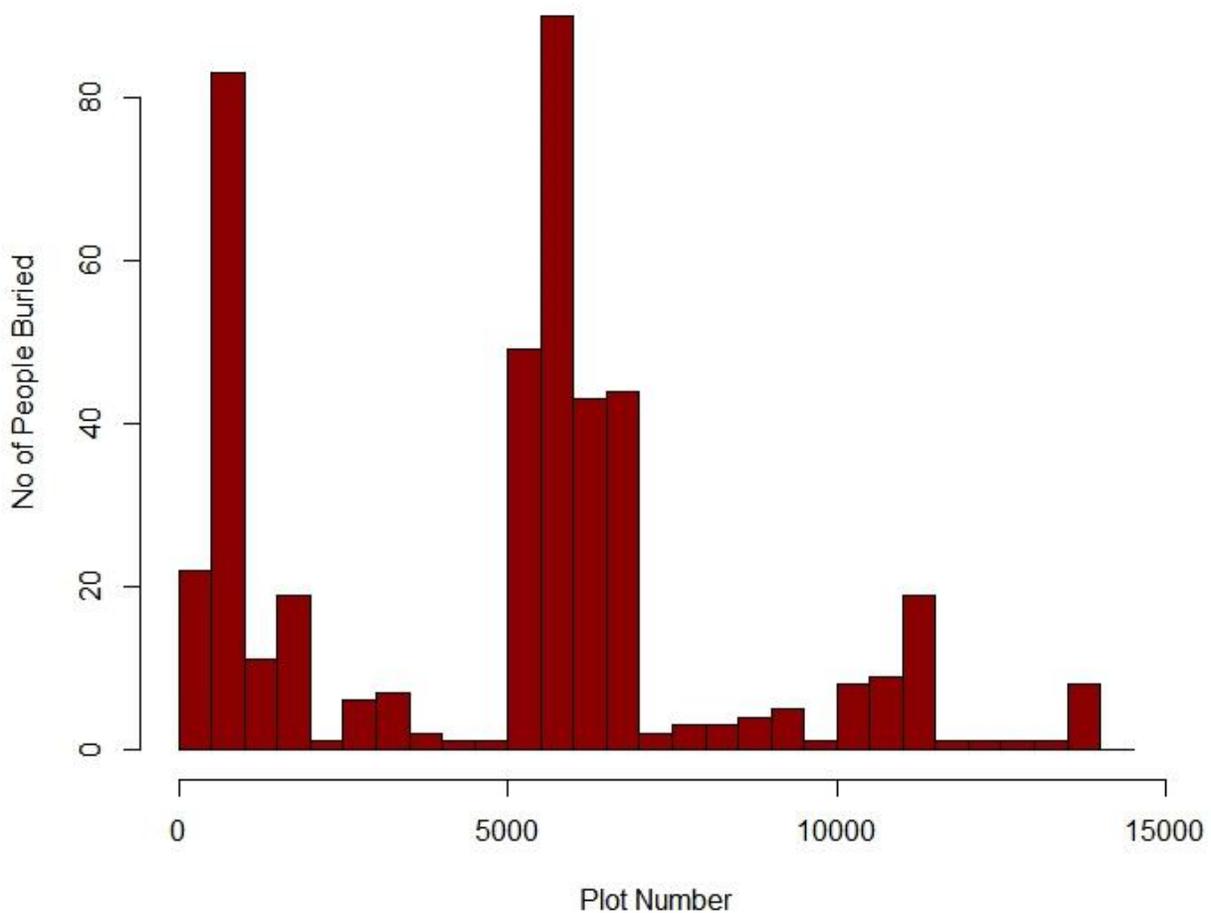
This histogram shows that labourers tended to be buried in the same areas as each other – represented by the chart's high peaks – and were only buried in other areas in very low numbers. Plot numbers 500-1000 and 5500-6000 are highly concentrated with labourers.
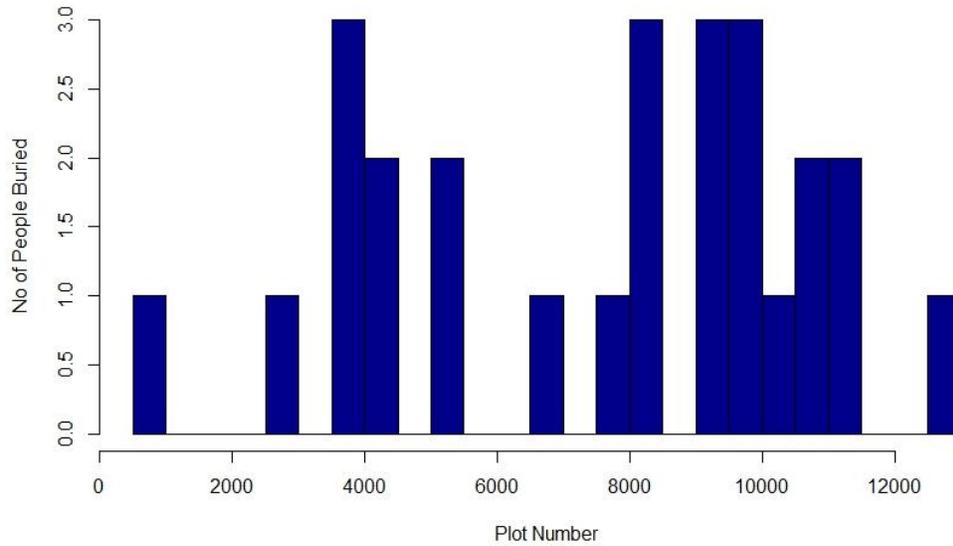
These two histograms show that the plot numbers of gentlemen and labourers buried in the LGC follow a distinctly different pattern to each other. The plot numbers with high numbers of gentlemen do not correspond with the plot numbers of high numbers of labourers.

We repeated this exercise looking at shorter time periods. We did this in order to check whether or not the differentiation between gentlemen and labourers in the pattern of their burial locations would still be evident within a shorter time frame. This was to rule
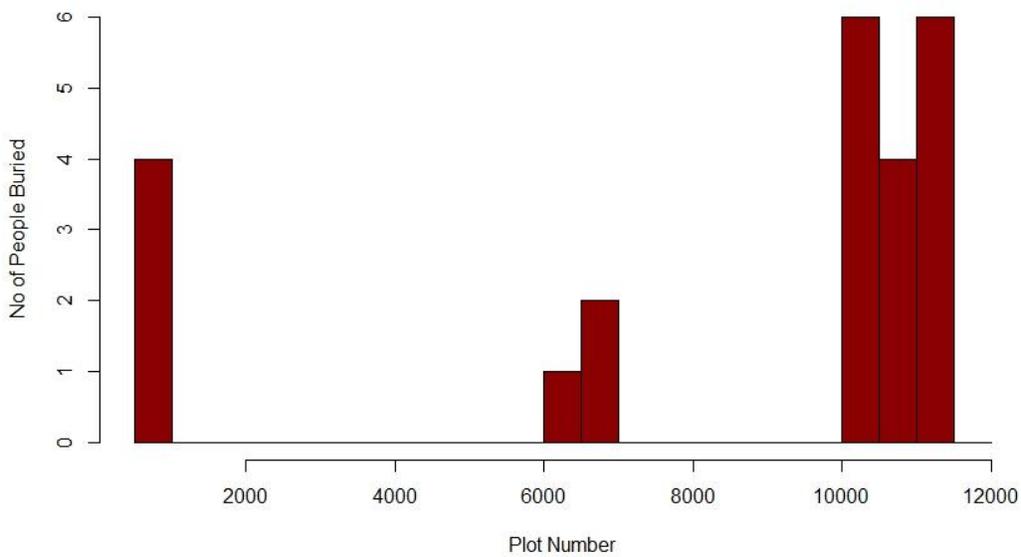
out the possibility that the distinction was only being caused by the different groups being buried at different times.

These two histograms show the plot numbers of gentlemen and labourers buried in the LGC between 1845 and 1850:

Plot Numbers of Gentlemen Buried Between 1845-50

Plot Numbers of Labourers Buried Between 1845-50

These two histograms show the plot numbers of gentlemen and labourers buried in the LGC between 1875 and 1880:

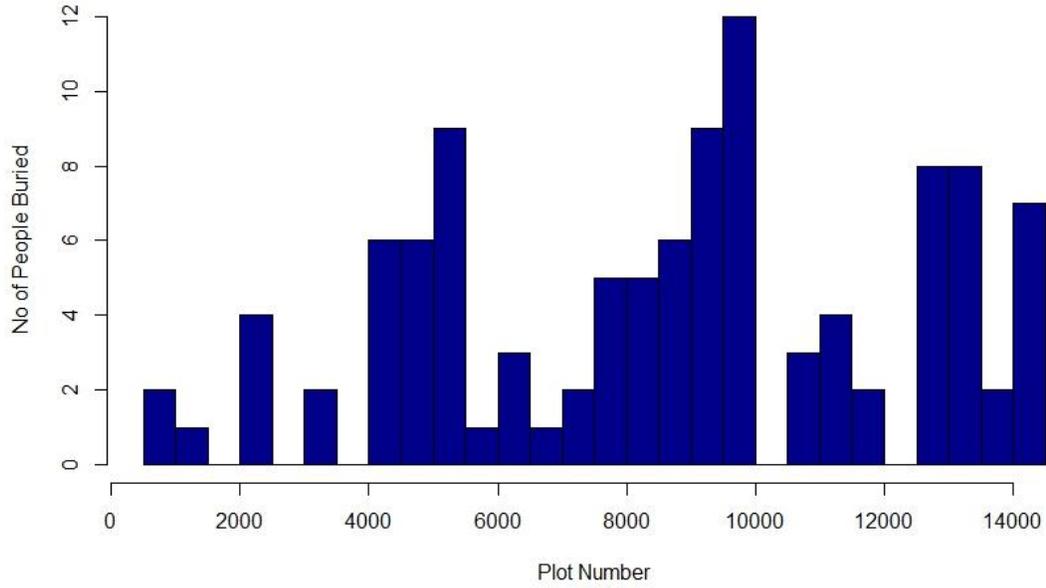**Plot Numbers of Gentlemen Buried Between 1875-80**
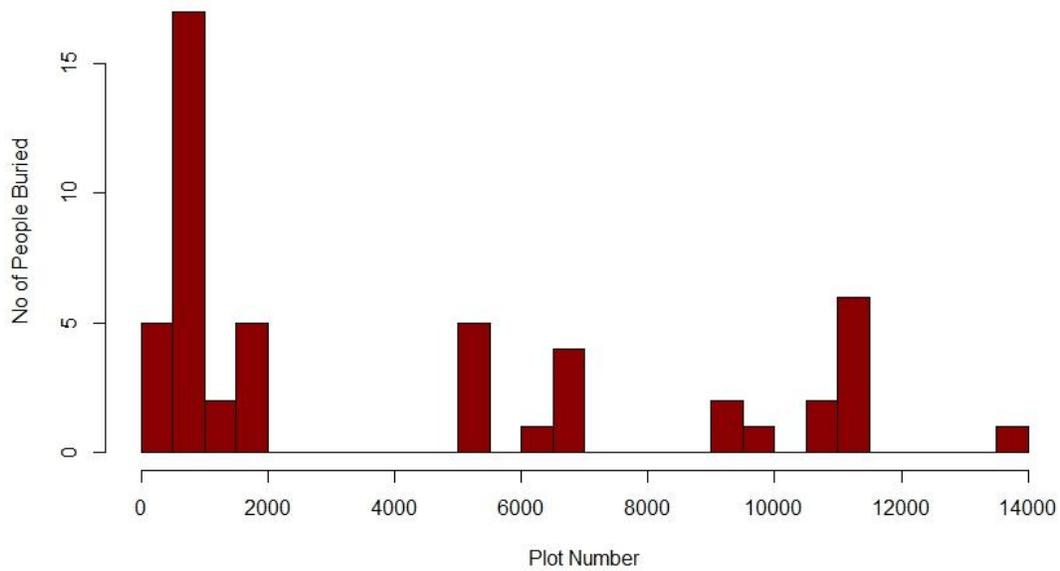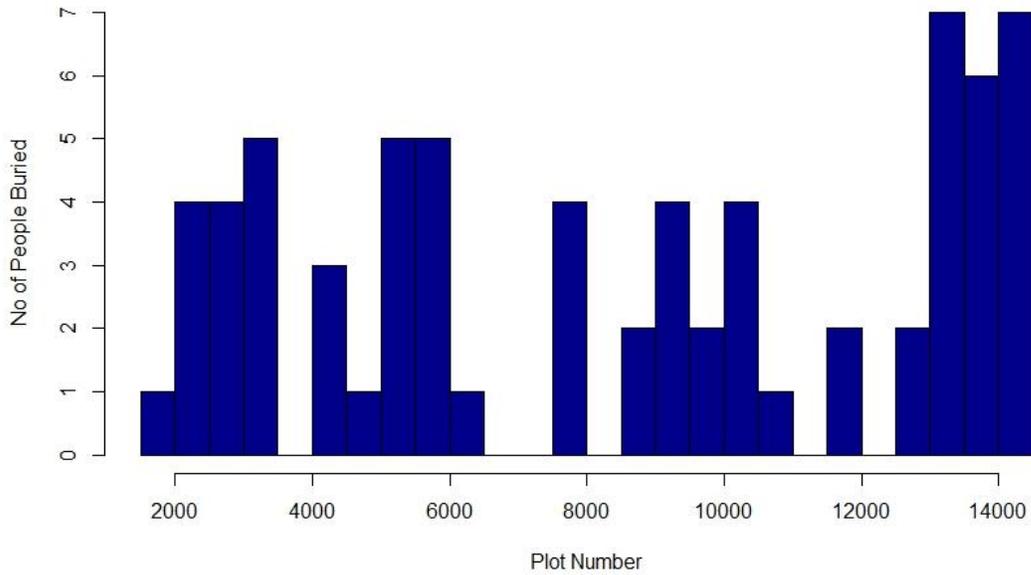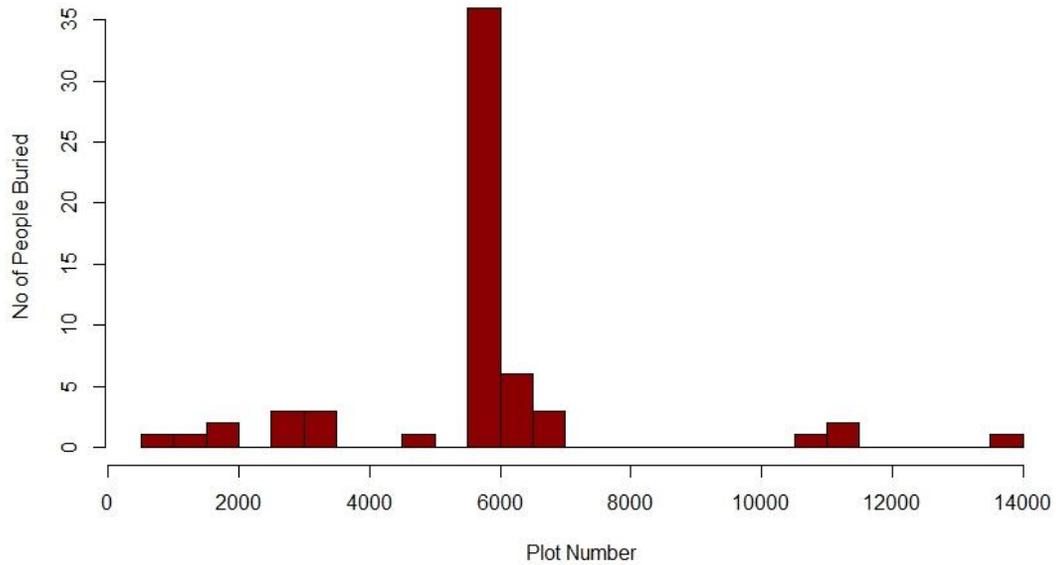

Plot Numbers of Labourers Buried Between 1875-80

These two histograms show the plot numbers of gentlemen and labourers buried in the LGC between 1895 and 1899:

Plot Numbers of Gentlemen Buried Between 1895-99

Plot Numbers of Labourers Buried Between 1895-99

This series of histograms shows that the gentlemen and the labourers occupied different plot numbers within each time frame. The charts show that even within shorter time periods there was little to no overlap in the plot number brackets of gentlemen and labourers. Throughout this period the gentlemen and the labourers were consistently buried in separate areas to each other.

We marked out the areas which are highly concentrated with gentlemen and labourers on a copy of the burial plot map of the cemetery.

This image shows the plot numbers 8500-9500 highlighted on the map. This is the area of the cemetery with the most gentlemen buried in it. (This was the highest peak on the histogram showing the plot numbers of the gentlemen buried in the cemetery during the first half of its life, before 1900, found on page 8).
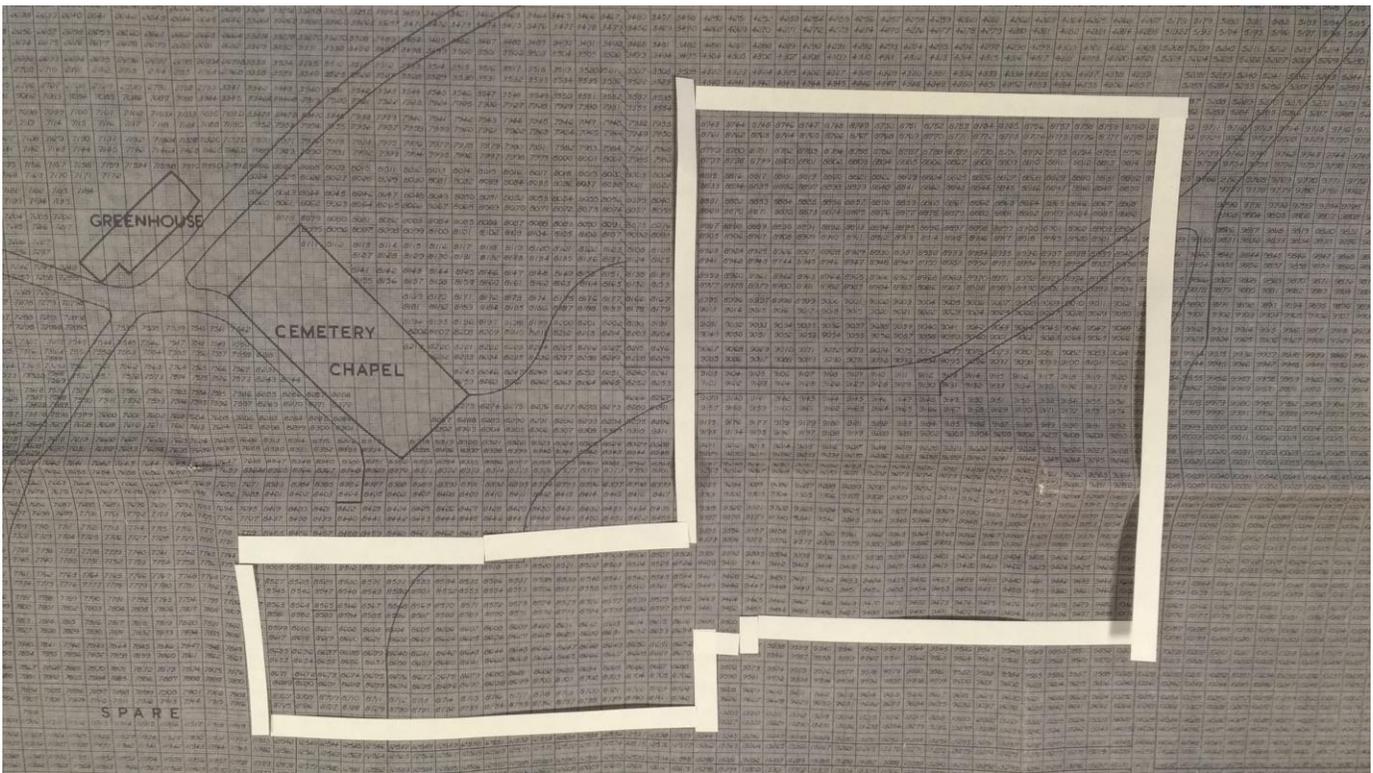
Reproduced with the permission of Special Collections, Leeds University Library. Item reference: MS 421/3/7, Burial Plot Map of the Leeds General Cemetery site: https://library.leeds.ac.uk/special-collections-explore/484414

The following image shows the plot numbers 500-1000 and 5500-6000 highlighted on the burial plot map of the cemetery. These are two areas that are highly concentrated with labourers. (These were the highest peaks on the histogram showing the plot numbers of the labourers buried in the cemetery during the first half of its life, before 1900, found on page 9.)
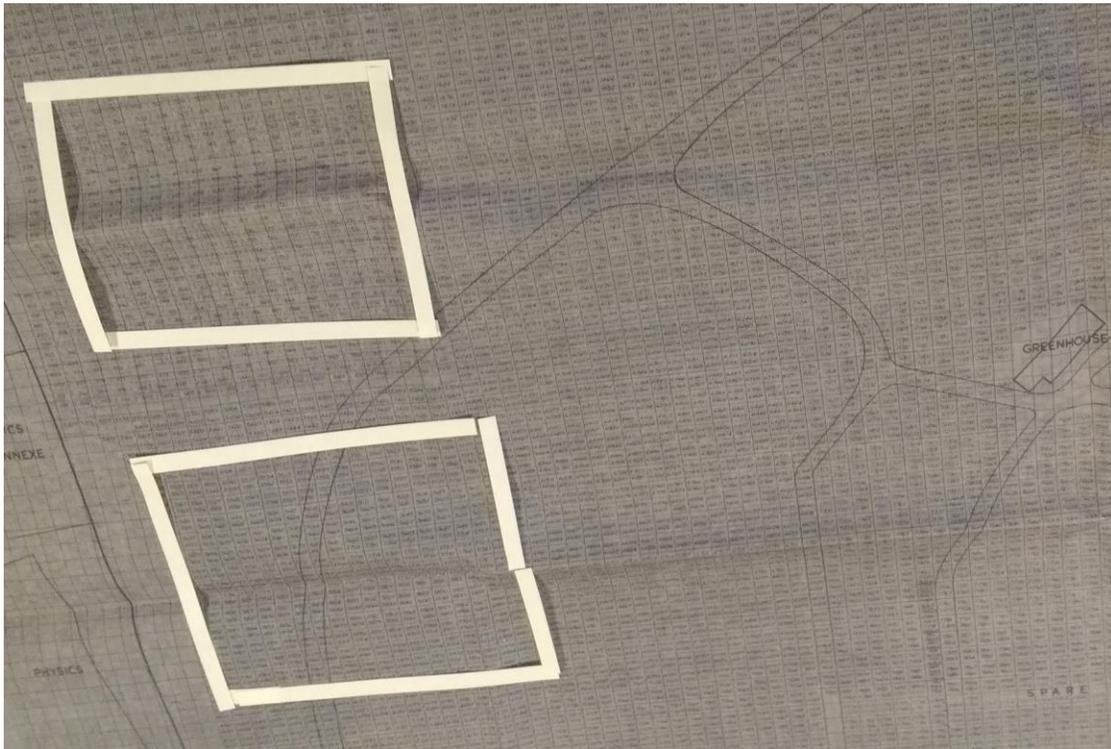
Reproduced with the permission of Special Collections, Leeds University Library. Item reference: MS 421/3/7, Burial Plot Map of the Leeds General Cemetery site: https://library.leeds.ac.uk/special-collections-explore/484414.

The following image of the plot map allows you to compare the location of the labourers' graves (shown on the left) compared to the location of the gentlemen's graves (marked on the right).
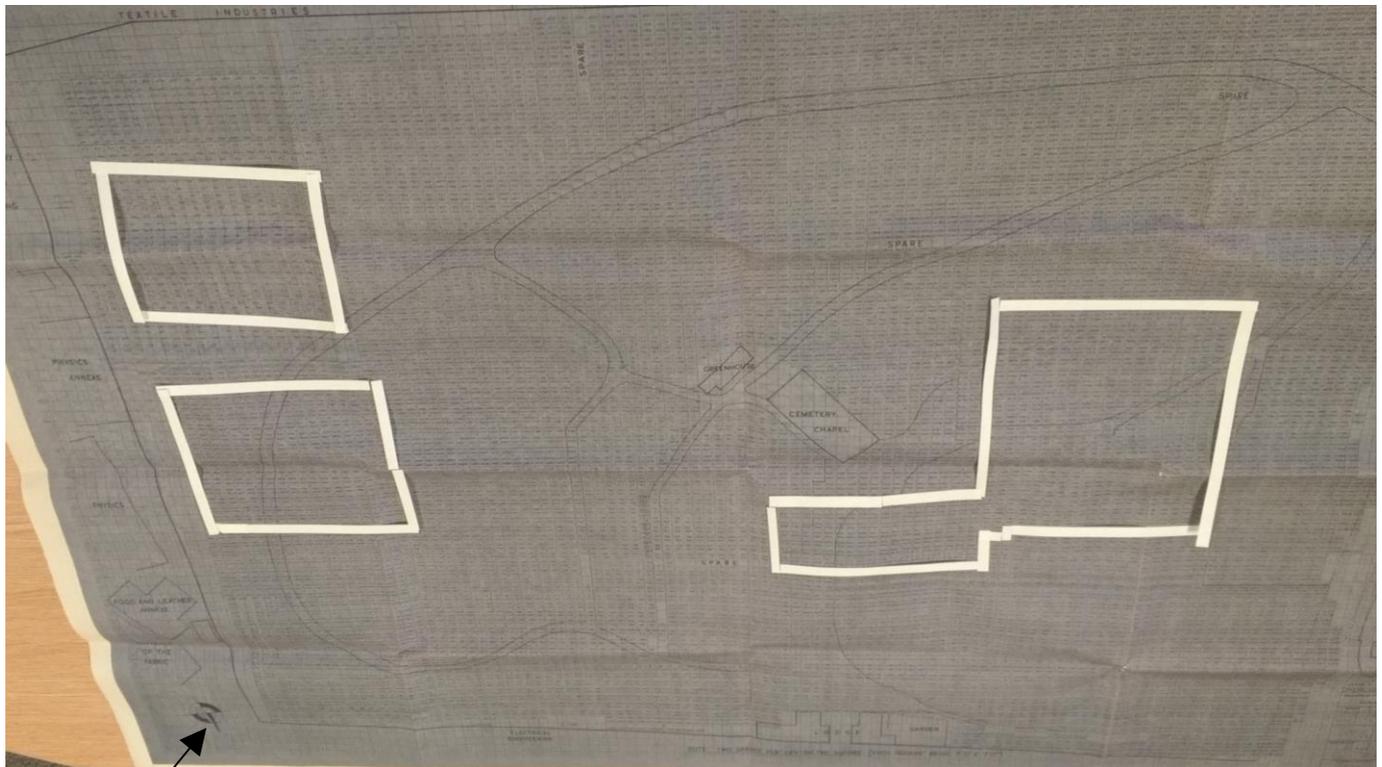

Compass

Reproduced with the permission of Special Collections, Leeds University Library.
Item reference: MS 421/3/7, Burial Plot Map of the Leeds General Cemetery site:
https://library.leeds.ac.uk/special-collections-explore/484414

Viewing the areas that the labourers and the gentlemen tended to buried within highlighted on this map allows us to make some interesting observations.

The gentlemen are buried much closer to the cemetery chapel. This suggests that plots closer to the chapel were more expensive and presumably more desirable, perhaps as these plots were furthest into the cemetery and were therefore some of the quieter plots available. Proximity to the vaults (the most expensive method of interment in the LCG) was perhaps also a factor in these plots being so popular with wealthier families. Furthermore, it is possible that some would have chosen to be closer to the chapel and/or main pathways in an effort to avoid their graves becoming the targets of "uncouth

behaviour" or grave robbers, both of which were topics of great discussion at the LGC's inception.[5]

This concentration of labourers in the southeast area of the cemetery indicates that the provision of graves for low cost burials was indeed acted upon following Rev. James Rawson's request. Further, the southeast section continued to be used for this purpose long after Rawson had been forced to resign, as indicated by the histograms for 1845 onwards.

However, this segregation may also have been self-perpetuating to a degree. We know that relatives often wished to be buried in close proximity to each other, so it might well be the case that once a few labourers were buried in the section, their relatives and friends may account for other labourers buried in the same zone. The more wealthy residents interred around the chapel, on the other hand, would have had more impressive monuments (and neighbours!) than other sections of the burial ground. This would have made the remaining empty plots in the area more attractive to wealthy and high status individuals.

The labourers are buried further away from the chapel and the cemetery's entrance, closer to the cemetery's outer edges. The compass shown in the bottom left corner of the burial map indicates that the labourers were buried in the cemetery's southeast corner. Interestingly, this is in line with Reverend James Rawson's request to designate the southeast area of the cemetery for the poorest patrons of the LGC to be buried: the very outermost corner (here shown above the highlighted areas) was to be set aside only for burying the poorest children.

What are the most common causes of death for people of certain occupations?

The LGC dataset records information about the deceased person's occupation and their cause of death, allowing us to investigate the relationship between these two variables. Different occupations will lead to different lifestyles and living conditions. Therefore we can expect to see a trend in cause of death and particular occupations.

Again, we have chosen to focus on the gentlemen and the labourers recorded in the LGC records because these are occupations which appear in high numbers in the dataset, and also reflect distinctly different lifestyles across the class divide.

---

[5] The photographs of the cemetery within Leeds General Cemetery Collection show protective metal cages on some of the graves, presumably to deter against the threat of grave robbers.

This is a chart generated by the LGC Data Index for all the gentlemen recorded in the registers between 1835 and 1899. It shows the top ten causes of death within this group of 606 gentlemen.

Top 10: Cause of death View details
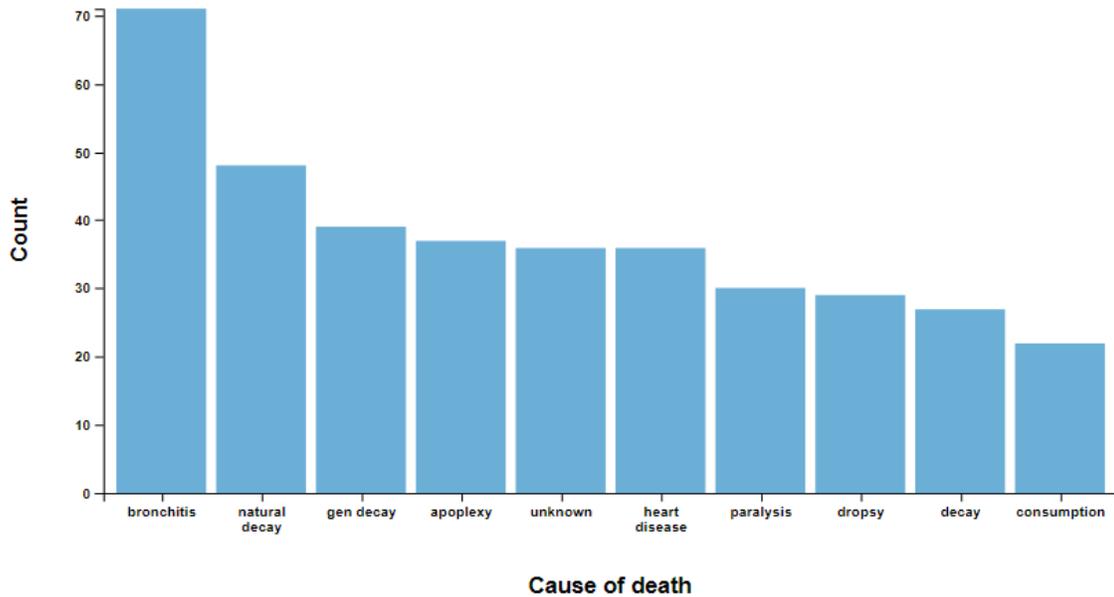
The following chart shows the top ten causes of death amongst labourers buried at the LGC between 1835 and 1899. 453 people recorded as labourers appear in the registers during this time period.

Top 10: Cause of death View details
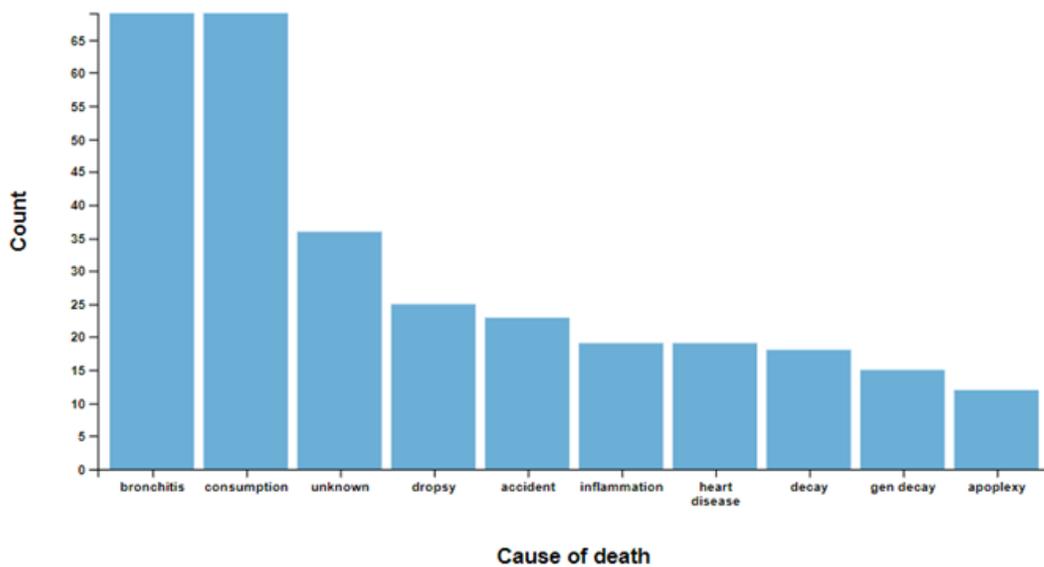
These two charts suggest that both labourers and gentlemen were badly affected by bronchitis. Throughout the 1800s Leeds was becoming increasingly industrial. The presence of mills and factories would have led to high levels of smoke, steam and air pollution in the city.[6] Moreover, both labourers and gentlemen may have smoked regularly.

The next two most frequently appearing causes of death for gentlemen are natural decay and 'gen decay', which presumably stands for general decay. The term decay is understood to have meant at the time, 'the gradual failure of health and strength incident to old age'.[7] This suggests that gentlemen were more likely to live long enough to die of old age. This is a trend we would expect to see, given the lifestyle, living conditions and resources that a member of the upper class would have enjoyed at this time.

On the other hand, labourers appear to have suffered greater levels of fatality caused by consumption, the archaic term for tuberculosis. The spread of tuberculosis is associated with overcrowding, damp, and poor hygiene, as well as damage to lungs from smoke or industrial pollutants.[8] As they likely had more limited means than the gentlemen in our sample, labourers were likely to have poorer housing and living conditions. This is a potential explanation for the higher number of fatalities caused by consumption amongst the labourers compared to the gentlemen.

Interestingly, accident appears in the labourers' top ten most common causes of death. Accident is a broad term that could apply to many situations. Nonetheless, we would expect that the working condition of labourers would put them at a greater risk of being in an accident compared to their upper class counterparts.

How can the data be used to evidence changes in medical terminology?

One interesting feature of the burial registers is the way they show the changing usage of "official" terminology, particularly in the way the cause of death is recorded. These

---

[6] Kasuga, Ayuka (2013) Views of smoke in England, 1800-1830. PhD thesis, University of Nottingham. http://eprints.nottingham.ac.uk/13991/1/Thesis_final_draft_after_viva_for_online.pdf

[7] Appleton 1900, 'Decay' in Glossary of Archaic Medical Terms, http://www.archaicmedicalterms.com/English/Senility.htm

[8] See: https://www.cdc.gov/tb/topic/basics/exposed.htm for a modern list of risk factors and methods of spreading tuberculosis. It is clear that many of these risk factors would also have been present in the homes of 19th century labourers (e.g. lack of ventilation, lack of access to clean water for sanitation purposes, etc).

registers could be a useful resource for exploring how and why medical terminology changes.

We used both R and Excel to isolate and analyse the categories that we were interested in. Three examples were chosen, primarily due to their frequency and relative simplicity. In this part of the analysis, our aim was to understand how this change in terminology is reflected in the burial registers. We focused on some of the clearer examples available: the varying terms for death in childbirth; consumption and tuberculosis; Bright's disease and nephritis. The tables below show the different terms used for the same conditions, as well as their earliest and latest dates of use. In doing so, these tables allow us track the usage of each term and suggest some possible explanations for these changes.

At present, we are somewhat unsure about how the cause of death was reported to the Registrar. With no other guidance or rules in place, it seems likely that the recorded cause of death would simply be the cause of death reported by the informant. Some informants may have given the cause as the diagnosis made by a visiting doctor prior to death, or simply their own conjecture. This is supported by the records made with no personal informant: many of these records state that the deceased was 'found dead' or are left blank. Whether the registrar recorded these reports verbatim or 'translated' the information given by the informant is unclear. Some records suggest that the information is exactly as the informant gave it, such as the lone record of death by 'childbearing', where the registrar's other records show a personal preference for 'childbirth' as the 'correct' term. Others suggest that the registrar was involved in 'translating' the informant's words, such as a number of entries which are signed with X, with the informant's name printed underneath in the registrar's handwriting.

However, some changes appear to be less personal and more legislative.
The 1881 amendment to the Burial Act required that an 'appropriate medical officer' sign a certificate before a body could be moved from the place of death. This certificate included a cause of death, with the legislation itself suggesting that this was to allow for better monitoring of infectious diseases. However, it also means that from the early 1880s onwards, the vast majority of those arriving for burial at LGC would already have been assigned an "official" cause of death. This is generally reflected within the burial registers, as the terms used tend to become less general (e.g. 'decline', 'weakness') and more specific.

Childbirth, etc.

The comparatively late appearance of the more technical term 'parturition' suggests that it is linked to the changes in legislation surrounding the registration of births, although

further research would allow us to see whether this is used regularly or was a quirk of a few particular informants or doctors.

It would also be possible to map whether the variants of the phrase are quirks of the individual registrar working at the time, as seems likely to be the case for some of the more similar variants ('Childbirth', 'Child-birth' and 'Child birth', for example, are all present in the registers with little overlap, suggesting they are the preferred term of individual registrars).

Term for childbirth	Frequency	Earliest use	Latest use
Childbed	110	15/11/1835	25/08/1889
Parturition	77	20/02/1865	14/12/1904
Childbearing	1	16/03/1862	16/03/1862
Child birth	118	03/05/1837	06/05/1949
Childbirth	90	27/12/1837	15/10/1928

Consumption and Tuberculosis

Consumption and tuberculosis account for a large number of deaths present in the records. The vast majority of these deaths are recorded as consumption, rather than the newer term of tuberculosis. There is also a degree of overlap between the continuing usage of consumption and the new usage of tuberculosis, suggesting that both terms were commonplace by the 1920s. Phthisis, again an older term, is the second most common term used to record this cause of death. Both of these older terms are used across a long period, but they both appear to fall out of use in late 1920s and early 1930s. It would be interesting to know why they both become eclipsed by the term 'tuberculosis' around the same time: was this perhaps due to a public health campaign, or education within schools? Or is this perhaps evidence that registrars were recording not from informants but from death certificates filled in by doctors?

It is possible that each term was preferred by a different demographic, though establishing whether this demographic changes on grounds of age or class or occupation would require further investigation. The variety of terms and the broad date

range they used over also suggests that the registrar had less control over what is recorded, instead simply writing down what he was told to. As the terms that prevail in the 1930s and onwards are the more modern, 'technical' terms, we can assume that at this point in time, the registrar is taking instruction from the death certificates, rather than the informants.

Term	Frequency	Earliest	Latest
Consumption	5637	27/08/1835	15/02/1927
Phthisis	454	05/01/1846	08/02/1935
Tuberculosis	174	03/08/1862	18/10/1949
TB	155	02/04/1931	27/12/1951

Bright's Disease and Nephritis

Bright's disease (named as such in 1827), is an archaic term for a condition we would now term nephritis, or inflammation of the kidneys. As the below table shows, Bright's disease was the preferred term for this condition for much of the life of the cemetery, despite nephritis appearing earliest in the records. Again, explaining this would require further research into the identity and role of informants in the period.
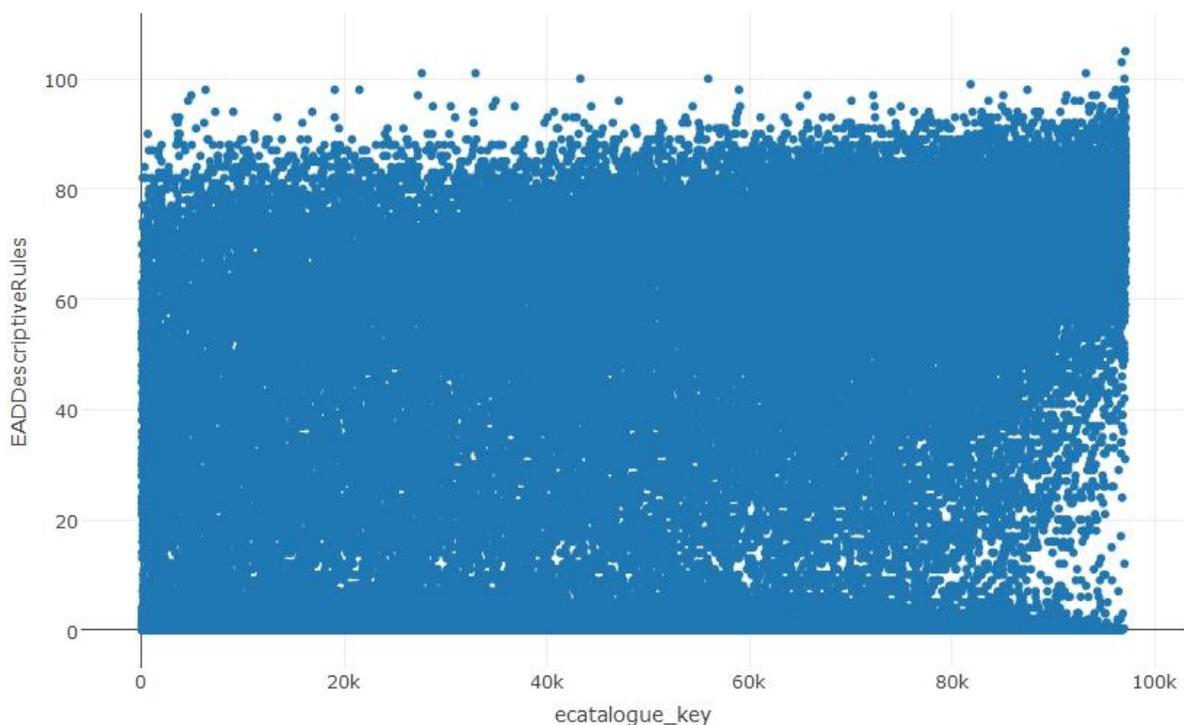
Term	Frequency	Earliest	Latest
Bright's Disease	258	02/03/1887	09/01/1946
Nephritis	24	04/02/1885	21/10/1933
Chronic Nephritis	18	11/04/1896	16/03/1936

How does age at death vary over time?

The LGC dataset spans the period of 1835-1969. During this period, huge progress was made in the general standard of living conditions in Britain. This was due to landmark Acts being passed such as the Public Health Acts of 1848 and 1875. These Acts introduced responsibilities for local authorities such as the appointment of Medical Officers, the regulation of sewers and the inspection of housing.[9]

Because the LGC's lifetime spans these innovations and improvements, we were interested to see whether or not the dataset evidenced any improvement in life expectancy over time.

To create a scatter point using R we plotted age at death (shown on the y axis) against the variable in the dataset, 'ecatalogue_key' (shown on the x axis). Ecatalogue_key reflects the date of burial: the first burial in the cemetery in 1835 was assigned the entry 1 in this field, while the most recent interment in 1969 is number 97,122. So moving from left to right, the x axis shows the burials over the time span of the cemetery's life.

---

[9] The Public Health Act 1848, Parliament UK, [http://www.parliament.uk/about/living-heritage/transformingsociety/towncountry/towns/tyne-and-wear-case-study/about-the-group/public-administration/the-1848-public-health-act/, accessed 11 Aug 2017]; and The Public Health Act 1875, Policy Navigator, [http://navigator.health.org.uk/content/public-health-act-1875, accesed 11 Aug 2017].

We created this scatter point to display a sense of how the age at death of the people buried in LGC varied over time. However, the sheer quantity of entries makes this a very dense and confusing scatter plot. It would be best to repeat this exercise using the average age at death.

Nonetheless, some interesting observations can be made from analysing this scatter point. There appears to be a slight general increase in the age at death over time, with more people living to be over 80 towards the end of the cemetery's working life than at the start. The white gap on the right hand side of the chart where less entries are plotted shows that less young people were dying during the cemetery's final years, compared to when it opened. This is the pattern we would expect to see given the advances made in the standards of living conditions in Britain during this time.

Furthermore, there is a consistent thick block of entries plotted at the bottom of this graph, representing the large quantity of children buried in the cemetery. In particular, there are many values plotted on the x axis at 0, indicating stillbirths. In this way, the infants who died at Leeds Maternity Hospital and were buried in the cemetery have had a visible impact on the data.

Conclusion

The LGC burial records have provided researchers with a vast dataset that has proven to be rich with information. The data's many variables, especially its records of 'cause of death' and 'occupation', make this a versatile and powerful source. The cemetery's working lifetime spans 1835 to 1969, covering the Victorian era and a great deal of the 20th century. This dataset offers insights which link to a variety of research topics, including the local history of Leeds and the study of disease.

Our project has shown that statistical analysis provides a useful and illuminating approach to the dataset. Our investigation has allowed us to compare different groups within the data, particularly considering the differences in both life and death that existed between the working class and upper class. We have also been able to trace the development of medical terminology within the context of the cemetery's lifetime.

With more time and statistical experience we believe this investigation could be developed. The potential scope of this project is large. We have outlined recommendations for further research in this report's close.

- We have taken into consideration how missing or unrecorded values may impact an analysis of the dataset. We have summarised the proportion of missing data within each variable or field. Administrative information which was essential in the keeping of the burial records – such as burial number, plot number and date of burial – has an extremely small amount of missing data. Additional information – such as details relating to the deceased's parents – has larger amounts of missing data.

- We have reflected on how we may be able to infer the social class of people represented in the burial registers from the information that we have available. We chose to use within this study a fairly crude comparison of labourers and gentlemen in order to investigate how class differences affect circumstances of death.

- We examined the plot numbers of labourers and gentleman buried in the cemetery to assess which parts of the cemetery were reserved for richer and poorer people. We found that gentlemen tended to be buried in the centre of the cemetery, in close proximity to the chapel. Labourers were buried in the southeast corner of the cemetery, further from the chapel. We theorised possible reasons why the central plots were more desirable.

- We compared the most common causes of death for gentlemen and labourers and considered how the differing lifestyles and living conditions of these two groups would have had an impact on the diseases they contracted and on their deaths.

- We examined the changes in terminology for consumption (which came to be known as tuberculosis) and Bright's disease (which came to be known as nephritis, as well as the many variant terms for childbirth. Our analysis confirmed that there were significant changes in the medical terminology that was used over time. There are a few possible causes of this: changes in registrar, changes in legislation demanding more 'official' terminology, or a variation in the terms known and used by those acting as informants (whether officials or relatives of the deceased). However, although the general frequency of older terms decreased over time, none of the examples we examined completely fell out of usage in the time the cemetery was in operation. This suggests that cause of death was not being reported by only one medical officer (as legislation stated), but by a variety of sources.

- We attempted to examine the average age of death of those interred in the LGC. The sheer size and breadth of the dataset made it difficult to make any clear conclusions. However, we can say that the age of death generally increased of the course of the LGC's operation, and that deaths of children (once they reached 12 months old) became less common in later years.

Further Research Recommendations

- Though not transcribed, the burial registers offer a wealth of information on migration into Leeds from 1835 to the early 1900s. This is because the burial registers have a field which records place of birth as well as residence of the deceased. If these addresses could be transcribed we would be able to trace individuals who migrated to Leeds and learn more about the diverse backgrounds of the inhabitants of Leeds.
    - Further research could be carried out into the prominent groups who came to be buried in the LGC, but also into groups who were *not* buried in the general cemetery: which religious or cultural groups sought to be buried elsewhere, and why?

- Comparing the information gathered in the burial registers with the information included on an individual's headstone or obituary may suggest to us what information the registrar valued and how this differed from how families remembered or memorialised the deceased's life.

- Similarly, it would be highly interesting to know more about the informants. Did informants tend to be family members and friends, or were many (any?) of them 'officials' (doctors, police officers, clergy, employers or colleagues)? Does this change over time as legislation becomes more restrictive and hospital deaths become more common? Knowing this would give us a clearer picture of the process of interring a body in the 19th century.

- It would be interesting to see a graph which mapped the frequency of different medical terms over time, rather than being limited only to the earliest and latest usages, as this would show if there was a gradual change or a sharp decrease.
    - We could also investigate whether any public health campaigns existed which may have contributed to either the reduced number of deaths from particular diseases, or contributed to a more widespread knowledge of the newer terminology.
    - It might also be interesting to follow up some examples where the recorded cause of death may be euphemistic or intended to protect

reputations. There are surprising few records of syphilis as a cause of death. How many of the 982 deaths attributed to 'brain affection' or 'paralysis' were in fact caused by 'general paralysis'; the late stages of syphilis?

- It would certainly be interesting to see what the links between gender and age of death are within the sample of those buried in the cemetery. Did the women buried in the LGC live longer than the men buried here? Did women listed as 'wife' live longer than women who worked outside the home? Are there any differences in the age and cause of deaths of those women who were widowed versus those who were spinsters?

- We would be interested in using statistical analysis to see if the parents' occupation affects the deceased's age at death. For example, if the deceased parents are 'Gentlemen', does that person tend to live longer compared with a person whose parents' occupation is recorded as 'labourer'?

- The LGC burial records would make an interesting point of comparison with other collections.
  - Special Collections' volumes on the Quarry Hill area (classmark Large Yorkshire H-Lee-4.1/LEE) provide detailed information about 'unhealthy' working-class housing in 1900 with an accompanying photo album. It would be worthwhile to check if any of these individuals were buried at the LGC. Using the two collections in conjunction could reveal more about the lives of the people represented within them.
  - Special Collections also hold the medical casebooks of surgeon William Hey (1736-1819). We would be interested in seeing how Hey's medical terminology and the causes of death recorded in his casebooks compares with the medical terminology used in the LGC burial records. This could reveal whether personal preference affected the use of terminology, or may suggest what terms were most commonly used.

- The investigation we carried out into class and the cemetery - and specifically where labourers and gentlemen tended to be buried - suggested there was a distinct pattern in the parts of the cemetery the upper class were being buried in, compared to the working class. We highlighted the different areas using archive paper placed on the map. To represent this burial pattern more clearly, this could be done digitally. Different sections of a digital copy of the burial plot map could be highlighted to show where people of different occupations were buried.

- ○ Similarly, an interactive version of the digital plot map would be really useful. This would be a large project, but if an online map could be tagged so that a user could easily see who was buried where, we could learn a lot more about the layout of the cemetery and its burial patterns.

Appendix

Data Dictionary

This is a descriptive guide to supplement the dataset, as it was in the form that we received it from Special Collections. The data had 18 variables, in 18 columns in the spreadsheet.

Column	Field Title	Description
A	ecatalogue_kE	This provides each entry with a unique **identifier** number. These numbers are ordered by the date of burial, with the first burial being 1 and the last scattering of ashes 97122.
B	EADUnitID	Similar to ecatalogue_kE , this field provides an **identifier** for each entry and facilitates the use of the data in the Collection Management System EMu used by Special Collections.
C	LeeAltReferenceCode_tab	**The burial number** that was assigned to each burial by the Leeds General Cemetery Company. Sometimes the use of certain burials numbers were repeated or duplicated across the burial registers, so the burial number does not always reflect the order of burial.
D	EADLocationOfOriginals	**Plot Number** or number of grave. The plot numbers correspond with the cemetery's plot map. A digital copy of the plot map is available on

		the Special Collections website.
E	EADUnitTitle	**Name** of the deceased. Surname, first name.
F	CreDateCreated	**Date of death**
G	EADUnitDate	**Date of burial**
H	EADBiographyOrHistory	**Age** at death
I	EADDescriptiveRules	**Decimal age** - whole numbers show the number of years old a person is. Months, weeks and days are expressed as decimals. For example, the age 1 year old would be 1. The age 6 months old would be 0.5.
J	EADPhysicalTechnical	**Gender**
K	EADScopeAndContent	**Cause of death** or disease
L	Occupation	**Occupation**, role, rank, trade or profession
M	Parent Occupation	Occupation, role, rank, trade or **profession of the deceased's parents**
N	EADOtherName_tab	**Father's name**
O	EADName_tab	**Mother's name**
P	EADAlternativeFormAvailable	Corresponding **image** of the relevant page in the burial register
S	NotNotes	**Notes** made by the archivist