

SCHOOL OF
ELECTRICAL AND
ELECTRONIC ENGINEERING



THE UNIVERSITY
of ADELAIDE

**CSI ADELAIDE: WHO KILLED THE SOMERTON MAN?
CIPHER CRACKING**

Honours Project 2018

Final Report (Draft)

Ashley Seroka a1687356

B.E. in Electrical and Electronic Engineering (Honours)

Date submitted: 1st of June 2018

Supervisor: Professor Derek Abbott

Abstract

The corpse of a mystery man's body was found at Adelaide's Somerton Beach, South Australia, Australia, on the 1st December 1948 and was hence referred to as the Somerton Man. Till this day the identity of the Somerton Man and the cause of his death is unknown.

This project will be broken up into three different tasks, with all contributing towards the unsolved case of the mysterious man.

The first task is based around the piece of paper that was found inside his trousers pocket. This piece of paper had five lines of capital letters, and is thought to be some kind of code or cipher. It was found that this mysterious code was a part of a poetry book, known as the '*Rubaiyat of Omar Khayyam*'. The letters on the mysterious code are thought to be first letters of words, based on previous year's project groups. The location of his death is near Morphettville Racecourse, which leads to belief that the Somerton Man's mysterious code are different horse names. However, using various statistical approaches, this was proven to be not true.

The second part of this project involved, using a mass spectrometer to analyse different isotopic signatures of the samples. More specifically, the samples used were the shaft of the hair, obtained from modern day people, which were burnt by the laser, and effectively recorded the various elements. This will be compared with the Somerton Man's hair, to identify specific elements present, as well as noting how long he was in Adelaide before his death.

Different DNA samples, were analysed in the final task. Using software tools the samples were degraded, until the DNA became unidentifiable.

Acknowledgements

Project supervisor Professor Derek Abbott, for the helpful and motivational advice, as well as the exceptional guidance, which was presented on each of the completed tasks.

Contents

Abstract.....	2
Acknowledgements.....	3
Contents.....	4
1. Introduction	6
1.1. Motivation.....	6
1.2. Previous Studies/ Related Work	8
1.2.1. Australian Department of Defence.....	8
1.2.2. The University of Adelaide project groups	8
1.2.3. Mass Spectrometer Pervious Work	9
1.3. Aims and Objectives.....	10
2. Technical Background	11
2.1. P-Value Theorem	11
2.2. Mass Spectrometer	12
2.3. Single Nucleotide Polymorphism (SNP)	12
3. Knowledge Gaps and Technical Challenges	13
4. Task 1: Code Analysis	14
4.1. Aim	14
4.2. Method	14
4.3. Results.....	15
4.3.1. Horse Names.....	15
4.3.2. Australian Beaches.....	15
4.3.3. South Australian Street Names.....	16
4.3.4. Australian City's	17
4.3.5. The Rubaiyat of Omar Khayyam book	18
4.4. Conclusion.....	18
4.5. Extension of Task 1	18
5. Task 2: Hair Analysis.....	20
5.1. Aim	20
5.2. Method	20
6. Task 3: DNA Analysis	21
6.1. Aim	21
6.2. Method	21
7. Project Management	22
7.1. Timeline.....	22

7.2. Budget.....	23
7.3. Risk Management	24
8. Conclusion.....	25
8.1. Future work.....	25
9. References	26

1. Introduction

1.1. Motivation

During this project, a study was undertaken on an unsolved murder case. On December 1st, 1948, an unknown man was found deceased, on Somerton Beach in Adelaide [1]. From then on, he was labelled, 'the Somerton Man'. There was no form of identification present on him, as well as little information surrounding his death. Therefore, the cause of his death is still unknown to this day [2]. Figure 1 shows the deceased man.



Figure 1: The Somerton Man [1]

A piece of paper with the words "Tamám Shud" printed on it, was found rolled up inside his trousers pocket, which can be seen in Figure 2. This statement can be translated from Persian to either "it is ended" or "it is finished". This piece of paper was found to be part of a poetry book, the '*Rubaiyat of Omar Khayyam*' [3]. The theme of the book is that, one should live life to the full and have no regrets when it ends [4].

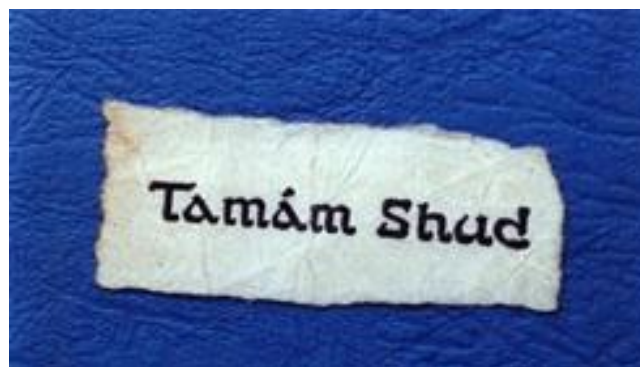


Figure 2: The Scrap of Paper [5]

The book is speculated to be related to the dead man, due to the parchment. Hence, the case being known as the Tamam Shud case. This has been considered, since the early stages of the police investigation, to be "one of Australia's most profound mysteries [6]."

Capital letters were found to be scribbled in the back cover of the *Rubaiyat of Omar Khayyam*, as seen in Figure 3. Thus, indicating that these letters are somewhat vital to the case at hand, as it is speculated that they may be a form of code or cipher.

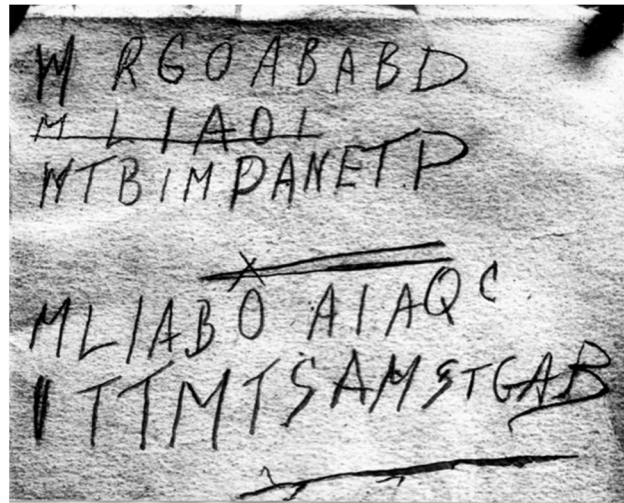


Figure 3: The Mysterious Code [5]

The code consisted of five various lines of capital letters, with a total of fifty letters all together. The second and fourth lines are very similar in the way in which they appear. It is therefore, believed that the Somerton Man may have made an error with the encryption, when writing the second line, hence why it is struck out. It is unclear whether some of the letters are in fact an “M” or “W”, therefore tasks referring to the code, result in one of two different versions. One being with an “M”, whilst the other being with a “W” [7].

His body was found to be located near the Morphettville Racecourse, this leads to the belief, that the Somerton Man’s mysterious code are related to different horse names. It was also noted that the people who found the deceased body of the Somerton Man, were racehorse jockeys [2].

1.2. Previous Studies/ Related Work

1.2.1. Australian Department of Defence

In 1978, a request was sent by journalist Stuart Littlemore to the Department of Defence cryptographers to analyse the code. Unfortunately, the cryptographers were unable to crack the code, as they could not provide a satisfactory answer. It was stated that the code had “insufficient symbols” or a “disturbed mind” generated the meaningless code [8].

1.2.2. The University of Adelaide project groups

There have been several Honours project groups at the University of Adelaide that have undertaken this project. The previous work the project groups have done include:

- Letter frequency analysis in different languages.
- Initial letter and sentence letter probabilities.
- The likelihood of the code being an initialism of a poem.
- Different cypher techniques.
- The design and implementation of a web crawler.
- 3D generated reconstruction bust of the Somerton Man.

Main conclusions that these project groups have come to are:

- It is unlikely that the mysterious code is created randomly.
- There is strong evidence to believe the mysterious code is most likely to be in English.
- It is not likely that the mysterious code are initialisms extracted from poems.
- The Rubaiyat of Omar Khayyam was not used as a straight substitution one-time pad for encryption.
- The Rubaiyat of Omar Khayyam was not created as a one-time pad for the mysterious code.

With these conclusions, this project will look into further detail of what the Mysterious Code is. [9] [10] [11] [12] [13] [14] [15]

1.2.3. Mass Spectrometer Previous Work

Previous years have also done study with Mass Spectrometer. The 2013 project group had some of Somerton Mans hair and plotted the different elements in the hair comparing with controlled samples. Analysis was conducted on different elements between the two samples. This was done using a glass slide, which have impurities in it [13]. In the 2016 project group, they recreated the analysis using a quartz slide, which does not have impurities. They concluded that Somerton Mans had some abnormally high readings, of some elements, one of which is strontium [15]. In this project, the strontium level will be looked in higher detail and this will be used to indicate how long the Somerton Man was in Adelaide before his death.

1.3. Aims and Objectives

The first task that will be done is to see if the mysterious code are a collective object (horse names, Adelaide street names, Australian beaches, etc.). This will be done using hypothesis testing. An extension of this task will also be completed, which involves the Rubaiyat of Omar Khayyam.

The second task that will be performed is using a mass spectrometer. Controlled sample hairs will be compared with Somerton man's hair, to see how long he was in Adelaide before his death, by finding different elements in the hair.

The third task that will be accomplished is using DNA data. The data will be degraded using software tools till it becomes unidentifiable. This then can be used to see how much DNA we need from Somerton man, where further research can undergo.

2. Technical Background

2.1. P-Value Theorem

A p-value is a recognised statistical probability, which acknowledges whether an equal or larger effect is present, in comparison to its observed counterpart. In statistics, the p-value helps you determine the significance of the statistical hypothesis by observing the results that were sampled. This determines the probability that the results are due to chance, rather than the experimental conditions. Thus, determining the strength and validity of the results against that of the null hypothesis [16].

In this project, the main focus with the p-values is to determine whether the mysterious code, represents local horse names. Where the null hypothesis is 'The group of letters are horse names' with the alternative hypothesis being 'The group of letters are not horse names.'

For null hypothesis to be accepted the p-value must be larger than 0.05, this indicates that the observed data point is located in the 'most likely observation' range, as seen in Figure 4. If the p-value is lower than 0.05, this indicates that the collected results are statistically significant and that the observed data point is located in the 'very unlikely observations' range. If the results are in the 'very unlikely observations' range, then the null hypothesis can be rejected, which means that the mysterious code is indeed, not horse names [17].

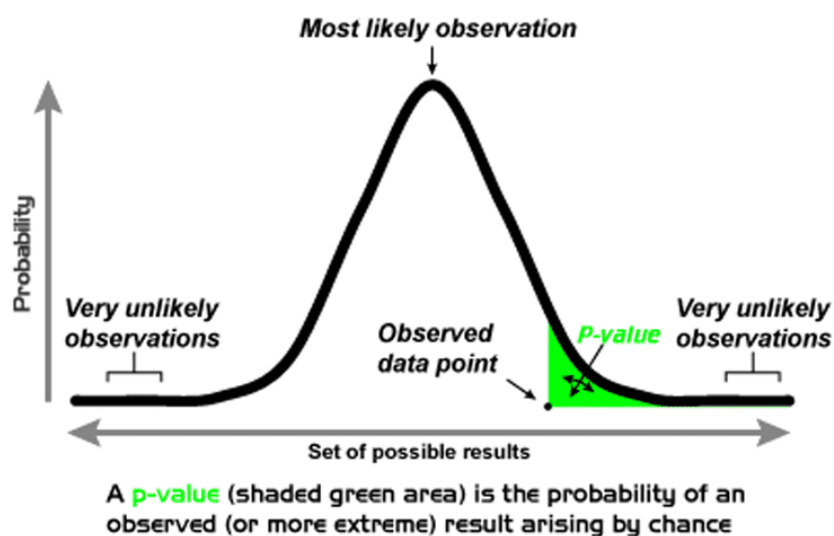


Figure 4: Computation of a p-value [18]

Figure 4 shows, the y-axis is the probability and the x-axis is the set of possible results.

2.2. Mass Spectrometer

A mass spectrometer is an analytical technique which, when given a foreign sample, it can detect unknown compounds within it. The mass spectrometer produces a multitude of ions in the sample, which then uses a mass to charge ratio of the different ions and records the quantity of each ion type [19].

The components of a typical mass spectrometer is seen in Figure 5. The three major components are the ion source, analyser and the ion detector system.

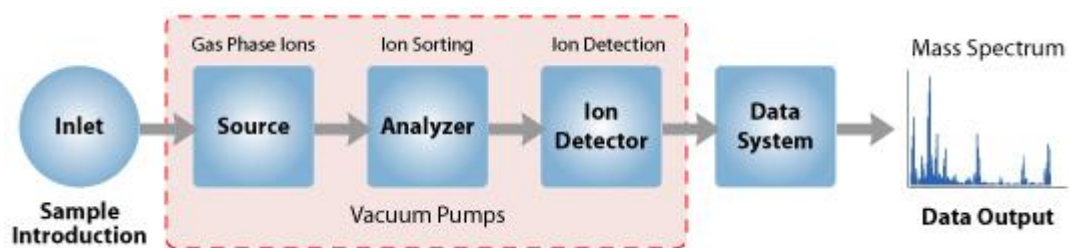


Figure 5: The Main Components of a Mass Spectrometer [19]

The ion source produces gaseous ions from the sample that was used. The analyser, then sorts the different ions by using the mass-to-charge ratio, according to their mass components. The ion detector system detects the different ions in the sample and records the quantity of each ion type and converts it into an electric signal. [20]

The Inductively Coupled Plasma Mass Spectrometer (ICP-MS) is the type of mass spectrometer that was used in this project. The ICP-MS is faster, more precise and sensitive at finding different ions compared to other types of mass spectrometer. [21]

In regards to the project, the sample being used is the shaft of the hair.

2.3. Single Nucleotide Polymorphism (SNP)

Single Nucleotide Polymorphisms (SNPs), are one of the most common and well talked about genetic variations, which can be present between humans [22]. These variations occur within a nucleotide (a single block of DNA) and happen roughly within 1 out of 300 base pairings [23].

In regards to the project, SNPs will be removed from a DNA sample.

3. Knowledge Gaps and Technical Challenges

The technical challenges that will be encountered in this project are related to all the knowledge gaps mentioned.

To complete each task within the project, further development for programming skills, such as Matlab were required. P-value calculation and hypothesis testing needed some revision, to ensure that a satisfactory level of understanding of the concepts was present. The skill to use Microsoft Excel to perform statistical analysis on the p-value, is required. It was also required to learn how to correctly use a mass spectrometer and interpret the results, this can be done by finding multiple ways to enhance knowledge, before trials.

4. Task 1: Code Analysis

4.1. Aim

The aim of this task is to see whether the mysterious code represents some collective object. The collective objects that will be used are horse names, Australian beaches and cities, South Australian street names, and The Rubaiyat of Omar Khayyam book. The assumption will be made that the letters in the mysterious code, are the initialism of a word. The Somerton Man had a lot of associations with horses, so further research will be going into the assumption that the mysterious code are 'horse names'. The null hypothesis is 'The group of letters are horse names' and the alternative hypothesis is 'The group of letters are not horse name.'

4.2. Method

The approach to determine if the mysterious code represents collective objects, will be done by calculating the p-value and implementing hypothesis testing. In the case of the horse names, there were no direct websites, which provided horse names in the year of 1948. This led to discovering evidence of these names within relevant newspapers and articles. This was done by using 'Trove', a search engine to help find resources in Australia. More specifically, in this case it was used to obtain articles and newspapers from 1948. The other collective objects, as mentioned above, were found using South Australian government websites. This led to an abundance of cross checking, to make sure that the list which was going to be utilised, was indeed correct.

Matlab was used as the software tool. The initial letter of the each collective object was all that was needed, where using code was implemented to perform this task. In the case, where a collective object had multiple words, both words were included. Also if there was any extra punctuation, it was removed. An example is shown below.

Input: Golden Bullet, Happy, Piano!, Crazy Super-Fast Monkey

Output: G B H P C S F M

Results were then produced in an excel worksheet, where the p-value test was performed and a comparison graph was completed.

4.3. Results

Each collective object was compared to the mysterious code by the frequency of each letter.

4.3.1. Horse Names

The comparison of horse names to the mysterious code is seen in Figure 6.

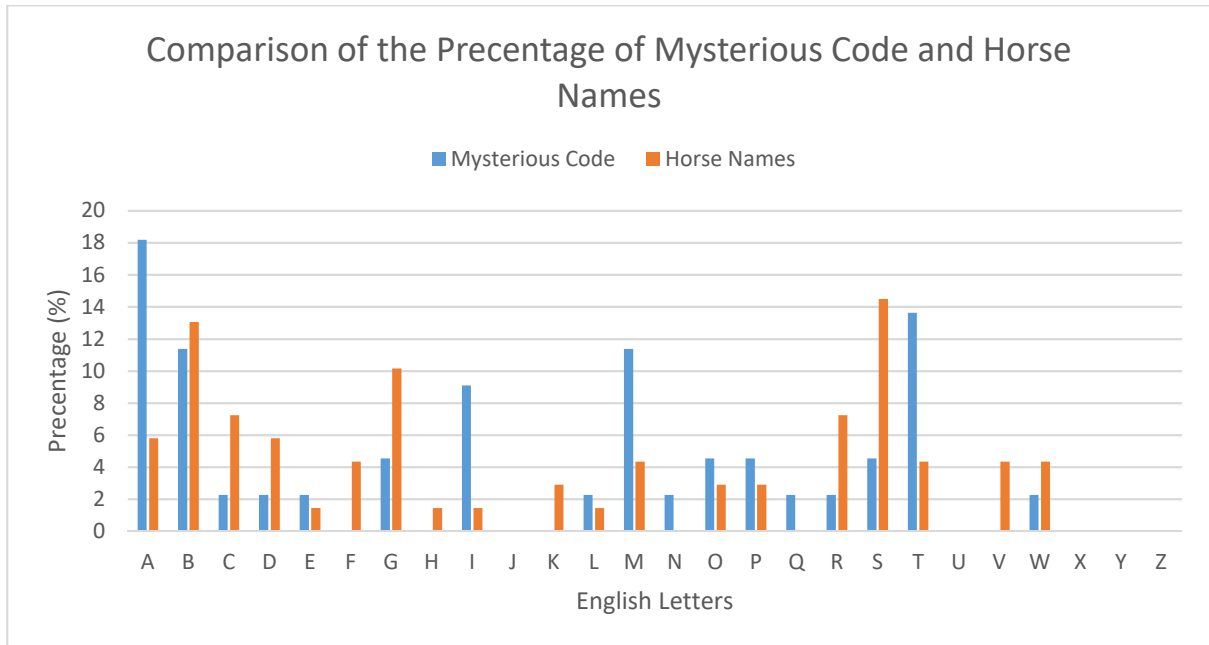


Figure 6: Comparison of Mysterious Code with Horse Names

There was a sample of 69 horse names and it can be seen on the graph that the horse names do not correlate with the mysterious code with many of the English letters. This was also proven by the p-value, as it was lower than 0.05, which means the null hypothesis is not accepted.

4.3.2. Australian Beaches

The comparison of Australian beaches to the mysterious code is seen in Figure 7.

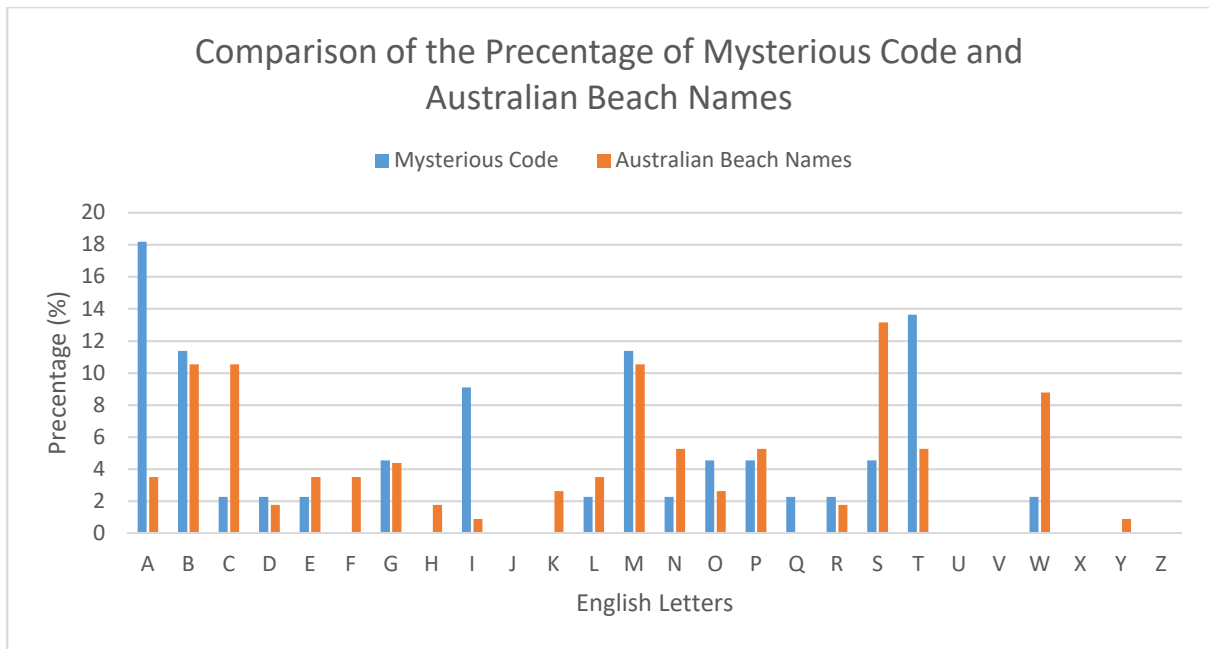


Figure 7: Comparison of Mysterious Code with Australian Beach Names

There was a sample of 114 beach names. Analysing the graph it be seen that the frequency of the letters do correlate with mysterious code. As the results seemed genuine a hypothesis test was done between this values. The results showed a p-value of greater than 0.05, which indicates that the mysterious code could be Australian beach names.

4.3.3. South Australian Street Names

The comparison of South Australian street names to the mysterious code is seen in Figure 8.

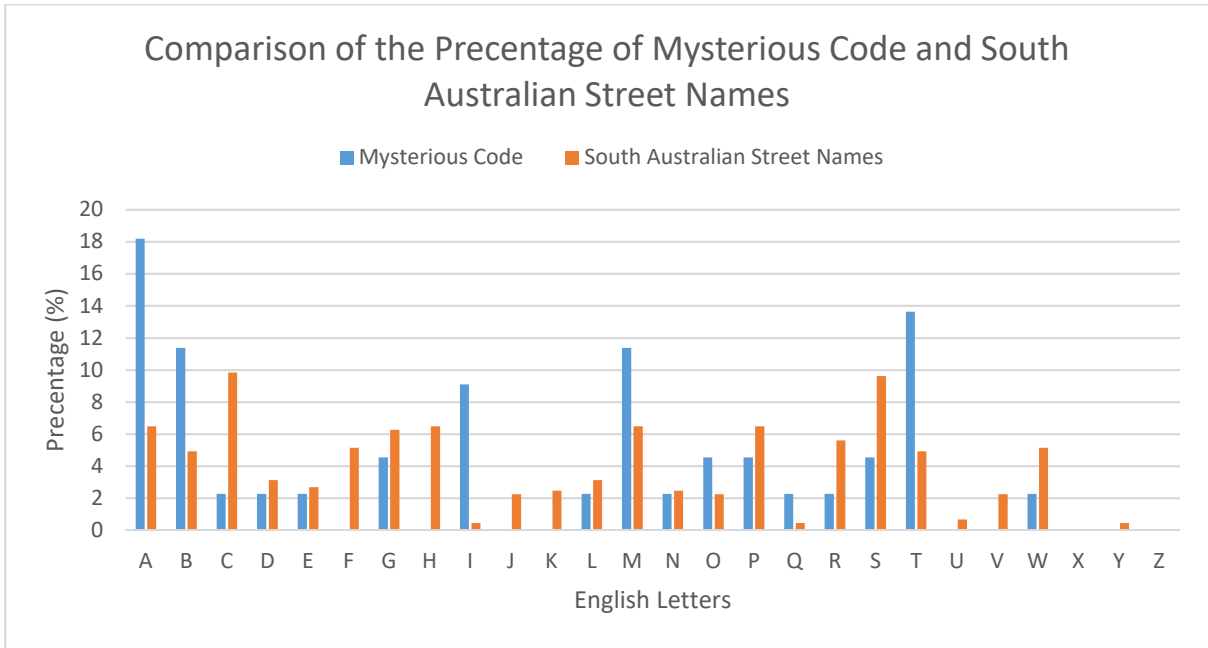


Figure 8: Comparison of Mysterious Code with South Australian Street Names

There was a sample of 447 South Australian street names. Observing the graph it can be seen that the frequency of the letter are not similar with the mysterious code.

4.3.4. Australian City's

The comparison of Australian city names to the mysterious code is seen in Figure 9.

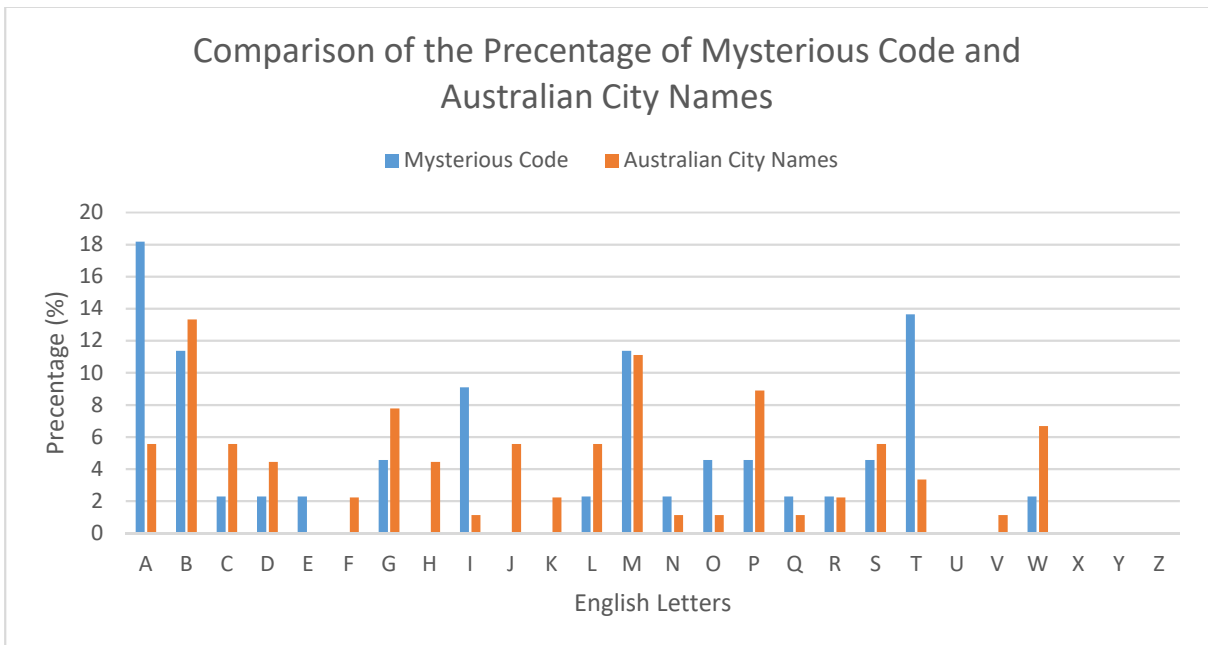


Figure 9: Comparison of Mysterious Code with Australian City Names

There was a sample of 90 Australian city names. Observing the graph it can be seen that the frequency of some letter are similar with the mysterious code. A hypothesis test was then done to check the results. The p-value that was obtained was less than 0.05.

4.3.5. The Rubaiyat of Omar Khayyam book

The comparison of the Rubaiyat of Omar Khayyam book to the mysterious code is seen in Figure 10.

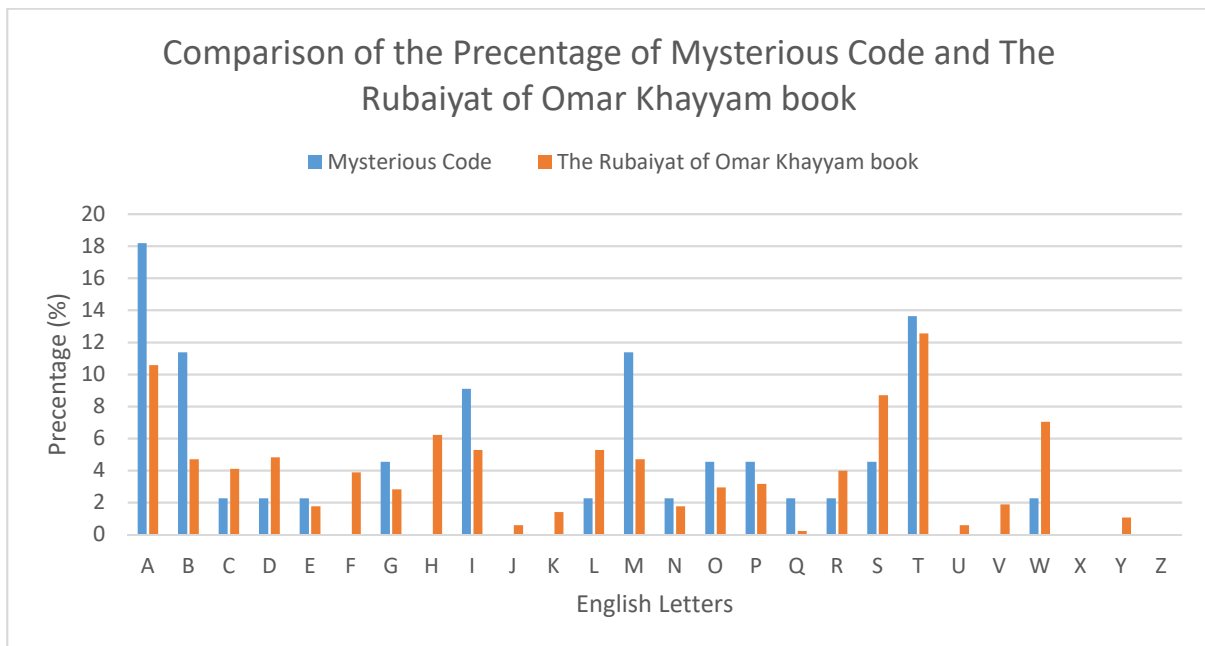


Figure 10: Comparison of Mysterious Code with The Rubaiyat of Omar Khayyam book

There was a sample of 852 words form the book. Observing the graph it can be seen that the frequency of the letter are not similar with the mysterious code.

4.4. Conclusion

Overall, the results have shown, that it is unlikely that the mysterious code represents horse names. It has also shown that it is unlikely to be South Australian street names, Australian city names or The Rubaiyat of Omar Khayyam book. But to some surprise it is possible that the mysterious code is Australian beach names.

4.5. Extension of Task 1

An extension of this task was also done. This includes analysing The Rubaiyat of Omar Khayyam book more carefully. Previous years stated that the mysterious code does not

correlate with the book. Each paragraph in the book has four lines of words, which compared with the mysterious code also has four lines. Still assuming that each letter in the mysterious code is an initial word, we can compare the two.

The task was to count how many words are in each line of the book and compare it with the mysterious code, this was done using Matlab. Then using excel, a graph was plotted with error bars to the number of letters in the mysterious code. This can be seen in Figure 11

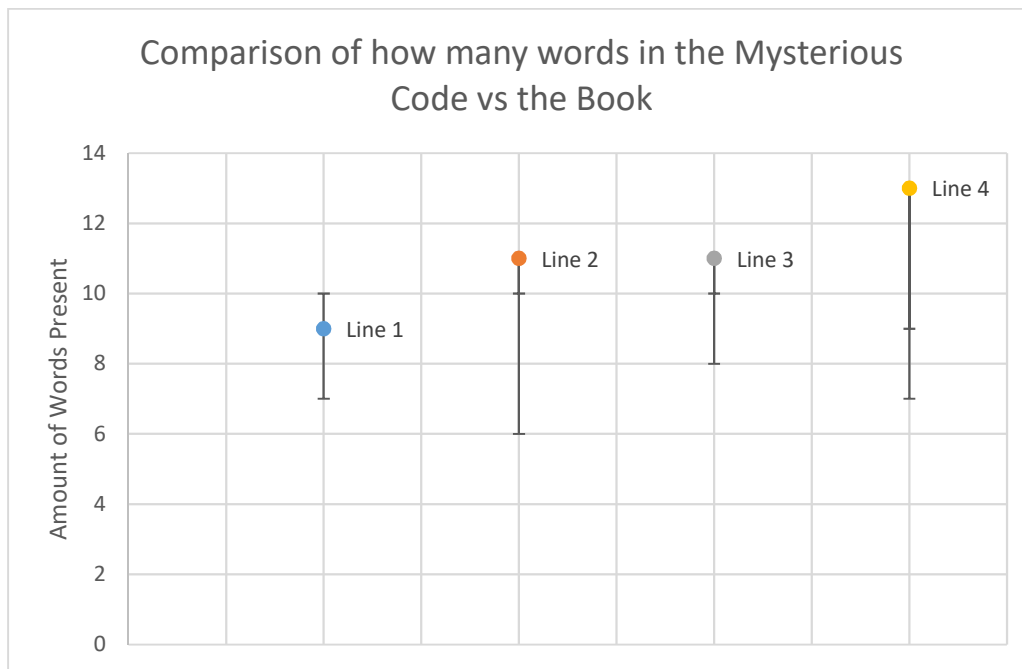


Figure 11: Error Bars against the book

It can be seen that on line one, the mysterious code is in the error bars. The rest of the lines are out of the error bars. This indicates that the mysterious code is not from The Rubaiyat of Omar Khayyam book and further proves the previous year's studies of the book not being part of the mysterious code.

5. Task 2: Hair Analysis

5.1. Aim

The aim of this task is to identify the different isotopes present in several different people's hair. More specifically the element of concern is strontium. Adelaide has high levels in strontium compared to the rest of Australia. With this knowledge, the task is to test various hair samples, which have left Adelaide, within the past year and compare it to that of hair samples that have not left Adelaide for at least a year, to see how the strontium values change. This will then be compared with the strontium levels in the Somerton Man's hair, which can determine how long he was in Adelaide before his death.

5.2. Method

The ICP-MS, is the approach to determine the different isotopes within the hair, which will then return a spectral analysis of the hair. This will then be placed on a slide, in this case the slide will be made of pure quartz. The reason a pure quartz slide is used, rather than an ordinary glass slide is that, glass slides have a lot of impurities that would contaminate the result, where as a pure quartz slide does not. The spectral analysis will be completed by laser ablation of the hair, where the hair will be burnt with a laser and the spectral elements are recorded.

6. Task 3: DNA Analysis

6.1. Aim

The aim of this task is to discover how much DNA can be degraded before it becomes unidentifiable. This information will also be used to indicate how much DNA is needed from Somerton Man.

6.2. Method

The approach to degrade DNA, will be done by removing SNP's from a DNA sample. A DNA sample was sent to an ancestry place called '23andME' where they tested the DNA sample and returned the results. With these results, SNP's will be removed at different percentages, which then can be utilised to discover how much DNA is required until it becomes unidentifiable.

7. Project Management

7.1. Timeline

The timeline for this project was made using a Gantt chart. The proposed Gantt chart can be seen in Figure 12.

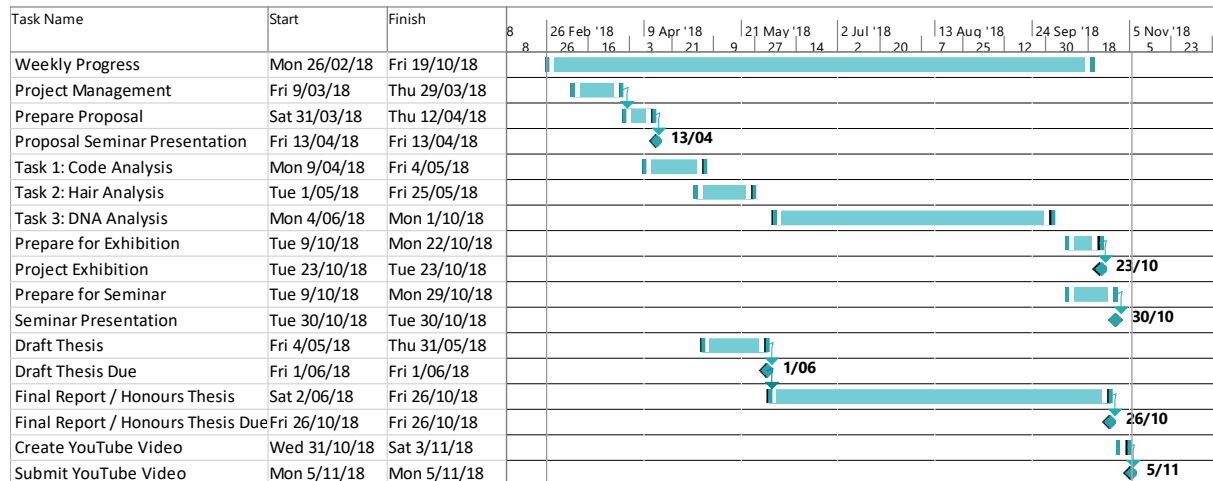


Figure 12: Proposed Gantt chart

The updated Gantt chart that will be used in Semester B is seen in Figure 13.

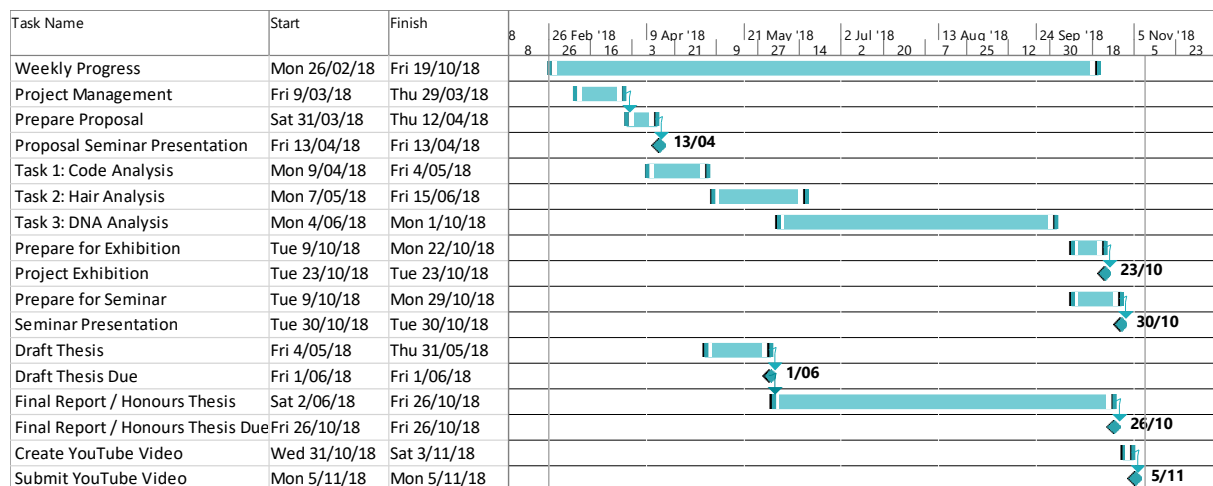


Figure 13: Semester B Gantt chart

The project timeline is behind of schedule, where task 2 was meant to be completed already. The Gantt chart is now updated, which if followed correctly, everything will be completed on time.

7.2. Budget

Each project member had \$250 for the project, hence we had \$500 for the entire project as two members were present in the group.

The pure quartz slide was purchased for the mass spectrometer

The DNA genetic testing kit was purchased so we could test our DNA.

A table of what items were purchased and how much each item cost is seen in Table 1.

Table 1: Budget Table

Item	Quantity	Cost
Pure Quartz Slide	2	\$40 each
DNA Genetic Testing Kit + Shipping	2	\$200 each
	Total Cost	\$480

7.3. Risk Management

A risk assessment table for the project can be seen in Table 2. One of the major risks that was encountered was bugs in the code. This caused lots of issues throughout the project, one of the main issues was that it delayed the project from proceeding further, this led to task completion delay. Another risk was the chance of having the Human Ethics not approved. This would mean that we could not do the mass spectrometer task and the DNA task, meaning we would have to do more in-depth analysis on the first task and find additional tasks to do. Slight misunderstanding of project tasks happened throughout the project, but these were explained through meetings with the supervisor.

Table 2: Risk Management

Risk	Likelihood	Rating	Risk Estimation
1. Bugs in code	Likely	Moderate	High
2. Data loss	Rare	Major	Medium
3. Illness/absent	Unlikely	Minor	Low
4. Misunderstanding project tasks	Rare	Minor	Low
5. Task completion time delay	Likely	Moderate	High
6. Human Ethics not approved	Unlikely	Moderate	Medium

8. Conclusion

This project has made some progress, with the code analysis task near competition.

Unfortunately, the hair analysis was meant to be completed by this semester. This means that during the break, the project team has to finish the hair analysis task by then, to be on schedule for Semester B.

The main task was to find out if horse names represent the mysterious code. Preliminary conclusions for the code analysis task, showed that the mysterious code is unlikely to be horse names. It was also discovered that is unlikely to be South Australian street names, Australian city names or The Rubaiyat of Omar Khayyam book. Even with further studies on the book, it was still unlikely to be the mysterious code.

There are no preliminary conclusions for either the hair analysis or DNA analysis tasks, as they both have not been started, but will be completed in Semester B.

8.1. Future work

The future work that will be completed in Semester B are:

- Collect controlled sample hairs and use the mass spectrometer on them. Then use mass spectrometer on the Somerton Man's hair.
- Analysis of the isotopes in the hair, followed by identification of how long he was living in Adelaide before his death.
- Begin work degrading the DNA samples to understand how much you can degrade DNA until it becomes unidentifiable, by removing SNPs.
- If time permits have more collective objects for code analysis.

9. References

- [1]. J. Bineth. "Somerton Man cold case," *ABC news*, 14th December 2017 [online]. Available: <http://www.abc.net.au/news/2017-12-14/somerton-man-cold-case-could-be-one-step-closer-to-solved/9245512>
- [2]. *The News*, "Dead Man Found Lying on Somerton Beach," 1st December 1948, p. 1 [online]. Available: <http://trove.nla.gov.au/ndp/del/article/129897161>
- [3]. *The Advertiser*, "Cryptic note on body," 9th June 1949, p. 1 [online]. Available: <https://trove.nla.gov.au/newspaper/article/36371152>
- [4]. *The Advertiser*, "New Clue in Somerton Body Mystery," 25th July 1949, p. 3 [online]. Available: <https://trove.nla.gov.au/newspaper/article/36677719>
- [5]. "Tamam Shud/ Somerton Man," [online]. Available: <http://ciphermysteries.com/tamam-shud-somerton-man>
- [6]. *The Advertiser*, "Tamam Shud," 10th June 1949, p. 2 [online]. Available: <https://trove.nla.gov.au/newspaper/article/36371416>
- [7]. Maguire, S. "Death riddle of a man with no name," *The Advertiser*, 9th March 2005, p. 28 [online]. Available: <http://www.eleceng.adelaide.edu.au/personal/dabbott/tamanshud/advertiser-mar2005.pdf>
- [8]. Inside Story, presented by Stuart Littlemore, ABC TV, screened at 8 pm, Thursday, 24th August, 1978
- [9]. A. Turnbull and D. Bihari. (2009). Final Report 2009: Who killed the Somerton man? [online]. Available: https://www.eleceng.adelaide.edu.au/personal/dabbott/wiki/index.php/Final_report_2009:_Who_killed_the_Somerton_man
- [10]. K. Ramirez and L-V. Michael. (2010). Final Report 2010 [online]. Available: https://www.eleceng.adelaide.edu.au/personal/dabbott/wiki/index.php/Final_Report_2010

- [11]. S. Maxwell and P. Johnson. (2011). Final Report 2011 [online].
Available: https://www.eleceng.adelaide.edu.au/personal/dabbott/wiki/index.php/Final_Report_2011
- [12]. A. Duffy and T. Stratfold. (2012). Final Report 2012 [online].
Available: https://www.eleceng.adelaide.edu.au/personal/dabbott/wiki/index.php/Final_Report_2012
- [13]. L. Griffith and P. Varsos. (2013). Semester B Final Report 2013 – Cipher Cracking [online]. Available:
https://www.eleceng.adelaide.edu.au/personal/dabbott/wiki/index.php/Semester_B_Final_Report_2013_-_Cipher_cracking
- [14]. N. Gencarelli and J. Yang. (2015). Semester B Final Report 2015 – Cipher Cracking [online] Available:
https://www.eleceng.adelaide.edu.au/personal/dabbott/wiki/index.php/Final_Report/Thesis_2015
- [15]. Y. Li and Y. Ma. (2016). Semester B Final Report 2016 – Cipher Cracking [online] Available:
https://www.eleceng.adelaide.edu.au/personal/dabbott/wiki/index.php/Final_Report/Thesis_2016
- [16]. B. David. “P Value and the Theory of Hypothesis Testing: An Explanation for New Researchers,” *Clinical Orthopaedics and Related Research*®, Vol.468 (3), pp.885-892 2010.
- [17]. G G. L et al. “What is the Value of a p Value?,” *The Annals of Thoracic Surgery*, Vol.87(5), pp.1337-1343 2009.
- [18]. T. Hung. “The p-value,” 21st March 2016 [online] Available:
<https://www.students4bestevidence.net/p-value-in-plain-english-2/>
- [19]. “Mass Spectrometer,” Premier Biosoft [online] Available:
http://www.premierbiosoft.com/tech_notes/mass-spectrometry.html
- [20]. R. Bakhtiar. “Biological mass spectrometry: A primer”. *Mutagenesis* pp.530-531 2000.

[21]. S. Jackson et al. "The application of laser ablation-inductively coupled plasma-mass spectrometry to in situ U–Pb zircon geochronology," *Chemical Geology*, Vol. 211, pp.47-69 2004.

[22]. "What are single nucleotide polymorphisms?," U.S National Library of Medicine [online] Available: <https://ghr.nlm.nih.gov/primer/genomicresearch/snp>

[23]. G. Shaw. "Polymorphism and Single nucleotide polymorphisms (SNPs)" *Science Made Simple*, Vol. 112, pp.664-665 2013.