Code Cracking: Who Murdered the Somerton Man

Group 142

Supervisors: Prof. Derek Abbot

Dr. Matthew Berryman

Students: Yami Li

Yifan Ma

Outline:

1. **Introduction**
2. **Motivations & Expectations**
3. **Previous Contributions**
4. **Specific tasks**
   - Code Analysis
   - Mass Spectrometer Data Analysis
5. **Future Works**
6. **Conclusions**

1. Introduction

Background: Who was the Somerton Man?

- A Man found died in Somerton Beach, Glenelg

- Time: 1st December 1948

- Identity remains unknown

Figure1. Tombstone of the Somerton Man [1]

Introduction

Background: Dubious facts in this case

- Victim was **NOT** a local resident.

- Victim dressed well in midsummer and met his demise while sleeping.

- Found a hidden Paper scrap in his trousers.

Figure2. Paper scratch found by Police [2]

Introduction

Background: Paper Scrap

- *Tamám Shud*: Means **finished or to the end** in English.

- Proved to be part of the 'Rubaiyat of Omar Khayyam' book.

- Related book was found later.

Background: Mysterious code

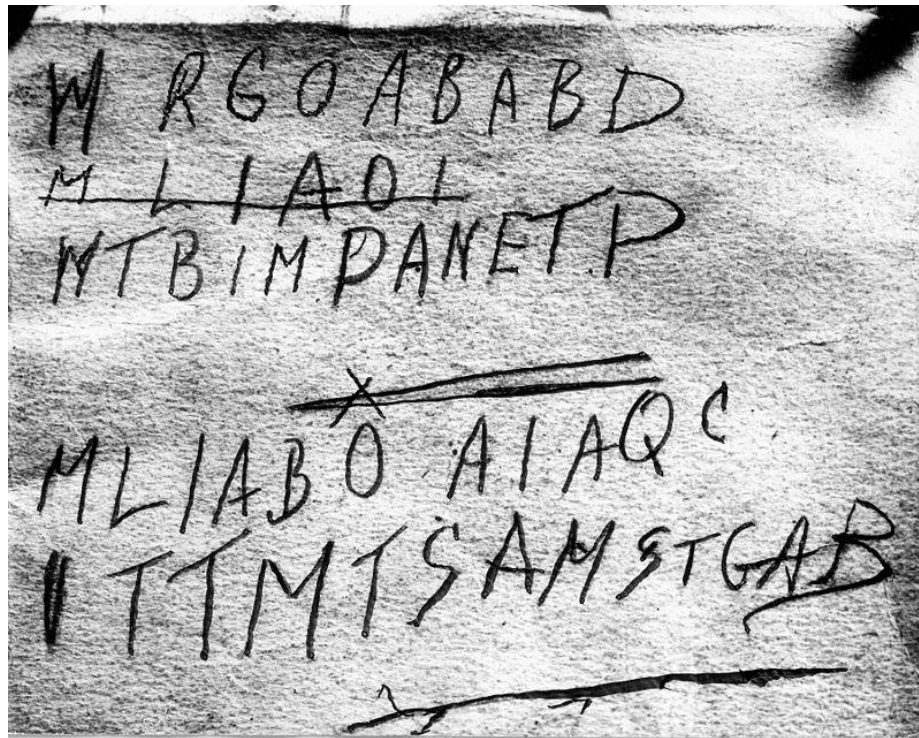The code was found written on the back of the book.

Figure3. The mysterious code[3]

Introduction

Background: Hair found accidentally

- A plaster model was made as per the Somerton man's upper body.

- One of his hair was found in the unformed plaster.

Figure4. Plaster mode of the Somerton Man[4]

Motivation of project

1. Show respect for the dead man.

2. Engineers' advantages in statistics, information theory and data processing.

3. It is cool acting like a detective.

Expectation of project

1. Determine the Language of the Code

2. Reveal the previous living conditions of the Somerton Man.

Previous Contributions

- SA Police: Provided initial investigations and related data.

- Australian Defense Force: Attempted to crack the code yet unable to crack.

- Final year project groups in UoA: Code consists of initialisms of a certain language[5].

Specific Tasks: Code Analysis

- Premise: The codes represents first letters of a series of words.

- Task: Determine the language of the code

- Method: Data reshaping, Similarity Check, Statistical Analysis

Code Analysis:

Step1: Generate the library for comparison use.

- *Criteria:*

    Having various translations in different languages.
    Long enough (at least 10k words.)

- *Selected Material:*

    The Declaration of Human Right (UDHR)
    War and Peace by Leo Tolstoy (excerpt)

Step2: Reshape to Initialisms

- Apply de-accent function to turn accent letters to English letters, remove all punctuations and blanks.

Hielten sich für
Captain Kirk, das gab
ein großes Feuerwerk

→

Hielten sich fur
Captain Kirk das gab
ein grobes Feuerwerk

- Extract first letters of every word and save as a string.

Text

Array

I confess that when first I
made acquaintance with
Charles Strickland…….

→

I, C, T, W, F, I, M, A, W, C, S…..

Step3: Implement Similarity Check Algorithms

1. Levenshtein Distance:

Calculate the minimum deletion, insertion and substitution times of turning one string into another

```
I N T E # N S I O N
| | | | | | | | | |
# E X E C U S I O N
| | | | | | | | | |
d s s — i s ————
```

Figure3. The mysterious code[6]

Step3: Implement Similarity Check Algorithms

2. Simhash Algorithm:

Developed by Moses S. Charikar in 2003.

Previously applied by Google for duplicative webpages removing.

Relatively simple but effective.

Simhash Algorithms

Simhash

feature, weight   hash, weight

Doc.

w_1 → 100110 w_1 → w_1 $-w_1$ $-w_1$ w_1 w_1 $-w_1$

w_2 → 110000 w_2 → w_2 w_2 $-w_2$ $-w_2$ $-w_2$ $-w_2$

w_n → 001001 w_n → $-w_n$ $-w_n$ w_n $-w_n$ $-w_n$ w_n

add

sign

110001 fingerprint ← 13,108,-22,-5,-32,55
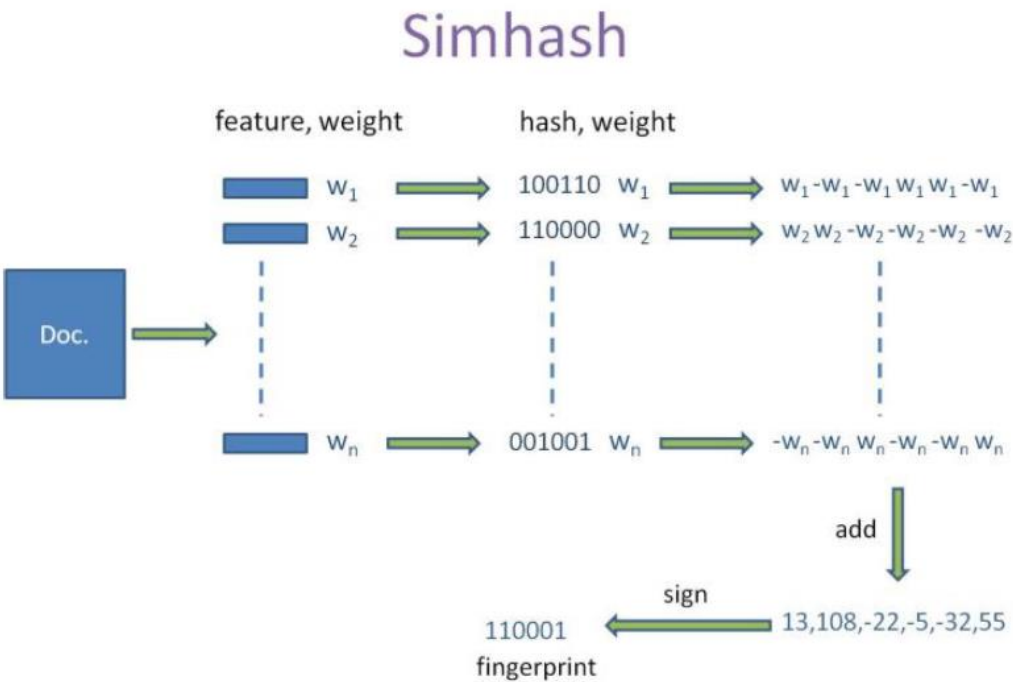
Figure4. Simhash Procedures[7]

1. Separate the string into n individual terms.
2. Calculate the weight of each term.
3. Apply hash mapping to each term, set 0 to -1, multiple with the corresponding weight.
4. Add all results in step3 together. Set positive bit to 1 and others to 0.

Code Analysis:

Test Layout:
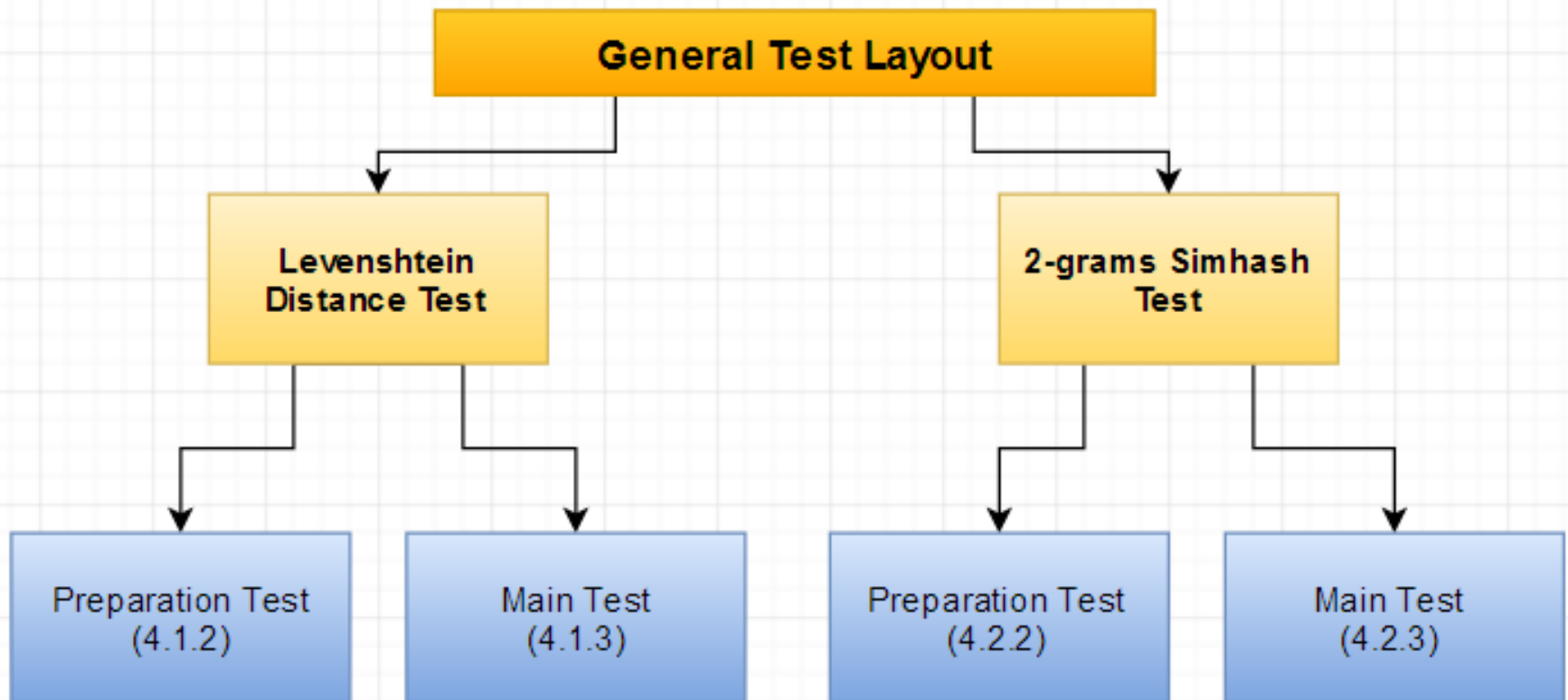
Levenshtein Effectiveness Test:
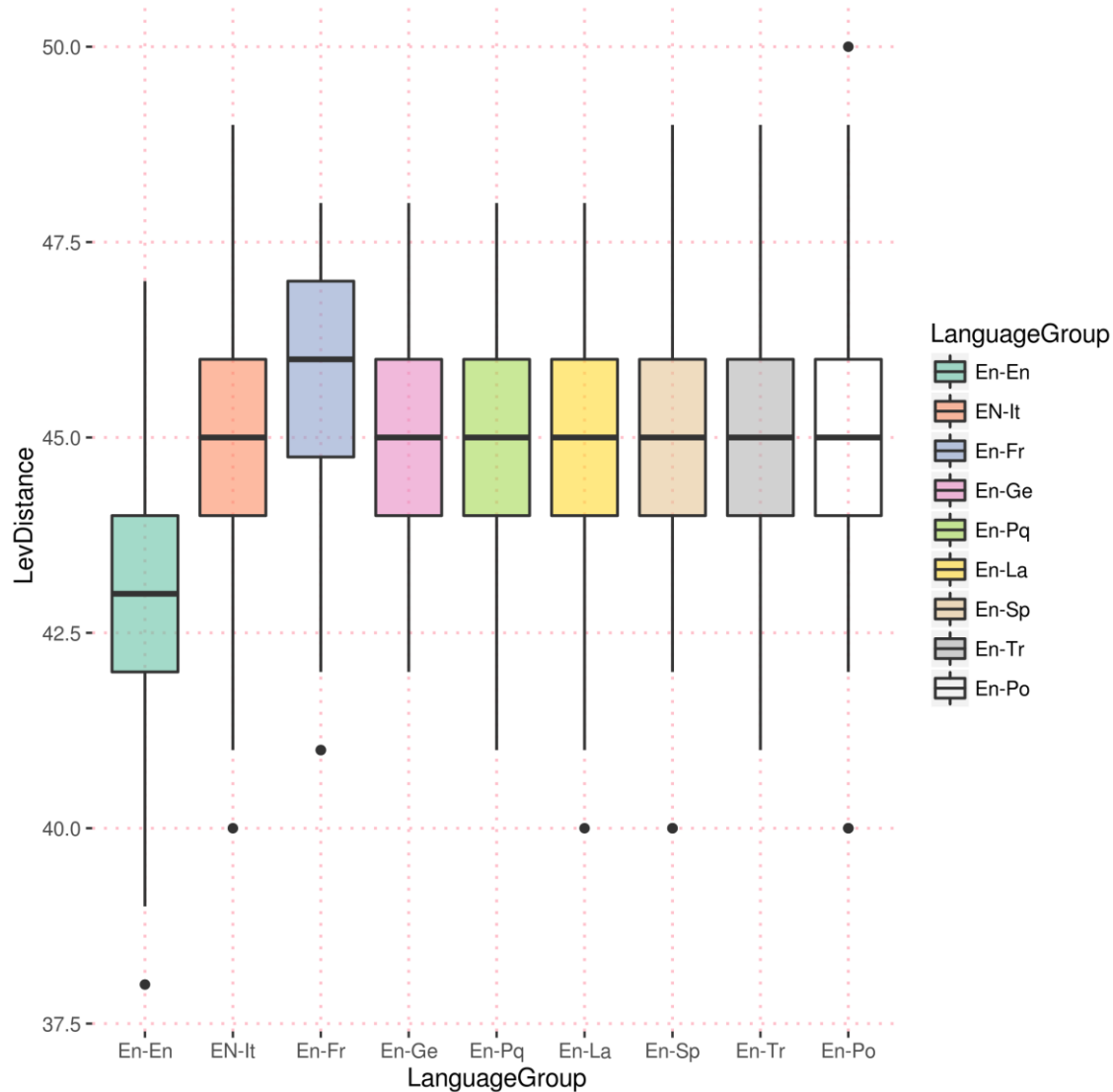
Each box represents a set of Levenshtein Distances between 50 string pairs from War and Peace.
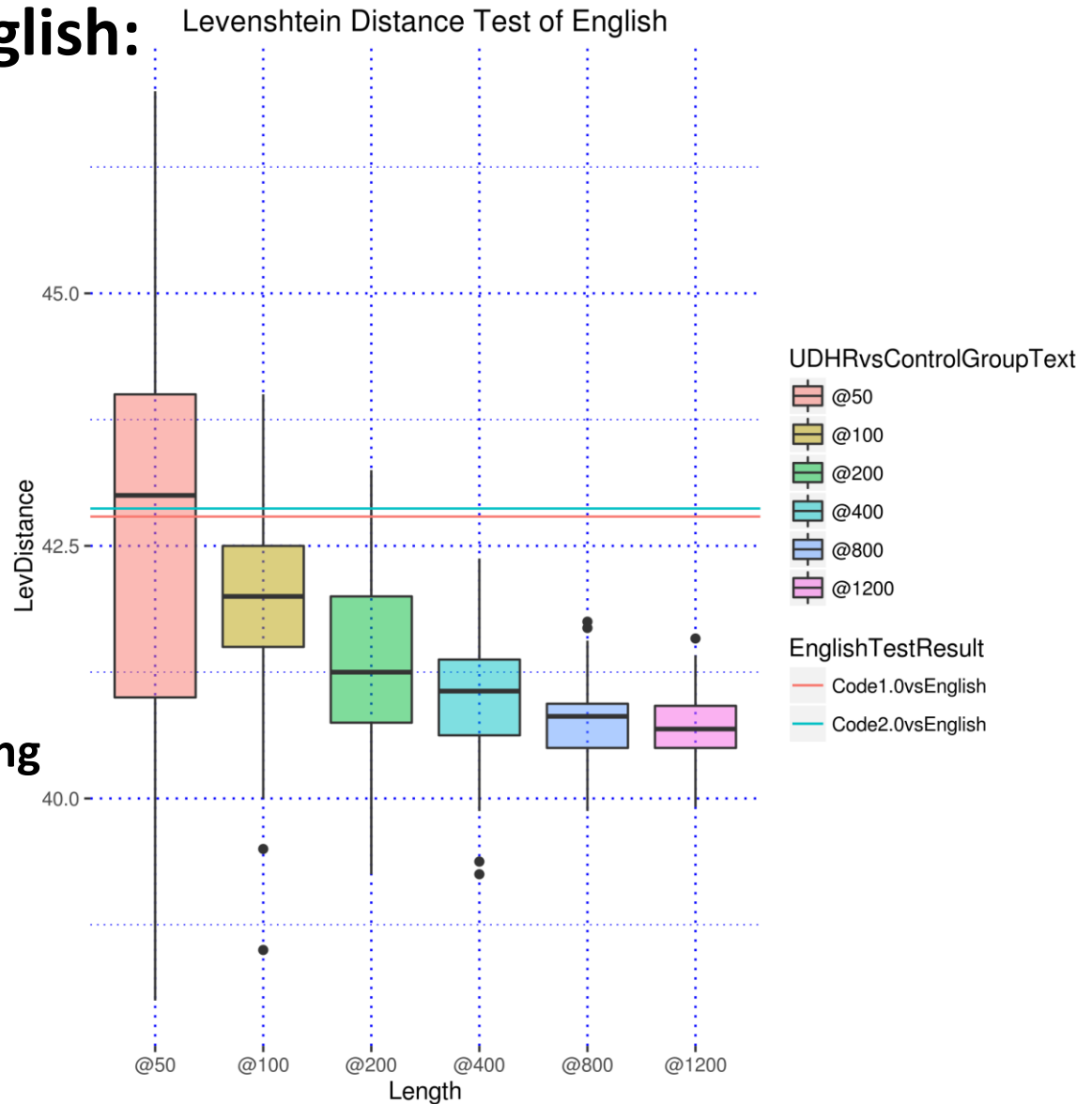
Left box was the lowest

Code Analysis:

Levenshtein Test, English:

Boxes: Levenshtein Distance between 50 UDHR and War and Peace pairs.

Line: Levenshtein Distance between code and War and Peace
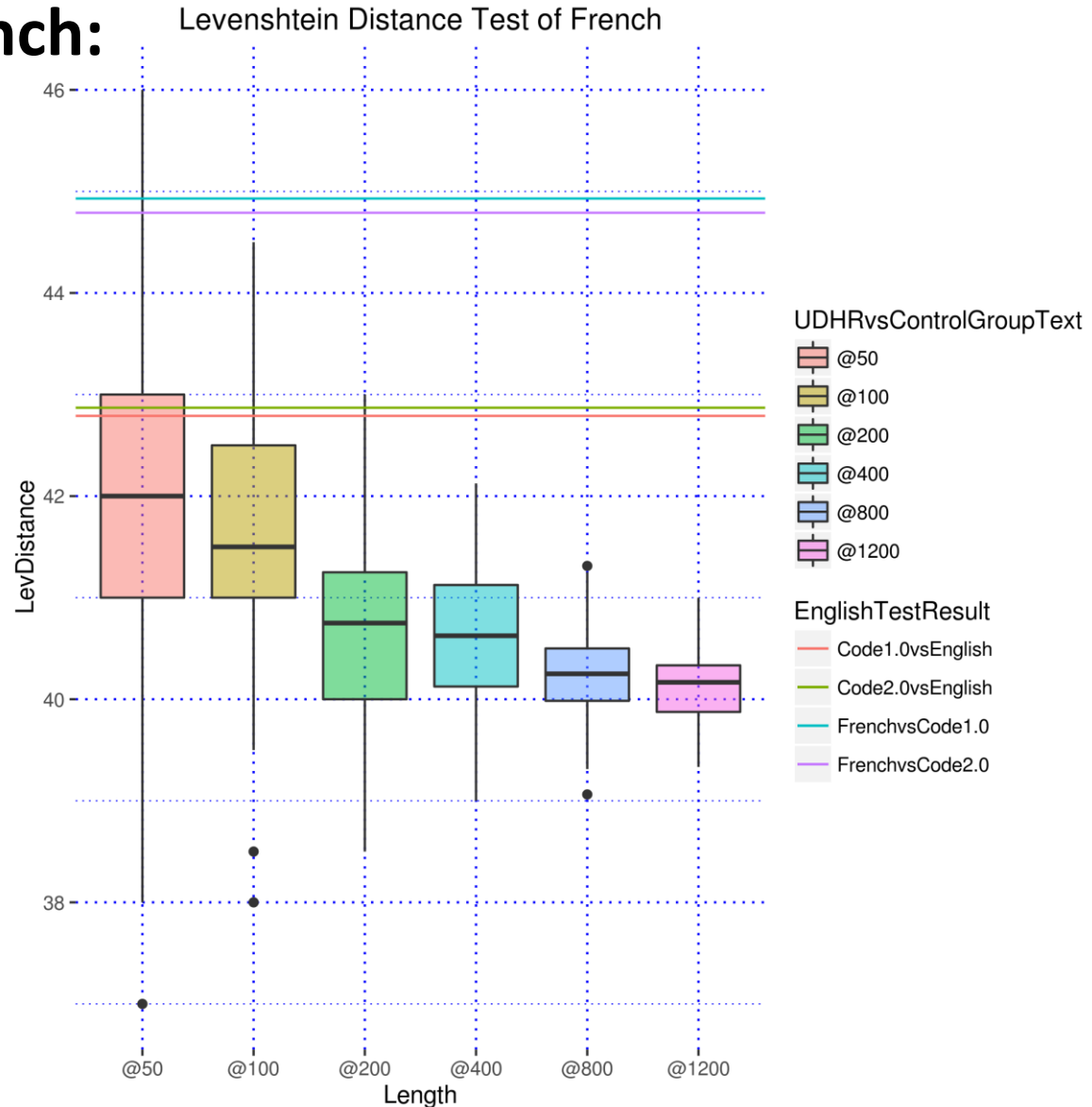
X-Axis: Length of string ranging from 50 to 1200 letters.

Levenshtein Distance Test of English

Code Analysis:

Levenshtein Test, French:

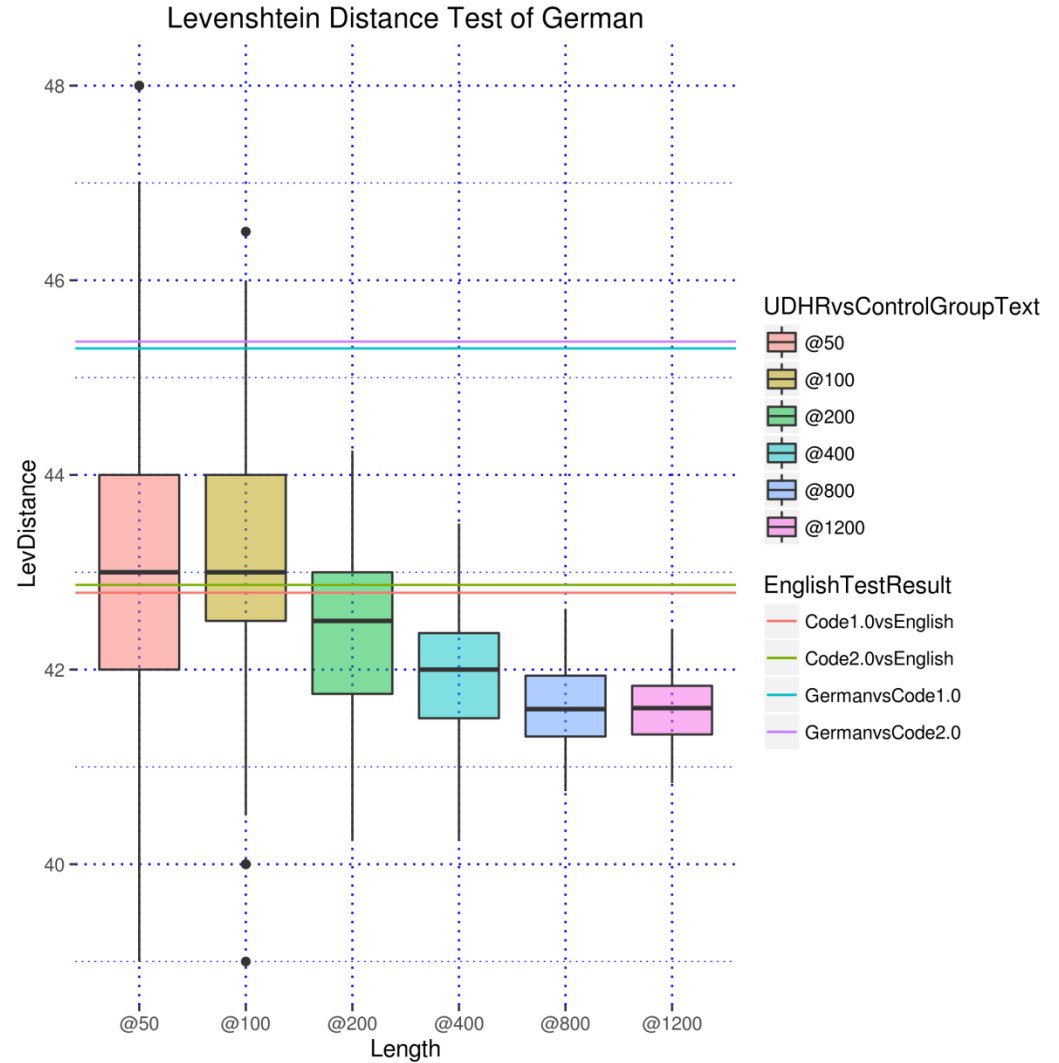**Boxes: Levenshtein Distance between 50 UDHR and War and Peace pairs in French.**

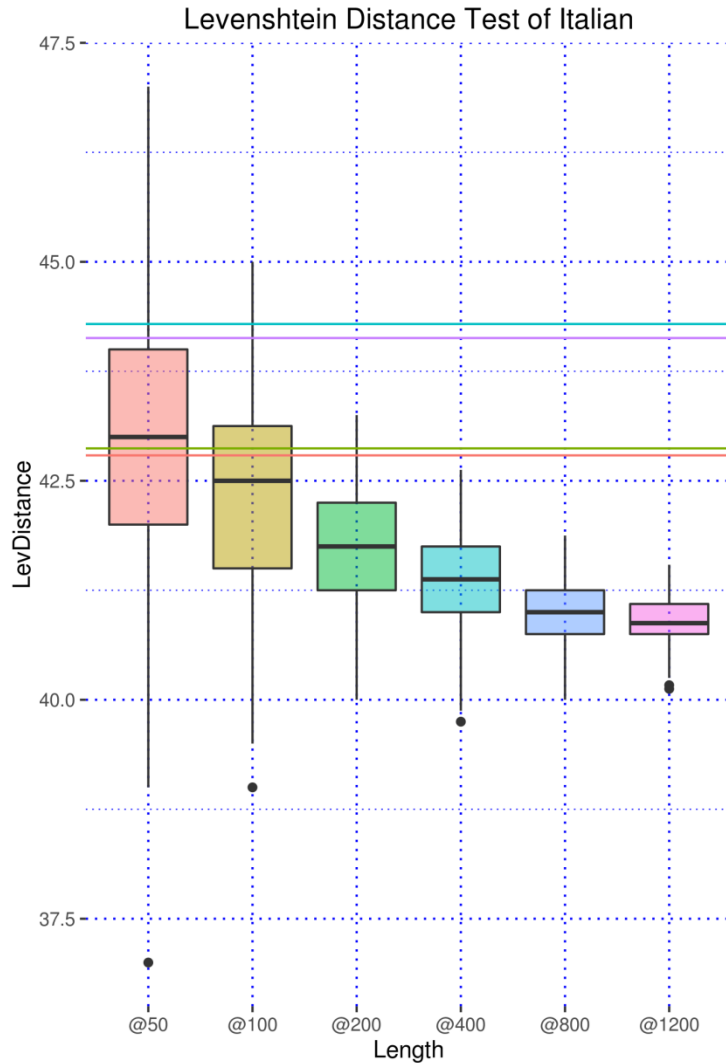**Red & Green Line: Codes vs War and Peace in English**

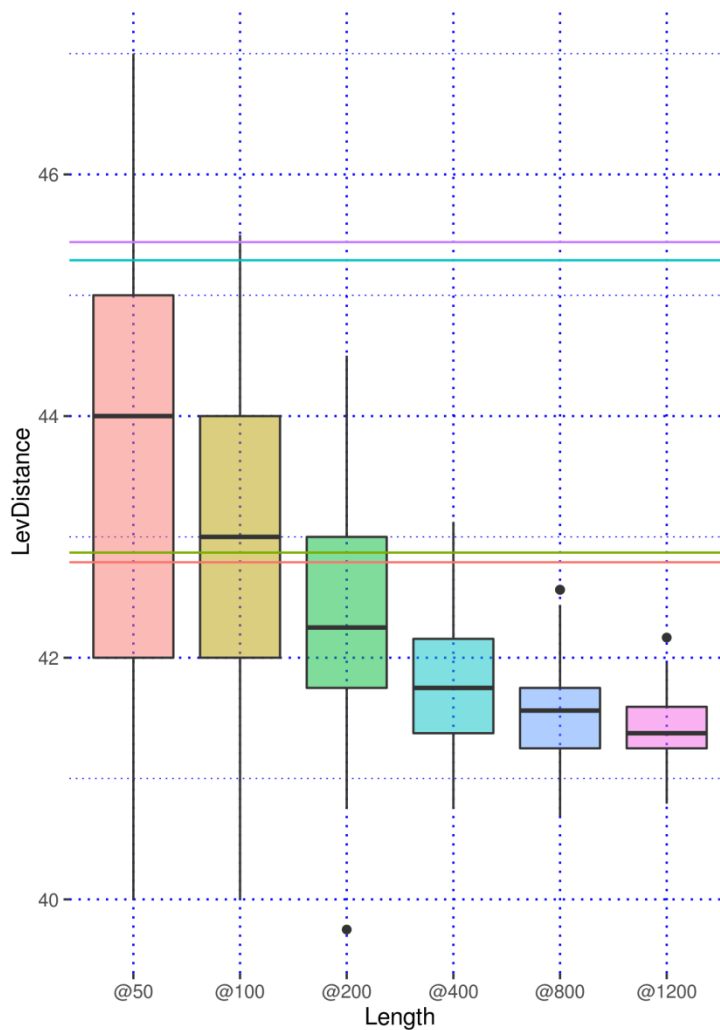**Blue & Purple Line: Codes vs War and Peace in French**

Levenshtein Distance Test of French

UDHRvsControlGroupText
- @50
- @100
- @200
- @400
- @800
- @1200

EnglishTestResult
- Code1.0vsEnglish
- Code2.0vsEnglish
- FrenchvsCode1.0
- FrenchvsCode2.0

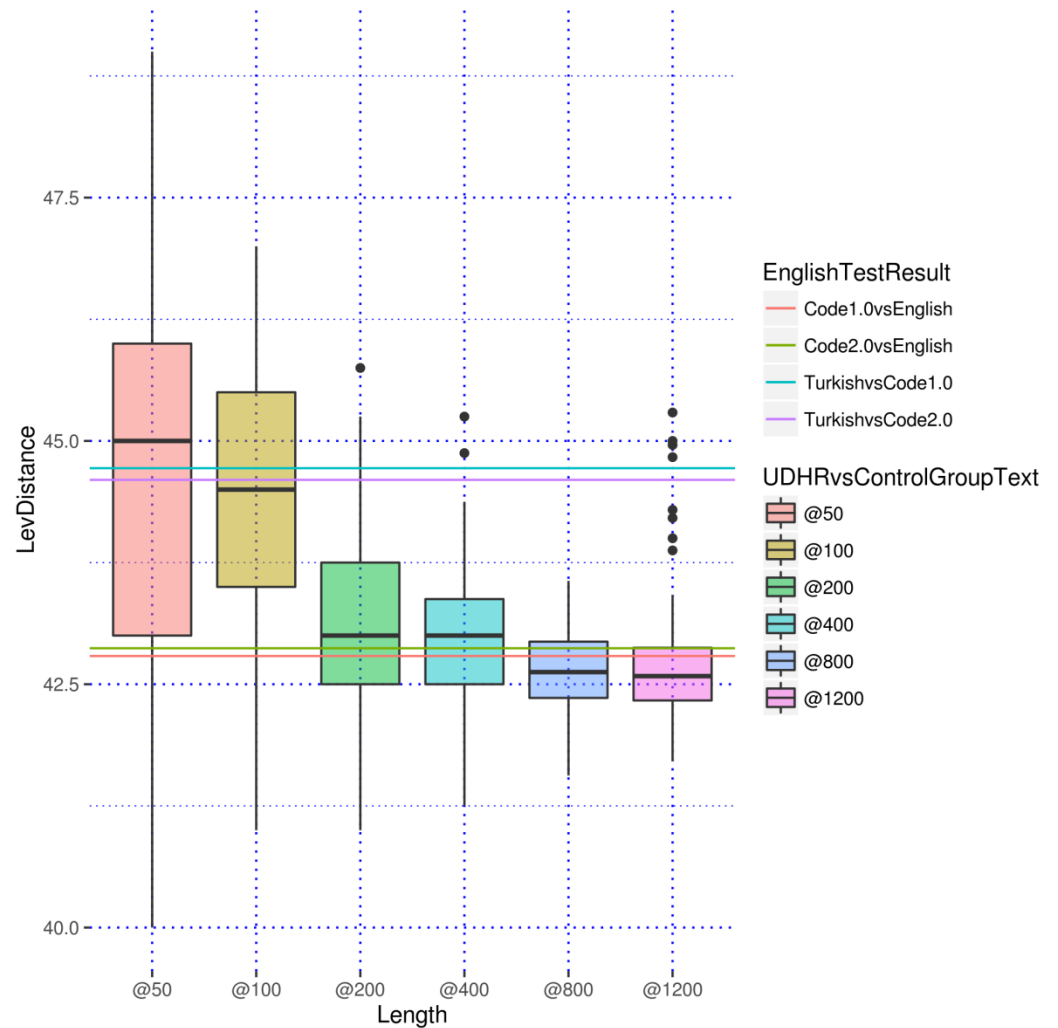# Code Analysis:

Italian and German:

Code Analysis:

Polish and Portuguese:
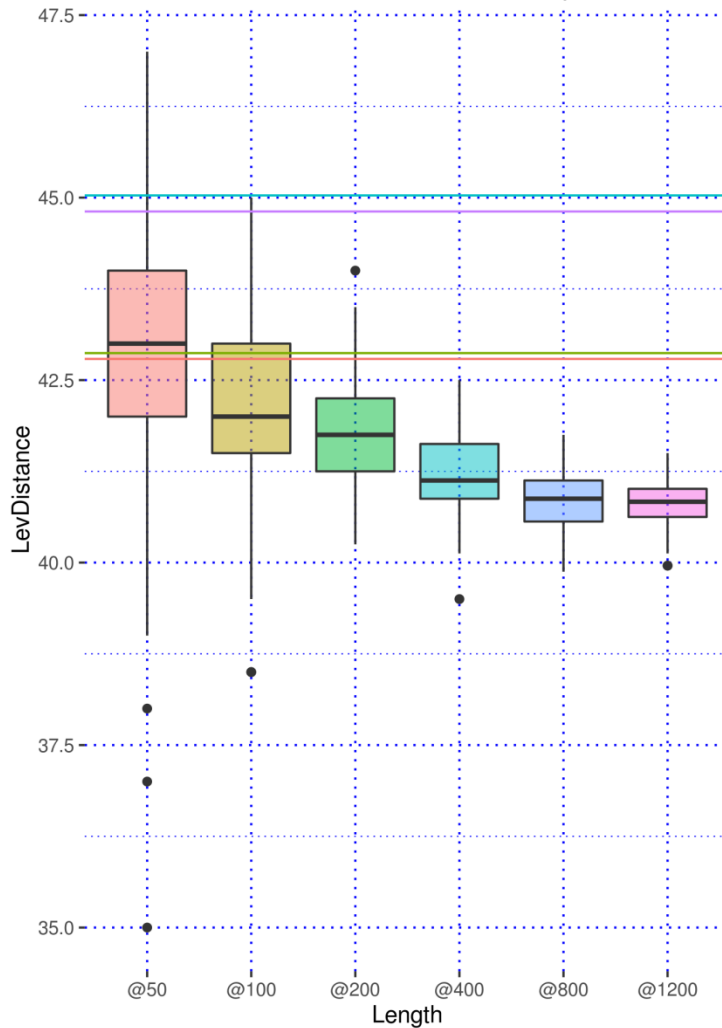
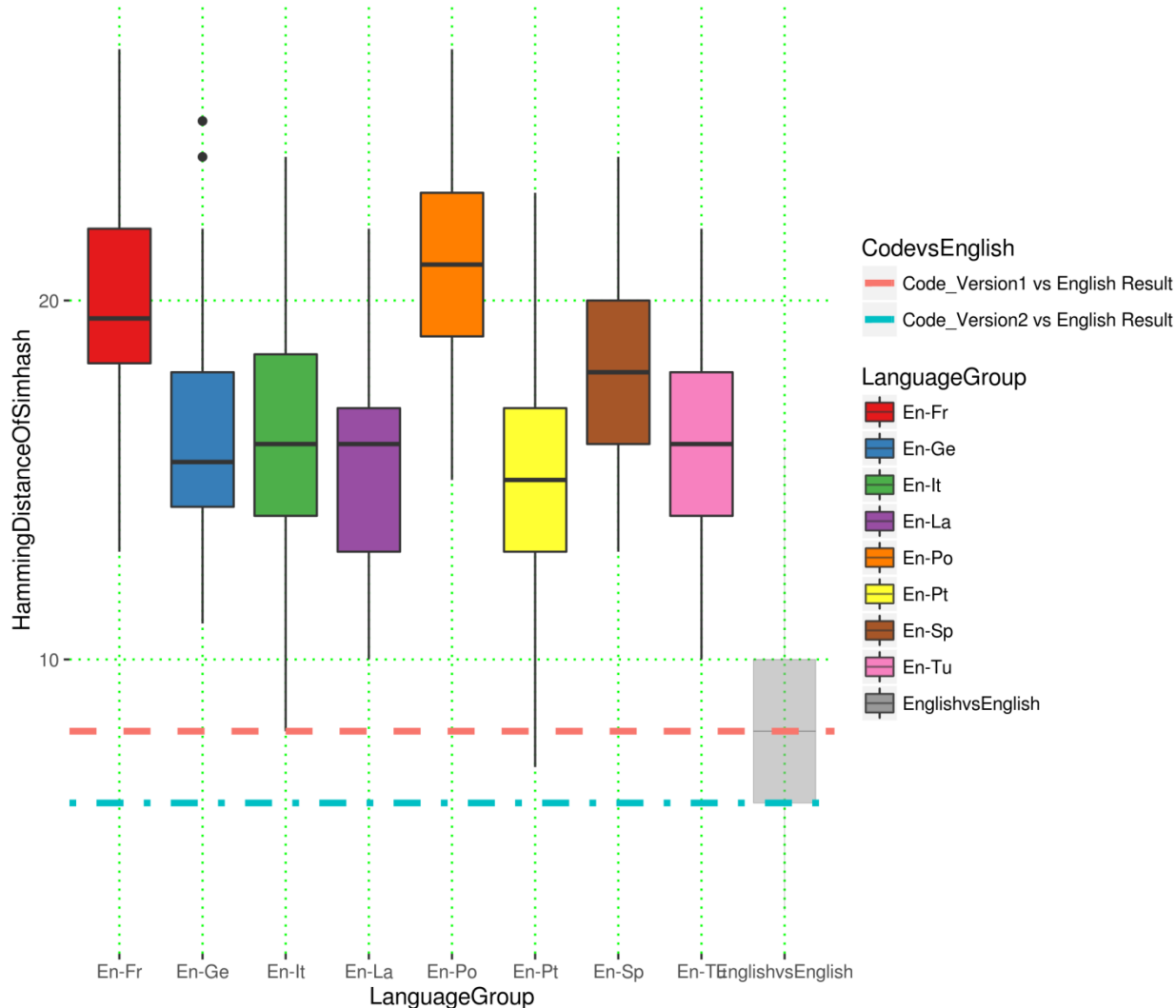# Spanish and Turkish:

Code Analysis:

Levenshtein Test Conclusion:

1. **Code vs English group has the minimum Levenshtein Distance.**

2. **Credibility of this test has been challenged by 50 – letters long group's SD being large.**

3. **Needs another new algorithm.**

Code Analysis:
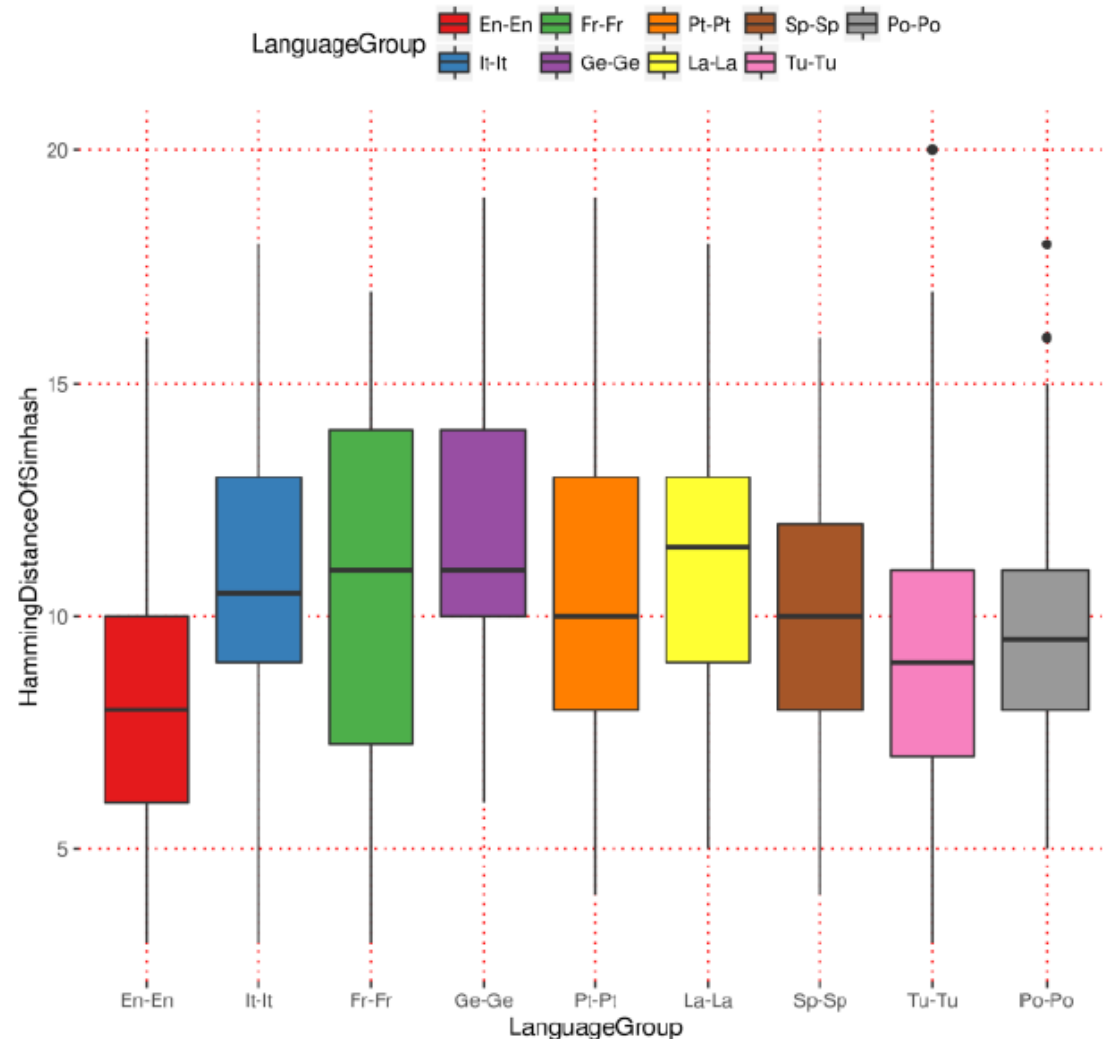
Simhash Effectiveness Test 1:

Each box represents a set of Hamming Distances between Simhash strings of 50 string pairs from War and Peace.

Transparent one indicates English vs English result. Colored boxes indicate test results of English vs other languages.
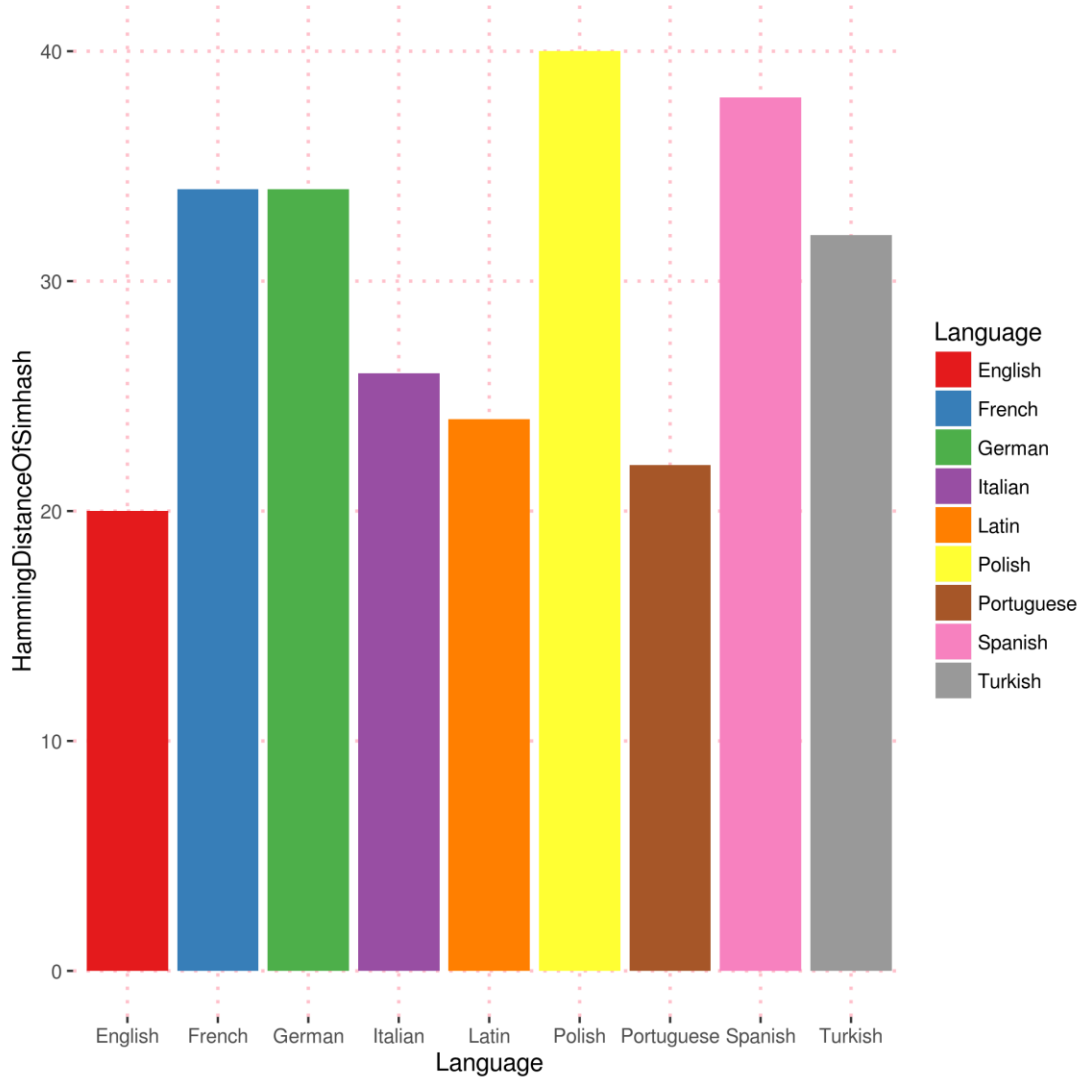
Simhash Effectiveness Test 2:

Same-Language SimHash Test

Each box represents a set of Hamming Distances between Simhash strings of 50 string pairs from War and Peace **in a same language**

Simhash Test:

Simhash Test of Mixed Code vs War and Peace in Different Languages

Hamming distances between Simhashed Code and Simhashed War and Peace strings in various languages.

Code Analysis:

Simhash Test Conclusion:

1. Code vs English group has the minimum Hamming Distance

2. No signs of significant defects found.

Final Conclusion:

Code consists of Initialisms of English.

Code Analysis:

Code Analysis Task Finished.

Next section: Mass Spectrometer Data Analysis

Mass Spectrometer Data Analysis related. Introduction:

- Plaster cast made in 1949

- Hair extraction of Somerton man
  - ✓ From plaster cast

- Comparison with other control samples
 - ✓ Selected at random from university students

<Source:www.sapolicehistory.org>

Inductively Coupled Plasma Mass Spectrometer ICP-MS

- Used for micro-sampling of solid material for trace element and predominantly cation analysis

- Adelaide Microscopy houses two Agilent ICP-MS instruments

  ✓ Measured the level of different isotopes present in a sample
  ✓ Burned the hairs with a laser and recorded the spectral elements

<source: www.adelaide.edu.au/microscopy/instrumentation/icpms>
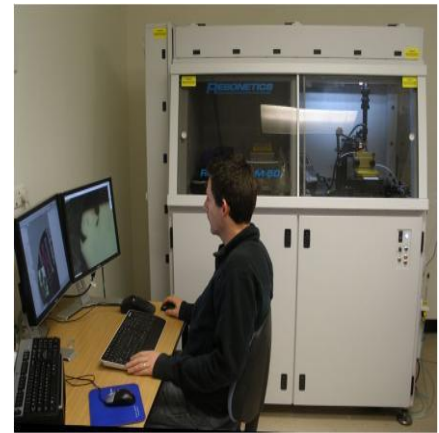
Mass Spectrometer Data Analysis
Glass test:

Laser scan speed: 5um/s

- ## Control samples
  - Scan length: 1000um
  - Method length: 230s(30s background)

- ## Somerton man's hair
  - Scan length: 6054.6um
  - Method length: 1240.9s(30s background)

Mass Spectrometer Data Analysis
Glass test:

- ## Hair growth rate: about 0.4mm/day
  - ### Control sample: 2.5 days
  - ### Somerton man: 2 weeks

root

<source:www.forensicmag.com>

Mass Spectrometer Data Analysis
Glass test:

These elements gained by laser ablation of hair mass spectrometer				
Li7	Ca43	Cu65	Zr90	Hf178
B11	Sc45	Zn66	Mo95	Au197
Na23	Ti47	Ga71	Ag107	Hg202
Mg24	V51	As75	Cd111	Tl205
Al27	Cr52	Se77	Sn118	Pb206
Si29	Mn55	Se82	Sb121	Bi209
P31	Fe57	Rb85	Te125	Th232
S34	Co59	Sr88	Cs133	U238
K39	Ni60	Y89	Ba137	

Placed on glass slide :

- Somerton man's hair with hair root

- Hairs of control samples

- Recorded 44 elements

Mass Spectrometer Data Analysis
Glass test:

Somerton man's hair data plot by Excel

- Loss of data

- Tendency unclear

- Some traces out of range

- Some traces changing curve are not obvious
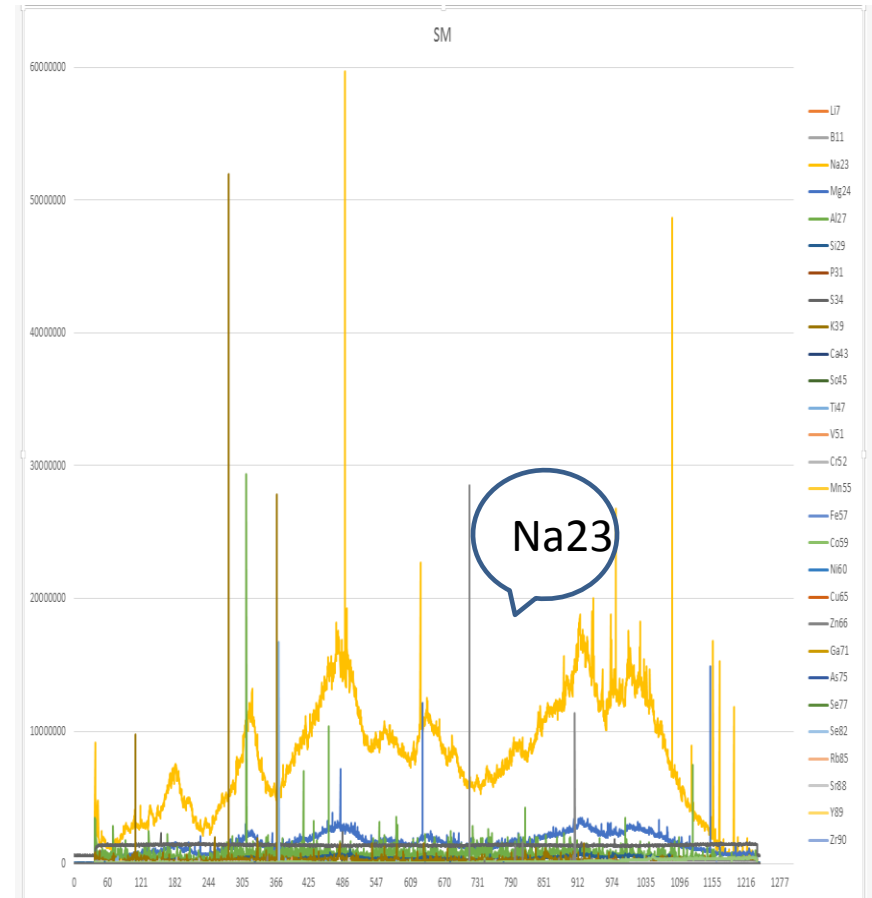
- Have difficulties in contrast

Figure: All elements comparison figure

Mass Spectrometer Data Analysis
Glass test:

Lead (Pb206)

- Beginning at the left side of hair (root) through to the end

- High level of Pb206

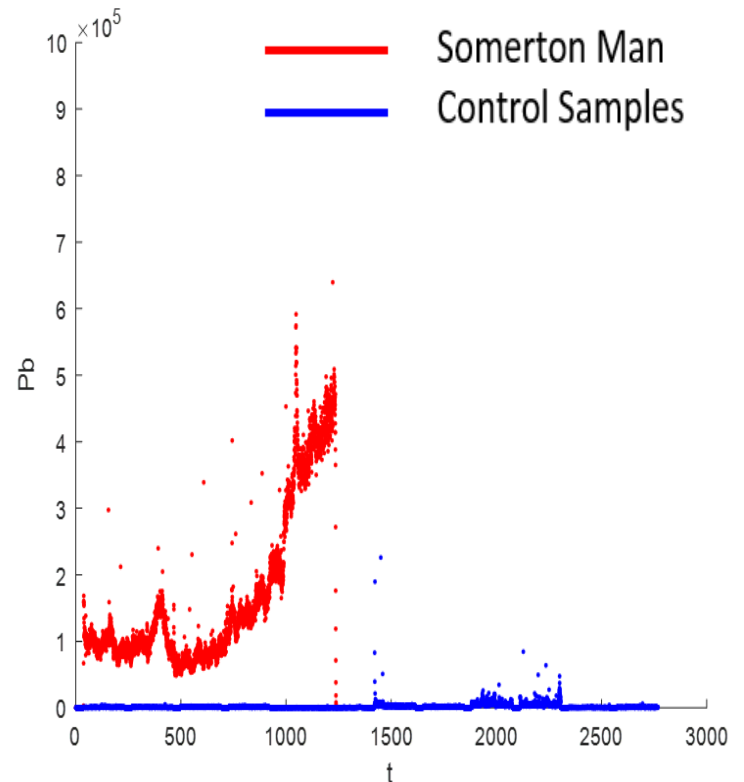- Reducing significantly near the time of death

- Not the cause of his death

Figure: Lead (Pb206) content comparison

Mercury (Hg202)

- Mercury is a highly toxic heavy metal pollutants of biologic toxicity

- The main pollution of mercury is from plastics, batteries, electronics and other industrial emissions.

- The higher level of mercury may be due to Somerton man's living area.

- Before dying there is not much change

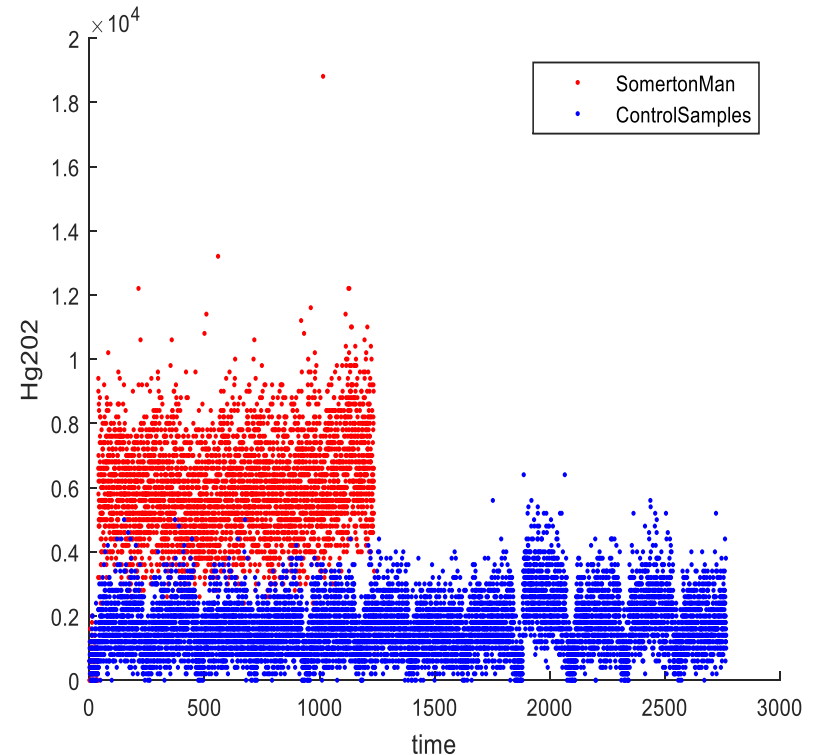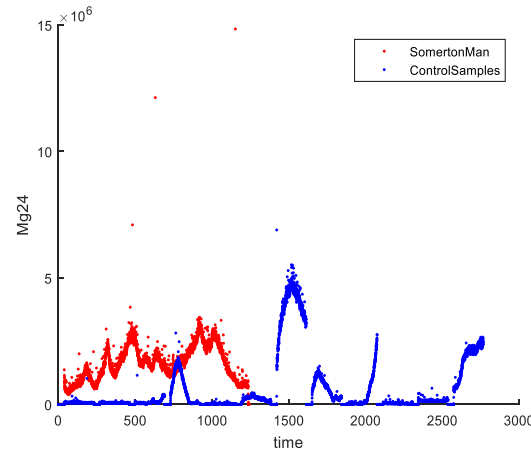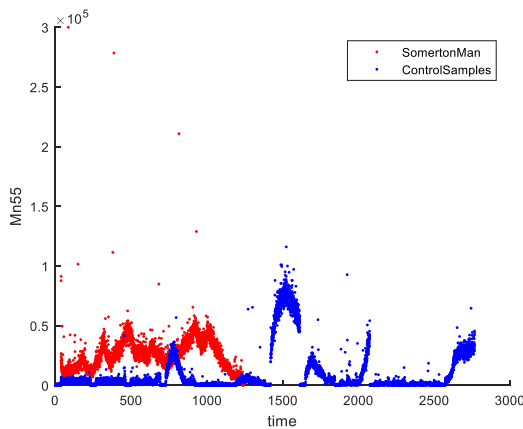- Thus, this should not be regarded as the cause of his death.
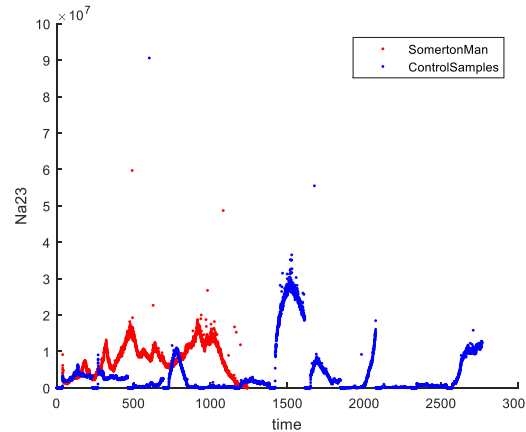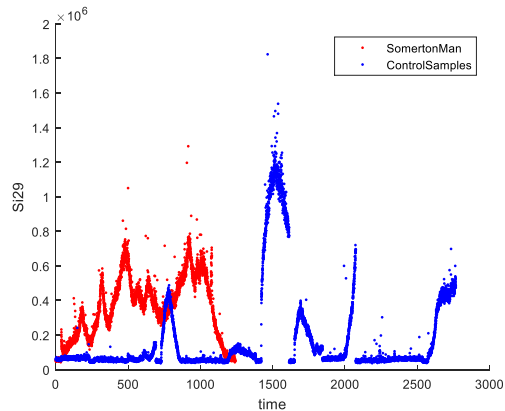

Figure: Mercury(Hg202) content comparison

Mass Spectrometer Data Analysis
Glass test:

- Si29, Na23, Mn55, Mg24 have the similar trend

- Impure background (glass) affect the data

- These elements contained in the glass

Figure: Silicon(Si29), Sodium(Na23), Manganese(Mn55), Magnesium(Mg24) content comparison

Mass Spectrometer Data Analysis
Quartz test:

Elements gained by laser ablation on quartz test

Li7	C13	Na23	Mg24	Al27	Si28
S34	K39	Ca43	Ti47	V51	Cr52
Mn55	Fe57	Co59	Ni60	Cu63	Zn66
As75	Se77	Se82	Sr86	Sr87	Sr88
Ag107	Cd111	Sn118	Au197	Hg202	Tl205
Pb206	Pb208	Bi209	Th232	U238	

Different elements table

Glass	Zr90	Hf178	B11	Sc45	Mo95	Ga71	Sb121
	Te125	Rb85	P31	Cs133	Y89	Ba137	
Quartz	Sr86	Sr87	Pb208	C13			

Placed on quartz slide :
- Somerton man's hair without hair root

- Hairs of control samples

- A bit rest of Somerton man's unburned hair (the glass test remainder)

- 35 elements recorded from quartz test

Mass Spectrometer Data Analysis
Quartz test:
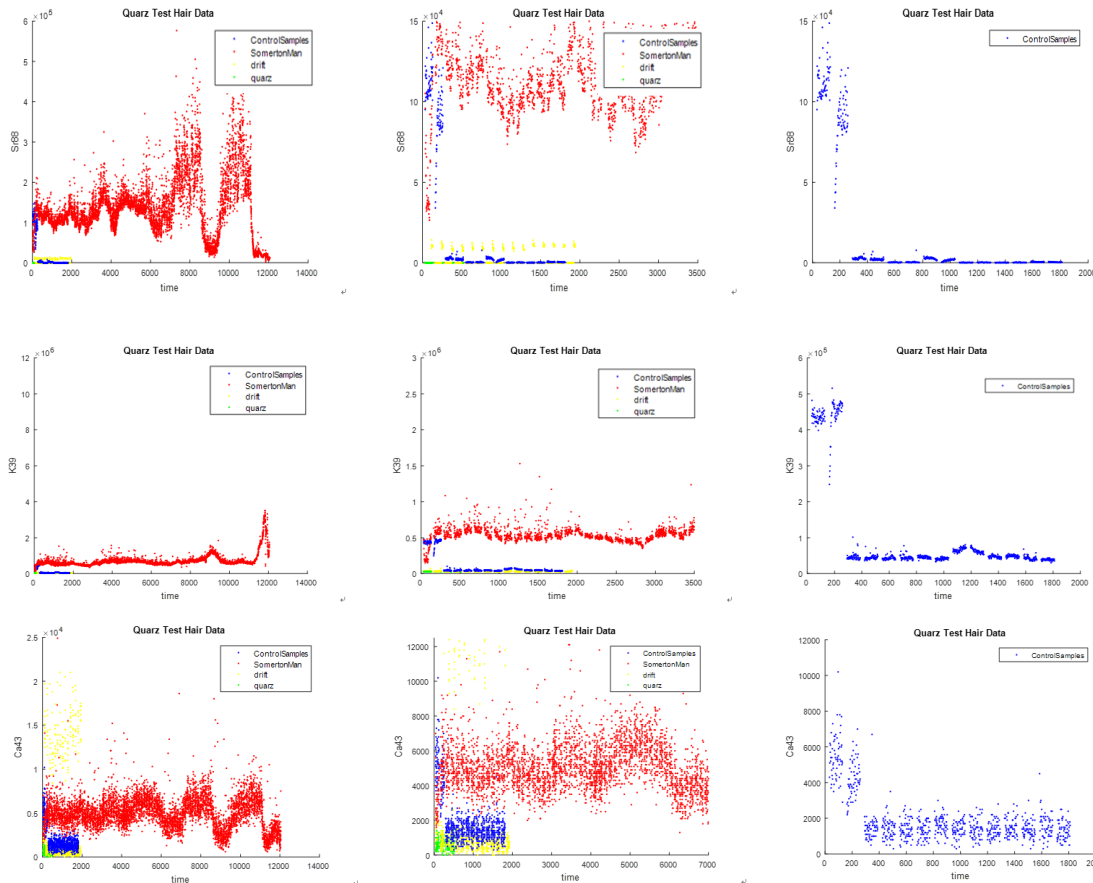

Figure: Sulphur(S34) content comparison
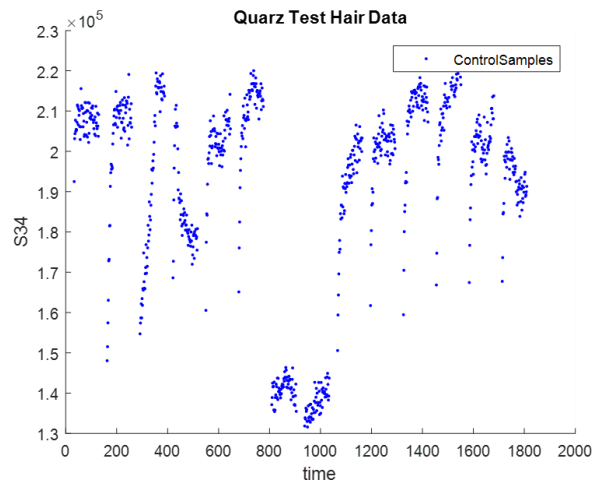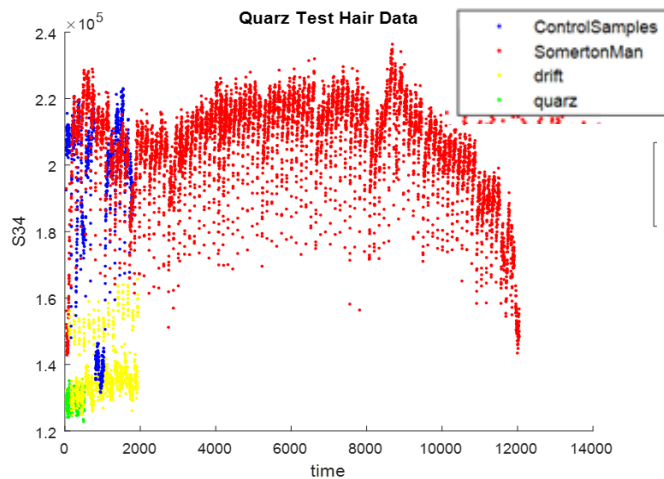
- By comparison, the sulphur relative values are similar among Somerton man and 12 control hairs.

- Only two samples have significant differences and these samples come from the same person's hair.

- Sulphur is an essential elements in human hair.

- The change in the content of the element can be a direct indicator of a person's living environment.

Mass Spectrometer Data Analysis
Quartz test:

Figure: calcium(Ca43), potassium(K39) and strontium(Sr88) content comparison

- The relative values of calcium, potassium and strontium in Somerton man' hair are higher than most control hairs.
- The content of these elements in two samples is close to the content of Somerton man' hair.
- Speculated that the content values of these three elements in the living environment for Somerton man and this sample are similar.

Future works

Data analysis

- Re-analysing the uncommon elements value of Somerton man's hair

- Find more clues related to Somerton man's living environment

- Find the same trend of element change in two Somerton man's hair tests.

- Gain the constant number to offset the difference between two test data.

Code

- Extend the Simhash test to more languages.

- Generate 50 letters long random sequences to see where the test result lies.

- Re-do the test with a different test material rather than War and Peace

- The algorithm can also be used to check other similarities.

Conclusions

Code analysis

- Code consists of Initialisms of English.

- Levenshtein Distance was not a good choice in this task.

- Simhash algorithm has a creditable performance on similarity check.

Mass spectral analysis

- Glass test data: The high content of these elements cannot be regarded as the main cause of Somerton man's death, only as a reference to the change of living environment.

- Quartz test data: Somerton man's living area is different with the fourth sample, but similar with the first sample.

- Two tests comparison: not achieve the desired results

Reference

- 1. < http://www.sapolicehistory.org/Somerton%20Beach%20Bust.jpg >
- 2. <http://news.wisc.edu/newsphotos/images/TESci_educ_program09_7801.jpg>
- 3. <http://www.forensicmag.com/sites/forensicmag.com/files/legacy/u730/hairDNA040313.jpg>
- 4. <http://www.forensicmag.com/sites/forensicmag.com/files/legacy/u730/hairDNA040313.jpg>
- 5. <https://www.adelaide.edu.au/microscopy/instrumentation/icpms.html#7700s>
- 6. Gencarelli. N & Yang. JK , 2015 'Final Report 2015' Derek Abbott's Wiki Project, viewed 24 March 2016,

<https://www.eleceng.adelaide.edu.au/personal/dabbott/wiki/index.php/Final_Report/Thesis_2015>

- 7. Griffith. L & Varsos. P, 2013 'Final Report 2013' Derek Abbott's Wiki Project, viewed 24 March 2016,

<https://www.eleceng.adelaide.edu.au/personal/dabbott/wiki/index.php/Semester_B_Final_Report_2013_-_Cipher_cracking>

- 8. 'Somerton man' Wikipedia, 2016

<https://en.wikipedia.org/wiki/Taman_Shud_Case>

Questions?