Exploring shape and size in British Neolithic Axes: new methods on old data

GTNV0

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Master of Science

of

University College London.

Institute of Archaeology

University College London

November 2020

I, GTNV0, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

British Neolithic axes, due to their numerousness and long survivability, provide a wealth of evidence with which to study non-subsistence production and trade in this period. They have long been ubiquitous in archaeological literature, however one aspect about these pervasive tools which has rarely been explored is their differing morphology. In addition, much archaeological data is paper-based or not digitised, and most likely lies unused in national or personal archives. This paper sets out to extract data from an old analogue dataset containing information about a large sample of British Neolithic axes, and attempts to use this to investigate potential groups based on their shape and size, in addition to carrying out everything reproducibly. Firstly data was extracted using crowdsourcing to transcribe record sheets and image processing techniques to digitise technical sketches. The sheet data, containing measurements, was then subjected to PCA and k-means clustering, and the processed images were Fourier-transformed and subjected to the same statistics. The results of these analyses suggest that groups of similarly shaped axes exist within this sample. Further, this dissertation demonstrates a reproducible workflow which can be used for large projects built with substantial software.

Acknowledgements

Many thanks to Andrew Bevan for providing mentoring and introducing me to this dataset, to Michael Pitts for creating it, and to the Implement Petrology Group for scanning it and making it publicly available.

Table of Contents

Introduction **15**

1 Background **17**

1.1 Archaeological context: British Neolithic axes 17

1.2 Methodological context in archaeology 21

1.2.1 Data extraction methods 21

1.2.2 Morphometric analysis 23

1.2.3 Reproducibility 25

2 Data and methods **27**

2.1 The dataset 27

2.2 Makefile and Dockerfile 28

2.3 Data extraction 29

2.3.1 Extracting record sheet data 29

2.3.2 Processing the axe drawings 33

2.4 Analysis 43

2.4.1 Size morphometrics with axe measurement data 43

2.4.2 Clustering shapes in the axe drawings 44

2.4.3 Drawing comparisons 46

3 Results **49**

3.1 Data collection outcomes 49

3.1.1 Image processing of axe drawings 49

3.1.2 Crowdsourcing with MicroPasts 50

3.1.3 Representativeness of data collected 50
3.2 Morphometric analysis using record sheet data 52
3.2.1 Sorting variables 52
3.2.2 Univariate 53
3.2.3 Principal components analysis 55
3.2.4 K-means cluster analysis 62
3.2.5 Within-cluster densities 64
3.2.6 Clusters on PCA 65
3.3 Outline morphometric analysis 67
3.3.1 Data sorting 67
3.3.2 PCA on elliptical-Fourier-transformed outlines 70
3.3.3 K-means analysis on the Momocs PCA results 72
3.3.4 Outline cluster membership 75
3.4 Comparisons 76

4 Discussion 81
4.1 Limitations and success of the data collection methods 81
4.2 Results from morphometric analyses 84
4.2.1 Measurements from the record sheets 84
4.2.2 Axe outlines 87
4.3 Comparisons 88

Conclusion 91

A List of appendix files 95

B List of abbreviations 97

References 99

List of Figures

2.1 Screenshot of an example transcription form with sheet. 32
2.2 Example drawing of axe 10031. 35
2.3 Dilated contours found from edge-detecting the drawing, then superimposed onto the drawing 37
2.4 Bounding boxes drawn around each view. 38
2.5 Filled contours. 40
2.6 Only the largest contour is now drawn. 41
2.7 Final filled axe shape. 42

3.1 Dotplot to visualise the amount of measurements for each feature and the conditions of the axes. 52
3.2 Density plots of the distribution of measurements for each feature. 54
3.3 Violin plots of measurement distribution for each feature. 55
3.4 Screeplots of eigenvalues and percentage variance explained for PCA. 56
3.5 PCA: individuals and variables for PC1 and PC2. 58
3.6 PCA: individuals and variables for PC3 and PC4. 59
3.7 PCA with 19 variables measured for 369 axes. 61
3.8 WSS and average silhouette plots for k-means clustering. 62
3.9 K-means analyses with 2, 3 and 5 clusters. 63
3.10 Density plots to show distribution of variables within 3 k-means clusters. 65
3.11 PCA of the first and second pairs of PCs with individuals coloured by k-means cluster. 66

3.12 All outlines imported into Momocs before inspection. 67
3.13 Final outlines for analysis with incorrect shapes removed. 68
3.14 Cumulated harmonic power without the first harmonic for three outlines. 69
3.15 Harmonic contribution plot to show how many harmonics can reconstruct a typical axe outline. 69
3.16 Variation in outlines explained by each principal component. . . 70
3.17 Scree plot of proportion of variance explained by each principal component. 71
3.18 PCA of elliptical Fourier-transformed outlines 72
3.19 Momocs k-means with 2 clusters 73
3.20 Momocs k-means with 3 clusters 74
3.21 Momocs k-means with 4 clusters 75
3.22 The 4 Momocs clusters colour coded onto a panel plot of axe outlines 76
3.23 Mean shapes for each of the 4 clusters 77
3.24 Mean shapes for the 3-cluster solution on the axe measurement data. 77
3.25 Mean shapes for 3 clusters on axe measurement data with 19 variables. 78
3.26 Find places of axes in each cluster plotted onto a map of Britain, with measurement clusters above and outline clusters below. . . 79

List of Tables

3.1 Summary statistics. 56
3.2 Variable loadings sorted by strongest correlation with PC1. . . . 57
3.3 Variable loadings for PCA with all rows containing NA values removed: 369 axes left. 61
3.4 Percent similarity between outline and measurement clusters . . 78

Introduction

This dissertation follows two main lines of investigation: how data can be extracted from legacy formats for use with computational methods, and whether size, shape or both can group Neolithic axes from Britain. The author firstly takes special interest in analogue datasets as they may be relatively unknown or unused for years, and also because they pose unique challenges in data extraction and cleanup. Extracting old archaeological data could help to give new insight into the past through use with modern methods, and helps to highlight that much paper-based information remains in recent years, if ever, which could be tapped into. Secondly, the differing morphology of British Neolithic axes is of particular interest as these tools are so ubiquitous in the archaeological record, yet the reasons behind their designs are still relatively ambiguous. Further, quantifying traits of artefacts to aid in grouping as an alternative to classical typological methods is of particular appeal for approaching the archaeological problem domain inferentially. In light of and in combination with recent research on British Neolithic axes, investigating this seldom-pursued aspect could help to reveal more about these tools and their manufacturers from a period which lasted thousands of years, with no surviving writing.

The data used in this project is a paper-based dataset compiled by Mike Pitts in the 1970s and scanned to JPGs by the Implement Petrology Group in 2016, which contains technical drawings of British Neolithic axes and record sheets cataloguing their measurements in addition to other information, such as their find places and materials. This dataset was originally publicly available on the Implement Petrology Group's website. Before any data extraction,

the literature of both the archaeological and methodological contexts relevant to this dissertation are reviewed in order to give context to this research. Next, to extract the analogue data, transcriptions of the record sheets were crowdsourced using the MicroPasts online platform, and the drawing images were processed using Python's (Van Rossum & Drake, 2009) OpenCV library (Bradski, 2000) into a format compatible with the R's (R Core Team, 2019) `Momocs` package (Bonhomme et al., 2014) for outline morphometric analysis. After the data was digitised, the two sets were then subjected to separate but analogous analyses. The measurements from the record sheet data were explored first with density and violin plots and then analysed with principal components analysis and k-means clustering, whilst the processed drawing outlines were elliptical Fourier-transformed and then subjected to the same multivariate analyses. The results of both sets of k-means clustering were then compared by plotting the find places of the axes from each on a map of Britain to see which clusters persisted geographically, in addition to comparing the similarity in axe membership between the two sets of clusters. All data extraction methods, analyses, and the building of this dissertation PDF were made reproducible using a `Makefile` (Stallman, McGrath & Smith, 2004), `Dockerfile` (Merkel, 2014), `renv` (Ushey, 2020) for R and `venv` for Python. The implications of the results from both sets of analyses and their comparisons were finally discussed in light of the initial literature review.

Chapter 1

Background

1.1 Archaeological context: British Neolithic axes

The history of stone implement study in archaeology had a turning point in the 1920s, up until which a so-called collector's approach had been dominant, which asserted that the preservation, appearance and integrity of axes were important above all else (Grimes, 1935, p.424). This meant that typologies were built predominantly upon appearance. In this decade, a move away from this school of thought was spurred on by such studies as that of Thomas, a geologist who inspected characteristic inclusions in the stonehenge 'foreign' blue stones and found the same pattern in boulders in Pembrokeshire, Wales and followed them into the Prescelly Mountains, where more of this type of rock was found *in situ* (Thomas, 1923, pp.247, 248). These stonehenge stones were called foreign as, until provenancing became mainstream, axes and other stone artefacts found in areas with no clear stone source were given this name (Grimes, 1979, p.1). Thomas was part of the British Geological Survey, started at the beginning of the 1900s to catalogue stone types in Britain, which would later be useful for further provenancing of stone implements. Another important development was that archaeologists in 1930s southern England began taking thin sections of axes to analyse, which kept the structural integrity of the axe but allowed the study of internal inclusions with a microscope (Grimes, 1979, p.1).

This technique was used by the Sub-Committee of the South-Western Group of Museums and Art Galleries on the Petrological Identification of Stone Axes, formed in 1936, which also named the first set of petrological groups, and eventually published a petrological report of more than 200 specimens. The main aim of the South-West Committee was essentially to analyse the petrography of stone artefacts (normally tools), collate previous work and publish the results (Keiller, 1937, p.485). Their first report in 1941 asserts that a previous hindrance in the study of stone artefacts was that their owners forbade any modification, however minute, which meant that petrologists were expected to identify the type of stone just from the surface. Due to the nature of axe manufacture, this was impractical, as much of the surface would have been worn away by grinding, polishing and so on (Keiller, Piggott & Wallis, 1941, p.50). This publication further cemented the idea that artefacts were no longer just pieces in a display cabinet but specimens for scientific investigation. Later, the South-West Committee's records were integrated with those from the rest of England and the national survey was created, the central committee of which became the Implement Petrology Committee. The Committee was responsible for publishing the largest stone implement survey so far, Stone Axe Studies ii, recording the petrology of 7625 axes, which ranged in period from the early Neolithic to the Bronze Age. This exhaustive survey covered all of the UK, with the largest number coming from south-west England (Clough, 1988, p.3). Important provenancing information was discovered and recorded for an unprecedented amount of artefacts in this survey, forming the basis for future research on production and exchange of these items, a topic that is once again prominent in the current literature. Later, the Implement Petrology Committee would become the Implement Petrology Group, which still exists today, and whose members were responsible for scanning the dataset used in this paper.

Shortly after its publication, however, reviews of Stone Axe Studies ii expressed disappointment that the volume mainly reported the results of petro-

logical analyses, and did not attempt to address such contentious topics as the distinction between petrological groups, or what may have led to the dispersal of these artefacts at such distances from their sources, and in such close proximity to other raw materials (Edmonds, 1989, p.633). However, it its defense, Stone Axe Studies ii did exactly what it set out to do: catalogue the provenances of as many stone axes as possible. Mike Pitts, the author of the dataset in this dissertation, cited such reviews in his 1996 paper as he too thought that the 1988 IPC volume left more to be desired. He states that because of the heavy focus on finding the origin of stone implements for the past few decades, there had been little interest in methods of manufacture, potential differing functions, or emergent styles of stone implements. Pitts also points out that there was often disagreement between petrologists on which stone group an axe belonged to, highlighting the subjective nature of previous categorisation methods (Pitts, 1996, p.311). His paper included PCA, k-means and linear discriminant analysis on the measurements in the axe dataset he compiled, in addition to creating six much broader axe groups (p313). There had been forays into similar morphometric analyses of stone tools in the past, however not from the Neolithic (Roe, 1969, p.23).

In recent years, Neolithic axes have received renewed attention, with investigations ranging from those concerning their potential meanings to those investigating why they were dispersed or manufactured in the way that they were. Stone Axe Studies III, the latest volume of the IPC's series, presented the latest findings about stone implement artefacts at home and abroad, with a number of the British articles concerning Neolithic axes. Within this volume, an article of a more speculative nature focuses on the meaning of the axes to Neolithic people. Examples include Darvill (p140) who suggests that the stone chosen for axe manufacture may have had symbolic significance, and Clarke (p319) who observes that Neolithic axes in Orkney tended to differ in size depending on the context they were found in, with larger, more complete specimens being found in graves (Darvill, 2011, p.140; Clarke, 2011, p.319). A

significant study is that by Walker, who posited that axe heads imported by incoming immigrant populations from continental Europe had influence on axe design in Neolithic Britain, and resulted in copying (Walker, 2015, p.105). She additionally suggests that the difficult-to-reach nature of many quarry sites in Britain could be part of the reason why these locations were chosen, perhaps as a form of pilgrimage or rite of passage (Walker, 2015, p.233). In addition to these discussions considering the structure and function of Neolithic axes in Britain, radiocarbon dating was being used increasingly in conjunction with provenance data in order to ascertain how tool use, manufacture and demography changed over time, and whether they were linked. One such example is the NEOMINE project undertaken by Shennan *et al.*, who were able to ascertain that mining activity in Britain coincided with a period of forest clearing that preceeded the introduction of farming by approximately 200 years (Shennan et al., 2017, p.78). The main impetus of this project was to collect radiocarbon dates for all of Britain's Neolithic mines. The authors also suggest that Neolithic populations from the continent immigrated to Britain and brought farming techniques with them, beginning to mine and clearing forests beforehand. In Scotland, it is also known that elm and birch pollen percentages rapidly declined in the Beinn Lawers palaeoenvironmental record around 3350 cal BC, which coincided with the uptake of stone axe manufacture at Creag na Caillich (Tipping, Edmonds & Sheridan, 1993, p.595).

Within the last year, three papers have been published from a project titled 'Supply and Demand in Neolithic Europe', which makes further use of the NEOMINE radiocarbon dates. The first of these supports the previous assertion that itinerant European Neolithic groups were mining for axe materials because these tools were needed to clear the land for farming. The aforementioned period of intensive axe production and forest clearing prior to farming appears to be unique to Britain as northwest Europe does not show similar behaviour (Schauer, Shennan, et al., n.d., p.p157). It is suggested that demand for axes here may instead be a response to the appearance of imported

Alpine Jade axes of high symbolic or status significance (Schauer, Shennan, et al., n.d., p.p158). Since Pitts' paper in which he devised new, broader stone groups, geochemical analyses have actually helped to assert that the original IPC groupings are useful after all, representing real groups of axes with homogenous provenance (Schauer, Andrew Bevan, et al., n.d., p.2). Spatial analysis of Neolithic axehead findspots has recently indicated that the majority of similar stone axe heads tend to be found near to their original source, however axes made from certain types of stone appear to be found further afield than others (Schauer, Andrew Bevan, et al., n.d., pp.15, 20). Perhaps this could be due to the strength of the raw material, which could either indicate preference or simply less likelihood of breakage in transit. Finally, 39 new results from two different radiocarbon calibration methods, using radiocarbon dates from the NEOMINE project, indicate that almost all of the quarries and mines, in both north and south England, were active contemporaneously (Edinborough et al., 2020, p.27). This then suggests that differing shape or size is not necessarily a product of change over time. The fact that the start dates of these mines are all very similar also shows the sheer scale and sudden onset of the demand for these axes (Edinborough et al., 2020, p.27). It has also been suggested that the immigrant Neolithic populations were from Brittany, as populations there had already began to manufacture high-quality axes similar to those made of jade from the Alps. In light of these recent findings, it appears as though Neolithic axes from Britain were mostly used for chopping trees.

1.2 Methodological context in archaeology

1.2.1 Data extraction methods

For text and data extraction of archaeological data from handwritten documents there appear to be two routes which are most pursued. These are transcription or handwriting text recognition (HTR) methods. The former appears to be more popularly used for analog datasets compiled by other ar-

chaeologists. Non-digital datasets come with their own problems, such as not being collected with the intention of others using it initially, or not having potential computational methods for extraction and analysis in mind (Faniel et al., 2018, p.106), particularly with legacy datasets. Much archaeological data is still paper-based (Emmitt et al., 2019, p.1), and as computational methods become more prevalent in archaeology, much of this data may need to be digitized. Even if archaeological data has been digitized, sometimes this is in the form of scanned images, for example the dataset in this project, which was scanned in 2016. This is more akin to data back-up. For very large digitization projects, crowdsourcing could be a favourable option, and the MicroPasts crowdsourcing platform has been used for this increasingly within archaeology. Released in 2013 and developed mainly by UCL and The British Museum, one of its first projects involved the transcription of over 30,000 artefact information cards about Bronze Age metal objects (Bonacchi et al., 2014, p.63). The platform was born out of concern for the lack of open data in archaeology, coupled with an ambition for fostering institutional and community collaboration. Current transcription projects include one which involves Leonard Woolley's field notes. Other crowdsourcing platforms include the more well-known Zooniverse, launched in 2007, which is much more general in subject area. HTR is an alternative to transcription, however archaeological applications in recent years tend to be on ancient data. Examples include translating and segmenting Akkadian inscriptions with a recurrent neural network amongst other methods (Gordin et al., 2020, pp.8–9).

In terms of extracting data from images prior to morphometric analysis, it appears that most studies have done this manually. This process usually necessitates the creation of a digitized outline and the placing of equally spaced coordinates or landmarks around it. One early example places landmarks around photographs of human phalanxes using a digitizer tablet (Palmqvist, 1996, p.105). A number of studies have used drawings of Bronze Age axes traced from a corpus of images which were scanned (Forel et al., 2009, p.722;

Monna et al., 2013, p.510; Wilczek et al., 2015, p.383). Forel et al. extracted coordinates by manually placing the start and end points, however Monna et al. used an unprovided MATLAB script, and and, whilst Wilczek et al. do not say how their points were extracted, they cite Monna et al. as an example extraction methodology. Some more recent studies use software called tpsDIG2 to extract landmarks, for example on Middle Palaeolithic bifaces (Iovita, 2010, p.243); on Acheulian handaxes (Iovita et al., 2017, p.10); and recently again on Acheulian handaxes by staff at the British Museum (García-Medrano et al., 2020, p.5), however this still appears to be manual, and slightly cumbersome as it deals with one image at a time, each of which needs to be inspected by the user for correctness. A very recent study, which lists a co-creator of `Momocs` as a co-author, uses Photoshop to create outline silhouettes of ancient barley grains and ImageJ to manually place landmarks (Wallace et al., 2019, p.1129).

1.2.2 Morphometric analysis

Morphometric analysis is the statistical treatment of size and shape (Claude, 2008, p.1). Applications of morphometric analysis in archaeology first dealt with size only, taking the form of multivariate analyses on artefact measurements under the name 'numerical taxonomy', which first arose in the 1960s. Taking inspiration from its uptake in biology (Doran & Hodson, 1975, p.159), (Shennan, 1988, p.192), proponents of this method hoped that it would be a good way of checking whether traditional classification was valid or not, and also be useful for analysing large amounts of artefacts (Doran & Hodson, 1975, p.186). In one of the very earliest instances of this method, numerical taxonomy is introduced as an objective method of study, proposed for use in grouping British lower and upper Palaeolithic handaxes (Roe, 1969, p.23). Other proponents of the method used this paper as a case study for how it should be carried out (Hodson et al., 1971, p.37). With this adoption came a number of publications also voicing concern that numerical typology may be reinforcing the negative aspects of archaeological classification, for example forcing the existence of predetermined types (Thomas, 1972, p.39), and that

results were validated on whether or not they fitted these types (Shennan, 1988, pp.193, 194). This is also the method used in Pitts' 1996 paper on the axe measurements from his compiled dataset. In recent years it appears that this method has fallen out of favour in archaeology.

Similarly to numerical taxonomy, archaeologists later took influence again from biology, or biometrics, but began to consider shape rather than size. These more complex morphometric methods do not appear to arise in archaeology until the mid 1990s (Wilczek et al., 2015, p.382). One such example is that of a study on Acheulian handaxes from Israel, where landmarks were placed on outlines of the axes where measurements were taken manually, and the landmark positions were then compared between the axes. This was done by using Hotelling's t^2 to test for statistically significant differences in mean landmark position on outlines drawn in two-dimensional cartesian space from their measurements (Brande & Saragusti, 1996, p.425). Another contemporary study used elliptical fourier analysis on a human phalanx to compare it to gorilla, Old World monkey, and other human phalanxes (Palmqvist, 1996, p.99). In recent years, morphometric analyses in archaeology have increasingly focused on comparing axe shapes from certain periods. One early example is Forel et al., who use discrete cosine transform, a Fourier-related method on lateral ribs of bronze age axes in the form of line drawings, implemented with MATLAB (Forel et al., 2009, p.723). Another study from around the same time examined and compared resharpened stone tools at different stages in their use-lives using elliptical Fourier analysis (Ioviţă 2010, p243). In the last decade, the `Momocs` R package for shape morphometrics has been increasingly used in archaeology. The impetus behind `Momocs` was to create one package for outline morphometric analysis in R, which as yet had not existed, and contemporaneous morphometry-adjacent packages focused primarily on placing landmarks (Bonhomme et al., 2014, p.4). Wilczek et al. use `Momocs` alongside a vast number of other methods to reveal groups in Bronze Age flanged axes (Wilczek et al., 2015, p.385). A 2017 paper uses elliptical Fourier analysis with

`Momocs` to investigate whether Acheulian handaxes became more symmetrical in design as time progressed (Iovita et al., 2017, p.10). Something that is apparent in the literature is that many morphometric analyses tend to focus on either shape or size but not both, perhaps with the exception of Brande & Saragusti (1996).

1.2.3 Reproducibility

Something many of the studies mentioned in this methodological background have in common is that their computational analyses or data extraction methods are not reproducible. This is true even for the more recent publications mentioned, for example those using `Momocs`, a package made to facilitate reproducibility in morphometric analyses, as they manually extracted points from the outlines before importing. The sharing and citation of data, as well as the sharing of reproducible methods used to arrive at published results, is still relatively rare in archaeology (Lodwick, 2019, p.14). Materials and methods sections of archaeological research papers can often be vague, sometimes not even citing all software used. Not only does this damage the credibility of published results as colleagues are unable to ascertain whether these methods actually do work, it also means that the methods cannot be improved upon or utilized by others, or perhaps even used again by their authors (Marwick, 2017, p.425). Even if methods are shared fully, often archaeologists are using proprietary software such as SPSS or MATLAB (Marwick, 2017, p.431), which either uses point-and-click methods or is behind a paywall. Furthermore, it prevents collaboration with experts from other fields as they are unable to share workflows. Meanwhile, discussions around open data are becoming more common in archaeology, particularly in archaeozoology (Eric C. Kansa & Sarah Whitcher Kansa, 2013, p.90) and in archaeobotany (Lodwick, 2019, p.14). There is also MicroPasts, which hosts publicly available crowdsourced datasets, and the new *Journal of Open Archaeology Data*, which publishes descriptions of datasets and information on how to access them.

Chapter 2

Data and methods

2.1 The dataset

This dataset consists of drawings and record sheets relating to 1919 Neolithic stone axes found in Britain, compiled by Mike Pitts in his spare time from 1977-1978. The record sheets contain 56 measurements, for example blade width, poll width, maximum length and weight, among others, in addition to provenance information, the museum collection which the axe is from, and general comments. It is a feat that one person was able to catalogue this much information, even if some is missing, as will become apparent in later sections. In addition, the associated drawings are very useful and rarely provided. Abbreviated variable names in the sheets were coined by Pitts and their definitions can be found in the glossary of this dissertation. Most record sheets have an associated drawing for the same axe. Both the drawings and the record sheets were hand-drawn or written, and seemingly in pencil. These were then scanned to JPG files in 2016 by the Implement Petrology Group. Mike Pitts used this dataset in his own set of 'numerical typology' analyses, similar to the workflow in this project, and he reported the outcome as inconclusive. This was because the analyses did not separate axes into known archaeological groups (Pitts, 1996, p.339). This raw data has its own biases and difficulties which are worth mentioning. Firstly, it was created and curated by one person. Pitts chose and traveled to several museums, at which he

recorded measurements for all of the complete axes, and some others of varying condition based on his choosing (Pitts, 1996, p.328). In addition, some record sheets for complete axes have blank fields for a number of measurements. The sheet template also differs in format and order of fields. Due to the nature of human handwriting, some words or numbers may be ambiguous, making data collection more difficult. Another point is that the axes are drawn rather than photographed. Cameras were good enough at the time to facilitate this, however it seems as though Pitts' aim in drawing the axes was to preserve more detail (Pitts, 1996, p.328). This does, however, introduce further bias, as obviously the images are one person's interpretation of a physical object. What is more, Pitts also says that the axes were sketched based on measurements in the record sheets taken when handling the artefacts, and not whilst in the museum observing them (Pitts, 1996, p.328). The drawings are also on the other side of the same sheet of paper that the record sheets are written on, which may cause problems with image processing as lines are faintly visible from the other side. A final point is that because these are drawn in pencil, there are gaps in the lines due to differing placement of pressure.

2.2 Makefile and Dockerfile

A `Makefile` (Stallman, McGrath & Smith, 2004) has been used to orchestrate the building of this project in a simple manner for easier reproducibility. This also proved to be invaluable in managing the complexity of the project which involved thousands of source files, multiple programming and command languages (R, Python, AWK, Bash, jq) alongside command-line programs (curl and xargs among others). The Makefile has 'targets', which are files it needs to create, 'dependencies', which are prerequisite files that are needed to make the targets, and 'recipes', which tell the Makefile how to make the targets. In this way, the syntax of the Makefile defines what needs to be done, but the order of particular commands is controlled dynamically by the GNU Make program. This declarative style complemented the procedural style used in the

project's scripts well. It eliminated the need to provide extensive, step-by-step, instructions on the order to execute the scripts and on which files, indeed the syntax of the `Makefile` in fact self-documents the dependencies in the project. For example, once ran, Make only create files which do not already exist, or updates those files whose dependencies have been modified, saving time on each execution (Marwick, Boettiger & Mullen, 2018, p.82). It also allowed the most computationally intensive part of the project, the image processing steps with OpenCV, to be ran in parallel (using the `-j` argument) to take advantage of the multiple cores on a modern CPU. The entire project can be built by running the command `make` inside the top level of the dissertation folder (but only if the required system dependencies are available).

The need for a reproducible Linux environment containing the extensive system dependencies led to the use of Docker. A `Dockerfile` has been provided to build a Docker image, it is based on the `rocker/geospatial:4.0.3` image (Carl Boettiger & Ross, 2020), with additional system packages, TeX Live packages (Berry, 2020), the Python `venv` and the `renv` R package. The `renv` itself is created when the Docker image is run so that the packages are downloaded once and cached within a mounted volume, saving time when the image is rebuilt. Recipes to build and run the Docker image are included within the `Makefile` targets: `make build` and `make run`. The recipe of the latter ensures the mounted volumes have the same user and group permissions as the Docker host, allowing for seamless access to and from the host's filesystem. The recommended way to reproduce this project is to first run `make clean` to remove the ready-made files provided, then `make build` and `make run`.

2.3 Data extraction

2.3.1 Extracting record sheet data

The method initially considered for digitizing the data from the record sheets was to use a handwriting recognition model. A handwriting recognition model

was found on GitHub, which was written in Python and implemented with TensorFlow, and took single segmented words or numbers (Scheidl, 2020). The model was chosen as it looked relatively straightforward, had a good README, and had been starred by other GitHub users 963 times. In order to get the record sheet data into this format, however, they needed to be split into as many images as there were measurement fields. For the more qualitative fields such as the comments, either something else would need to be considered, or perhaps they would be left out. Segmentation was attempted by using `HoughLines` from the `opencv-python` (Bradski, 2000) library in `Python` (Van Rossum & Drake, 2009) to draw lines onto the grids, the coordinates of which would then be used to crop the sheets, so that each image was split into 56 squares (please see `appendices/hough-transform.ipynb` and `appendices/hough-transform-copy1.ipynb`). This method successfully found the grid lines, however after this point was reached, HTR was decided against due to the expected time and effort it would take to ensure it worked satisfactorily. This would have required transcribing a large number of the record sheets to use as labels for the training data. In addition, everything returned by the model would need to be checked, which would be a mammoth task.

After considering this deep learning method, it was decided that transcriptions of the record sheets would be crowdsourced instead. This would enable more time to be spent on processing the axe drawings, as this would take far longer to do manually using a raster graphics editor. The crowdsourcing platform chosen was MicroPasts as it is specifically for archaeological projects. MicroPasts uses Pybossa for crowdcrafting, which uploads projects and tasks to the platform, and has an API which can be accessed via the command line. The tasks, in this case the record sheet images, needed to be hosted on another website and linked to using URLs imported to MicroPasts with Pybossa. To do this, the sheets first needed to be separated from the drawings, and any sheets which did not have a drawing associated, or had duplicated axe IDs in their

filenames, were omitted. This left 1825 pairs of record sheets and drawings, each pair representing one axe. These were placed into separate folders and their filenames and paths were saved to csv. The images were then uploaded to a GitHub repository and the filenames from the csv were imported into R and joined with the GitHub URL to make their individual URLs, which were then exported to csv for Pybossa to import.

To create the project, the template HTML from the Bronze Age Index Card transcription project on MicroPasts (Pett, 2018) was modified to show the 63 fields in the record sheets and a tutorial. This template was chosen as it had a similar input form which could be easily modified for this project. Below is a screenshot of the transcription form, alongside one of the record sheet images, or 'tasks' (Figure 2.1). Please see the `micropasts-project` folder for the files mentioned in this section, and the HTML used to create the form and tutoral (please note that original sheets are not provided due to file size).

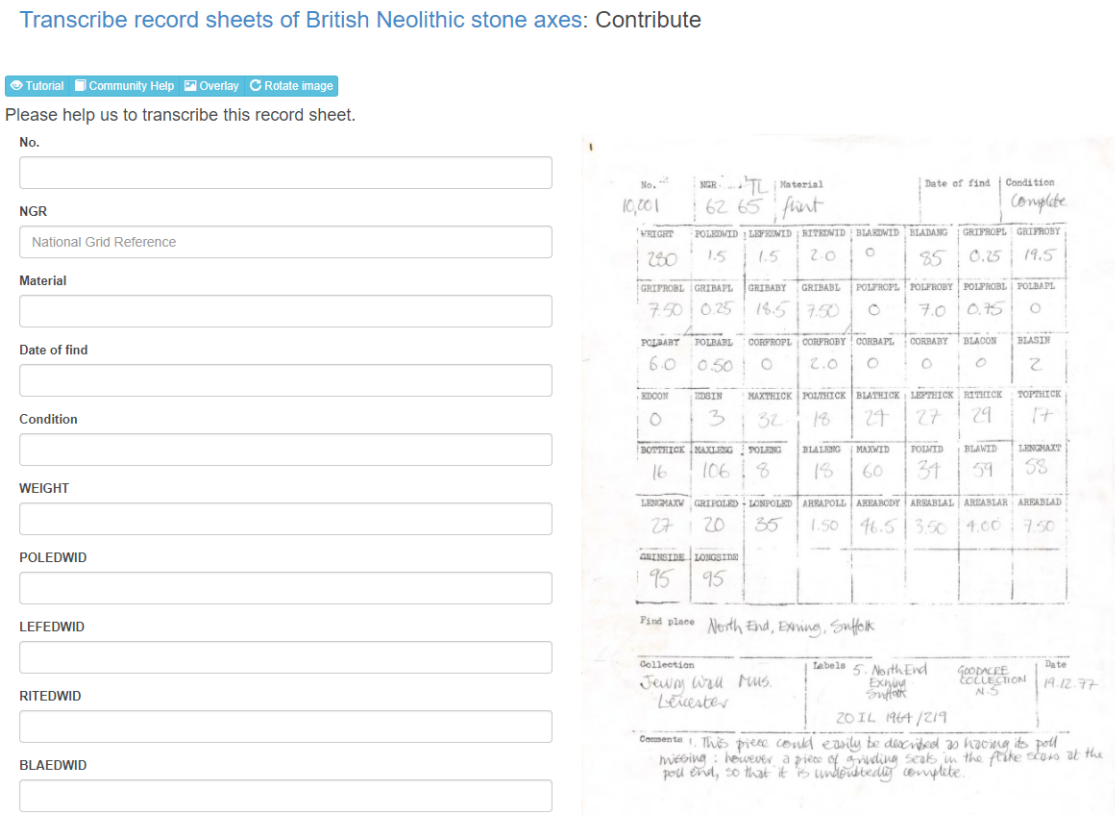

Figure 2.1: Screenshot of an example transcription form with sheet.

It was then important to provide detailed instructions in the tutorial on how the transcribers should enter the data as some aspects of the sheets were ambiguous. Firstly, the field name abbreviations were defined to prevent confusion, and the units of measurement were provided to aid in ascertaining correct data entry. In addition, volunteers were asked to enter an ellipsis in square brackets next to any dubious transcribed text so that it could later be found during data cleanup. The instructions also asked volunteers to make sure all characters were transcribed, even if it did not make sense, for example brackets around numbers. This was because this project is aiming to preserve as much of the original information as possible for potential future use as training data.

To retrieve the data from MicroPasts, a separate Makefile (`micropasts.mk`), which the main Makefile also runs, was used to download 100 tasks at a time into jsons in order to bypass Pybossa's pagination limit. This communicates with the API and gives it the instructions to limit downloads to 100

and 'offset' them, so that each 100 is really 0-99 and so on. This Makefile also runs the R script (`scripts/sheets.R`) responsible for creating the final `data/micropasts-neoaxes1.csv` data for import into the analysis. This script removes all non-numerical characters from the numerical fields and imports the data entry error corrections from `data/correction-sheet.csv`, which is explained below.

Due to time constraints and the volume of axes in the dataset, the redundancy was set to 1 for all tasks. This means that each sheet was only transcribed once, by one person, so there was only one chance to correctly enter the data on the crowdsourcing platform. Once all sheets were transcribed, a period of time was spent making corrections to data entry errors and missed data. The first part of this cleanup process involved creating a csv of all axes with missing national grid references. This was done by attempting to convert them into latitudes and longitudes using the `osgparse` function in the `rnrfa` (Vitolo et al., 2018) R package and creating a dataframe with their latlongs and axe Ids. Any which did not successfully convert were found using `anti_join` from `tidyverse` (Wickham, 2019) to see which axe IDs appeared in the MicroPasts data but not in the coordinates dataframe. These were then exported to csv and checked against their original record sheets, with 50 of these being actually incorrect, and the rest having been missing from the original sheets initially. The second strategy for identifying data entry errors was to examine each column in `data/micropasts-neoaxes1.csv` and order them by ascending and descending to spot atypically low or high values. The `data/correction-sheet.csv` file was made with the axe number, feature to change, and the new value, which was then imported and joined back onto the imported Micropasts data to correct it. This way the corrections are logged and reproducible.

2.3.2 Processing the axe drawings

Since the majority of analyses involving calculations performed on the geometry of shapes in archaeology extracted their data manually, this methodology

sets out to not only automate these processes, but also make them reproducible. `Momocs` requires raster silhouettes - black shapes on white backgrounds - for input, and takes one shape per image. In addition, these shapes need to be similar enough to compare (Bonhomme et al., 2014, pp.5, 11). Each axe drawing usually shows 3 views: plan, profile and top (Figure 2.2), so to compare effectively, each type would need to be analysed with others of the same view. The filled silhouettes have not been included in this project due to file size, however please make these using the reproducibility instructions if you wish. Please also see `appendices` for jupyter-notebooks used to devise this workflow.

The workflow for image processing from pencil outline to filled binary image is as follows.

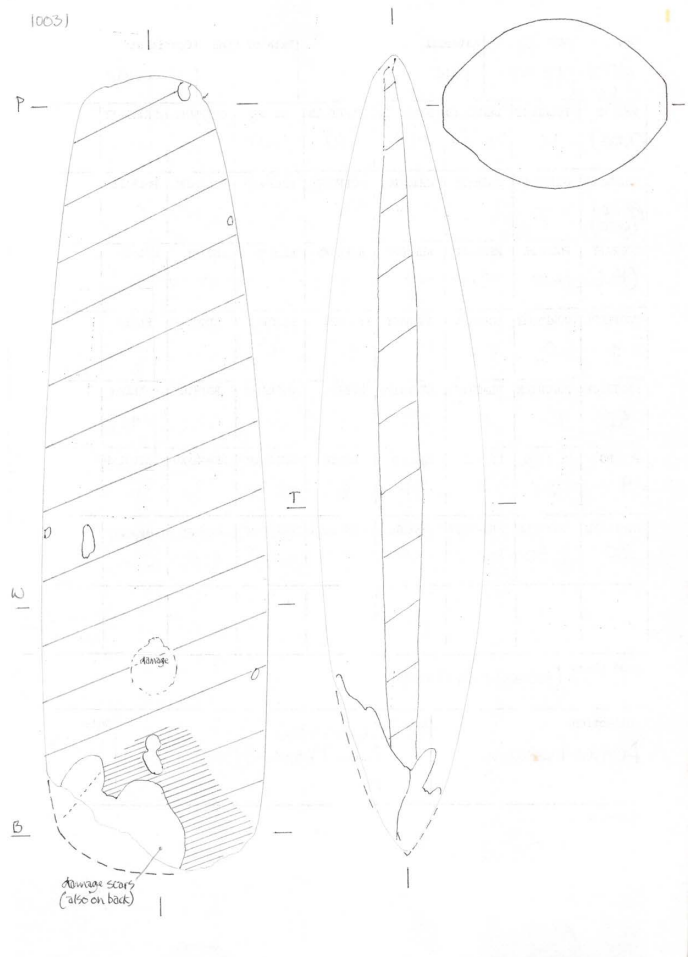

Figure 2.2: Example drawing of axe 10031.

The first step was to crop the drawings into single views, which is carried out by `scripts/view-finder.py`. This was done by first finding the contours of the drawing outlines. Contours are found by using edge detection on an image, then finding the contours of the edge-detected image. Because the pencil drawings are faint in some areas, the lines have gaps, which creates problems finding the entire shape. To mitigate against this, the edge-detected image was dilated so that the edges were thicker, forcing them to join, and making the contours larger (code snippet; Figure 2.3).

```
#find the edges of the drawing
edge_detected_image = cv2.Canny(imgray, 1, 100)
#increase width of edges forcing them to join
kernel = cv2.getStructuringElement(cv2.MORPH_ELLIPSE,(9,9))
```

```
dilated = cv2.dilate(edge_detected_image, kernel)
#find contours of dilated edge-detected image
contours, hierarchy = cv2.findContours(dilated.copy(),
                    cv2.RETR_EXTERNAL, cv2.CHAIN_APPROX_SIMPLE)
outlines = segments.copy()
#append curved contours to list
contour_list = []
for contour in contours:
    approx = cv2.approxPolyDP(contour,0.01*cv2.arcLength
            (contour,True),True)
    area = cv2.contourArea(contour)
    if ((len(approx) > 4) & (area > 60) ):
        contour_list.append(contour)
#not included in final script, used in development to see contours
cv2.drawContours(outlines, contour_list, -1, (255,0,0))
```

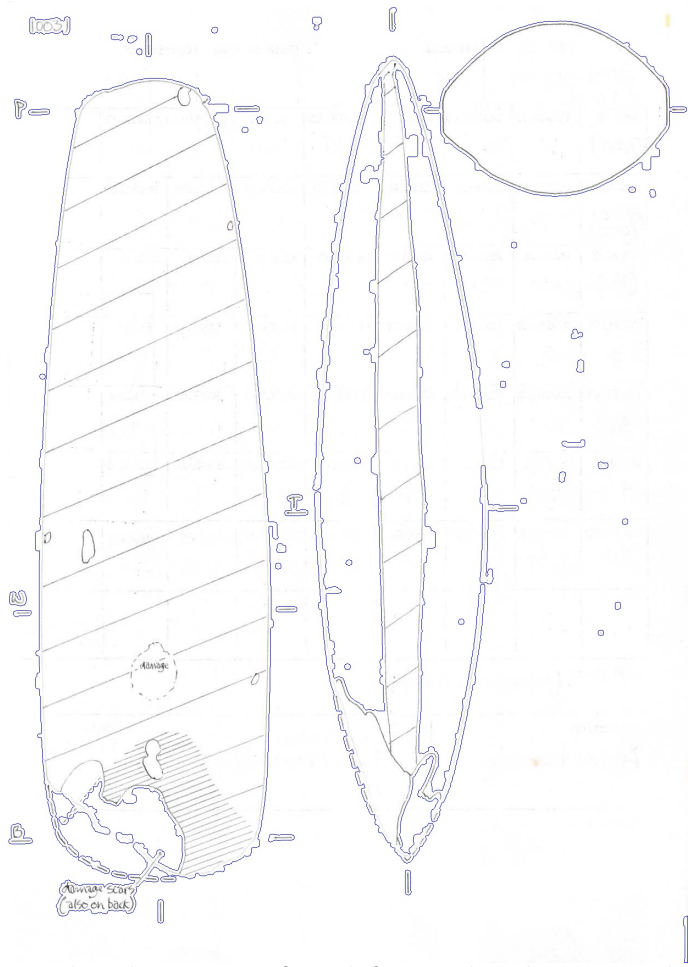

Figure 2.3: Dilated contours found from edge-detecting the drawing, then superimposed onto the drawing

Once the contours were found, it was possible to find the bounding boxes of each view (code snippet; Figure 2.4). The coordinates for each box were appended to a list of lists, which was then converted to a `pandas` dataframe for export to csv.

```
for c in contours:
    rect = cv2.boundingRect(c)
    if rect[2] < 100 or rect[3] < 100: continue
    print(cv2.contourArea(c))
    x,y,w,h = rect
    cv2.rectangle(outlines,(x,y),(x+w,y+h),(0,255,0),2)
```

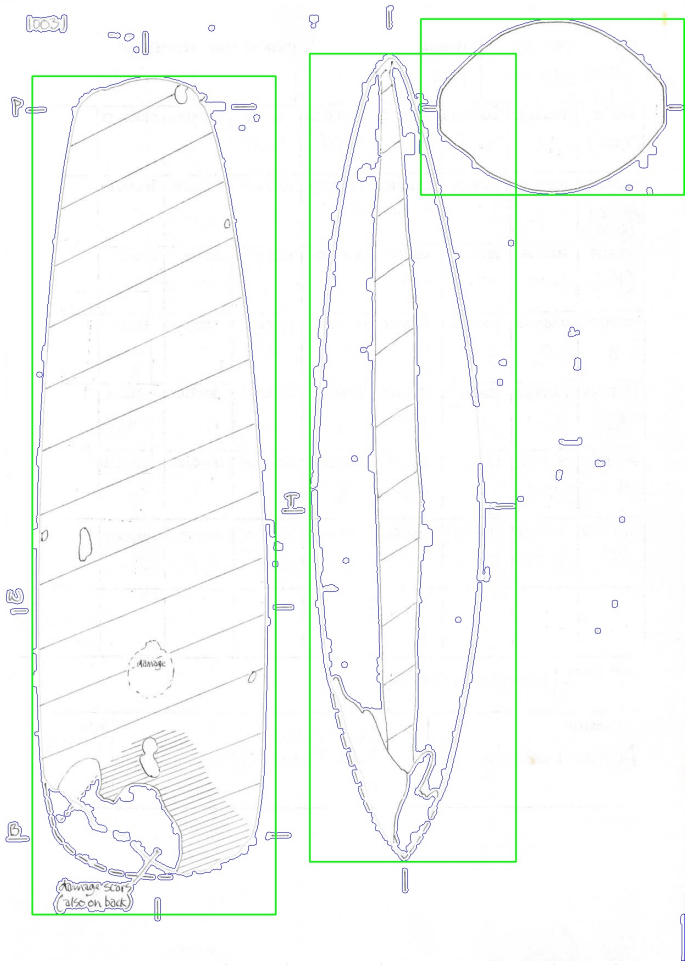

Figure 2.4: Bounding boxes drawn around each view.

This dataframe also included the area and aspect ratio of each view, which was used to determine which view had been cropped, and this was used in the filename (code snippet). The dataframe is exported to csv for each axe, which the Makefile looks in for recipe instructions.

```
#function to name views based on logic
def nameview(ca, max_ca, min_ca, ratio, max_r, min_r):
    if ca == max_ca:
        return "plan"
    elif ratio == max_r:
        return "profile"
    elif ca == min_ca and (min_r > ratio < max_r) :
        return "top"
    else:
        return "tbc"

#need to vectorise above function as pandas deals in vectors
vectfunc = np.vectorize(nameview, cache=False)

df = pd.DataFrame(boxes, columns=['x', 'y', 'w', 'h'])
df['axe'] = df.index
df['drawing'] = image
df['area'] = df['h'] * df['w']
df['aspect_ratio'] = df['h'] / df['w']

#using vectorised function above to name the views
df['view'] = vectfunc(df['area'], max(df['area']), min(df['area']), \
            df['aspect_ratio'], max(df['aspect_ratio']), \
            min(df['aspect_ratio']))

df['filename'] = df.drawing.str.replace('d\.jpg','').str.cat \
            ('-'+df.view.astype(str)+'-'+df.axe.astype(str) \
            +'.jpg')
```

The `Makefile` crops the views using the `convert` program from `imagemagick` and arguments are given to it which are replaced with the co-ordinates for the bounding boxes (please see `Makefile`). Finally, the cropped axe views are filled in by `scripts/fill_view.py` in three steps. The outline is first filled by finding the contours again, but this time filling them when drawing over the image (code snippet; Figure 2.5).

```
cv2.drawContours(outlines_copy, contour_list, -1, (0,0,0), \
                 thickness=cv2.FILLED)
```

Figure 2.5: Filled contours.

The largest contour was then found, which is the axe shape, and only that was drawn onto the image, removing other contours which were not connected to it (code snippet; Figure 2.6).

```
#using contour area to find largest contour
maxcont = max(contours, key = cv2.contourArea)
large_cont = []

#if contour area is larger than or equal to max, append
for cont in contours:
    if cv2.contourArea(cont) >= cv2.contourArea(maxcont):
        large_cont.append(cont)

outlines_copy() = outlines.copy()

#creates image with largest contour only
cv2.drawContours(outlines_copy, large_cont, -1, (0,0,0), \
thickness=cv2.FILLED)
```

Figure 2.6: Only the largest contour is now drawn.

Finally, the now-filled largest contour was thresholded so that `cv2.MORPH_OPEN` could be used on it to dilate the contour and remove the connected nearby shapes from its outline (code snippet; Figure 2.7).

```
#thresholding the filled largest contour
thresh, outlines_thresh = cv2.threshold(outlines4, 250, 255, \
                                cv2.THRESH_BINARY_INV)

#dilating the thresholded contour
closing = cv2.morphologyEx(outlines_thresh, cv2.MORPH_OPEN, \
            kernel, iterations=10)

#final silhouette colours inverted
silhouette = cv2.bitwise_not(closing)
```

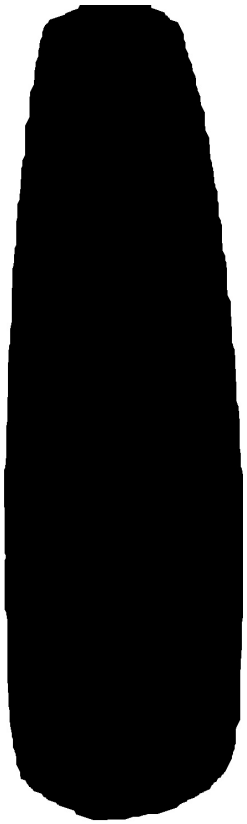

Figure 2.7: Final filled axe shape.

2.4 Analysis

2.4.1 Size morphometrics with axe measurement data

The statistical treatment of the axe measurements involved methods which were trialed on 100 randomly selected axes from this dataset for a previous assignment toward the module ARCL0087 earlier this year. A diagram with these variables labeled on an axe can be found in Pitts' paper (Pitts, 1996, p.331). First, a dotplot was used to visualise the frequency of measurements for each feature and the amount of axes with each condition type. After this, features which were not recorded for most of the axes were removed. Density and violin plots were then used to preliminarily explore the distributions of the variables and observe the influence of outliers. Boxplots were initially considered, however as they can be misleading when used alone on multimodal or very leptokurtic distributions (Baxter, 2015, p.30), violin plots were chosen to ensure clarity for other types of distribution. This is especially useful as it is hoped that some distributions will be multimodal since this investigation is looking for groups.

After this, principal components analysis (PCA) was performed to reduce dimensions and reveal patterns in the axe measurement data. PCA was chosen as it shows relationships between different variables (for example short and squat versus tall and narrow), which is helpful for shape analysis and grouping. PCA is especially useful for multivariate datasets where many of the variables correlate, as groups of correlated variables can be separated, which in turn can help to distinguish groups and discern which variables define groups, if any exist (Shennan, 1988, p.297). The `prcomp` function in the R stats package was used for direct comparison with Momocs. Because PCA does not work with missing values, any axes with measurements still missing after only keeping the most measured variables will have them imputed using the `missMDA` package (Josse & Husson, 2016). This R package was chosen as its `impute_pca` method is best for highly correlated variables, which is the case with this dataset (Dray & Josse, 2015, p.665). The method for calculating the variable loadings from a

`prcomp` object has been taken from Rhys (2019) and uses the `map_dfc` function from the `tidyverse` package to return a tibble (Rhys, 2019, p.332). Scree plots were also examined to determine the number of components which explain the most variance.

Once the measurements responsible for the most variation have been identified in PCA, k-means clustering was used to help reveal any existing groups. K-means was chosen instead of other clustering methods as it is nonhierarchical and does not depend on ordinal or presence/absence data, neither of which this apply to this dataset. To help decide the optimal amount of clusters to set in the k-means analysis, the total within cluster sum of squares (WSS) and average silhouette were plotted. Often it is best to compare multiple k-means solutions as it can be difficult to discern the optimal amount of clusters to use (Baxter, 2015, p.147), so all options suggested by the WSS and silhouette plots will be explored. Combining these analyses can give a good insight into potential groups which may exist in the data. This workflow was also chosen for comparison with Pitts' analyses on the same data.

2.4.2 Clustering shapes in the axe drawings

The Keras library, which allows Python to interface with the deep-learning library Tensorflow, was initially considered for grouping the axe drawings with unsupervised clustering. Deep learning in recent years has been trialed in archaeological remote sensing to help automate the identification of features in large image datasets (Küçükdemirci & Sarris, 2020, p.108). This made Keras seem particularly attractive for revealing patterns in the axe drawing dataset which is too large to inspect manually. However, deep learning methods are best suited for highly dimensional datasets, which remote sensing data is a good example of, whereas the axe drawings are relatively simple outlines. The ability of deep learning models to deal with such complexity can make them susceptible to overfitting (Patel, 2019 iii 8), which would most likely happen with a dataset of simpler images such as this one.

The `Momocs` package (Bonhomme et al., 2014) for `R` was instead chosen to

compare the axe drawings by analysing their outlines morphometrically, since this was more comparable with the morphometric analysis using the measurements. As morphometric analysis is the statistical treatment of size and shape (Claude, 2008, p.1), and the size of axes has already been compared, this analysis deals with shape, providing a more complete morphometric analysis of the axes.

Momocs uses Fourier transforms to quantify shapes with closed outlines for statistical analysis. Closed outlines, as opposed to open outlines which have a definitive start and end point, are continuous, and because of this can be described by periodic functions (Bonhomme 2013 p5). Fourier transform methods break down these periodic functions into a sum of simpler, weighted functions (Claude, 2008, p.213). It is similar to reducing dimensions in PCA. Of the three different types of Fourier transform offered by `Momocs`, elliptical Fourier transform, which fits curves to closed outlines, will be used. It was chosen as it does not require equally spaced points or a regular outline, can be fitted to almost any shape (Bonhomme et al., 2014, p.7), coefficients can be made regardless of outline position (Claude, 2008, p.221), and the coefficients are normalized meaning that prior alignment steps are usually not necessary (Rohlf & Archie, 1984, p.315). `Momocs`' `efourier` function uses the algorithm for elliptic Fourier analysis developed by Giardina & Kuhl (1977) and Kuhl & Giardina (1982) (p.277 and 238-9, respectively). `efourier` transforms each outline into a list of four coefficients that describe different aspects of the shape, which are then multiplied by the number of harmonics chosen (Bonhomme et al., 2014, p.14). These outlines are then in a quantified form suitable for multivariate analyses.

All axes were aligned the same way before importing into `Momocs`, with their blades at the bottom and their polls at the top. Before perfoming any analyses on the shapes, outlines were checked for correctness by using the `Momocs` function `panel` and displaying the names as just the ID numbers from each filename over each shape, extracted from the character vector of filenames

using regex. All erroneous shapes were then removed from the list of filenames to import into `Momocs`.

`efourier` was used with its default normalize setting, which is not appropriate for shapes with strong bilateral symmetry or that are very circular, however the shapes in this dataset are reasonably irregular and more elliptical, so this option should be safe in this instance (`Momocs efourier` documentation). The method of deciding how many harmonics to use is up to the user and is a topic of debate in the literature (Bonhomme et al., 2014, p.12). The number of harmonics represents the number of iterations around the shape it takes to reconstruct it, the first harmonic in elliptical Fourier analysis always being an ellipse. To determine the number of harmonics to choose, firstly `harm_pow` was used to plot the harmonic power of the first, middle and last axes in the dataset, and secondly `hcontrib` on the overall dataset. The former function shows a line plot of how much of the shape has been reconstructed by each harmonic, and also shows at which point this stops improving, which is where the line plateaus. `hcontrib` shows the appearance of a mean axe shape when reconstructed from an increasing number of harmonics.

For analysing the Fourier-transformed outlines, `PCA` and `KMEANS` were used, which are `Momocs` functions that use the R stats library functions `prcomp` and `kmeans` but are also able to plot mean shapes onto a morphospace. In PCA, these mean shapes show a typical axe at intervals on each principal component, and in k-means, these show mean shapes for each cluster. Finally `panel` will be used with the k-means clusters colour-coded onto the shapes to show the membership of the clusters in more detail.

2.4.3 Drawing comparisons

Once the analyses on both types of data have revealed an ideal number of clusters, these two sets of clusters will be compared to see if there is overlap in membership, and also if they persist geospatially. The `MSHAPES` function from `Momocs` will be used to plot the mean outline from each record sheet data cluster, using the outlines from the drawings associated with the axes in each.

A comparison table will also be used to see the percentage of record sheet cluster axes which are present in the outline clusters. Finally, `ggmap` (Kahle & Wickham, 2013) was used to plot the find places on a map of Britain, which was faceted to show one plot for each cluster, with record sheet clusters on top and outline clusters underneath. `geom_density_2d` from `ggplot2` was used to visualise areas which were highly populated.

Chapter 3

Results

3.1 Data collection outcomes

3.1.1 Image processing of axe drawings

The cropping was mostly successful, and often produced the three intended images: plan, profile and top. Exceptions appear to occur for some where the original drawings had been labeled, as bounding boxes were drawn around writing, identifying it as shapes, so sometimes writing was included in the crop, or cropped instead of one or more of the views. Another exception occurs when views are drawn too close together or are overlapping, as these are then all included in the same cropped image. Additionally, some of the profile views were cropped in two as many of these have an edge drawn through the centre, which tended to be bolder than the outer lines. Another point of concern is that the crops do not leave much room around the outlines, which sometimes led to problems later with the filling. Despite mostly succeeding in cropping three views from each drawing, a decision was made to only focus on the plans as that is where most of the variation seems to occur between each axe. Only the plans which had two other associated images were chosen, as if there were less or more than three, the wrong view may have 'plan' in the filename and would be included in error. Of the 1825 axe drawings, this brought the number to import into `Momocs` down to 1198.

The filling method was overall successful, and the shape of many axes

was retained, but some data was lost on the outlines, causing them to become jagged. In addition, due to the lack of room between the bounding box and the outlines, some silhouettes were incorrect as they became joined to nearby shapes. This was also true for any views which overlapped in the original drawings.

3.1.2 Crowdsourcing with MicroPasts

All 1825 record sheets were transcribed, so the project is now listed as complete on MicroPasts, and the raw data is publicly available. This data collection method took by far the longest as volunteer uptake was slow, and individual tasks took a while to fill out, with 66 fields to complete. Mostly authenticated users took part in the project, however interestingly, some anonymous users also contributed a number of transcriptions. Investigation of the recovered data showed almost all of the features had impossible outliers due to data entry errors at the highest and lowest values. In the NGRs, a common error was that often, the two letters at the beginning of the NGRs were transcribed as numbers. In addition, a number of transcriptions had data missing, or were completely blank – this was found to be due to a glitch on the MicroPasts website which caused the transcription to automatically submit when the enter key was pressed on a single-line field. Some volunteers had also added in their own notes in square brackets about the sheet they were transcribing, which was not in the instructions. These also needed to be removed as they were not part of the original dataset. Overall, however, the vast majority of data appears to have been entered correctly.

3.1.3 Representativeness of data collected

Flaws in the cropping and filling processes to produce the silhouettes entailed that of the 1198 outlines imported into Momocs, 1137 could be used in the final analysis. The 61 outlines which were dropped were unrecognizable from their original drawings. The 1137 usable outlines have had some slight data loss, for example their edges have become slightly more jagged, however not

enough to truly affect the overall shape. For the record sheets, however, as much data was retained as possible, as all 1825 sheets were transcribed. This will help to account for the axe drawings which cannot be included in the morphometric analysis with Momocs, as some educated guesses can be made about their shapes based upon analysis of their measurements and comparison with other similar axes which had usable silhouettes.

3.2 Morphometric analysis using record sheet data

3.2.1 Sorting variables

The data cleaning step after downloading the data from MicroPasts revealed that many axes did not have measurements for all 56 features. The dotplot shows a significant decrease in frequency of measurements for features below bladang (Figure 3.1).

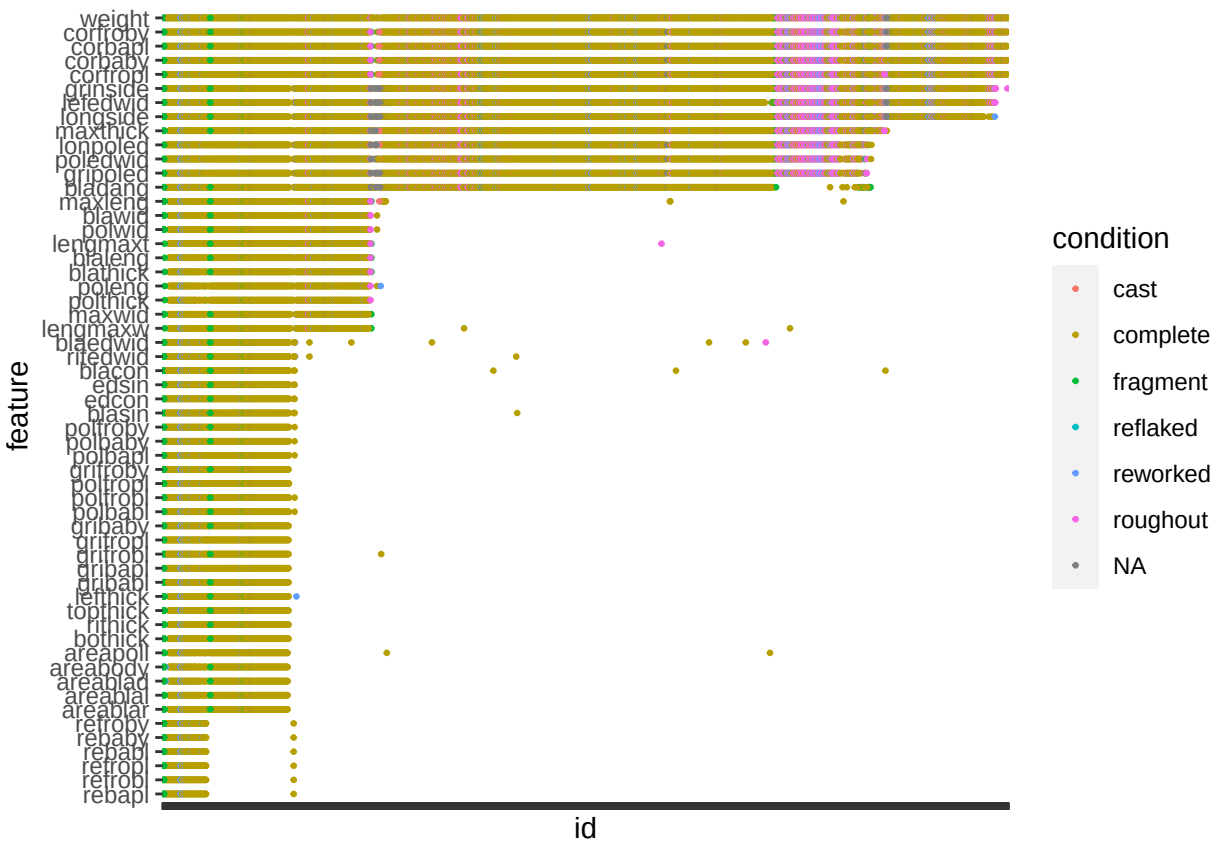

Figure 3.1: Dotplot to visualise the amount of measurements for each feature and the conditions of the axes.

A tally was created in order to choose a cut-off amount for the most frequently measured features, which showed that bladang was recorded 1202 times, whereas the next feature below it, maxleng, is recorded just 407 times. Using 1207 occurrences as the cut-off left 13 features, which helped to reduce dimensions considerably even before PCA. Next, the conditions of the axes were considered. By far, the vast majority of axes were in complete condition. Roughout and cast condition axes were also kept as these retain the original or intended shape. Axes in other conditions were excluded. After these exclusions, 125 axes were omitted, leaving 1700 axes in the analysis.

3.2.2 Univariate

At first glance, it is evident that many of the variables' densities show a fair amount of variation either side of the mean/median, they are mostly skewed, and mostly unimodal (Figure 3.3). Because some of these distributions are skewed due to outliers, I have provided lines indicating the mean and median for each. Distance between the mean and median lines grows depending on the leptokursis of the variable. This is evident in gripoled, lefedwid and weight. The two grinding measurements are bimodal, whilst bladang and lonpoled are potentially trimodal. Bladang appears to have three modes very close together at more acute angles, with a rise in its tail at larger angles. The mean and median are unhelpful measures of central tendency for these multimodal distributions, and the peaks should be paid attention to instead (Shennan, 1988, p.39). Another noticeable outcome is that the cortex measurements seem to mostly be 0, with very little variation around the central values. Weight has the most skewed distribution of all, with values reaching more than 2000g, and the rug beneath the density showing that most of the weights fall below 1000g.

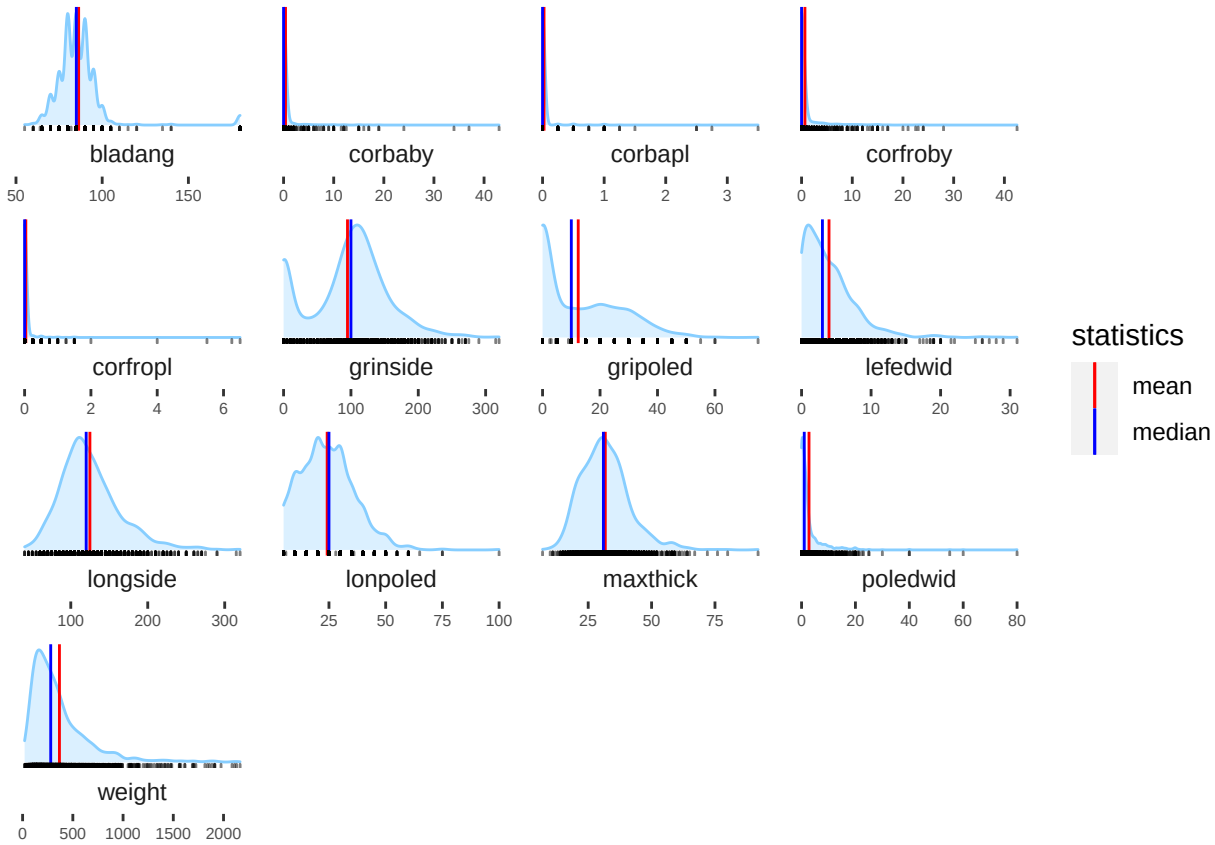

Figure 3.2: Density plots of the distribution of measurements for each feature.

The violin plots show that all features have at least some outlier measurements (Figure 3.3). As some of the distributions are skewed, their interquartile ranges (IQRs) appear misleadingly tiny, which reaffirms that boxplots would have been a bad choice here. For example, poledwid, whose IQR is a thin line just above 0, has outliers up to around 80mm. The cortex measurements are similarly difficult to see. However, corbaby and corbapl do appear to be mostly 0 as their outliers are not as extreme, so do not skew the axis. Violin plots have also proved to be a better choice for visualising bimodal distributions as we can see that often the IQRs only spread around the most frequent mode, and also that the median is misleading. This is evident for grinside and lonpoled. As expected, the variation in bladang appears to concentrate at values just before 100, which is where the three peaks fall, however there are outliers at lower and higher values. We can also see that the most common values for weight are actually lower than 500, and that there are many outliers. Gripoled

shows the most variation in most common values.

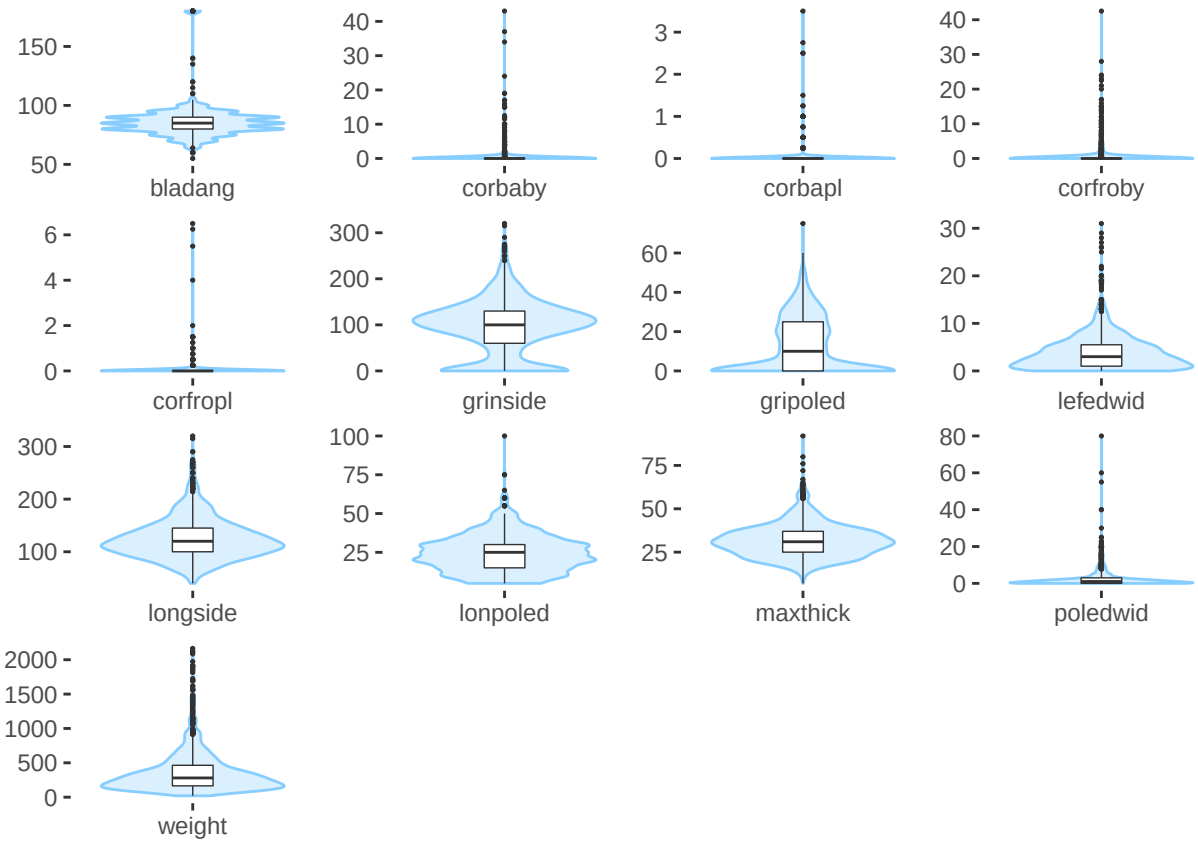

Figure 3.3: Violin plots of measurement distribution for each feature.

The summary statistics further indicate skewed variables, such as poledwid, whose typical values fall between 0 and 3mm, whilst the maximum value is 80. Lefedwid is also similar with most values falling between 1 and 5mm, and a maximum value of 30. As expected, the variable with the largest range is weight, with an interquartile range of 297, its lowest value being 20g, with its maximum at 2165g. Another observation is that a typical blade angle lies between 80 and 90 degrees, and the largest is 180. The 1st and 3rd quartiles for all cortex measurements are both 0. After these univariate results, a decision was made to omit the cortex measurements as they were mostly 0 and only refer to the area of unpolished surface on the axe.

3.2.3 Principal components analysis

The first PC explains the most variance, however this is still just 38.8% of the overall variance; the second PC accounts for 16.8%, and the rest of the

Table 3.1: Summary statistics.

bladang	grinside	gripoled	lefedwid	longside	lonpoled	maxthick	poledwid	weight
Min. : 55.00	Min. : 0.0	Min. : 0.00	Min. : 0.000	Min. : 40.0	Min. : 5.00	Min. : 7.00	Min. : 0.000	Min. : 20.0
1st Qu.: 80.00	1st Qu.: 60.0	1st Qu.: 0.00	1st Qu.: 1.000	1st Qu.:100.0	1st Qu.: 15.00	1st Qu.:25.00	1st Qu.: 0.000	1st Qu.: 165.0
Median : 85.00	Median :100.0	Median :10.00	Median : 3.000	Median :120.0	Median : 25.00	Median :31.00	Median : 1.000	Median : 280.0
Mean : 86.35	Mean : 94.9	Mean :12.41	Mean : 3.949	Mean :124.8	Mean : 24.21	Mean :31.73	Mean : 2.764	Mean : 366.9
3rd Qu.: 90.00	3rd Qu.:130.0	3rd Qu.:25.00	3rd Qu.: 5.500	3rd Qu.:145.0	3rd Qu.: 30.00	3rd Qu.:37.00	3rd Qu.: 3.000	3rd Qu.: 463.8
Max. :180.00	Max. :320.0	Max. :75.00	Max. :31.000	Max. :320.0	Max. :100.00	Max. :92.00	Max. :80.000	Max. :2165.0
NA's :466	NA's :52	NA's :318	NA's :67	NA's :73	NA's :318	NA's :286	NA's :316	NA's :7

variance is split over many subsequent PCs (Figure 3.4). The first two PCs account for 55.6% of the variance. PC3 and PC4 account for 10.9 and 10.1% respectively, in total explaining 21%, so these were also kept for analysis. The first 4 PCs account for 76.6% of the variance. Likewise, the eigenvalues for PC1 and PC2 are 3.5 and 1.5, which equates to 38.8% and 16.8% of the sum total of eigenvalues, which is 9.

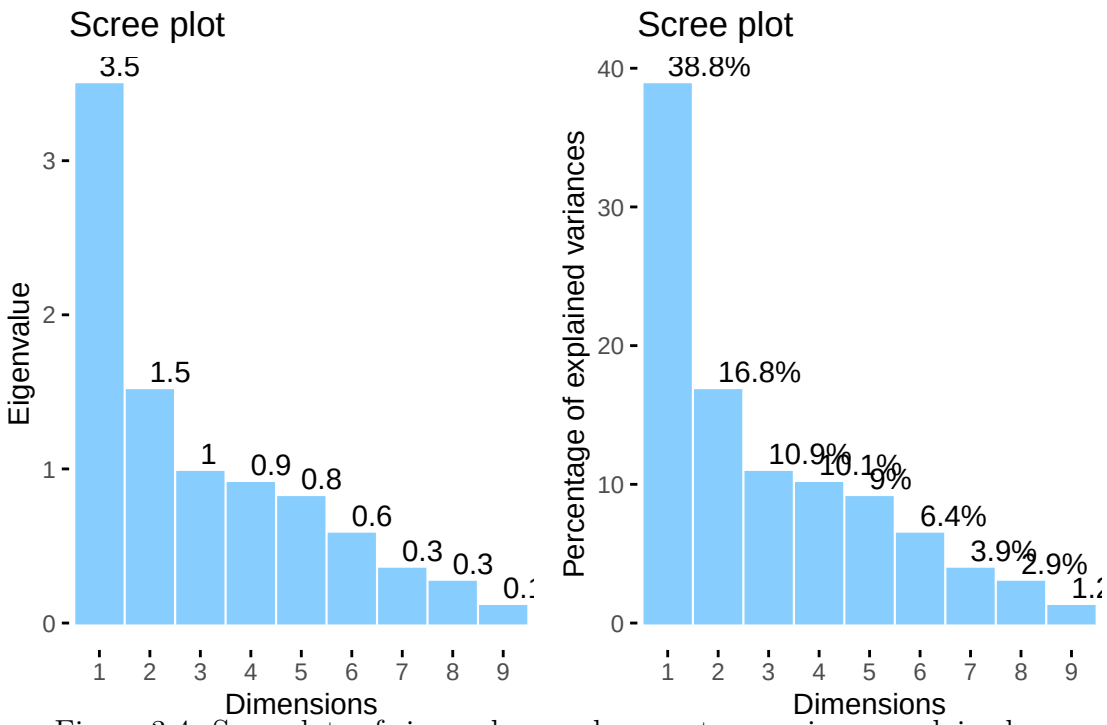

Figure 3.4: Screeplots of eigenvalues and percentage variance explained for PCA.

Looking at the variable loadings, there are 4 variables which are strongly correlated with PC1, which are weight, maxthick, grinside and longside at 0.89, 0.83, 0.74 and 0.71 respectively (Table 3.2). Gripoled, lefedwid and lonpoled are also moderately correlated. For PC2, its strongest correlation is with poledwid at -0.69, and is moderately correlated with longside, bladang and lefedwid. PC3 has no particularly strong correlations, however it mainly relates to lonpoled and gripoled. Finally, PC4's correlations are weaker still, with its strongest being grinside and lefedwid at 0.47 and 0.46.

The biplot for PC1 and PC2 shows the 4 variables which are strongly correlated with PC1 correlating positively with one another, and likewise for those correlated with PC2 (Figure 3.5).

Table 3.2: Variable loadings sorted by strongest correlation with PC1.

	PC1	PC2	PC3	PC4
weight	0.89	0.25	0.12	-0.20
maxthick	0.83	0.19	0.13	-0.30
grinside	0.74	0.01	0.01	0.47
longside	0.71	0.55	0.12	-0.02
gripoled	0.58	-0.35	-0.54	0.28
lefedwid	0.50	-0.40	0.34	0.46
lonpoled	0.50	-0.35	-0.57	-0.36
bladang	0.31	-0.48	0.31	-0.35
poledwid	0.08	-0.69	0.32	-0.15

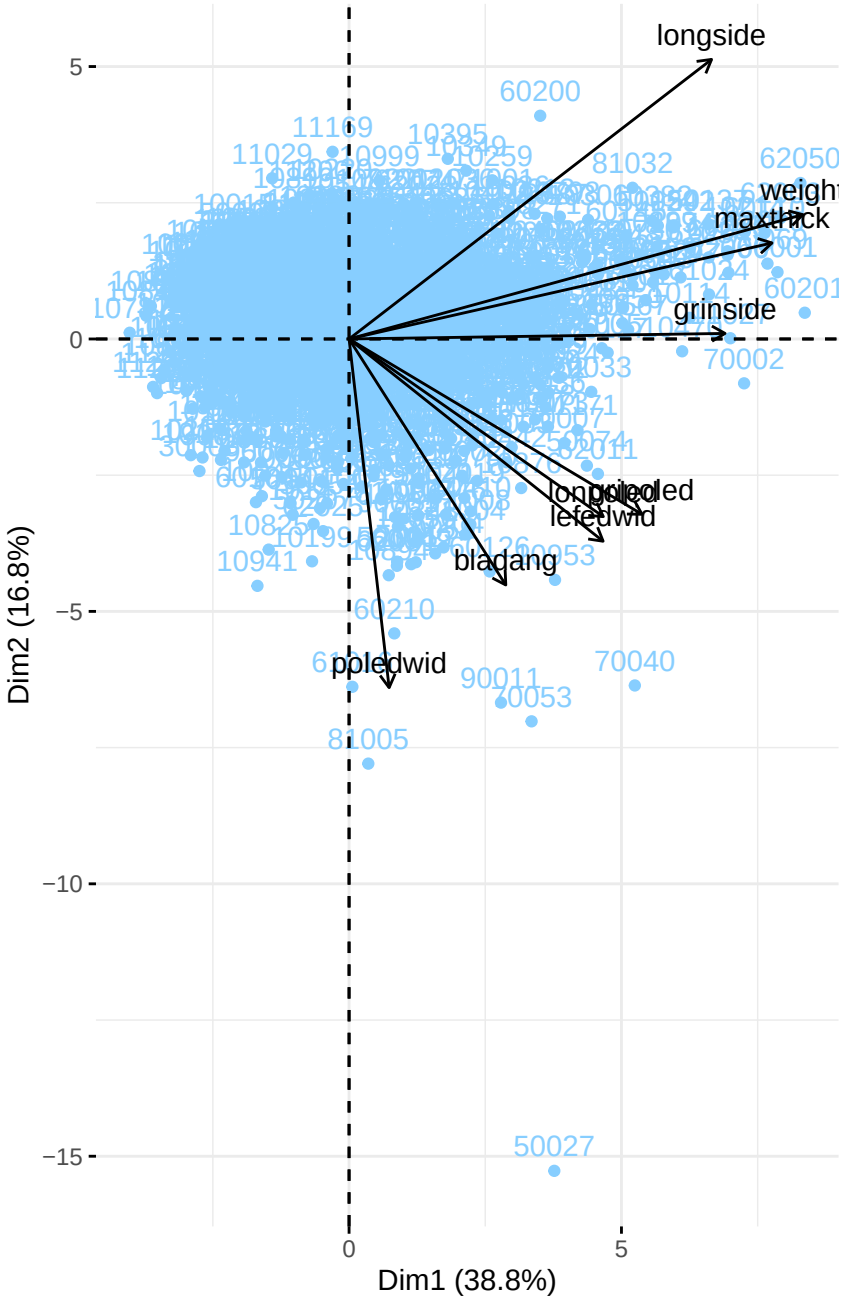

Figure 3.5: PCA: individuals and variables for PC1 and PC2.

Weight and maxthick are strongly correlated with one another, as are gripoled, lonpoled and lefedwid. The arrangement of the individuals on PC1 and PC2 shows long, heavy, thick axes with substantial grinding on their left side edges on the right, and the shortest, lightest, thinnest axes with smaller values for grinside on the left. For example, 62050 on the far right weighs 2085g, has a longside of 320mm, and a maxthick of 58mm. Similarly 60201 weighs

2110g with a longside of 270mm and a maxthick of 58mm. Meanwhile on the far left (which is difficult to discern as so many axes fall here), examples include 10787 which weighs just 25g with a longside of 45mm and 70006 weighing 30g with a longside of 70mm. Axes which are plotted further down PC2 may have larger polls, for example 50027 and 70053 which have poledwid measurements of 80mm and 55mm.

For PC3 and PC4, the biplot reveals some interesting inverse relationships between variables (Figure 3.6).

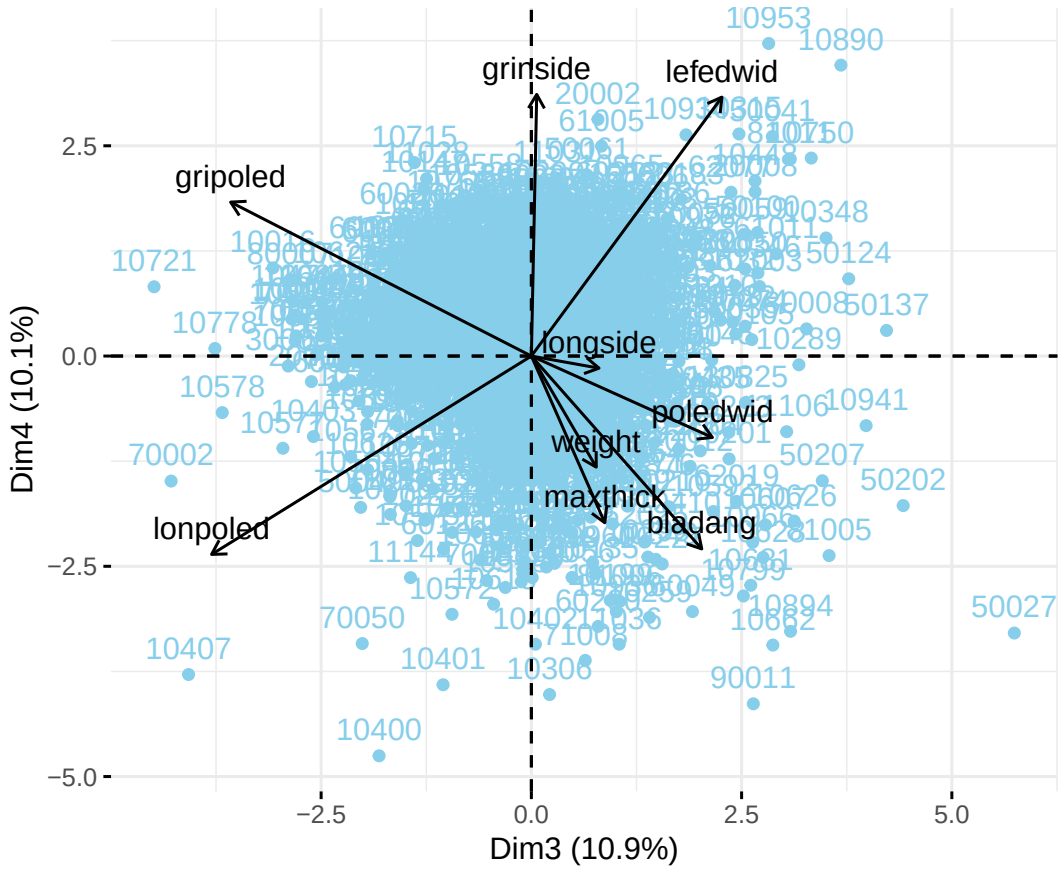

Figure 3.6: PCA: individuals and variables for PC3 and PC4.

Axes with larger lefedwid measurements tend to have smaller measurements for lonpoled, and vice versa; the same is happening with poledwid and gripoled, and potentially with grinside and maxthick. Axes further to the left of PC3 tend to have longer poll edges with more grinding, and to the right smaller measurements for both. For example, 70002 is 75mm for both, whereas on the opposite end, 50202 is 5mm for both. 50027 appears out of place on

the far right of PC3 as it has 45mm for lonpoled and gripoled, however it has an outlier measurement for the more weakly correlated variable poledwid at 80mm, so this is deciding its placement. As the variable loadings are relatively weak for these PCs, in addition to how little variance they account for, it is slightly more difficult to discern the relations between axes on either side of PCs 3 or 4 (Shennan, 1988, p.267). PC4 appears to mainly be characterised by lefedwid and grinside. Axes at the top of this component tend to have larger measurements for both, for example 10953 has a grinside of 165cm^2 and lefedwid of 31mm, whilst on the other end, 10400 has a grinside of 0cm^2 and a lefedwid of 0mm.

Dropping all NA values from the dataset before PCA left 369 axes which had measurements for most variables. Since the 19 variables from maxwid upwards included features which were more descriptive of shape, PCA was attempted again with these 19 variables but a reduced dataset of 369. Since these all had measurements for 19 variables, imputations were also not needed. For brevity, only PCs 1 and 2 will be analysed here. The variable loadings show that 9 variables are correlated with PC1, with weight and maxwid being the strongest at 0.92 and 0.89 respectively; these are followed by maxthick, blawid, maxleng, blathick, longside, grinside and lengmaxt (Table 3.3). For PC2, the strongest correlations are with polwid and polthick at -0.71 and -0.77, followed by moderate correlations with lonpoled, poleng, maxleng and poledwid. The biplot shows the variables which were most correlated with PC1 correlating with one another, and likewise for those which correlate the most with PC2 (Figure 3.7). Interestingly, PC2 is characterised by all of the poll measurements, whilst PC1 relates to measurements which define the overall size of an axe. This is reassuringly similar to the PCA results from the larger dataset with 9 variables, as poledwid and lonpoled were correlated with PC2 and the overall size features correlated with PC1.

Table 3.3: Variable loadings for PCA with all rows containing NA values removed: 369 axes left.

	PC1	PC2	PC3	PC4
weight	0.92	0.18	-0.06	-0.06
maxwid	0.89	-0.05	0.19	-0.07
maxthick	0.87	0.05	0.08	-0.12
blawid	0.83	-0.14	0.27	0.00
maxleng	0.82	0.42	-0.29	-0.04
blathick	0.75	-0.23	0.42	-0.21
longside	0.75	0.52	-0.28	-0.03
grinside	0.74	0.38	-0.01	0.29
lengmaxt	0.73	0.27	-0.42	-0.18
blaleng	0.67	-0.20	0.33	-0.22
lengmaxw	0.53	0.19	-0.20	-0.32
polwid	0.51	-0.71	-0.27	0.21
gripoled	0.50	-0.10	0.25	0.51
lonpoled	0.50	-0.58	-0.24	0.21
lefedwid	0.42	0.07	0.34	0.44
bladang	0.27	-0.31	0.44	-0.04
polthick	0.27	-0.77	-0.28	-0.21
poleng	0.23	-0.53	-0.52	0.26
poledwid	-0.07	-0.40	0.09	-0.51

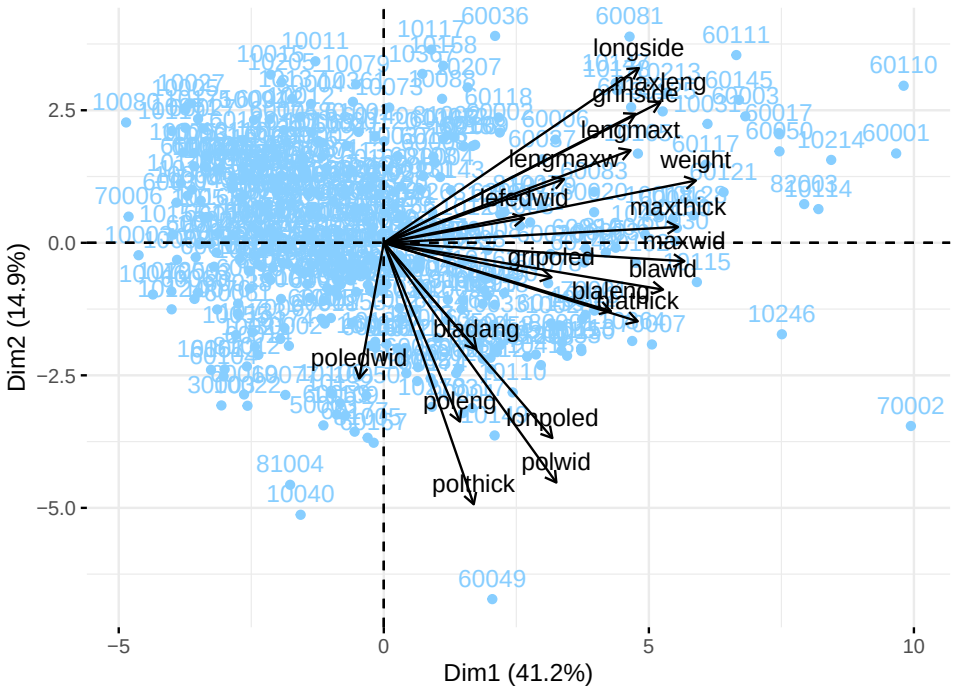

Figure 3.7: PCA with 19 variables measured for 369 axes.

3.2.4 K-means cluster analysis

The WSS plot displayed an 'elbow' or sudden change at 2, 5 and 8, whilst the average silhouette overwhelmingly indicated 2 as the optimal number of clusters with subsidiary peaks at 3 and 7 (Figure 3.8).

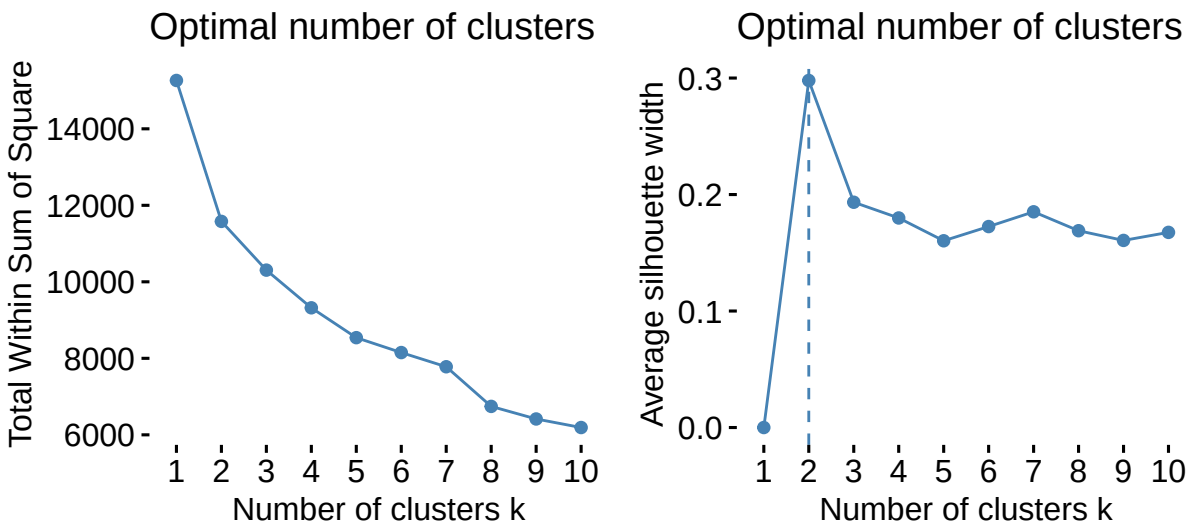

Figure 3.8: WSS and average silhouette plots for k-means clustering.

In light of these results, k-means analyses were performed with clusters set to 2, 3 and 5 (Figure 3.9). The axes visible on the far right of the plot in cluster 1 of the 2-cluster arrangement are the same as those which were on the far right of PC1 in the PCA, which have the largest measurements for weight, maxthick, longside and grinside. These are lumped in with smaller axes, such as 81005, weighing 285g, with a longside of 95mm. Cluster 2 has the same smaller axes on its far left which were on the far left of PC1.

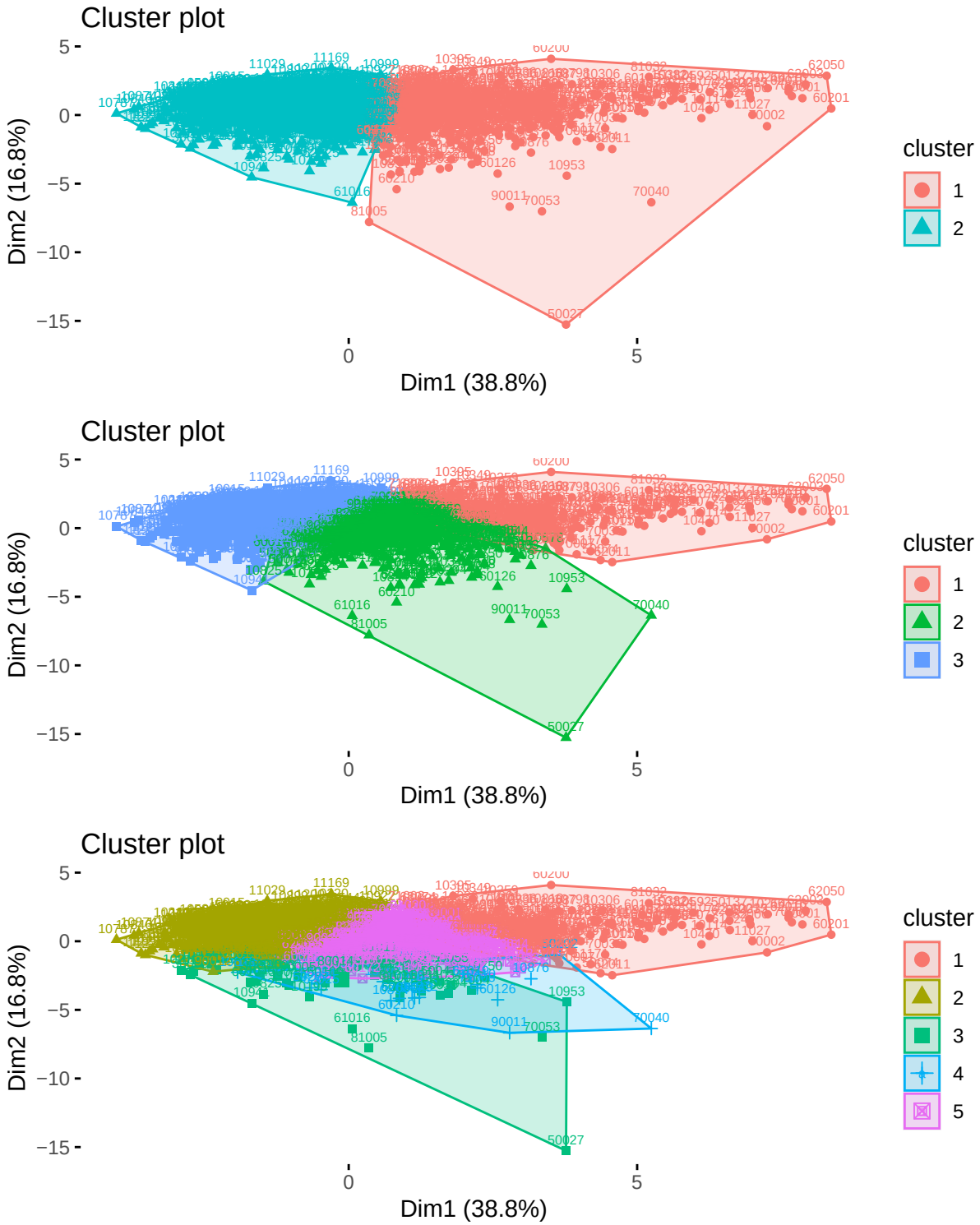

Figure 3.9: K-means analyses with 2, 3 and 5 clusters.

Moving onto the 3 cluster plot, the axes on the extremities of the plot have remained the same, however many of the axes which were on the borders between clusters 1 and 2 previously have now been relocated to cluster

1. Most of the axes tend to fall into clusters 1 and 2, which are similar sizes. This solution has removed some of the smaller axes from the cluster containing the largest axes, which is an improvement, however all clusters are partitioned through where most of the axes fall, almost appearing to overlap, which suggests poor partitioning or overfitting (Rhys, 2019, p.403). Cluster 2 contains the axes with the largest measurements for bladang, gripoled and lonpoled.

The 5-cluster plot has reassuringly made negligible changes to the previous cluster 1 membership from the 3-cluster analysis, which has now become cluster 4. Cluster 2 also appears to have stayed more or less the same as cluster 3 from the 3-cluster plot, but overlaps considerably with the other three clusters, which were previously one middle cluster. Because this potential solution has made little changes to clusters 1 and 3 from the 3-cluster solution, plus clusters 2, 3 and 4 appear to be splitting one cluster, the 3-cluster solution was selected as the best of these options.

3.2.5 Within-cluster densities

It is clear that cluster 3 mostly contains the axes with the largest values for grinside, longside, maxthick and weight (Figure 3.10). Clusters 3 and 2 have similar modes for lonpoled and gripoled. Cluster 1 is clearly characterised by the smallest values for most variables, however it overlaps considerably with cluster 2's distribution for longside. The bimodal distributions of grinside and gripoled are still evident here and it is clear which clusters cause these. For grinside, the subsidiary mode is due to one of cluster 3's modes, and the largest mode consists of the most common values for cluster 1, 2 and the other mode for cluster 3. The main mode for gripoled is caused by cluster 3, whilst its subsidiary mode consists of clusters 1 and 2. Bladang and poledwid do not show much difference in distributions between the clusters.

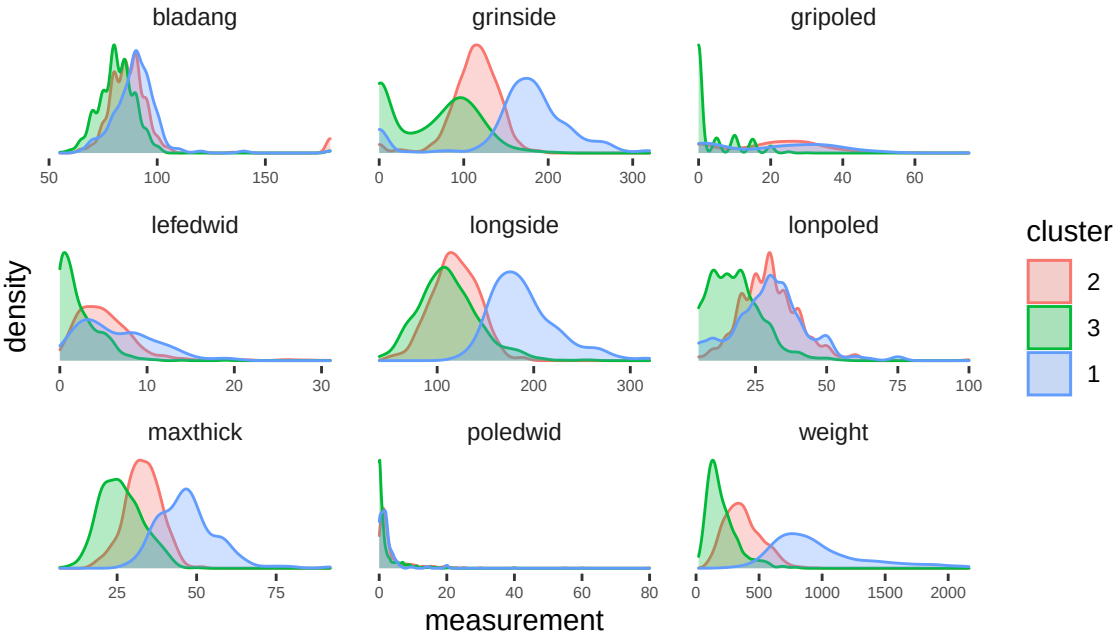

Figure 3.10: Density plots to show distribution of variables within 3 k-means clusters.

3.2.6 Clusters on PCA

The clusters are very clearly visible on PC1 and PC2, but are amalgamated on PC3 and PC4 (Figure 3.11). All clusters overlap, with cluster 1 trailing off to the right, containing the axes with larger measurements for PC1-correlated variables. Cluster 2 contains all of the axes with large measurements for variables correlated with PC2, and is slightly lower on PC2 than the other two clusters. Some axes in cluster 2 are placed similarly on PC1 to cluster 1 axes, but lower on PC2. PCs 3 and 4 do not show much difference between any of the clusters.

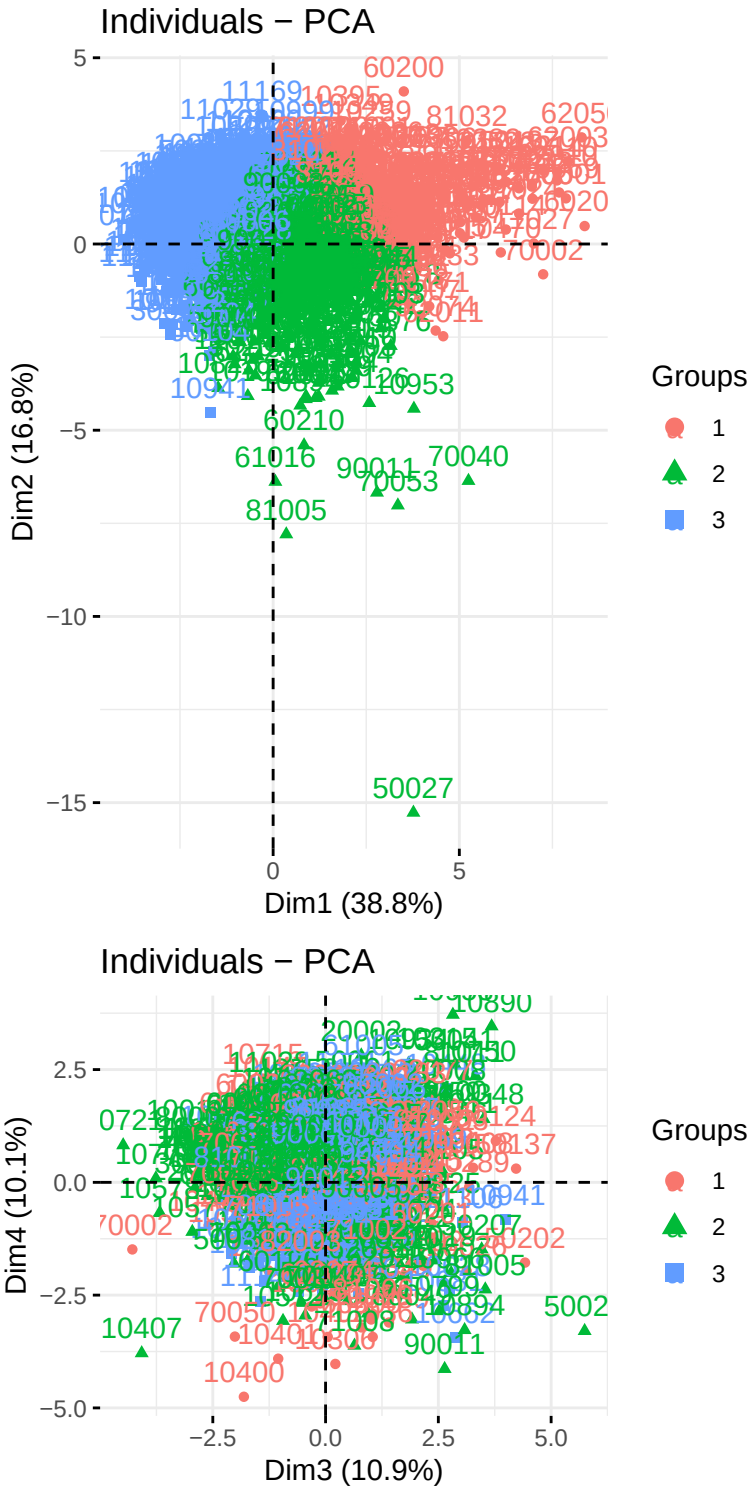

Figure 3.11: PCA of the first and second pairs of PCs with individuals coloured by k-means cluster.

3.3 Outline morphometric analysis

3.3.1 Data sorting

The panel plot of the 1198 outlines imported into Momocs revealed a number of incorrect shapes resulting from the cropping and filling process (Figure 3.12).

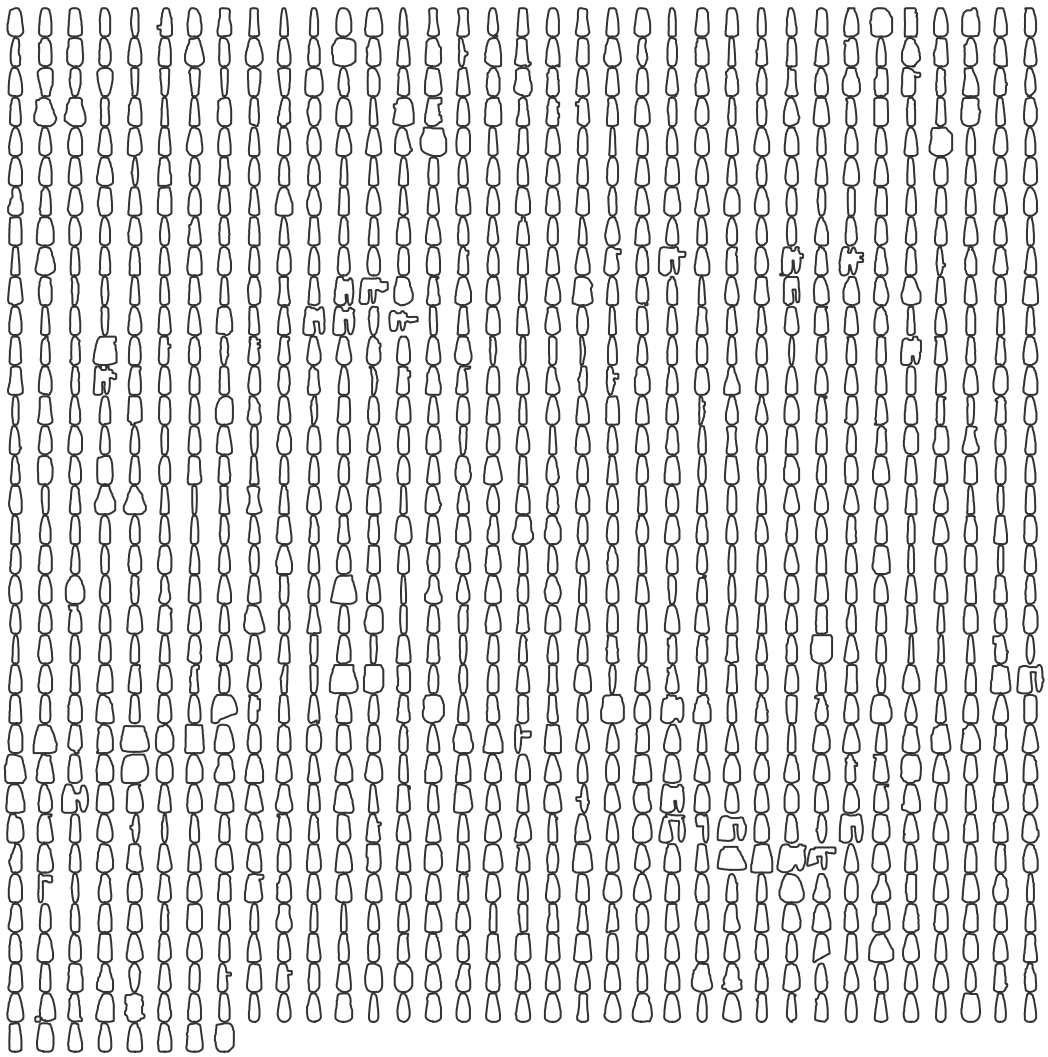

Figure 3.12: All outlines imported into Momocs before inspection.

The final dataset is below (Figure 3.13). Some of the wider, more square outlines are peculiar, however these do exist in the original drawings. In total, 61 outlines were removed.

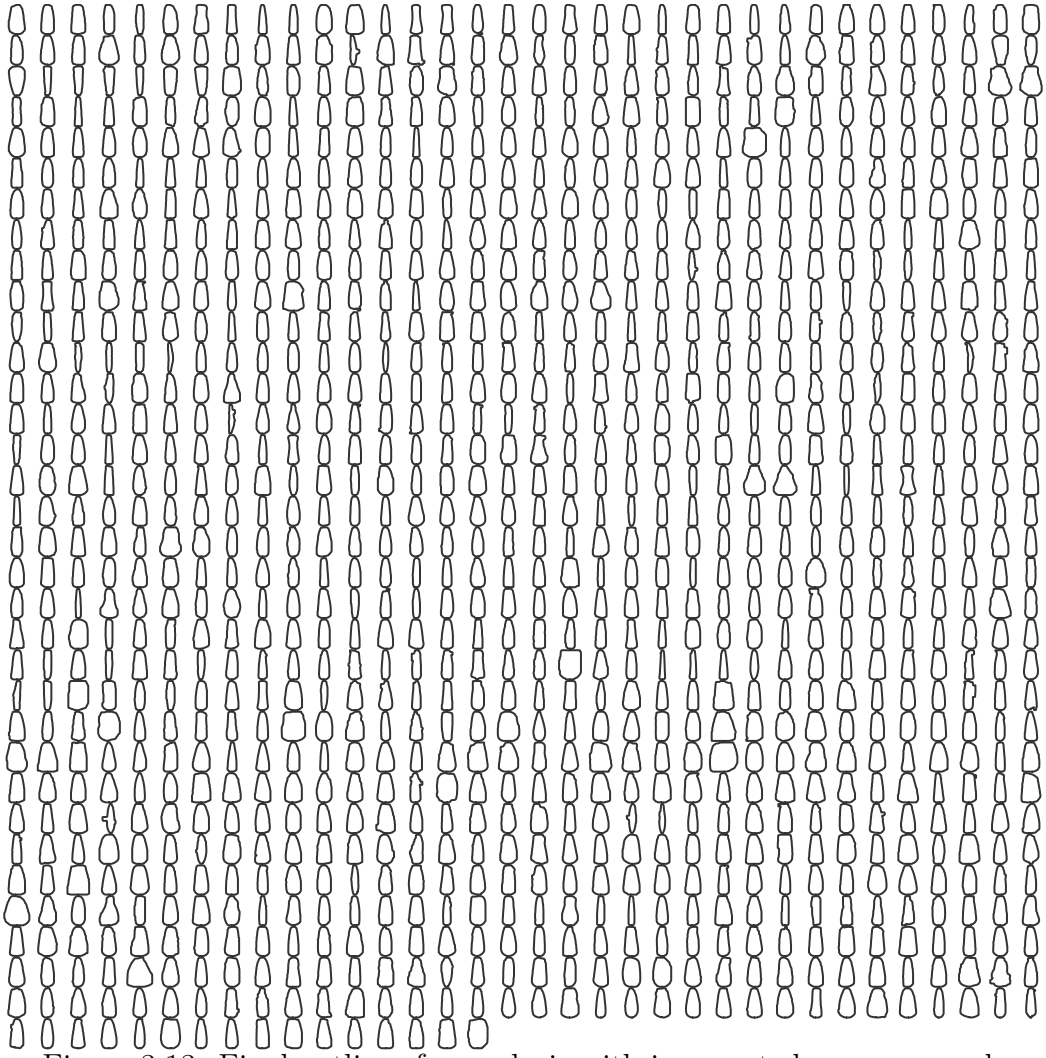

Figure 3.13: Final outlines for analysis with incorrect shapes removed.

For the first, middle and last axes in the dataset, the harmonic power plots showed that there was no more noticeable improvement on the reconstructed shape after 12 harmonics (Figure 3.14). This was also the case when running the `hcontrib` function on the entire dataset passed to `efourier` with 20 harmonics (Figure 3.15). As a result of these checks, 12 harmonics was settled on for the elliptical Fourier transformation.

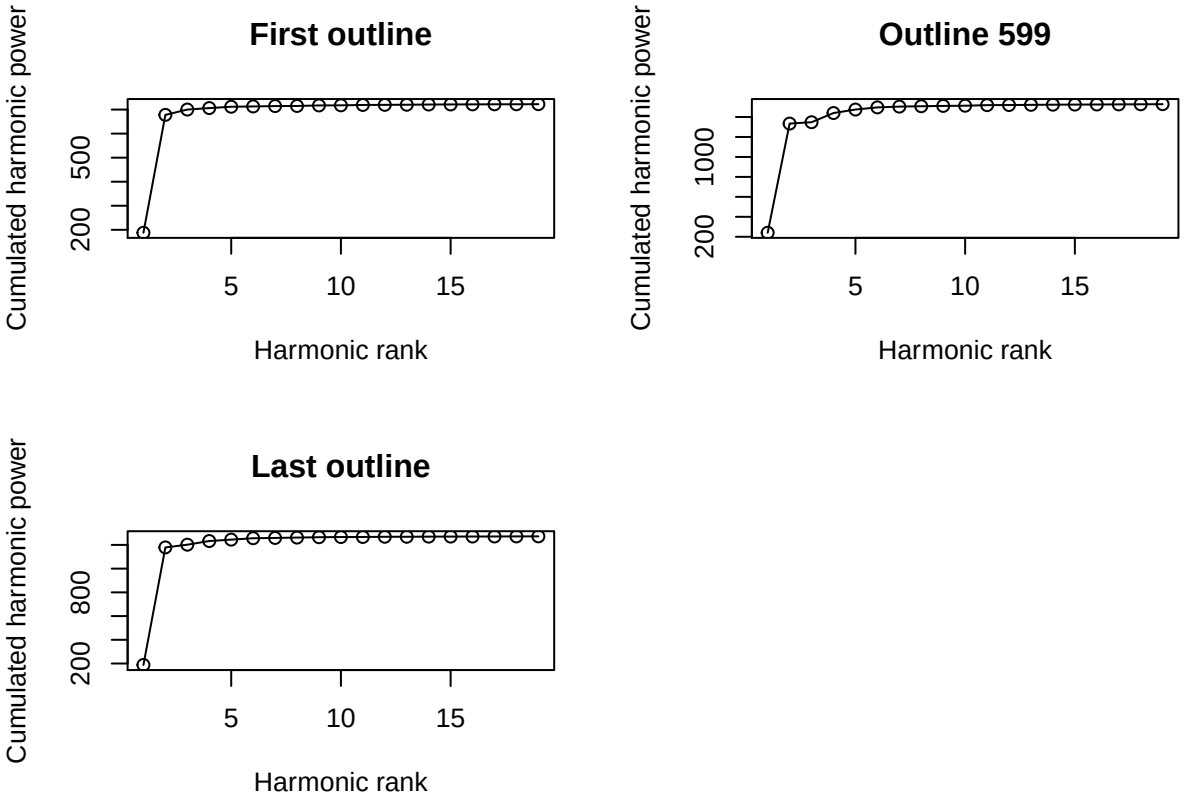

Figure 3.14: Cumulated harmonic power without the first harmonic for three outlines.

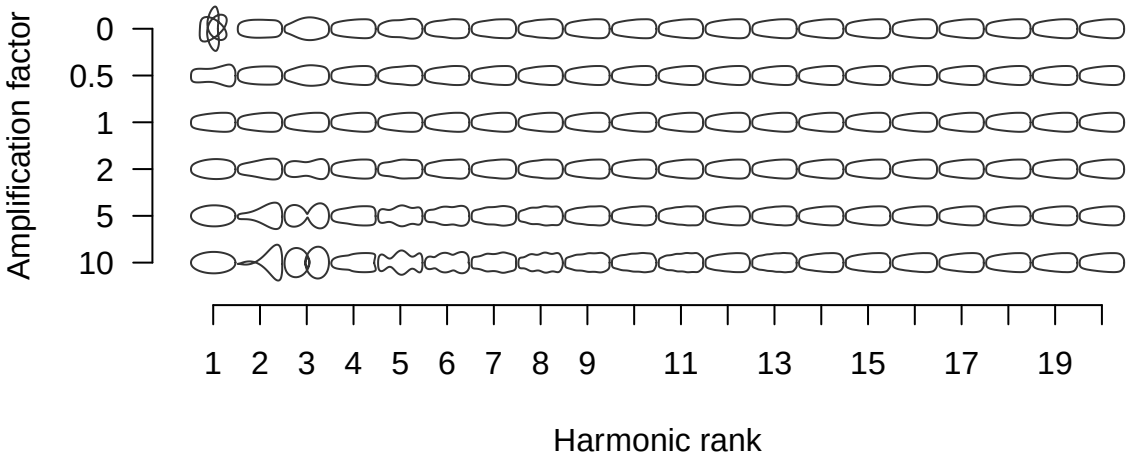

Figure 3.15: Harmonic contribution plot to show how many harmonics can reconstruct a typical axe outline.

3.3.2 PCA on elliptical-Fourier-transformed outlines

The PC contribution plot indicates that only the first three PCs explain any variation in the outlines (Figure 3.16). Here, we can see that the first PC accounts for width, and the second shows the ratio of blade width to poll width. PC3 appears to distinguish rounder blades and edges from straighter blades and edges, but seems to contain shapes which do not exist. The rest of the PCs do not appear to show much variation at all.

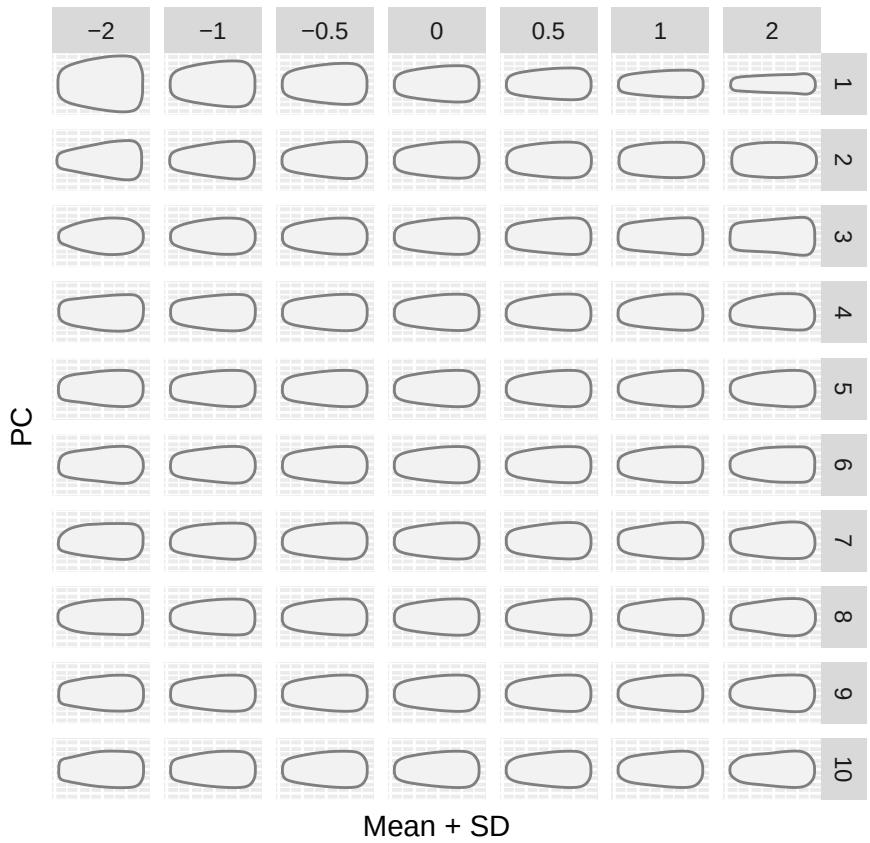

Figure 3.16: Variation in outlines explained by each principal component.

Although some variation is shown in PCs 3 and 4, the proportion of overall variance they explain is negligible (Figure 3.17). In light of this, only PC1 and PC2 were used for PCA.

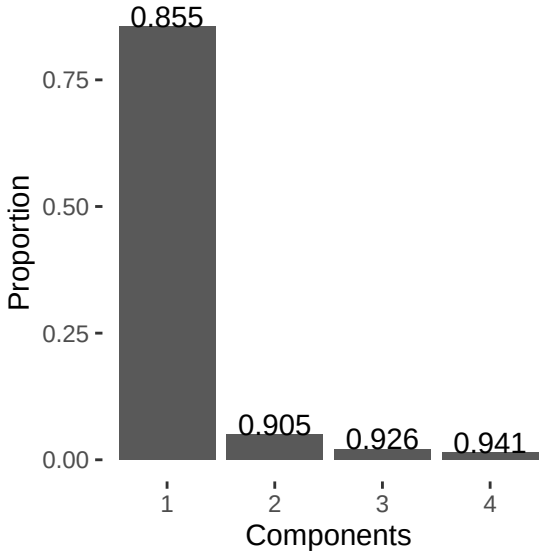

Figure 3.17: Scree plot of proportion of variance explained by each principal component.

The mean shapes drawn onto the morphospace are wider toward the left of PC1 and thinner toward the right (Figure 3.18). Mean shapes at the top of PC2 have much wider blades than polls, and wider polls than blades toward the bottom, with straight axe shapes appearing toward the origin. The majority of axe shapes fall around and just beyond the origin, represented by the mean shapes in the centre. The mean shapes on the morphospace show some odd, seemingly predictive shapes at the far edges where no axes are plotted, for example the almost bowling-pin-like shapes at the top and bottom right corners. At first glance it would appear that the shapes are facing two different directions, which was initially assumed to be due to incorrect alignment, however this has been observed in other studies which use Momocs (Bourcy, 2018, p.49). As PC2 appears to represent the ratio of blade to poll width, these odd shapes indicate an attempt to show thinness at opposing ends of the outline. On the far left of PC1 but in the centre of PC2, individuals are plotted where the mean shapes are wide and square. The vast majority of individuals appear to fall where the mean shapes are slightly narrower toward the poll edge and in the mid-range of thickness, represented by the shapes just to the left of and at the origin. In addition, some straight oblong shapes seem to be present in the dataset, plotted just to the right of the origin. Some less common shapes have

also become evident here, with individuals plotted on the far left end of PC1 which are wide and square, and to the far right, which are slightly bulbous at the blade end. There are also a small amount of outliers at the top and bottom of PC2 just to the left of the origin, showing there are axes in the dataset with much larger blades than polls and vice versa.

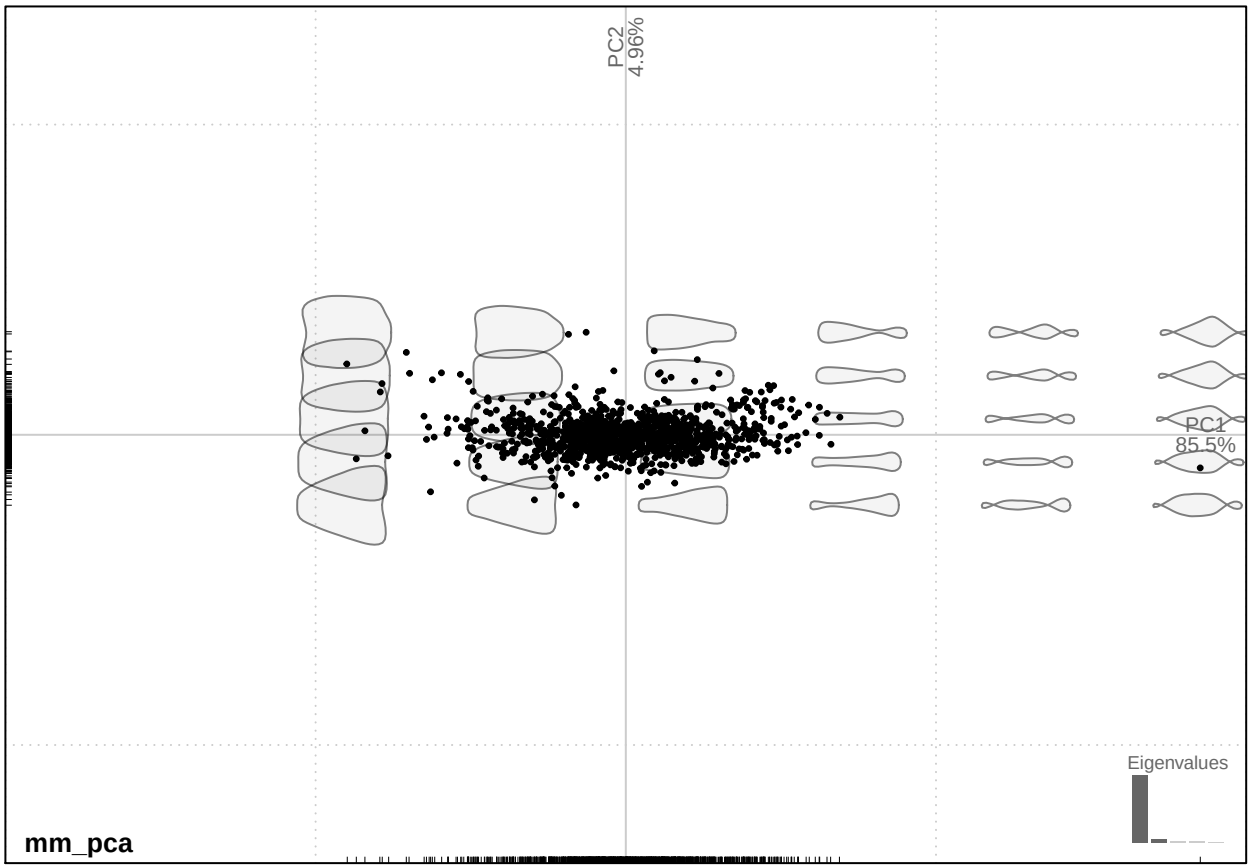

Figure 3.18: PCA of elliptical Fourier-transformed outlines

3.3.3 K-means analysis on the Momocs PCA results

K-means objects for up to 10 clusters were created and their total within sum of squares were inspected, which showed that the greatest change was at 2 and 3 clusters, so these were chosen for analysis. A 4-cluster plot was also trialed to see if any of the clusters would retain their membership when a larger number of clusters was imposed.

Firstly, the 2-cluster solution shows more narrow, straighter axes in the right cluster, and wider, slightly more trapezoidal axes in the left. Both clusters

appear about the same size. There is potential overfitting here as the groups split the densely populated area of the plot in two, and include outliers in with these more typical axe shapes (Figure 3.19).

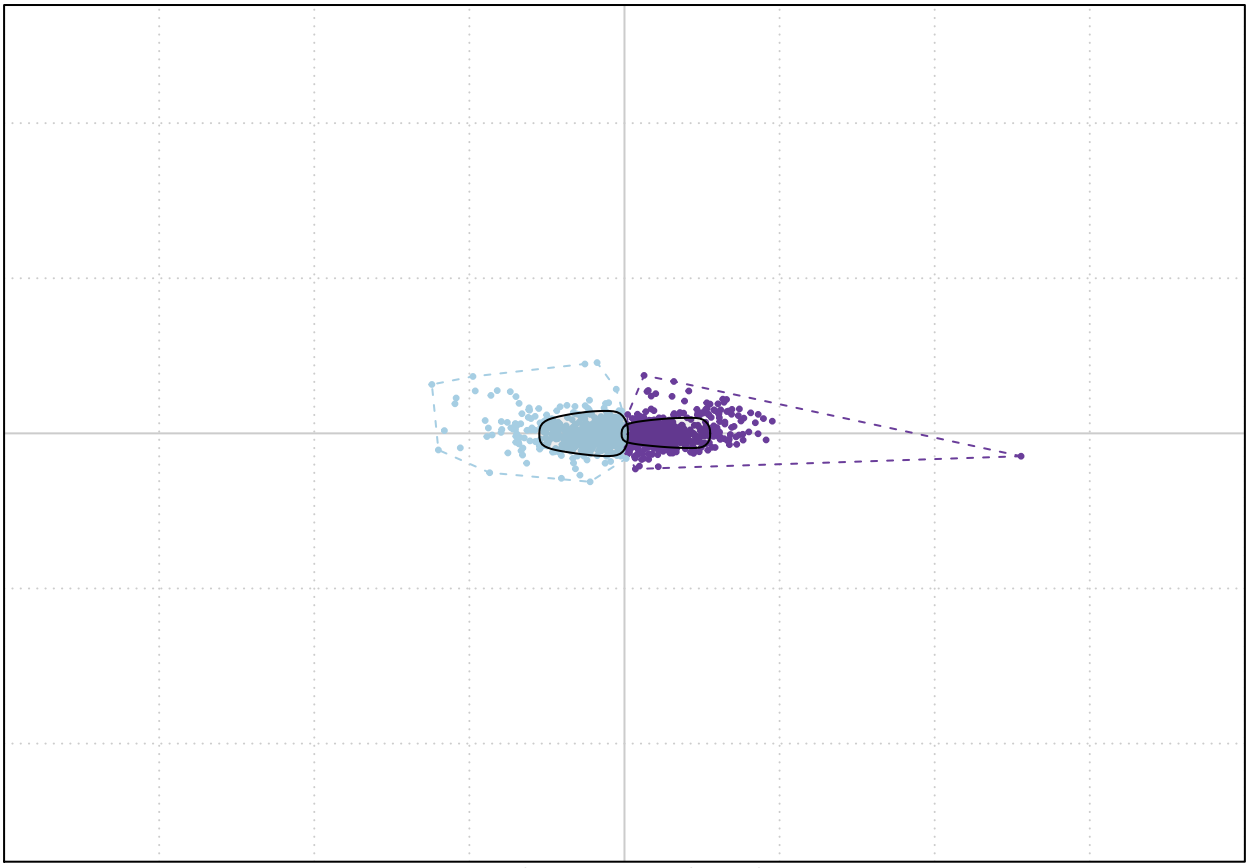

Figure 3.19: Momocs k-means with 2 clusters

For 3 clusters, there is now a wider, more trapezoidal mean shape cluster in purple, a narrower, straighter shape cluster in red, and a shape which appears to be between these two extremes in the central blue cluster. The clustering of individuals around the origin has mostly been contained in the blue cluster, however some are still being split into the neighbouring red cluster with the most narrow shapes. The purple cluster contains the most extreme outliers (Figure 3.20).

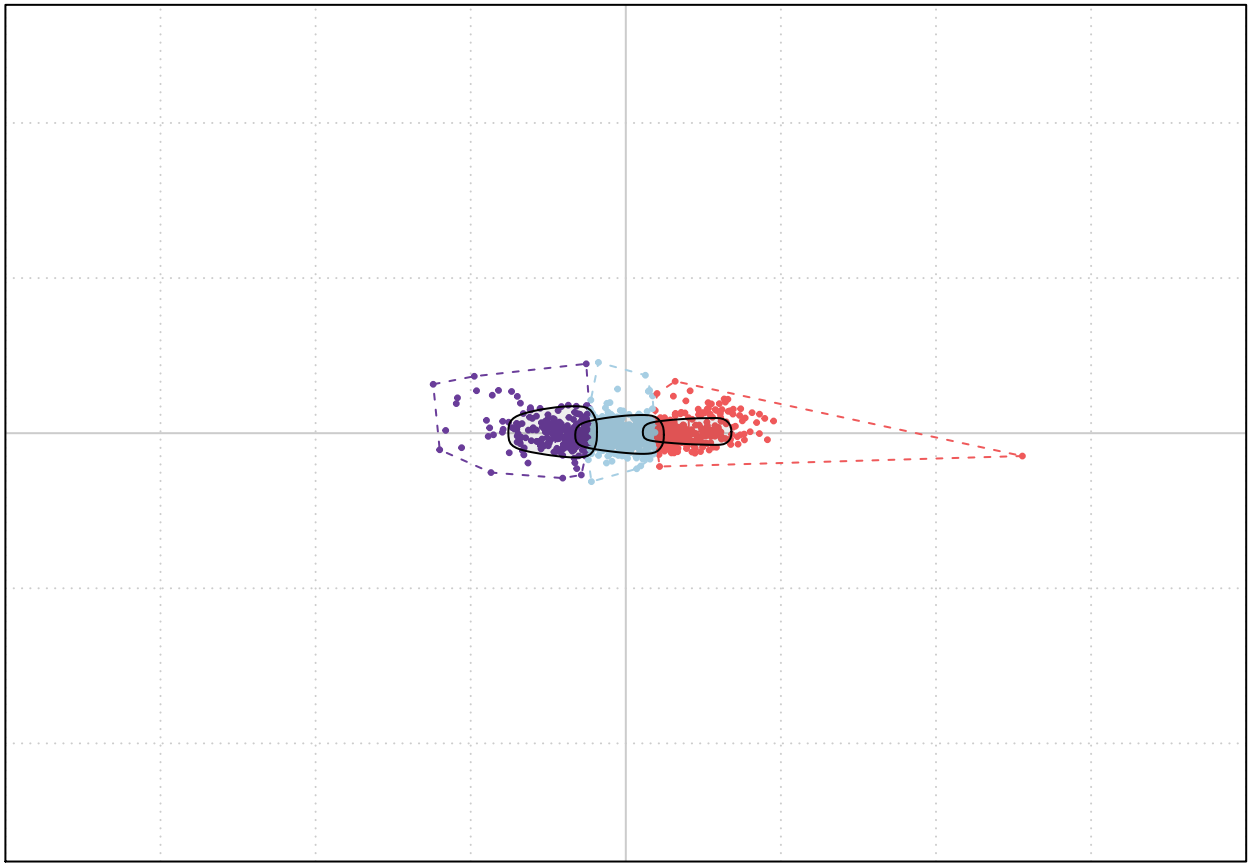

Figure 3.20: Momocs k-means with 3 clusters

When attempting to separate the shapes into 4 clusters, the plot shows two similar looking mean axe shapes for the central clusters which partition the majority of axes, however the thinnest group and the widest group now have more exaggerated mean shapes and smaller cluster membership (Figure 3.21). The wider axe group's mean shape now looks more squat, and the narrower axe group's mean shape is now even more narrow and straight. These probably account better for the axe shapes at both extremes. Despite the WSS not indicating much change at 4 clusters, this option does seem to be the most appealing, as although there may be overfitting of the two central clusters, the most narrow and most wide axes are best represented here. In addition, whilst the mean shapes of the blue and yellow clusters are similar, they do differ in width and straightness.

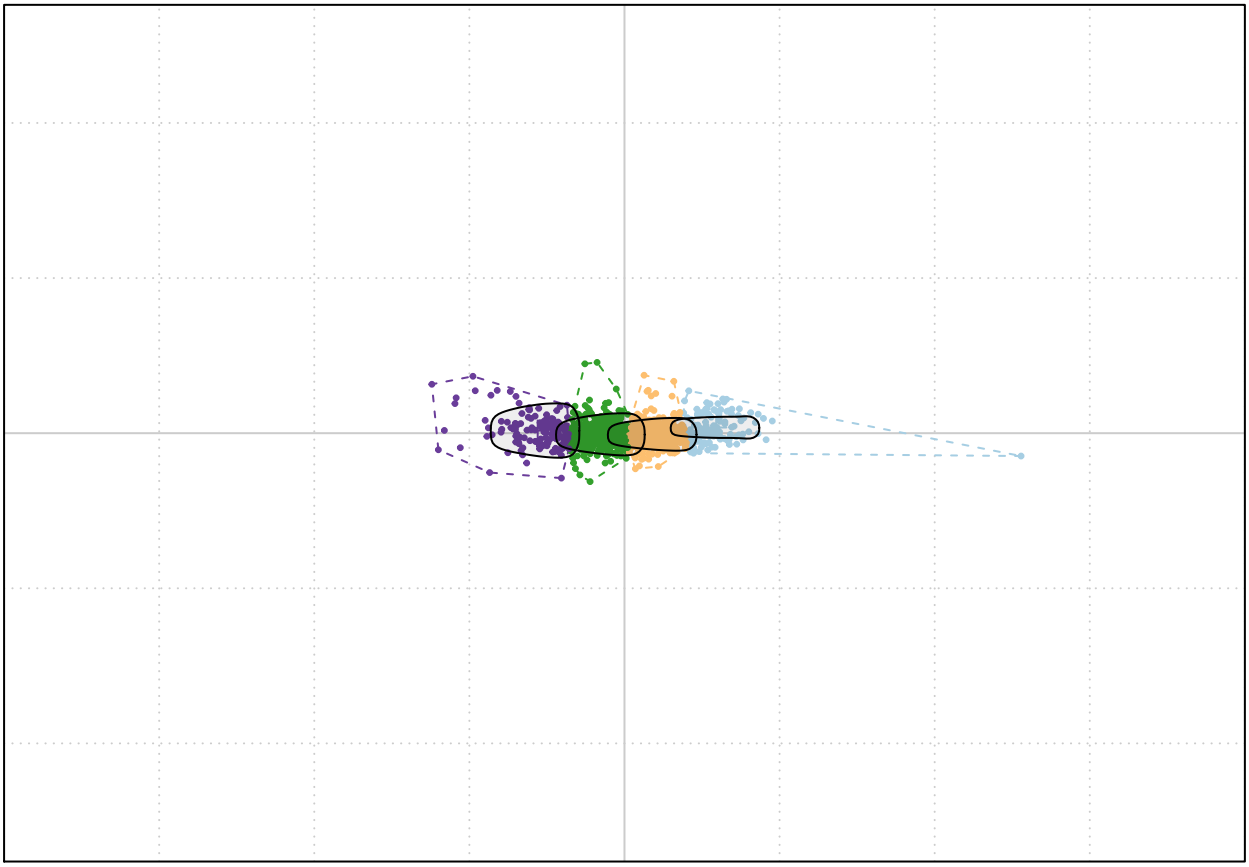

Figure 3.21: Momocs k-means with 4 clusters

3.3.4 Outline cluster membership

In the panel plot with the clusters numbered and colour-coded, the widest axes are clearly in cluster 4, and the most narrow in cluster 1 (Figure 3.22). The two similar clusters, 2 and 3, which were in the centre of the k-means plot, do appear slightly different here, with the axes in 2 mostly being wider than those in 3. Axes in cluster 2 also appear more trapezoidal than cluster 3. Overall it seems that the aspect ratio of the outlines is the deciding factor in the groupings here, however the two wider groups tend to have more axes with wider blades than polls compared to the two narrower groups. Cluster 4 also contains all of the most square axes in the dataset, further suggesting that width of shapes is the cause for this partitioning.

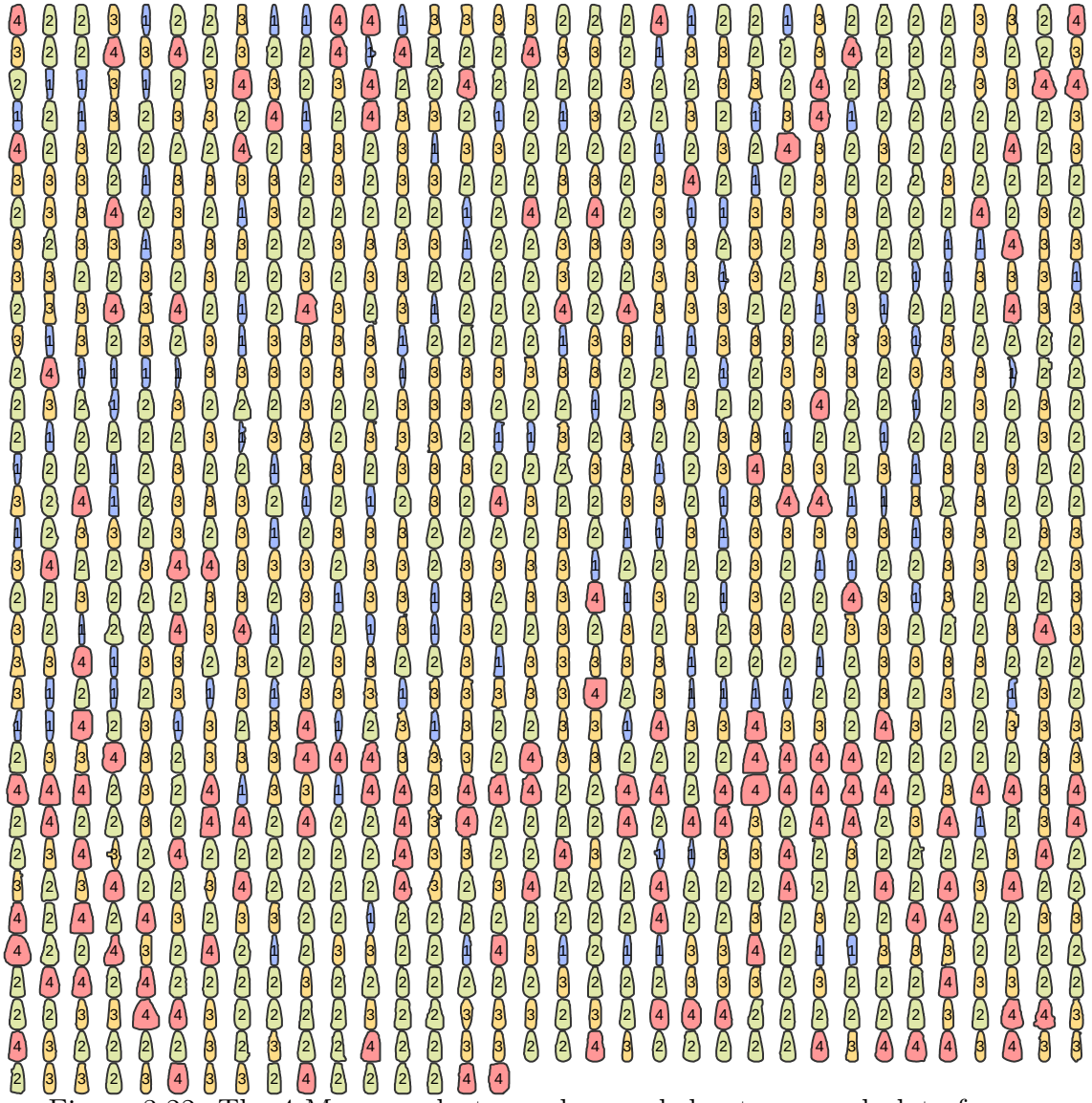

Figure 3.22: The 4 Momocs clusters colour coded onto a panel plot of axe outlines

3.4 Comparisons

The mean shapes of the 4 clusters revealed by using Momocs on the axe outlines show distinct shapes for each group (Figure 3.23). A typical axe from clusters 1 and 2 tends to be wider at the bottom than they are at the top, and is wider overall than those in clusters 3 and 4. Cluster 4 axes are almost entirely straight, and cluster 4 axes appear very squat.

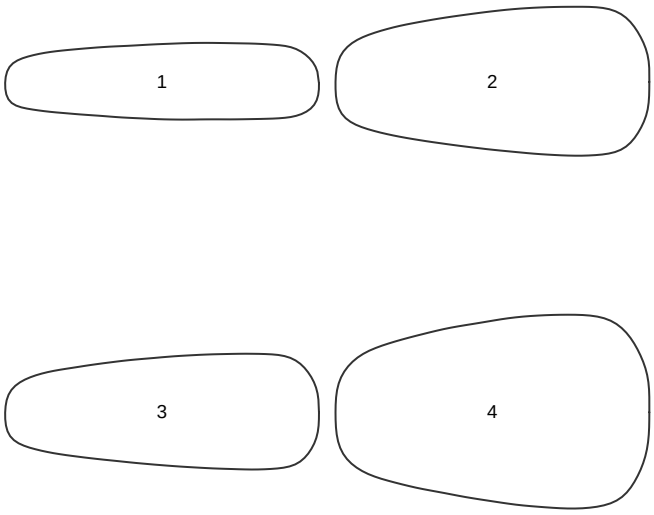

Figure 3.23: Mean shapes for each of the 4 clusters

Meanwhile, it is difficult to distinguish any difference between mean shapes of outlines within the 3-cluster solution from the k-means analysis of the axe measurements (Figure 3.24). This is also the same for the mean shapes of outlines within 3 clusters on the reduced dataset with 19 variables (Figure 3.25).

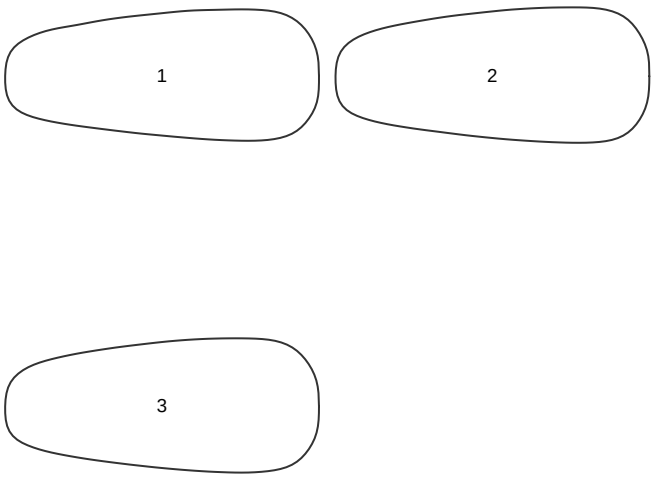

Figure 3.24: Mean shapes for the 3-cluster solution on the axe measurement data.

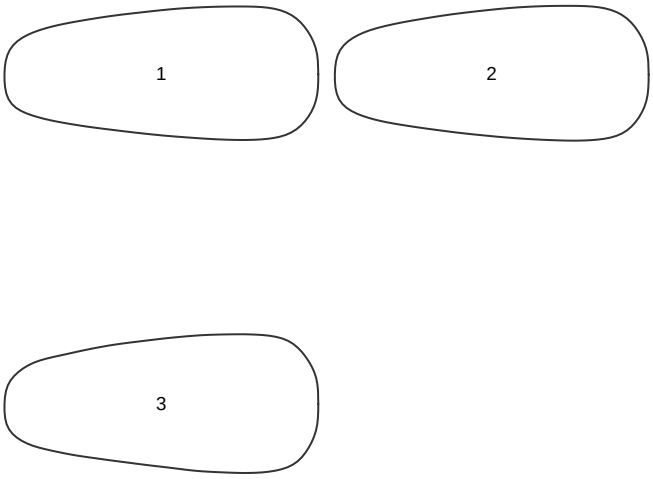

Figure 3.25: Mean shapes for 3 clusters on axe measurement data with 19 variables.

In terms of overlapping clusters between the two sets of k-means analyses, Momocs' group 1 membership is 74.6% the same as measurement group 3, the smallest axes (Table 3.4). Measurement group 1, the largest axes, appears the least numerous, and tends to have either Momocs group 1 or Momocs group 3 shape. The widest group in Momocs is split almost evenly between the mid-size and smallest axe groups from the measurements. Moving onto the geographical distributions of the clusters, we can see that all groups concentrate around the Norfolk area (Figure 3.26). Measurement group 1 appears to hint at concentration in north-west England. Measurement group 3 has the most axes in south-west england out of the size groups, and appears to concentrate equally in three areas. Momocs group 4, the widest, has no axes in the north-west. Narrow axes are least common in Wales.

Table 3.4: Percent similarity between outline and measurement clusters

mm_cluster	1	2	3	total
1	14.0350877	7.894737	78.07018	114
2	5.2272727	51.590909	43.18182	440
3	17.4731183	29.301075	53.22581	372
4	0.7462687	46.268657	52.98507	134

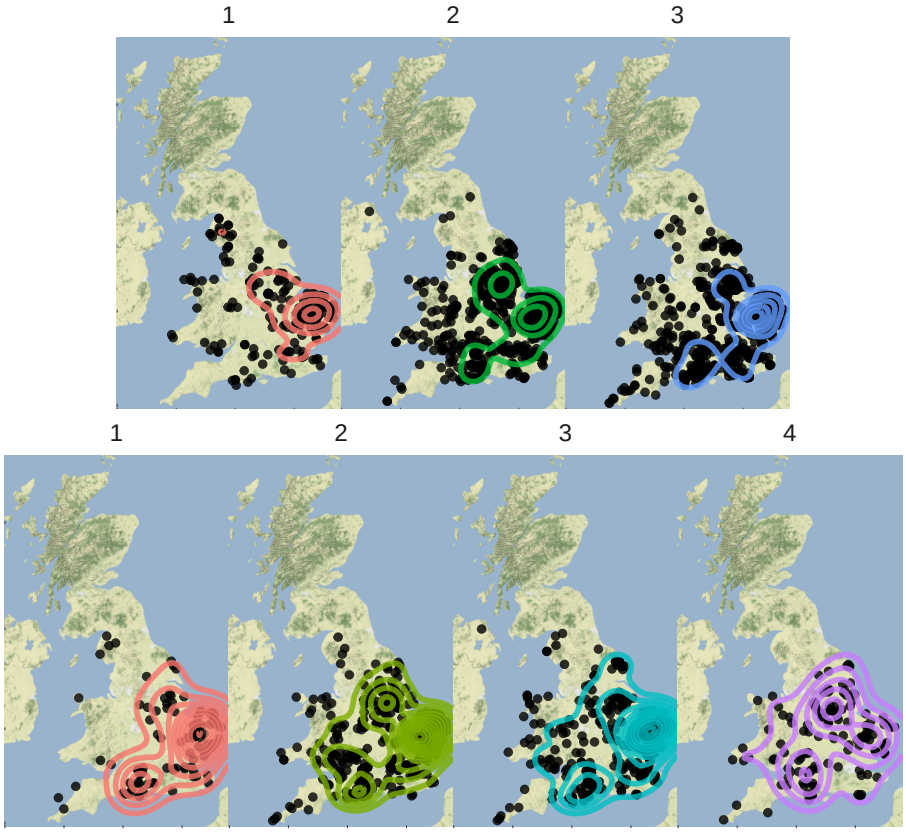

Figure 3.26: Find places of axes in each cluster plotted onto a map of Britain, with measurement clusters above and outline clusters below.

Chapter 4

Discussion

4.1 Limitations and success of the data collection methods

Because this project works with old technical drawings which were never intended to be read by computer software, there was always going to be a degree of difficulty when processing them into the desired format. Faint pencil lines, labels close to outlines and lack of space between views or bounding box borders created most of the problems in the cropping and filling processes. One of the most common issues was that nearby writing would become joined to the largest contour and filled as one shape with the rest of the view. As the filling method removes any contours which are smaller than the largest, if these are connected to the largest they become amalgamated into one shape. Most notably, some of the axes which either overlapped or were drawn very close to one another also happened to be among the largest, which is due to the fact that Pitts was drawing to scale. This unfortunately meant that the final outline dataset was biased against the largest axes, which later became apparent when comparing the analysis results.

Although some of the original drawings are just unusable with this method due to overlapping, there are ways in which it could be improved to successfully crop and fill more of the drawings. One way to improve the method would be to increase the size of the bounding boxes used to find the coordinates

for cropping the views. This could be done by further dilating the binary thresholds of the drawings that are used to find the contours, which would increase the size of the contours, and thus space within the bounding boxes. After this step it would then be acceptable to increase the contrast despite noise coming through from the page behind, as increasing the space around the view would make it the largest continuous contour in the image, so these would be ignored and removed. Another possible improvement to this dataset would be to manually inspect the cropped drawings with more or less than three views associated as some may have filled correctly.

As MicroPasts uses Pybossa, which has a page limit to prevent their server from being overloaded with requests, the option to export task runs on the MicroPasts web interface limits export to the first 20 completed tasks. Additionally it appears as though Pybossa has not been maintained, and some of its documentation is no longer available. As a result there were no instructions on how to bypass the page limit and download data from a project, which necessitated writing a reasonably complicated R script and `Makefile` to overcome. In addition, a potential pitfall of the MicroPasts website is that because it is specific to archaeology and heritage, it most likely does not get as much traffic as other multidisciplinary crowdsourcing platforms such as Zooniverse, which has incidentally been around for longer. This could have contributed to the relatively slow uptake of tasks. Conversely, because of the subject-specific nature of the platform, the volunteers may have particular interest in and enthusiasm for contributing helpfully to a project. In this project's case, it does seem as though volunteers took special care in data entry, as outlined below.

Once the data was downloaded and inspected, there were a number of common mistakes which repeatedly arose in the transcribed data. One example was that because the position of lefedwid moved depending on the sheet format, it would usually have swapped measurements with bladang, as this was the field next to it. Fortunately, many data entry errors were reasonably easy to detect, as they usually came in the form of extremely high or low

outliers. Some aspects of the sheets were ambiguous, such as numbers which were clearly written, but crossed out in a line. There were also numbers written outside of fields on the record sheets. Because of unclear meanings, some volunteers included messages in square brackets in the 'Comments' field indicating that the number was crossed out. This suggests that perhaps an 'Any other comments' field, not to be used in the final data but to help with data cleaning, would have been useful here. Additionally, a number of volunteers had entered values in the wrong fields due to the sheet fields in the image changing from the order of the fields in the form. However, some of these also appeared to have been because there would be an empty field in the record sheet after a series of consecutively completed fields, and the transcriber would enter the previous field's value in the one which should be empty. This could potentially have been mitigated against by changing the layout of the forms to imitate that of the record sheets, however the format would still change from sheet-to-sheet. Numbers instead of letters transcribed at the beginning of the UK NGRs could have been a result of the decline in use of NGRs compared to lat-longs in recent years, so perhaps this should have been explained in the instructions. This would also have been alleviated if data validation was employed in the NGR field so that anything other than A-Z could not be entered as the first two characters. In addition, the glitch which caused the transcription to be submitted when the enter key was pressed on a single-line field is present in other MicroPasts projects, as the template for this project was taken from its GitHub and has been used by others. The glitch did, however, result in surprisingly few completely blank submissions. On balance, volunteers for this project appeared to take care with data entry, and the mistakes made were usually due to ambiguous handwriting.

Some axes still appeared to have confusingly low or high values for some measurements even in the original data, especially when taken in context with some of the maximum measurements such as maxleng or maxwid. For example, in the original record sheet, axe 60117 had a longside of 190mm, but a

maxleng of 21.3mm. Not only does it not make sense that the length of the left side edge would be larger than the maximum length, but an axe with a maximum length of just over 2cm cannot be possible. When inspecting the cleaned MicroPasts data it was clear that 22 other axes had measurements for maxleng which were far too small. These were present in the original data, and it appears as though Mike Pitts had been using centimetres and and millimetres interchangeably. In addition, many of these axes had measurements with decimal points which looked as though they had been slightly scribbled out, however still looked like a decimal point, so the transcriber had still included them. These measurements were multiplied by 10 to give the value in milimetres. Another ambiguity was that some sheets had 'NFR' written above 'Condition', rather than underneath, and the meaning of this was unclear, so it was left out.

4.2 Results from morphometric analyses

4.2.1 Measurements from the record sheets

That there were 56 features on each record sheet, but most axes only had 19 of these recorded, and most of these axes were complete, suggests potential data collection bias. This is further suggested by the fact that the first few axes with the earliest collection dates, from 10001 to 10033, had large record sheets and were completely filled, but sheets recorded after those were not. Perhaps these record sheets were created before Pitts knew the sample size, with the shorter versions created after. Nonetheless, even these shorter versions were not consistently completed. Although it is most likely not possible that these features did not exist on the axes, this considerable reduction of variables did simplify the analysis, perhaps for better or worse. The 13 most recorded variables were not particularly the best for describing shape or defining groups. The only poll variables which made the cut were poledwid and lonpoled, which refer to very specific, tiny parts of the poll, and not the overall size of that end of an axe. It is interesting and perhaps peculiar that useful measurements to

discern size, such as maxleng, were measured for a surprisingly small amount of axes, and that small measurements such as poledwid were among the most frequent. According to Pitts, every complete axe in each collection he visited had their measurements taken (Pitts, 1996, p.328). Cast represents replica axes which were created from casts taken of the originals, so here weight cannot be recorded as the material the casts were made from will be different to the original axe. Reflaked and reworked are taken to mean that an axe had been repurposed later in its life cycle, and the shape had been changed, so the design may not be the same as when it was first manufactured. The distinction between reflaked and reworked is that reflaked is taken to be specific to flint implements. There were also fragments, which are either broken or incomplete axes. Finally, roughout axes are those which have complete shapes, but are not polished. It was based on this information that the decision was made to only keep complete and roughout axes, since these would retain the shape. Interestingly, out of Pitts' larger sample of 1919 axes – 124 of these were excluded from this investigation due to problems matching up axe IDs – only 1638 were used in his analysis. This is said to be because those with unique shapes and those which were unground were removed (Pitts, 1996, p.330).

The most interesting distributions shown in the univariate analyses were the bimodal grinding variables, as the two groups arising here may indicate differing manufacuring practices. The three modes in bladang are not really significant as these are all acute angles, and the real difference is between these and the 180 degree angles, which are comparatively rare. Both the 9 and 19 variable PCAs had similar PC1 correlations with variables that related to overall size of axes such as weight and maxthick, and also with grinside. Interestingly though, PC2 on the 19-variable dataset showed considerable correlation with all of the poll features compared to its PC1, whereas the 9-variable data PC2 only strongly correlated with poledwid. From the results of the 19-variable PCA, it could be possible to predict the shape of an axe based upon where it is positioned on PC1 and PC2. Because the blade is not usually

smaller than the poll, the width of the poll can suggest how square an axe is, as very large polls will most likely be a similar size to the blade. Axes plotted at the bottom of PC2 are most likely to be straight rather than trapezoidal, since they have large poll measurements. Those which are further to the left of PC1 but far down PC2 are likely to be more square since they are shorter, and likewise far down PC2 but to the right of PC1 will be more rectangular since they are longer. Quickly examining the axe drawings of 10040 and 60049, which are toward the bottom of PC2 in the 19-variable PCA, reveals that these are indeed quite square. If an axe is plotted near the top of PC2, but to the far left of PC1, it could be more triangular, and to the far right perhaps teardrop shaped, again due to the length of the axe. PC3 and 4 of the 9 variables are also worth mentioning as PC3 indicates an interesting negative correlation between the amount of grinding on the poll edge and poll edge width, which suggests that more grinding makes the poll width smaller. PC4 also shows grinside negatively correlating with lonpoled and bladang, which makes sense, as this suggests more grinding on the left side edge will reduce the length of the left side edge and the blade angle.

Whilst both the WSS and silhouette plots suggested that 2 was the optimum number of clusters for this dataset, this was shown to be an oversimplification, as extreme cases on the far ends were lumped in with the more typical axes around the origin. As the axes on both clusters' extremeties were the same as those on the edges of PC1, the axes were clearly just arranged from largest on the right to smallest on the left, and split at some arbitrary point into two. The three-cluster solution was a slight improvement on this as some of the more typical axes were moved from cluster 1, however not completely separating the very largest axes from the more typical ones. Membership of cluster 2 appears to be based on having large bladang, gripoled and lonpoled measurements, in addition to mostly having middle-of-the-road values for other variables. Because the 5-cluster solution showed little change in cluster 1 and created overlapping clusters it helped to confirm that any more than 3 clus-

ters would not make sense. Still, however, the 3-cluster solution left many of the typical axes intermixed with extreme cases in each cluster, particularly for cluster 1. The density plots of the variable distributions within each cluster confirmed that cluster 1 had the axes which are the heaviest, thickest, longest and most ground. These also reaffirm that cluster 2 membership appears to be based on having smaller measurements for PC1-correlated variables than cluster 1, but larger than cluster 3, but also having the highest for PC2-correlated variables. Additionally, the cluster densities show that the smallest axes were also ground the least. Because cluster 2 is plotted further down PC2 than the other two clusters, but some of its individuals are on the same areas of PC1 as the others, it could be suggested that variables correlated with PC2 are defining this cluster. This in conjunction with the 19-variable PCA results potentially hints at what could be possible if more variables were involved, as perhaps the aforementioned square and triangle shapes would emerge in clusters. Even with the reduced variables, PCA has picked out the poll measurements for component 2, indicating that there is perhaps a pattern happening with these measurements.

4.2.2 Axe outlines

PC1 appears to show small to large aspect ratio of the axe shapes, and PC2 seems to explain the ratio of blade width to poll width, starting from straight oblong shapes at one end to triangular shapes on the other. Wilczek et al.'s study on Bronze Age axes which also uses Momocs had similar results from their first two components, with the first representing aspect ratio and the second showing the blade to body width ratio (Wilczek et al., 2015, p.385). On the morphospace, the majority of axes were arranged around four mean axe shapes; a narrower and a wider trapezoidal shape, and a narrower and a wider oblong shape, just to the left of and at the origin. It is also surprising that two axes with smaller blade ends than poll ends are present, and the fact that there are so few brings into question whether they really are complete, or perhaps whether it is an artefact of the silhouette produced from the drawing.

It would be fair to say that the PCA results indicate the presence of at least three distinct shapes in the dataset: straight and narrow oblongs, trapezoidal shapes, and squat square shapes. These squat shapes could be axes which have similar poll widths to blade widths, which were spotted at the bottom of PC2 in the 19-variable record sheet data PCA.

Now that the existence of some more uncommon shapes was clear, it would be fair to hope for a satisfactory k-means result to separate these into their own groups, rather than diluting them with more of the typical axe types partitioned into the same group. The latter seems to be the result for the 2-cluster k-means analysis, which sees an already-existing cluster being split into two groups along with outliers at either end. The 3-cluster solution was more satisfactory in that many of these typical axes were contained in the central cluster, with a narrower and a wider group at either end, but these still contain some of the typical axes. The best option was 4 clusters, as many of the more average axes were partitioned between the wider and narrower typical axe shapes in the centre, which appear much like the mean shapes in PCA where most of the axes fell. There were also less axes in the widest and narrowest groups of the 4-cluster arrangement which makes sense as it did not appear that there were many of these extreme cases in the PCA.

4.3 Comparisons

The first two components in the PCA analysis on the elliptical-Fourier-transformed outlines appear remarkably similar to the first two components in the PCA of the axe measurement data with 19 variables.. Additionally, PC2 in the 19-variable reduced dataset relates to all of the poll features, whilst `Momocs` PC2 relates to ratio of blade to poll width. Meanwhile, PC1 is characterised by thickness, weight, length and grinding, and PC1 in Momocs by width, or aspect ratio. Rather square shapes, as predicted from the 19-variable PCA of the axe measurements, are indeed present in the outlines dataset. However, the mean shapes of the outlines from both the 9-variable and 19-variable clusters

of the MicroPasts data show little or no distinction from one another, whilst Momocs was able to pick out differing shapes in the outline dataset. This is not too surprising since the results of the k-means on the MicroPasts data showed typical axes lumped in with extreme cases in all three clusters, which most likely resulted in diluted mean shapes. In addition, not only were there quite a lot less outlines than there were axes in the measurement analysis, but there were only 9 variables to analyse the measurements with. Further, these 9 variables were not especially descriptive of shape, as useful variables such as those relating to width and maximum length were recorded for 407 axes or less. Another unsurprising outcome is that measurement group 1, the largest axes, are the least present in the outlines dataset; this is most likely a result of the failed silhouettes mostly being for the largest axes. The most narrow axes tend to also be the smallest axes, whilst the other shape groups tend to be split evenly between the middle and largest size groups. Overall, shape and size do not appear to necessarily dictate one another here, however this could be down to the significant amount of missing data from both sides.

Looking at the distributions of where these axes were found across Britain, the most concentrated areas could potentially be related to ancient mine locations in south-west England, Norfolk, Wales and Cumbria (Schauer, Andrew Bevan, et al., n.d., p.151). Four mines were located in southernmost England, with one in Norfolk, one in Cumbria and two in Wales. It is particularly interesting that there is a small concentration of the largest axes around the Cumbria area. Of course, that the find places tend to accumulate in certain areas mostly in southern England could also be a consequence of Mike Pitts being able to travel to these areas more easily. Similar axes do not tend to cluster near each other geospatially, and generally speaking it appears as though all shapes and sizes were intermingled across Britain. This could further illustrate the highly interwoven trade network of stone tools during this period (Schauer, Andrew Bevan, et al., n.d., p.20).

Conclusion

Successfully extracting and analysing the data from 1825 record sheets and 1137 technical drawings has not only provided further insight into Neolithic production and trade, but also highlighted the potential contribution to research that digitising legacy data can make. Automating the production of binary masks for use with morphometric packages has demonstrated that this too can be made reproducible, in addition to the analysis itself, which removes the subjective aspect of placing landmarks. In addition, with some improvements, there is the possibility that this method could be used on other analog datasets of drawings or photographs, whether black and white or in colour. The outline analysis of the processed drawings revealed that there are reasonably distinct groups of shapes within this sample of Neolithic axes from Britain, and this could give useful insight for future research. Taken alongside their stone groups, these broad shape categories could be used to determine whether similar shapes originate from certain industrial areas. Crowdsourcing transcriptions of the record sheets called attention to the difficulties which can arise if initial data collection is not carried out with potential future users or analyses in mind. If archaeology is pursuing a philosophy of open data, then rigorous and logical data collection strategies must be made mainstream. It should however be stressed that for the time, Mike Pitts was doing something different, and also catalogued all of this data voluntarily in his free hours. This is no mean feat, and the axe drawings have proved especially insightful here. Further to this, trialing PCA with a smaller, non-imputed version of the record sheet dataset with no NAs did suggest the potential patterns

which could emerge had the entire dataset been complete. The vast amount of missing measurements from the wider dataset most likely contributed to the inconclusive results. Comparing the two analyses showed some correlation between narrow shapes and small measurements, however was largely inconclusive due to missing data. The lack of similarity between axes found near to one another further supports recent research which suggests that axes often traveled far from their sources, and probably changed hands more than once.

The reproducible workflow of this project has utilised tools such as GNU Make and Docker which are typically used for building software, in order to not only make the analyses reproducible, but also the data extraction and cleaning, and to facilitate the use of large swathes of packages without compromising reproducibility. The author hopes to see similar reproducible strategies being utilized within archaeology in the coming decades. This is especially pertinent given that most articles consulted in the methodological literature review, bar the few which were solely about reproducibility, were not only not reproducible but also difficult to establish a workflow for. This became especially evident when attempting to find example workflows for `Momocs`, as many methodologies included manual steps or simply did not explain certain methods. It thus becomes difficult to learn from others who have published.

A possible future direction for the automated mask processing method would be to improve it so that common issues are mitigated against, for example by increasing the size of the bounding box used to crop the images. Since `Momocs` was so useful in revealing shapes within the axe drawing dataset, this reproducible methodology could be used to extract data for other artefact images, and it would be good to develop this method for transferability. This is especially useful in light of the recent newly released version 2 of `Momocs`, now part of `MomX`, which may result in the uptake of more morphometric analyses in the coming years, hopefully in archaeology. In addition to this, because the volunteers on MicroPasts mostly adhered to the instructions and transcribed that was written in the record sheets, this dataset could be used for

HTR training labels and data in future, and this is partly why the instructions were written this way. The transcribed record sheet data is now publicly available for download in JSON files from the MicroPasts website, however the author plans to make available the enclosed cleaned CSV version in the coming months.

Appendix A

List of appendix files

`10078-form.jpg`

`fill-axe-outlines.ipynb`

`hough-transform-copy1.ipynb`

`hough-transform.ipynb`

`opencv-cropping.ipynb`

Appendix B

List of abbreviations

Measurement	Description (unit)
Poledwid	Width of poll edge (g)
Lefedwid	Max width of left side edge (mm)
Ritedwid	Max width of right side edge (mm)
Blaedwid	Width of blade edge (mm)
Bladang	Angle at blade edge (degrees)
Grifropl	Area of grinding on poll, front face (cm^2)
Grifroby	Area of grinding on body, front face (cm^2)
Grifrobl	Area of grinding on blade, front face (cm^2)
Gribapl	Area of grinding on poll, back face (cm^2)
Gribaby	Area of grinding on body, back face (cm^2)
Gribabl	Area of grinding on blade, back face (cm^2)
Blacon	Concavity of blade (mm)
Blasin	Blade sinuosity (mm)
Edcon	Right edge concavity (mm)
Edsin	Right edge sinuosity (mm)
Maxthick	Maximum thickness (mm)
Polthick	Thickness of poll (mm)
Blathick	Thickness of blade (mm)
Lefthick	Thickness left of centre, opposite point of max thickness (mm)
Rithick	Same as LEFTHICK, but right of centre (mm)

Measurement	Description (unit)
Topthick	Thickness of top half (mm)
Bothick	Thickness of bottom half (mm)
Maxleng	Maxmimum length (mm)
Poleng	Poll length (mm)
Blaleng	Blade length (mm)
Maxwid	Maximum width (mm)
Polwid	Poll width (mm)
Bladwid	Blade width (mm)
Lengmaxt	Length from blade to point of max thickness (mm)
Lengmaxw	Length from blade to point of max width (mm)
Gripoled	Length of poll edge ground (mm)
Lonpoled	Total length of poll edge (mm)
Grinside	Extent of grinding on left side edge (mm)
Longside	Length of left side edge (mm)
Areapoll	Area of poll (cm^2)
Areabody	Area of body (cm^2)
Areablal	Area of left segment of blade (cm^2)
Areablar	Area of right segment of blade (cm^2)
Areablad	Total area of blade (cm^2)

Variables beginning with POL, REF and COR are the same as GRI measurements but instead of grinding replace with polish, reworking and cortex respectively, and also in cm^2

References

Baxter, M.J. (2015). *Exploratory multivariate analysis in archaeology.* Clinton Corners, New York: Percheron Press, a division of Eliot Werner Publications, Inc.

Berry, K. (2020). 'The tex live guide—2020'. Available [Online] at: `https://www.tug.org/texlive/doc/texlive-en/texlive-en.pdf`. [Last accessed 30 November 2020].

Bonacchi, C., Bevan, A., Pett, D., Keinan-Schoonbaert, A., Sparks, R., Wexler, J. and Wilkin, N. (2014). 'Crowd-sourced Archaeological Research: The MicroPasts Project'. *Archaeology International*, 17, pp.61–68.

Bonhomme, V., Picq, S., Gaucherel, C. and Claude, J. (2014). 'Momocs: Outline Analysis Using R'. *Journal of Statistical Software*, 56 (13).

Bourcy, S.M. (2018). 'The shape of diversity: A morphometric analysis of late archaic bifaces from lamoka lake'.

Bradski, G. (2000). 'The OpenCV Library'. *Dr. Dobb's Journal of Software Tools.*

Brande, S. and Saragusti, I. (1996). 'A Morphometric Model and Landmark Analysis of Acheulian Hand Axes from Northern Israel'. In L. F. Marcus and N. A. T. Organization (Eds.), *Advances in morphometrics.* NATO ASI series. New York: Plenum Press, pp.423–435.

Carl Boettiger, D.E. and Ross, N. (2020). 'Rocker/geospatial'. Available [On-

line] at: `https://hub.docker.com/r/rocker/geospatial`. [Last accessed 30 November 2020].

Clarke, A. (2011). 'Does size matter? Stone axes from Orkney: Their style and deposition'. In V. Davis and M. R. Edmonds (Eds.), *Stone axe studies III*. Oxford ; Oakville, CT: Oxbow Books, pp.309–322.

Claude, J. (2008). *Morphometrics with R*. New York, NY: Springer.

Clough, T.H. (1988). 'Introduction to the regional reports: Prehistoric stone implements from the British Isles'. In T. H. Clough (Ed.), *Stone axe studies. 2: The petrology of prehistoric stone implements from the British Isles*. CBA research report. London:

Darvill, T. (2011). 'Misty mountain hop: Prehistoric stone working in south-west Wales'. In V. Davis and M. R. Edmonds (Eds.), *Stone axe studies III*. Oxford ; Oakville, CT: Oxbow Books, pp.131–146.

Doran, J. and Hodson, F.R. (1975). *Mathematics and computers in archaeology*. Edinburgh: Edinburgh University Press.

Dray, S. and Josse, J. (2015). 'Principal component analysis with missing values: A comparative survey of methods'. *Plant Ecology*, 216 (5), pp.657–667.

Edinborough, K., Shennan, S., Teather, A., Baczkowski, J., Bevan, A., Bradley, R., Cook, G., Kerig, T., Parker Pearson, M., Pope, A. and Schauer, P. (2020). 'New Radiocarbon Dates Show Early Neolithic Date of Flint-Mining and Stone Quarrying in Britain'. *Radiocarbon*, 62 (1), pp.75–105.

Edmonds, M. (1989). 'Reviews - T. H. McK. Clough & W. A. Cummins (ed.). Stone axe studies 2: The petrology of prehistoric stone implements from the British Isles. 279 pages, 42 *figures* , 240 *tables* , 3 *maps* . 1988. London: Council for British Archaeology [Research Report 67]; ISBN 0-906780-52-7 *paperback* £35.'. *Antiquity*, 63 (240), pp.633–634.

Emmitt, J., Littleton, J., Young, R. and Phillipps, R. (2019). 'Digitizing Roonka: The creation of a 3D representation from archival records'. *Digital Applications in Archaeology and Cultural Heritage*, 13, p.e00094.

Eric C. Kansa and Sarah Whitcher Kansa (2013). 'Open Archaeology'. *Journal of Eastern Mediterranean Archaeology & Heritage Studies*, 1 (1), p.88.

Faniel, I.M., Austin, A., Kansa, E., Kansa, S.W., France, P., Jacobs, J., Boytner, R. and Yakel, E. (2018). 'Beyond the Archive: Bridging Data Creation and Reuse in Archaeology'. *Advances in Archaeological Practice*, 6 (2), pp.105–116.

Forel, B., Gabillot, M., Monna, F., Forel, S., Dommergues, C.H., Gerber, S., Petit, C., Mordant, C. and Chateau, C. (2009). 'Morphometry of Middle Bronze Age palstaves by Discrete Cosine Transform'. *Journal of Archaeological Science*, 36 (3), pp.721–729.

García-Medrano, P., Maldonado-Garrido, E., Ashton, N. and Ollé, A. (2020). 'Objectifying processes: The use of geometric morphometrics and multivariate analyses on Acheulean tools'. *Journal of Lithic Studies*, 7 (1).

Giardina, C.R. and Kuhl, F.P. (1977). 'Accuracy of curve approximation by harmonically related vectors with elliptical loci'. *Computer Graphics and Image Processing*, 6 (3), pp.277–285.

Gordin, S., Gutherz, G., Elazary, A., Romach, A., Jiménez, E., Berant, J. and Cohen, Y. (2020). 'Reading Akkadian cuneiform using natural language processing' M. Lippi (Ed.), *PLOS ONE*, 15 (10), p.e0240511.

Grimes, W.F. (1935). 'Recent Books on British Archaeology'. *Antiquity*, 9 (36), pp.424–434.

Grimes, W.F. (1979). 'The history of implement petrology in Britain'. In T. H. M. Clough, W. A. Cummins, and C. for British Archaeology (Eds.), *Stone axe studies: Archaeological, petrological, experimental and ethno-*

graphic. CBA Research Report. London: Council for British Archaeology, pp.1–3.

Hodson, F.R., Kendall, D.G., Tăutu, P., Britain), R.S. (Great and România, A.R.S. eds. (1971). *Mathematics in the archaeological and historical sciences: Proceedings of the Anglo-Romanian Conference, Mamaia, 1970*. Edinburgh: Edinburgh University Press.

Iovita, R. (2010). 'Comparing Stone Tool Resharpening Trajectories with the Aid of Elliptical Fourier Analysis'. In S. J. Lycett and Parth R. Chauhan (Eds.), *New perspectives on old stones: Analytical approaches to Paleolithic technologies*. New York: Springer.

Iovita, R., Tuvi-Arad, I., Moncel, M.-H., Despriée, J., Voinchet, P. and Bahain, J.-J. (2017). 'High handaxe symmetry at the beginning of the European Acheulian: The data from la Noira (France) in context' N. Bicho (Ed.), *PLOS ONE*, 12 (5), p.e0177063.

Josse, J. and Husson, F. (2016). 'missMDA: A package for handling missing values in multivariate data analysis'. *Journal of Statistical Software*, 70 (1), pp.1–31.

Kahle, D. and Wickham, H. (2013). 'Ggmap: Spatial visualization with ggplot2'. *The R Journal*, 5 (1), pp.144–161.

Keiller, A. (1937). 'Petrological Analysis'. *Antiquity*, 11 (44), pp.484–485.

Keiller, A., Piggott, S. and Wallis, F.S. (1941). 'First Report of the Sub-Committee of the South-Western Group of Museums and Art Galleries on the Petrological Identification of Stone Axes'. *Proceedings of the Prehistoric Society*, 7, pp.50–72.

Kuhl, F.P. and Giardina, C.R. (1982). 'Elliptic Fourier features of a closed contour'. *Computer Graphics and Image Processing*, 18 (3), pp.236–258.

Küçükdemirci, M. and Sarris, A. (2020). 'Deep learning based automated

analysis of archaeo-geophysical images'. *Archaeological Prospection*, 27 (2), pp.107–118.

Lodwick, L. (2019). 'Sowing the Seeds of Future Research: Data Sharing, Citation and Reuse in Archaeobotany'. *Open Quaternary*, 5.

Marwick, B. (2017). 'Computational Reproducibility in Archaeological Research: Basic Principles and a Case Study of Their Implementation'. *Journal of Archaeological Method and Theory*, 24 (2), pp.424–450.

Marwick, B., Boettiger, C. and Mullen, L. (2018). 'Packaging Data Analytical Work Reproducibly Using R (and Friends)'. *The American Statistician*, 72 (1), pp.80–88.

Merkel, D. (2014). 'Docker: Lightweight linux containers for consistent development and deployment'. *Linux journal*, 2014 (239), p.2.

Monna, F., Jebrane, A., Gabillot, M., Laffont, R., Specht, M., Bohard, B., Camizuli, E., Petit, C., Chateau, C. and Alibert, P. (2013). 'Morphometry of Middle Bronze Age palstaves. Part II – spatial distribution of shapes in two typological groups, implications for production and exportation'. *Journal of Archaeological Science*, 40 (1), pp.507–516.

Palmqvist, P. (1996). 'Comparative Morphometric Study of a Human Phalanx from the Lower Pleistocene Site at Cueva Victoria (Murcia, Spain), by means of Fourier Analysis, Shape Coordinates of Landmarks, Principal and Relative Warps'. *Journal of Archaeological Science*, 23 (1), pp.95–107.

Patel, A.A. (2019). *Hands-on unsupervised learning using Python: How to build applied machine learning solutions from unlabeled data.* First edition. Sebastopol, CA: O'Reilly Media.

Pett, D. (2018). 'MicroPasts/bronzeAgeIndex'. Available [Online] at: `https://github.com/MicroPasts/bronzeAgeIndex`. [Last accessed 29 November 2020].

Pitts, M. (1996). 'The Stone Axe in Neolithic Britain'. *Proceedings of the Prehistoric Society*, 62, pp.311–371.

R Core Team (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available [Online] at: `https://www.R-project.org/`.

Rhys, H. (2019). *Machine Learning with R, the tidyverse, and mlr*.

Roe, D.A. (1969). 'British Lower and Middle Palaeolithic Handaxe Groups'. *Proceedings of the Prehistoric Society*, 34, pp.1–82.

Rohlf, F.J. and Archie, J.W. (1984). 'A Comparison of Fourier Methods for the Description of Wing Shape in Mosquitoes (Diptera: Culicidae)'. *Systematic Zoology*, 33 (3), p.302.

Schauer, P., Bevan, A., Shennan, S., Edinborough, K., Kerig, T. and Pearson, M.P. 'British Neolithic Axehead Distributions and Their Implications'. *Journal of Archaeological Method and Theory*.

Schauer, P., Shennan, S., Bevan, A., Cook, G., Edinborough, K., Fyfe, R., Kerig, T. and Parker Pearson, M. 'Supply and demand in prehistory? Economics of Neolithic mining in northwest Europe'. *Journal of Anthropological Archaeology*, 54, pp.149–160.

Scheidl, H. (2020). 'Githubharald/SimpleHTR'. Available [Online] at: `https://github.com/githubharald/SimpleHTR`. [Last accessed 29 November 2020].

Shennan, S. (1988). *Quantifying archaeology*. Edinburgh: Edinburgh University Press.

Shennan, S., Bevan, A., Edinborough, K., Kerig, T., Parker Pearson, M. and Schauer, P. (2017). 'Supply and Demand in Prehistory? Economics of Neolithic Mining in NW Europe (NEOMINE)'. *Archaeology International*.

Stallman, R., McGrath, R. and Smith, P.D. (2004). *GNU Make: A program*

for directing recompliation ; GNU make version 3.81. Boston, Mass: Free Software Foundation.

Thomas, D.H. (1972). 'The use and abuse of numerical taxonomy in archaeology'. *Archaeology and physical anthropology in Oceania*, 7 (1), pp.31–49.

Thomas, H.H. (1923). 'The Source of the Stones of Stonehenge'. *The Antiquaries Journal*, 3 (3), pp.239–260.

Tipping, R., Edmonds, M. and Sheridan, A. (1993). 'Palaeoenvironmental investigations directly associated with a neolithic axe 'quarry' on Beinn Lawers, near Killin, Perthshire, Scotland'. *New Phytologist*, 123 (3), pp.585–597.

Ushey, K. (2020). *Renv: Project environments.* Available [Online] at: `https://CRAN.R-project.org/package=renv`.

Van Rossum, G. and Drake, F.L. (2009). *Python 3 reference manual.* Scotts Valley, CA: CreateSpace.

Vitolo, Claudia, Fry, Matthew, Buytaert, Wouter, Spencer, Michael, Gauster and Tobias (2018). *Rnrfa: An r package to retrieve, filter and visualize data from the UK National River Flow Archive.*

Walker, K. (2015). *Axe-heads and identity: An investigation into the roles of imported axe-heads in identity formation in Neolithic Britain.* Oxford: ProQuest Dissertations Publishing.

Wallace, M., Bonhomme, V., Russell, J., Stillman, E., George, T.S., Ramsay, L., Wishart, J., Timpany, S., Bull, H., Booth, A. and Martin, P. (2019). 'Searching for the Origins of Bere Barley: A Geometric Morphometric Approach to Cereal Landrace Recognition in Archaeology'. *Journal of Archaeological Method and Theory*, 26 (3), pp.1125–1142.

Wickham, H. (2019). 'Welcome to the tidyverse'. *Journal of Open Source Software*, 4 (43), p.1686.

Wilczek, J., Monna, F., Gabillot, M., Navarro, N., Rusch, L. and Chateau, C. (2015). 'Unsupervised model-based clustering for typological classification of Middle Bronze Age flanged axes'. *Journal of Archaeological Science: Reports*, 3, pp.381–391.