Geometric Morphometrics (GMM) and Archaeological Science

Dr. Christian Hoggard

July 2020

Welcome to these workshops!

Instructor

Dr. Christian Steven Hoggard (Visiting Fellow, University of Southampton)

@cshoggard / @cshoggard / www.sites.google.com/site/christianhoggard

C.S.Hoggard@soton.ac.uk

Research Interests:

1. Quantitative Archaeology in the R Environment
2. The European Lower and Middle Palaeolithic
3. Cultural Evolution and Taxonomies
4. Open Science Approaches to Archaeology

Previous Workshops:

- September 2019: Sendai (Japan)
- December 2019: Cologne (Germany)
- May 2020: #StayHomeButStudy (Online)

Publications: Click here

Workshop Structure

There are **three workshops** in total...

- **Workshop One** (now): Introduction to GMM and Archaeology
- **Workshop Two** (20th July): Landmark-based approaches to GMM (R practical: Geomorph/Tidyverse)
- **Workshop Three** (27th July): Outline-based approaches to GMM (R practical: Momocs/Tidyverse)

About these workshops...

- No prior experience in R is necessary 😄
- All data and R Markdowns (documents and presentations) are on GitHub
- Additional software demonstrations: IDAV Landmark Editor (Checkpoint) and TPSdig2/TpsUtil
- Breaks throughout the workshops will facilitate additional questions and guidance! ✔

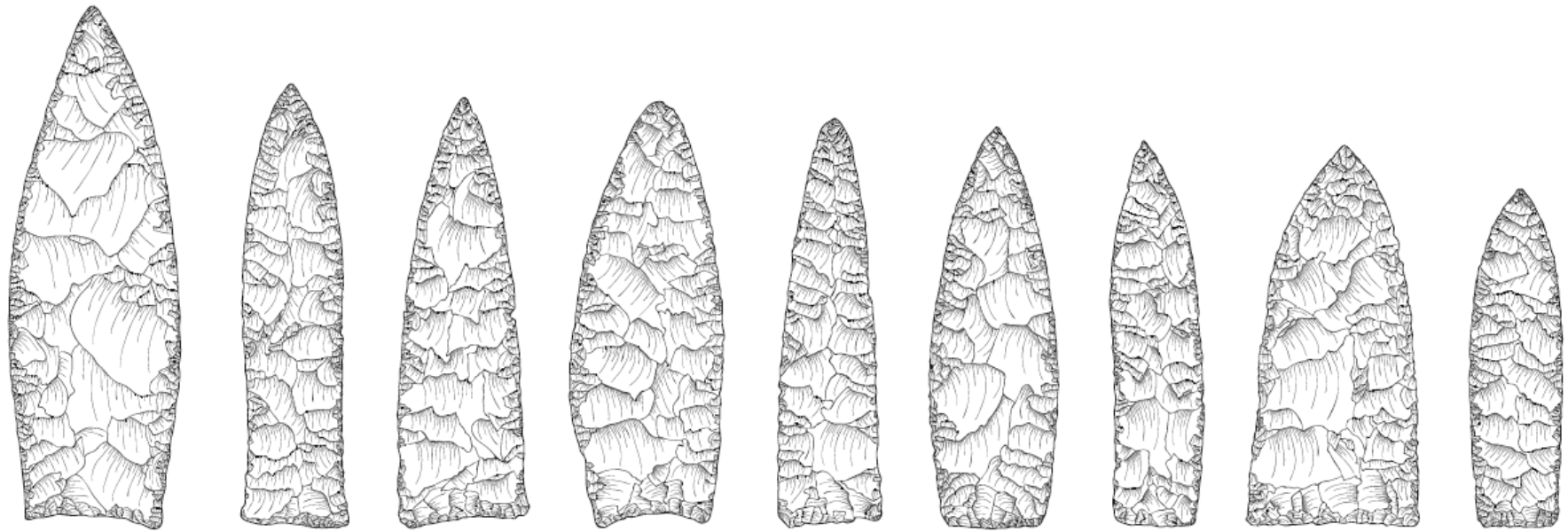
What are we going to do today?

I will introduce the subject of geometric morphometrics (GMM), specifically:

1. Shape and shape theory
2. Advantages (and disadvantages) of using GMM
3. History of GMM
4. How we 'do' GMM (data creation, transformation, analysis and visualisation)


$\frac{dN}{dt} = \frac{1}{q V_{act}} - g_0 (N - N_0)(1 - \varepsilon S) S + \frac{N_e}{\tau_n} - \frac{N}{\tau_p}$
$\frac{dS}{dt} = \Gamma_b g_0 (N - N_0)(1 - \varepsilon S) S + \frac{\beta_{sp} N}{\tau_n} - \frac{S}{\tau_p}$
$\frac{S}{P_t} = \frac{\Gamma \varepsilon p \lambda^2}{V_{act} \eta_c} = \Theta$
$S < \frac{1}{\varepsilon}$

GMM & Archaeological Science ✨



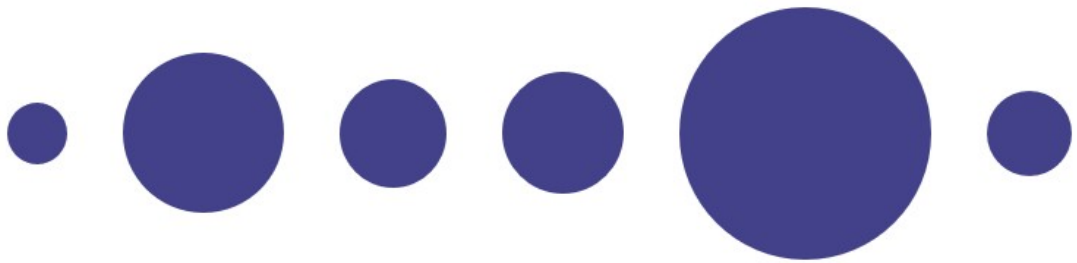
Selden, R.Z. et al. (2018). Lithic morphological organisation: Gahagan bifaces from the Southern Caddo Area. *Digitial Applications in Archaeology and Cultural Heritage* 10: e00080.

Mounier, A. et al. (2018). Who were the Nataruk people? Mandibular morphology among late Pleistocene and early Holocene fisher-forager populations of West Turkana (Kenya). *Journal of Human Evolution* 121: 235-253.

But what is shape?

Shape as...

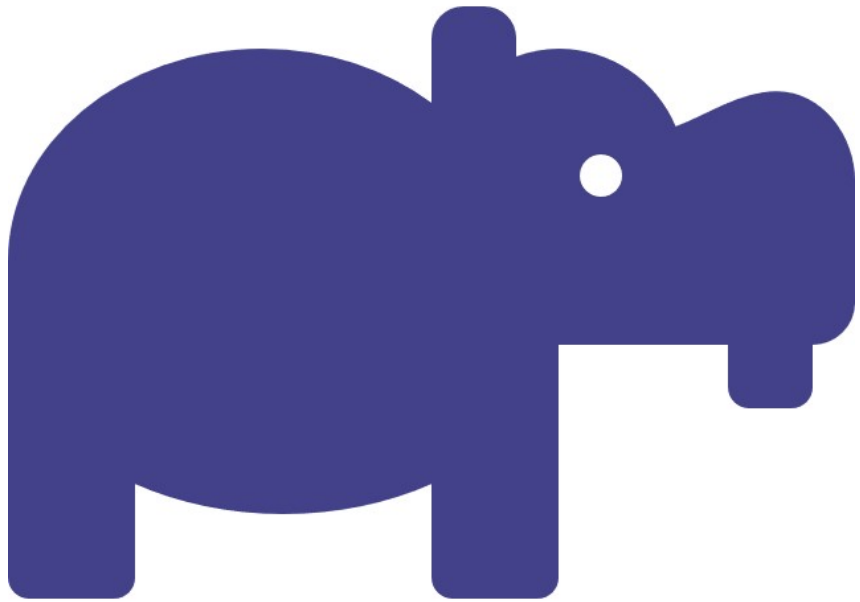
“In general terms, the shape of an object, dataset or image that can be defined as the total of all information that is invariant under **translation**, **rotation**, and **isotropic rescalings**” (Small, 1996: 6)

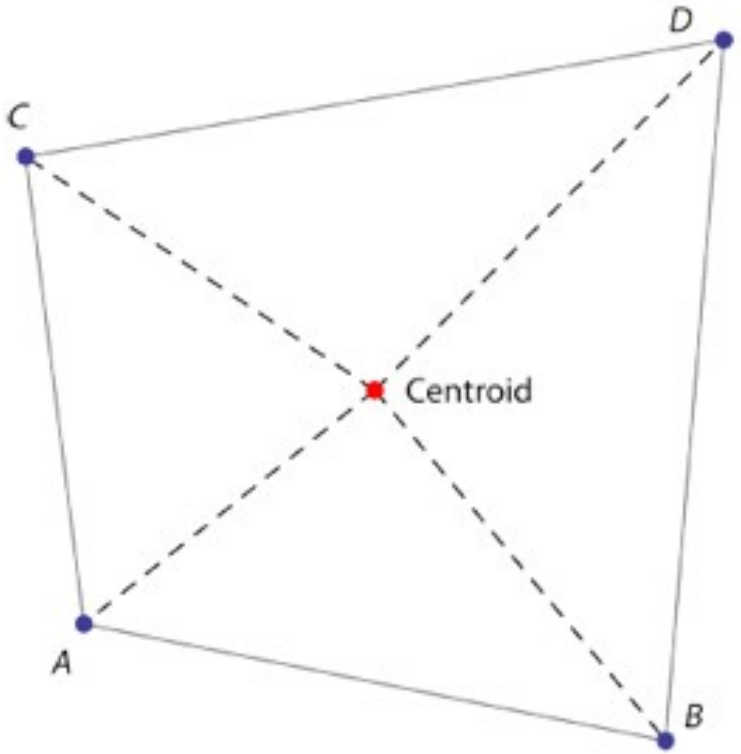


Small, C. (1996). *The statistical theory of shape*. New York: Springer.

But what about size?

Do we use the length? the volume of the object? the mass of an object?





Mitteroecker et al. (2013). A brief review of shape, form, and allometry in geometric morphometrics, with applications to human facial morphology. *Hysterix, the Italian Journal of Mammalogy*: 59-66.

Centroid size: square root of the summed squared lengths of the dashed lines

Shape + Size = Form

Morphometrics 101

- Term: Professor of Zoology Robert Blackith (1957)
- Quantitative study of shape, shape variation and shape covariation
- Two types of morphometric studies:
  - Traditional morphometrics (length measurements, angles, ratios...)
  - Geometric morphometrics or GMM (landmarks, outlines, curves, surfaces...)

Benefits and Drawbacks

👍 Benefits 👍

- Degree of **shape resolution**
- **Information loss** (vs. linear measurements)
- **Data can easily be collected** from methods including photographs, drawings and scans
- Abstraction and registration procedure allow an analysis of **shape *sensu stricto***
- Useful **visualisation** tool

👎 Drawbacks 👎

- Required **skill level** (software, statistics…)
- **Proceduralisation** (R is helping!)
- **Landmark subjectivity**

We can use GMM to determine whether...

two or more assemblages are different in shape?

shape is related to: size? sex? raw material? hominin?

shape differences correspond to a hypothesis or a model?

mean or median shapes are representative of a site?

there are biologically inherent relationships?

networks of specific shapes can be identified? (over several sites)

A short history of Geometric Morphometrics...

Albrecht Dürer (1471-1528)

- Renaissance Paint, printer, theorist and **founder of descriptive geometry**
- Worked on helices, conchoids, epicycloids and the Delian Problem (doubling the cube)
- **Grid transformations** to catalogue and investigate morphological variation

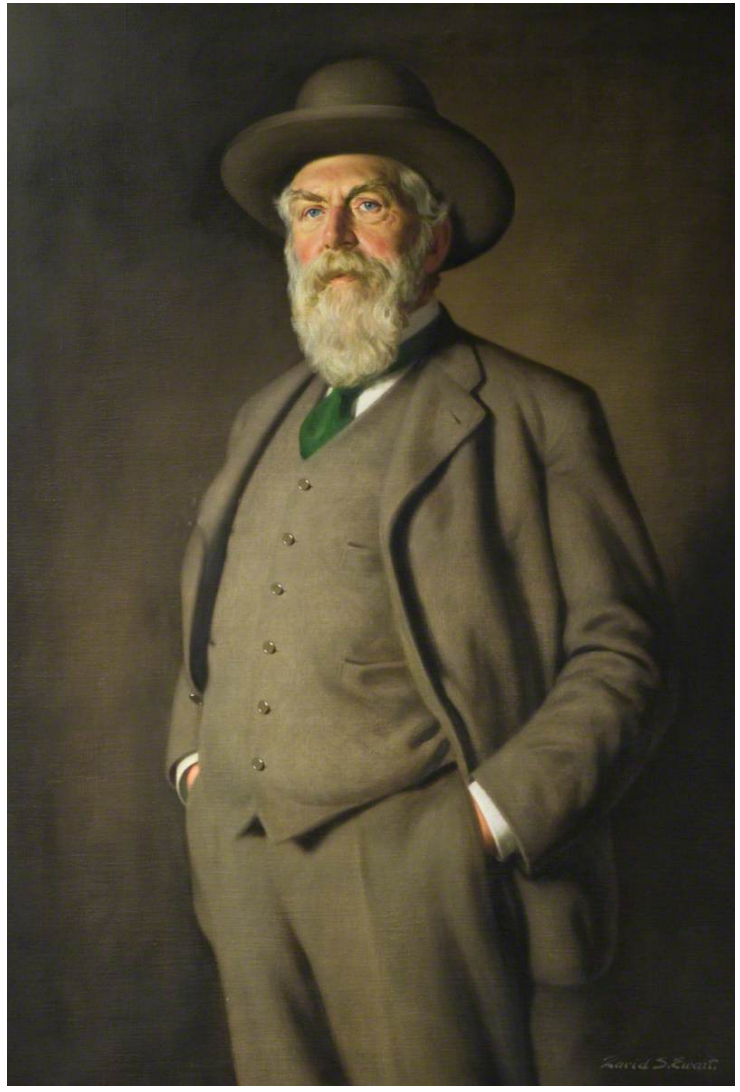
Draughtsman Making a Perspective Drawing of a Reclining Woman c. 1525 Albrecht Dürer (The Met Museum: Public Domain).

Sir D'Arcy Wentworth Thompson (1860-1948)

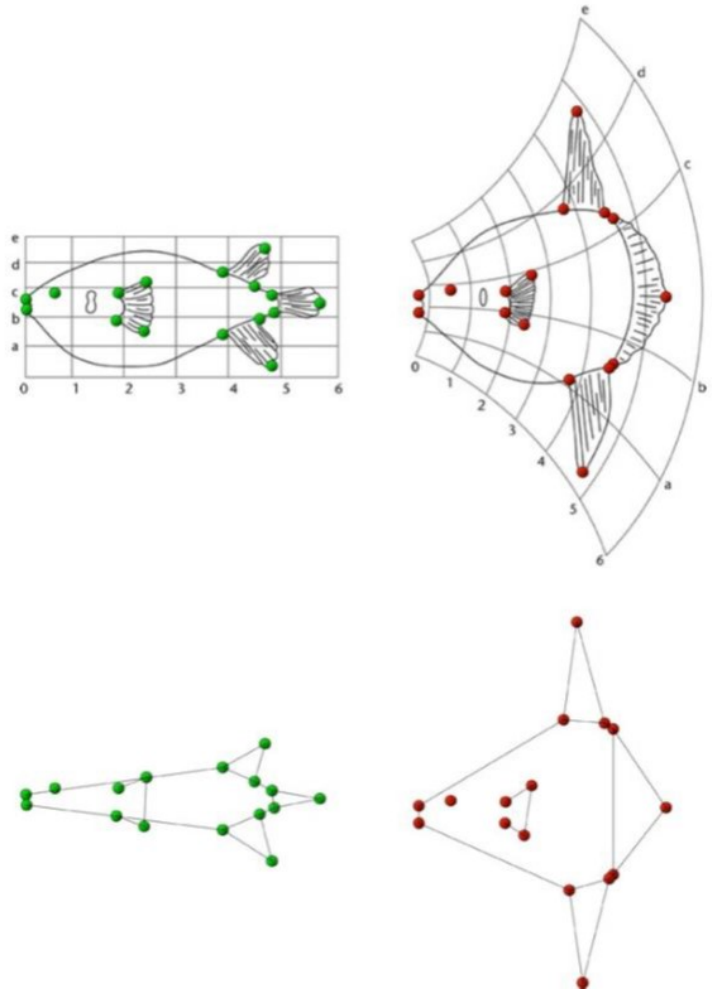
- Biologist, mathematician and classics scholar
- Famous for his quotes on the mathematical beauty of nature (inspiring Huxley, Turing, Lévi Strauss and van der Rohe)

On Growth and Form (1916)

- **Fundamental book on morphological variation in nature**
- Emphasis on **mathemtical structures** accounting for diversity
- Counter-argument to **vitalism** (while challenging **natural selection**)



Professor Sir D'Arcy Wentworth Thompson (after a 1938 original).
David Shanks Ewart (1901-1965). University of Dundee Fine Art Collections. CC BY-NC-ND.



Thompson, D.W. (1917). *On Growth and Form*. Cambridge University.

Transition to Multivariate Morphometrics

- **Fred Bookstein**: Cartesian transformations, Bookstein transformations, etc.
- **David George Kendall (1918-2007)**: Objects of the same shape as seperate points in a geometric space
- **Miriam Zelditch**: Complex shapes, morphological evolution and ecology
- **Ian Dryden and Kanti Mardia**: Development of Fourier-based outline analysis
- **Norman MacLeod**: Eigenshape analysis and palaeontological applications
- **James Rohlf**: Statistical developments, biological applications and software

Another Ten Minute Break! 

10:00

Any Questions?

Slack Workspace ("workshop_1") and Zoom!


BEGIN.

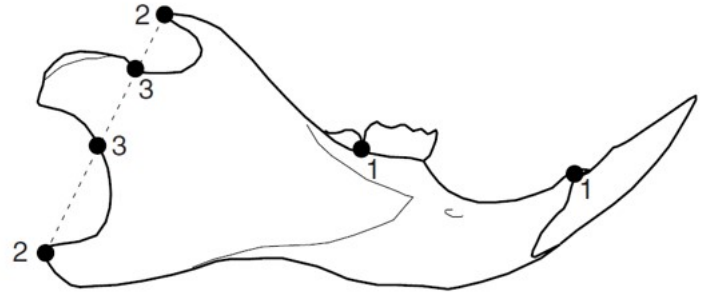
Stage 1: Collecting Input Data

Methods include:

- **CT Scanning**
- **Photogrammetry** and **Structure-from-Motion**
- **Structured-light** Scanning
- **Mono IR** (e.g. Kinect) and **Stereo IR** (Intel RealSense) Scanning
- **MicroScribe** digitisers
- **3D scanners** e.g. NextEngine Desktop
- **Photographs**
- **Professional Illustrations**

Stage 2: Landmark Choice

- Landmarks are central to GMM
- Definition: a coordinate point used to represent a shape and/or a homologous point on a structure
- Quantifiable as Cartesian coordinates (x, y / z coordinates)
- Flexibility in what types of landmark are required
- Can be treated as individual points [workshop two] or converted into curves and outlines [workshop three]



Claude, J. (2008). *Morphometrics with R*. Springer Publishing.

Stage 2: Landmark Choice

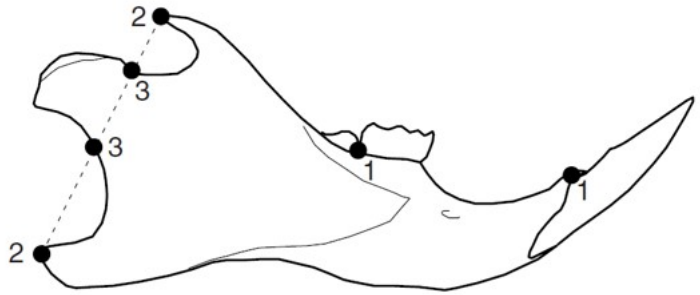
Various types of landmarks

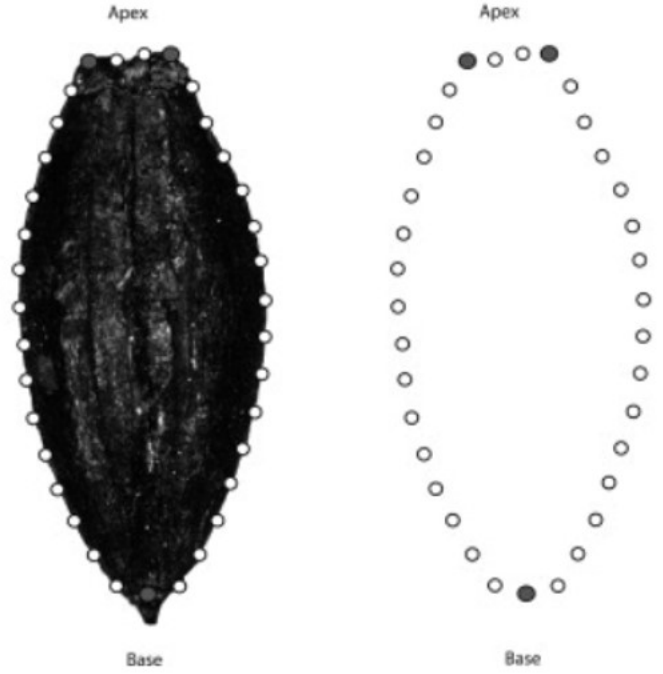
- Type I: Homologous biological structures
- Type II: Geometric definition e.g. greatest curvature
- Type III: Point with reference to another point

A special case: semi-landmarks

- Placed using an algorithm
- Can cover the whole or part of a shape
- Equidistant and placed between one or two end points
- A special Type III landmark (*sensu* Bookstein)

Ros et al. (2013). Geometric morphometric analysis of grain shape and the identification of two-rowed barley (Hordeum vulgare subsp. distichum L.) in southern France. *Journal of Archaeological Science*. 41.



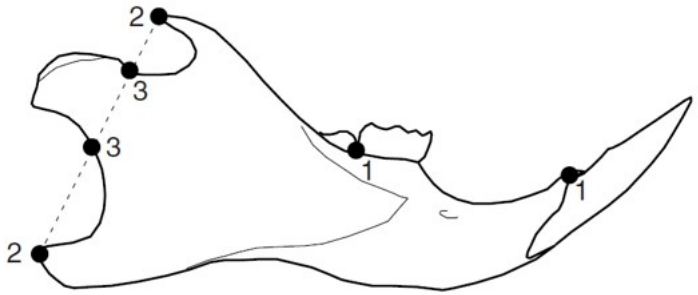


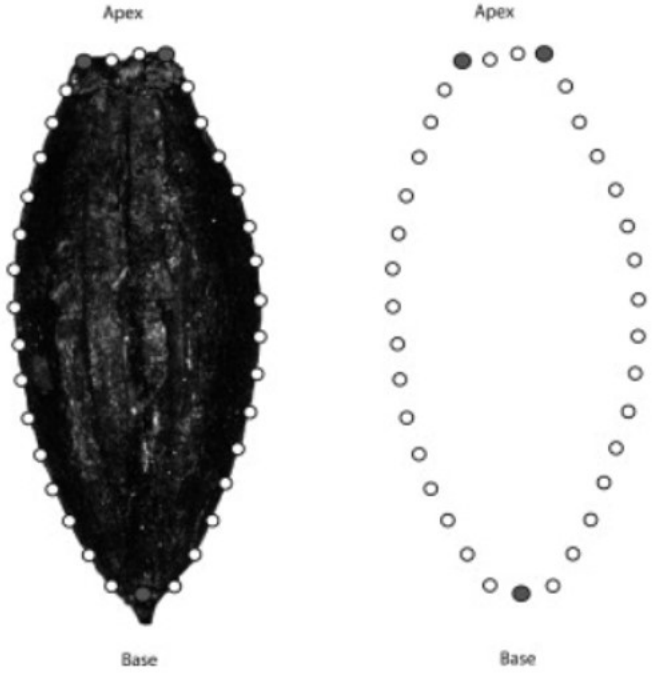
Stage 2: Landmark Choice

Landmarks should...

- sample aspects which are of archaeological interest
- be repeatable and identifiable on all examples (if possible)
- cover as much of the shape as possible
- be sufficient as to not increase the ‘weighting of areas’
- always be plotted in the same order!

Ros et al. (2013). Geometric morphometric analysis of grain shape and the identification of two-rowed barley (Hordeum vulgare subsp. distichum L.) in southern France. *Journal of Archaeological Science*. 41.



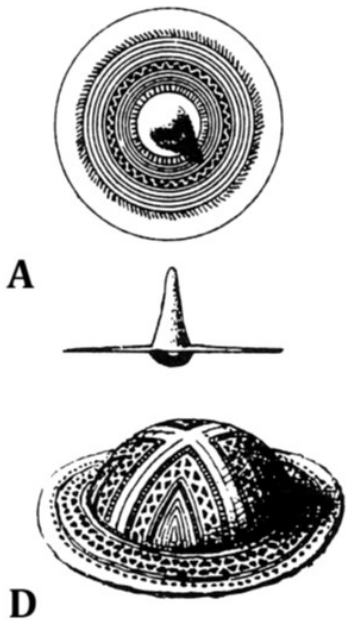


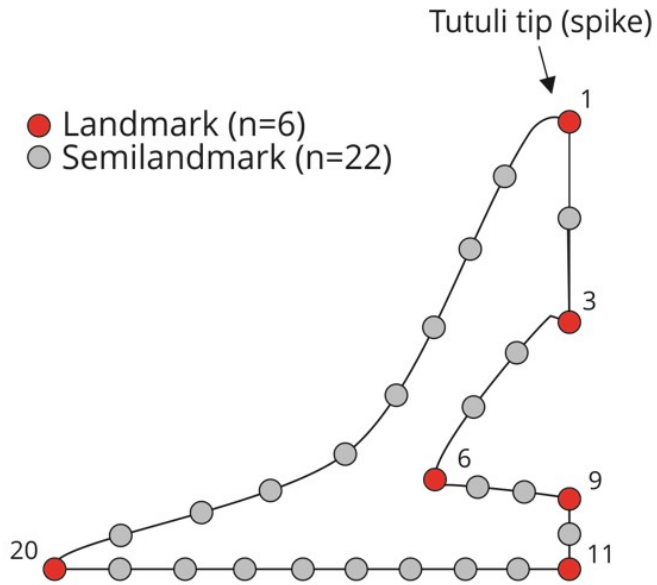
Some observations...

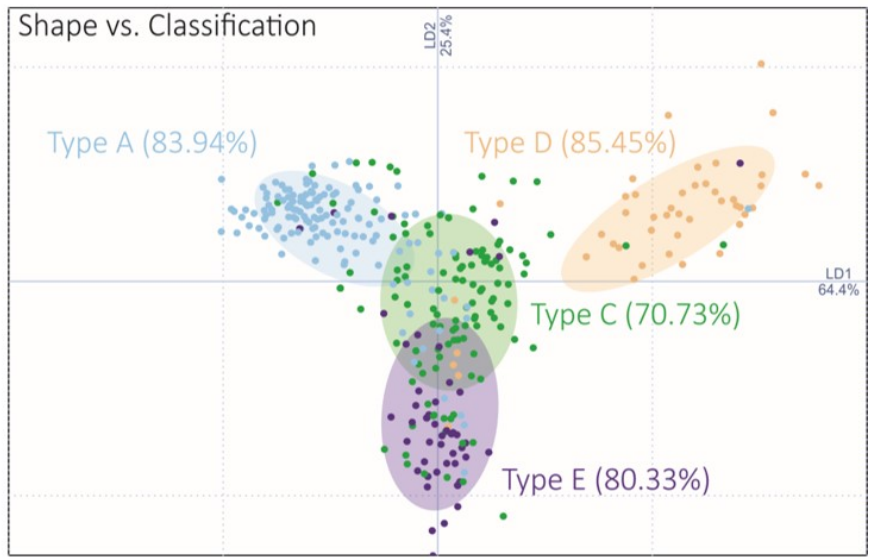
For archaeological material...	For bioarchaeological material...
Poor morphological correspondence	Greater morphological correspondence
Less intuitive to examine in 3D	More intuitive to examine in 3D...
Fewer number of examples to compare with	Greater number of examples to compare with
Fragmentation: major issue	Fragmentation: major issue
Higher sample size	Lower sample size is necessary (esp. prehistory)
Often outline-centric	Often landmark-centric

A Novel Geometric Morphometric (GMM) Application to the Study of Bronze Age Tutuli

Christina Vestergaard[1,3] and Christian Steven Hoggard[1,2]







Stage 3: Landmark digitisation

A variety of software is available, including:

- TPS Suite (**TpsUtil** and **TpsDig2**)
- CRAN-certified R Packages e.g. **geomorph**, **Momocs**, **StereoMorph** and **shape**
- GitHub-based R Packages e.g. **GUIMorph**
- **SlicerMorph** and **3DSlicer**
- **PhyloNimbus**
- **Stratovan Checkpoint** (£)
- **EVAN Toolkit** (£)
- and many, many more...

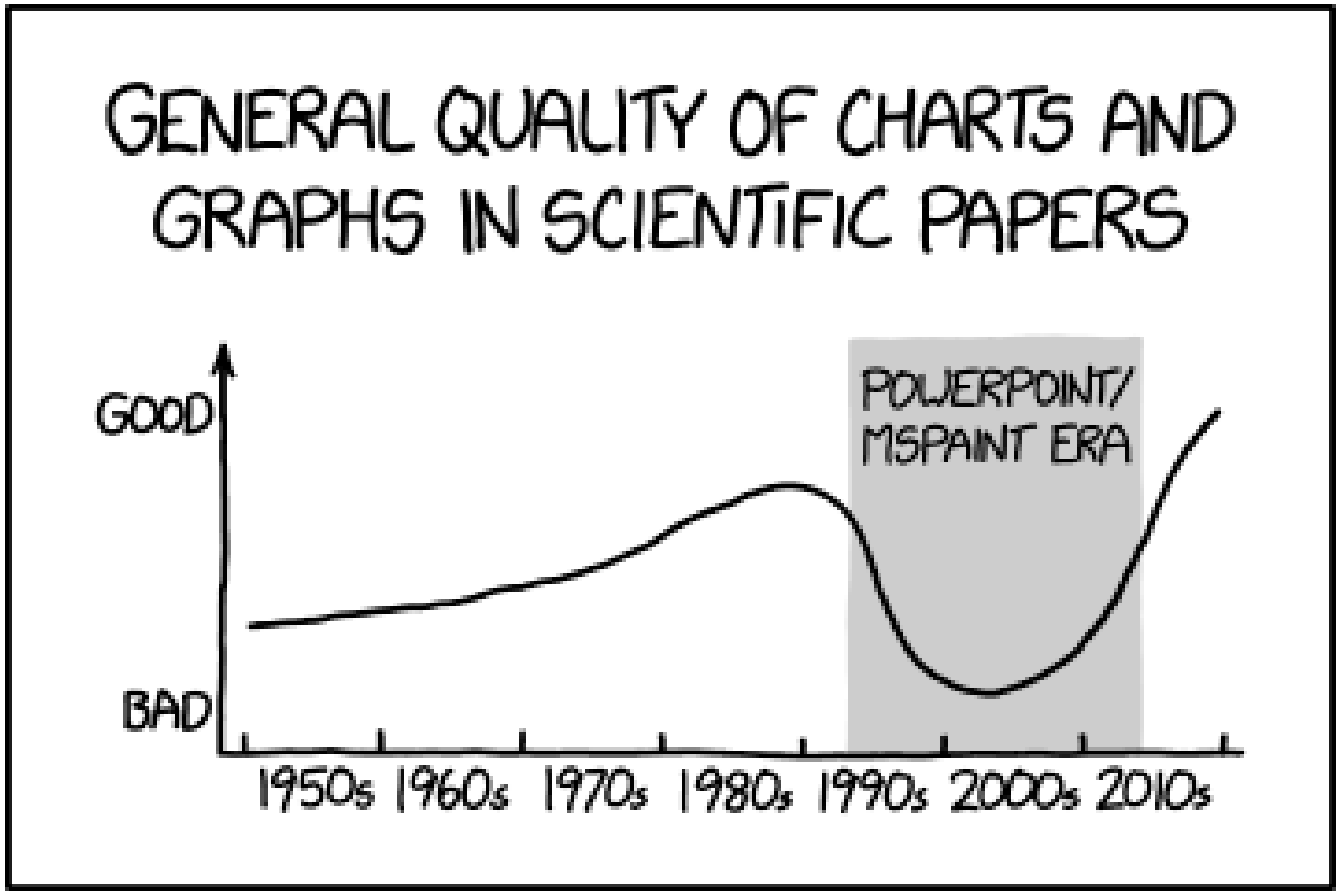
Note: IDAV Landmark Editor (no longer available)

Output: various file types including **.tps**, **.nts**, **.csv** and **.txt** formats

GMM (and Archaeology) in the R Environment

- **Increasing number of packages** for shape (and archaeological) data
- **Powerful** and **fairly straight-forward** to use
- Last decade: **Archaeology as a more code-literate discipline**
- **Advantages in data visualisation**
- Product: **reproducible** and **replicable Open Science** approach

https://media.giphy.com/media/l0K4hqqqwgFijgVLa/giphy.gif



https://xkcd.com/1945/

Stage 4: Data screening

- Do all my specimens have the **correct number of points**?
- Are all my landmarks in the **correct order**?
- Are the **ID labels correct**?
- Are they to **scale**? (for size-integrated analyses)
- Are **sliders** defined? (if using sliding semi-landmarks)

Importance of a code-book

- Necessary to **'know'** your meta-data (e.g. #ID, technology, sex or raw material)
- Ensure all your meta-data is **clear**, **easy to understand** and **formatted appropriately**

Know your research question!

Ten Minute Break! 

10:00

Any Questions?

Slack Workspace ("workshop_1") and Zoom!


BEGIN.

Stage 5: Data Registration (isolating shape)

A variety of analyses are available in GMM:

1. **Landmark Analysis** (using landmarks)
2. **Outline Analysis** (using semi-landmarks)
    - **Elliptic Fourier Analysis** (semi-landmarks of a closed outline)
    - **Radii Fourier Analysis** (semi-landmarks of a closed outline)
    - **Discrete Cosine Transform** (semi-landmarks of an open outline)
3. **Miscellaneous**
    - **Eigenshape Analysis** (semi-landmarks of a closed outline)

> The registration method changes depending on if you're analysing landmarks or outlines!
>
> Landmarks: **Generalised Procrustes Analysis** | Outlines: **Fourier-based transformation**

Generalised Procrustes Analysis (GPA)

(Procrustes Superimposition / Procrustes Fitting / Generalised Least Squares)

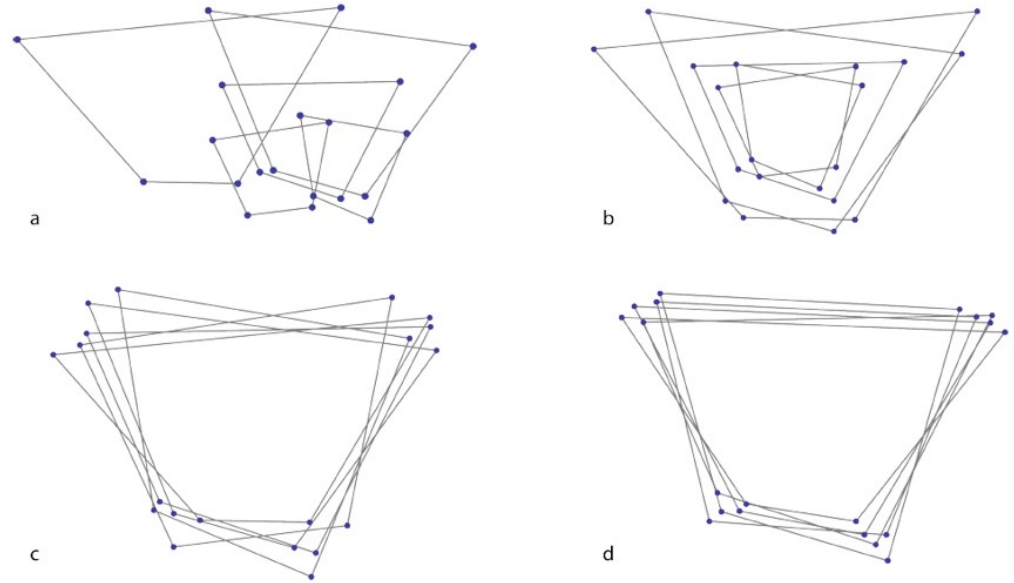
A procedure to isolate shape from a number of variables: rotation, size and translation.

https://media.giphy.com/media/11DFuwckOK9mdG/giphy.gif

Generalised Procrustes Analysis

The raw coordinates (for all shapes) are...

- Translated to a **common centroid** (A-B)
- Scaled to a **common centroid size** (B-C)
- Rotated to **minimise the sum of squared-distances between landmarks** (C-D)
- Outcome: **Procrustes coordinates** (shape)



Mitteroecker et al. (2013). A brief review of shape, form, and allometry in geometric morphometrics, with applications to human facial morphology. *Hysterix, the Italian Journal of Mammalogy*. pp. 59-66.

Why Procrustes?

- Damastes was a son of Poseidon
  who lived on a sacred way (Attica)
- There he had a bed, in which he invited every passer-by to spend the night. He would set to work on them with his blacksmith's hammer, to stretch them to "fit" the bed
- Nickname: Procrustes ("The Stretcher")
- If the guest proved too tall, Procrustes would amputate the excess length - nobody ever fitted the bed exactly! (**Procrustes Fitting**)
- Procrustes continued his reign of terror until he was "fitted" to his own bed by Theseus!
- Link to source: click here

Fourier-based transformation

In these methods, the **semi-landmarks are converted into curves**,
and the **coefficients which quantify these curves** are what we examine.
For example, in **Elliptic Fourier Analysis (EFA)** *sensu* Kuhl and Giardina (1982)...

$$A_n = \frac{T}{2n^2\pi^2}\sum_{p=1}^{k}\frac{\Delta x_p}{\Delta t_p}\left[\cos\left(\frac{2\pi n t_p}{T}\right)-\cos\left(\frac{2\pi n t_{p-1}}{T}\right)\right]$$

$$B_n = \frac{T}{2n^2\pi^2}\sum_{p=1}^{k}\frac{\Delta x_p}{\Delta t_p}\left[\sin\left(\frac{2\pi n t_p}{T}\right)-\sin\left(\frac{2\pi n t_{p-1}}{T}\right)\right]$$

$$C_n = \frac{T}{2n^2\pi^2}\sum_{p=1}^{k}\frac{\Delta y_p}{\Delta t_p}\left[\cos\left(\frac{2\pi n t_p}{T}\right)-\cos\left(\frac{2\pi n t_{p-1}}{T}\right)\right]$$

$$D_n = \frac{T}{2n^2\pi^2}\sum_{p=1}^{k}\frac{\Delta y_p}{\Delta t_p}\left[\sin\left(\frac{2\pi n t_p}{T}\right)-\sin\left(\frac{2\pi n t_{p-1}}{T}\right)\right]$$

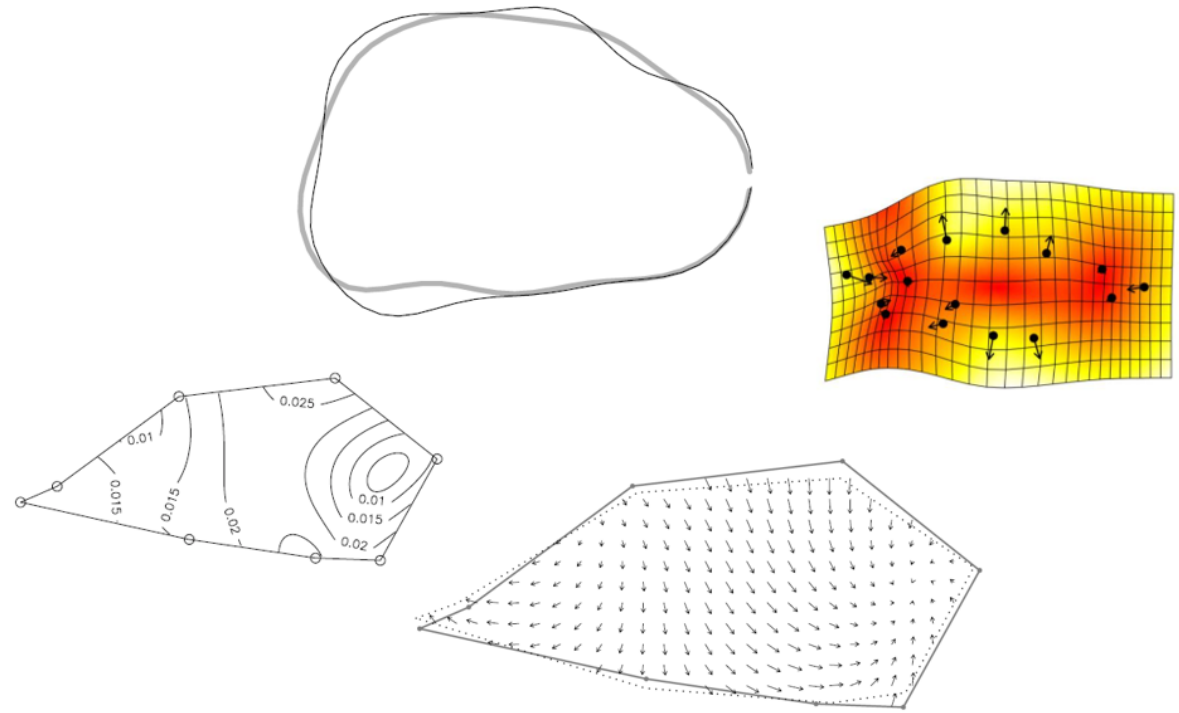
k = the **total number of steps around the outline**
n = the **harmonic number**
Δx = the **displacement** between point p and p+1
Δt = the **length of the step** between point p and p+1
tp = **accumulated length** of step segments at point
pT = **sum of lengths** of all steps around outline

Stage 6: Some Exploratory & Analytical Procedures

Method	Data Input
Visualise Shape Change	Procrustes coord. / Fourier coeff.
Principal Component Analysis (PCA)	Procrustes coord. / Fourier coeff.
Discriminant Analysis (DA/DFA/LDA/CVA)	Procrustes coord. / Fourier coeff. / PC Scores (+ Factor)
Procrustes ANOVA / MANOVA	Procrustes coord. / Fourier coeff. / PC Scores (+ factor)
Correlation & Regression	Procrustes coord. / Fourier coeff. / PC Scores (+ quant. variables)
Tree-building / Cluster Exercises	Procrustes coord. / Fourier coeff. / PC Scores (+ optional factor)

Visualising Shape Change

- Useful for comparing **individual** and **mean** shapes
- Shape change can be represented as:
  - Deformation grids
  - Contours
  - Outlines
  - Lollipop sticks
  - Vectors



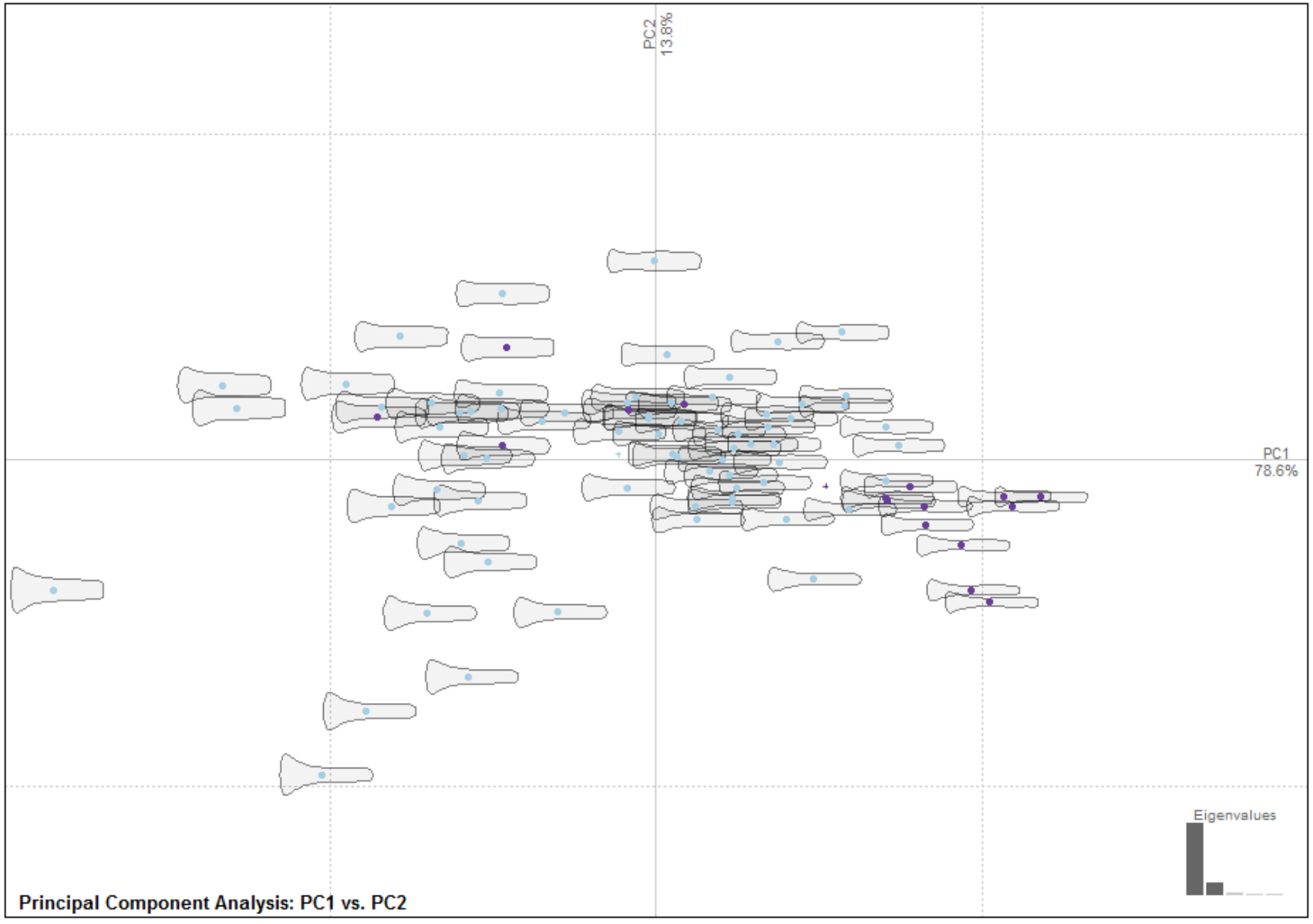
Claude, J. (2008). *Morphometrics with R*. Springer Publishing.

Scholtz et al. (2020). D'Arcy W. Thompson's Cartesian transformations: a critical evaluation. *Zoomorphology*. (https://doi.org/10.1007/s00435-020-00494-1)

Principal Component Analysis (PCA)

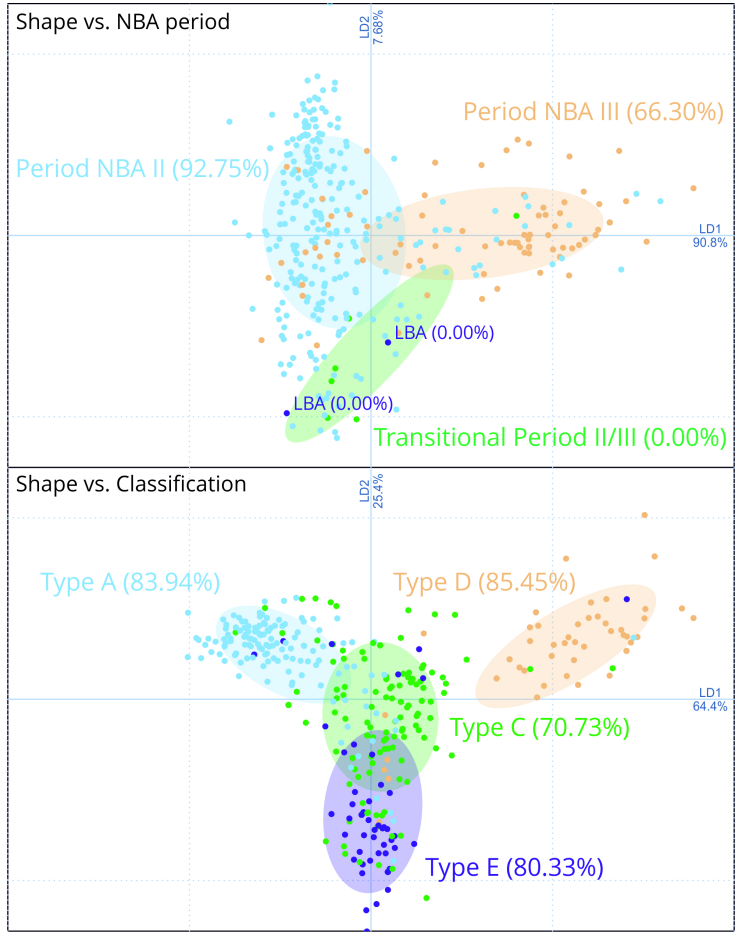
- Often the **first exploratory procedure**
- **Ordination (multi-dimensional)** method
- Each coordinate configuration = a shape
- **Principal components = sources of shape variation**
- Axis origin (0,0) = **mean shape**
- Scree plots (for examining PC contributions)
- **Important: total within-group variance is examined and not between-group variance! (factors overlay the plot!)**

Hoggard et al. (2019). The Potential of Geometric Morphometrics for Danish Archaeology: Two Case Studies. *Arkæologisk Forum*, 40: 30-42. (http://www.archaeology.dk/16738/Nr.%2040%20-%202019). OSF: https://osf.io/en5d2/


PC2
13.8%
PC1
78.6%
Eigenvalues
Principal Component Analysis: PC1 vs. PC2

Discriminant Analysis (DA)

- Otherwise known as **Discriminant Function Analysis (DFA)** or…
  - **Linear Discriminant Analysis** (LDA)
  - **Canonical Variate Analysis** (CVA)
- Investigates *a priori* classification through **maximum between-group classification**
- Percentage = success of classifier (as based on group data)
- **Jacknifed percentage** 👈👈👈
  - **Leave-one-out cross-validation**
  - **Success with which a random shape can be correctly classified**



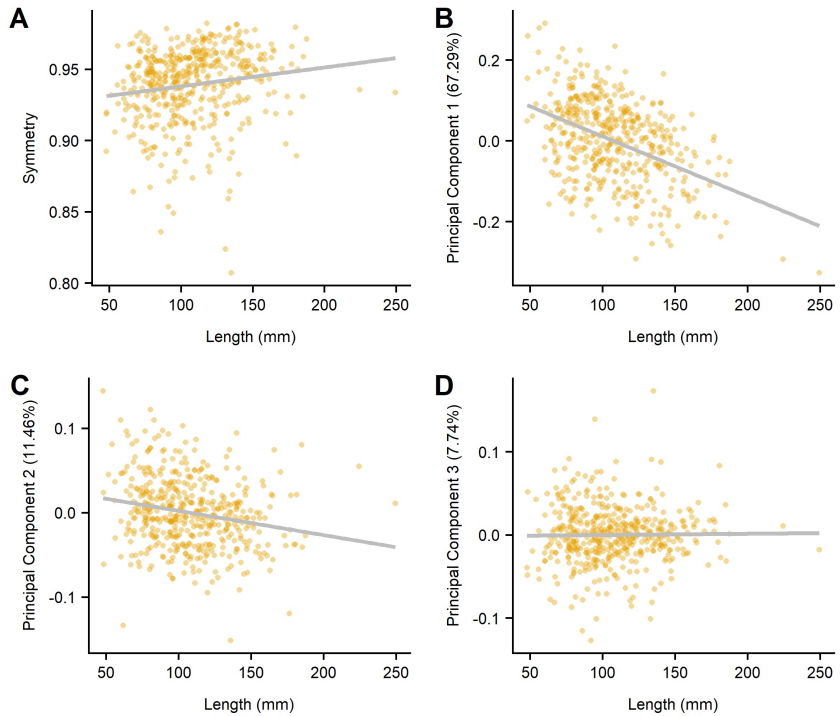
Vestergaard, C. and Hoggard, C.S. (2019). A Novel Geometric Morphometric (GMM) Application to the Study of Bronze Age Tutuli. *Danish Journal of Archaeology*, 8: 5-28.

Procrustes ANOVA / MANOVA

- PCA and DA act as **exploratory devices** for looking at shape difference
- MANOVA and Procrustes ANOVA provide a **statistical framework** for examining shape
- Procrustes ANOVA: performed in `R::Geomorph`
- MANOVA: performed in `R::Momocs`
- **Null Hypothesis: same populations / same shape**

Correlation and Regression

- Useful for shape vs. quantitative data.
- **Measure of association (impact of unit change)**
- Possible variables:
  - **Centroid size**
  - **Symmetry**
  - **Latitude**
  - **Shape score**
- Example methods: `base:: lm ()` or `tidyverse`



Hoggard et al. (2019). The application of elliptic Fourier analysis in understanding biface shape and symmetry through the British Acheulean. *Journal of Paleolithic Archaeology*, 2 (2): 115-133.

Cluster & Tree Exercises

- Useful for examining *a posteriori* classifications
- A variety of cluster analyses are available:
  - `Momocs::CLUST` (Hierarchical Clustering)
  - `Rphylip::contml` (Maximum Likelihood)

Ivanovaite et al. (2020). All these fantastic cultures? Research history and regionalisation in the Late Palaeolithic tanged point cultures of Eastern Europe. *European Journal of Archaeology*, 23 (2): 162-185.

Technical Notes and Suggestions

Principal Component Analysis: consider **between-groups PCA (bgPCA)**

Discriminant Analysis: consider **Kovarovic et al. 2011** (notes on sample size!)

Other cool methods:

1) **Machine Learning** (Unsupervised vs. Supervised classificatory methods)

2) **Bayesian Approaches** to GMM

3) **Partial Least Squares (PLS) and Modularity Studies**

Kovarovic et al. (2011). Discriminant function analyses in archaeology: are classification rates too good to be true? *Journal of Archaeological Science*, 38: 3008-3018. doi: 10.1016/j.jas.2011.06.028

Concluding Remarks

Geometric Morphometrics = 😍😍😍

In the future we will see...

a wider application of GMM in Archaeological Science!

a more powerful GMM analyses in Archaeological Science!

a more open, replicable and reproducible Archaeological Science!

Thank you for listening!

Slides created via the R package **xaringan**.

The chakra comes from remark.js, **knitr**, and R Markdown.