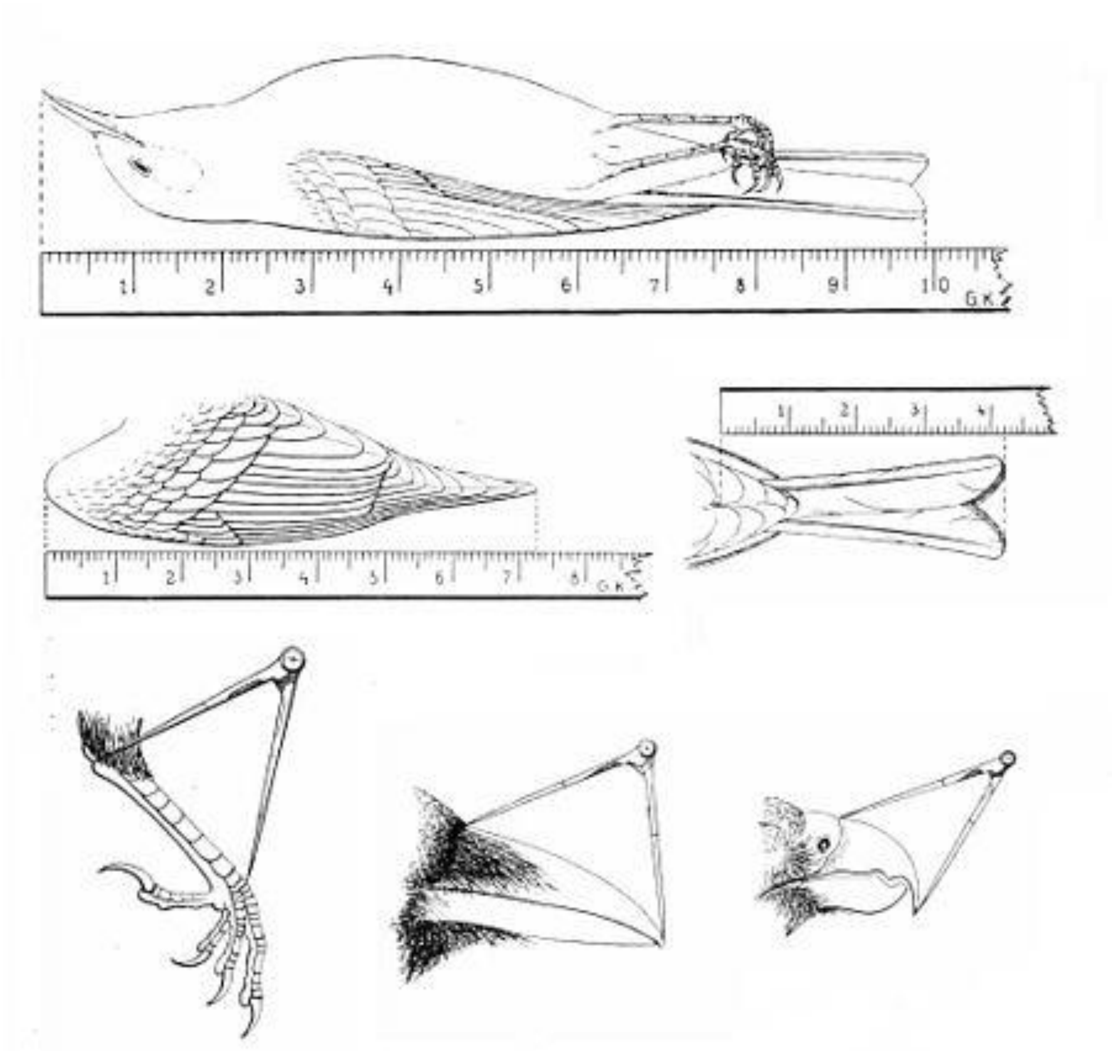
Statistical Shape Analysis #1 – Concepts and Landmarks

A. Murat Maga, PhD
Associate Professor
University of Washington
Division of Craniofacial Medicine
&
Seattle Children's Research Institute
Center for Developmental Biology and Regenerative Medicine

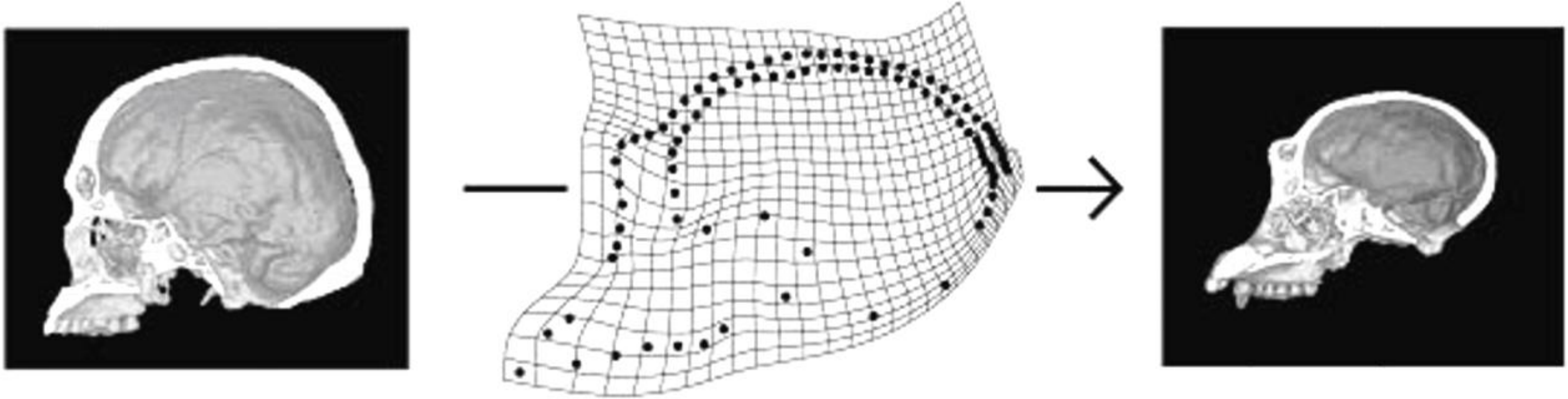
Traditional Morphometrics



Traditional morphometrics analyzes: **lengths**, **widths**, **masses**, **angles**, **ratios** and **areas**.

These data are also useful when size measurements are of theoretical importance such as body mass and limb cross-sectional area and length in studies of functional morphology. However, these measurements have one important limitation: **They contain little information about the spatial distribution of shape changes across the organism.**

Thin-Plate Spline (TPS) deformations

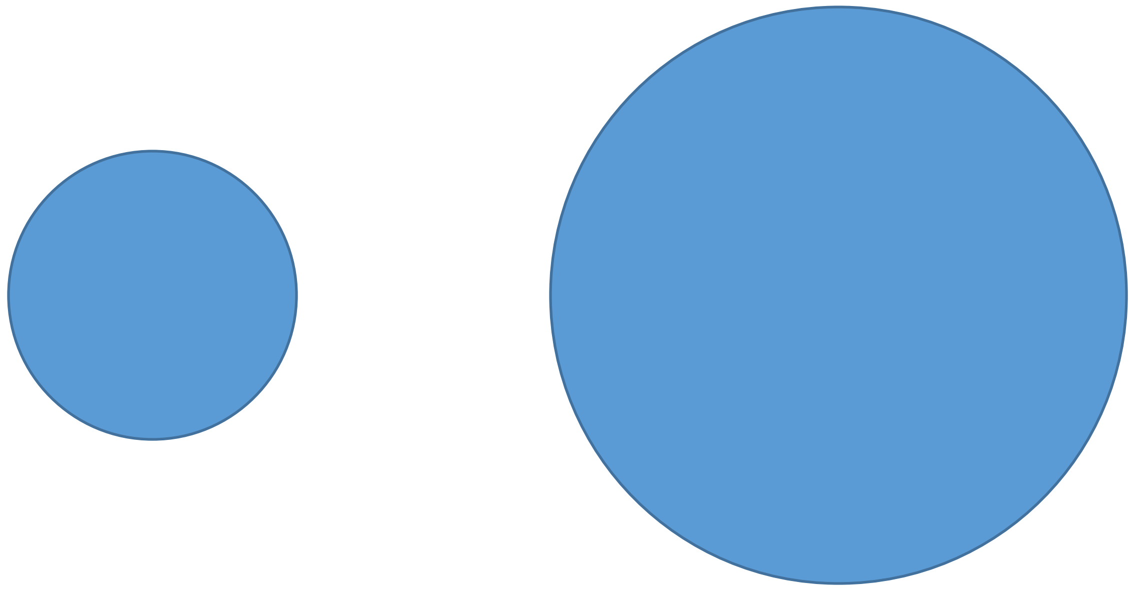


Bookstein FL. 1989. Principal warps: thin-plate splines and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence 11:567–585.

What is biological shape/form?

- **Biological shape** is the residual information left after differences in translation, rotation and **uniform size** is removed from the data.
- **Biological form** is the residual information left after **only** differences in translation and rotation is removed from data. (Size becomes part of the analysis).

Shape vs form



- Same **shape** (circle), but different **forms** (small vs big)**.**

Shape vs form

- What about these?

Shape vs form

- To decide, we have to establish some sort of correspondence (or mapping) between these shapes.

Homology and types of landmarks

Type I: juxtaposition of tissue or foramina

Type II: self-evident geometry (tips of prominences or notches)

Type III: geometric construction (e.g., deepest point in the notch)

Semi-landmarks (typically equidistant along a curve)

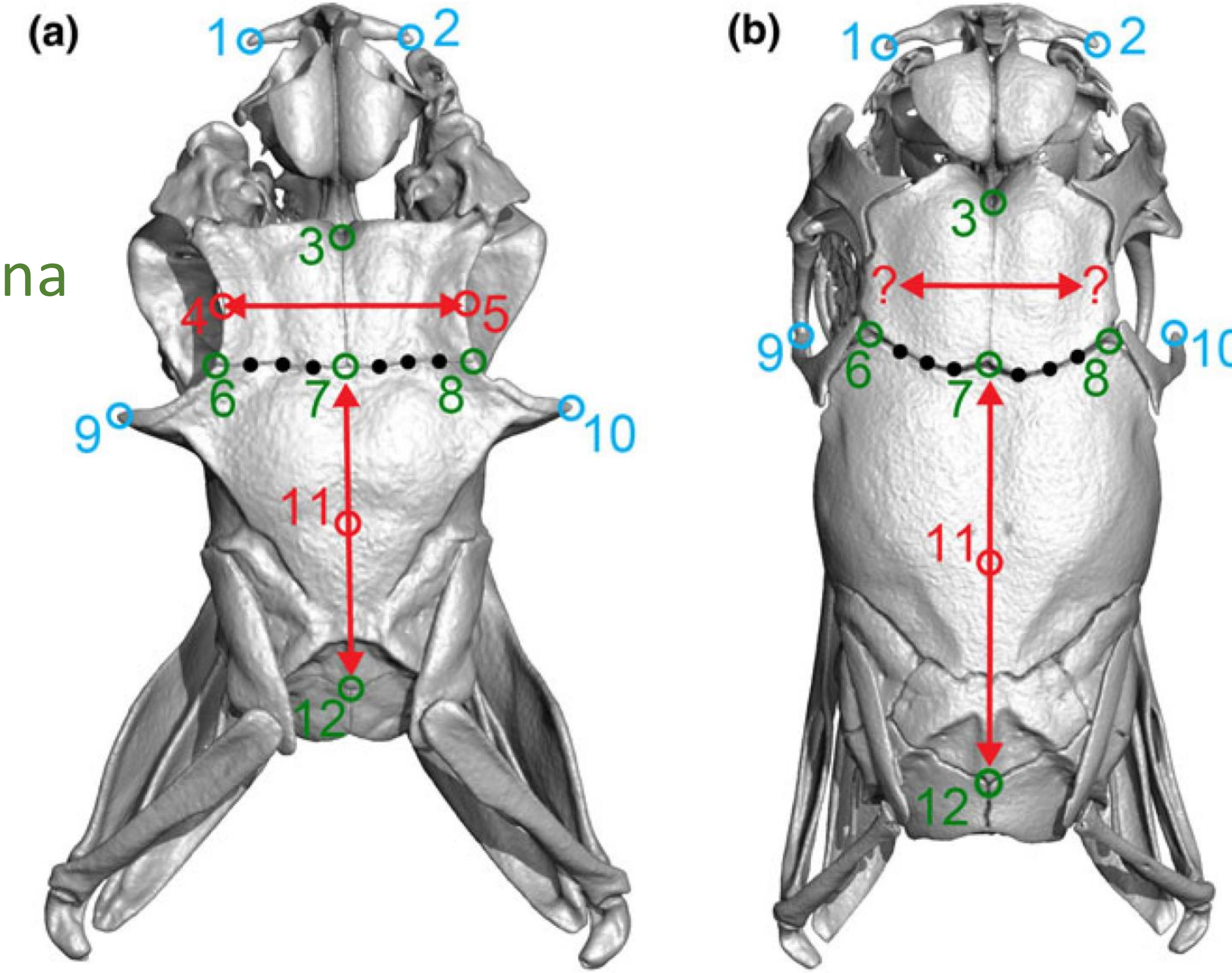
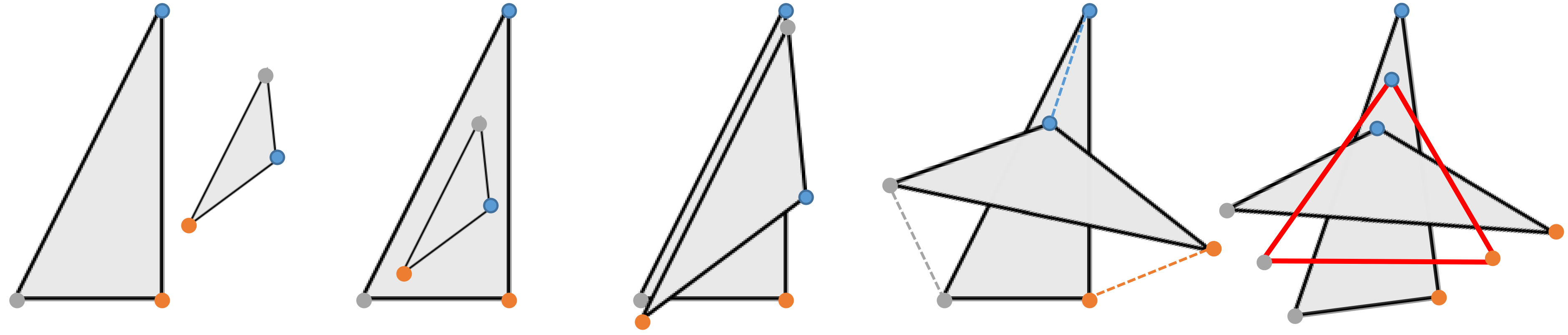


Figure from: Palci, Alessandro, and Michael S. Y. Lee. "Geometric Morphometrics, Homology and Cladistics: Review and Recommendations." *Cladistics* 35, no. 2 (2019): 230–42. https://doi.org/10.1111/cla.12340.

Putting all together



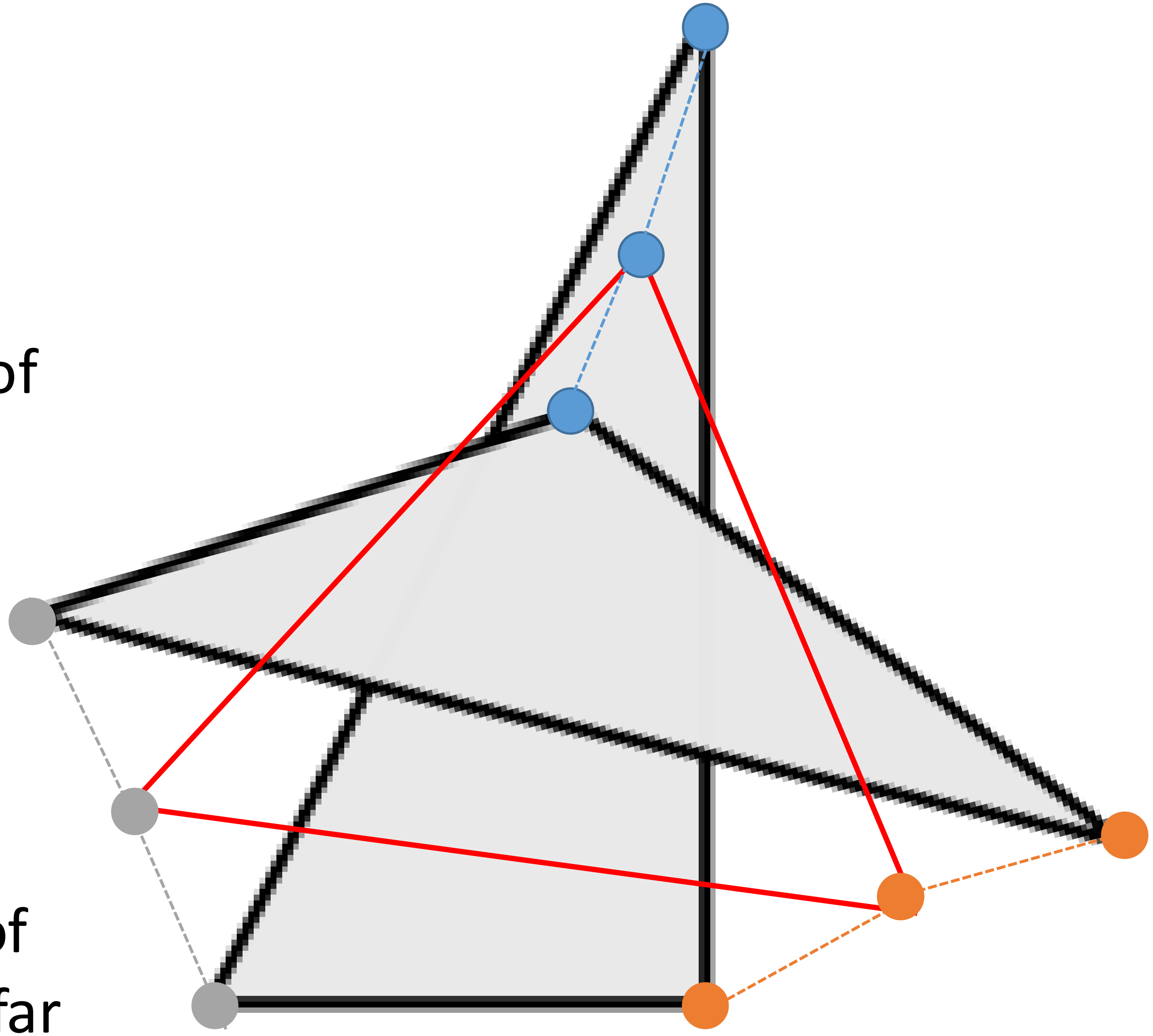
Original Configurations | Superimposition | Scaling | Rotation | Iterations to minimize the difference

Generalized Procrustes Analysis (or GPA)

- Gower JC. 1975. Generalized procrustes analysis. Psychometrika 40:33–51.
- Kendall DG. 1984. Shape Manifolds, Procrustean Metrics, and Complex Projective Spaces. Bull London Math Soc 16:81–121.
- Rohlf FJ, Slice D. 1990. Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks. Systematic Zoology 39:40–59.

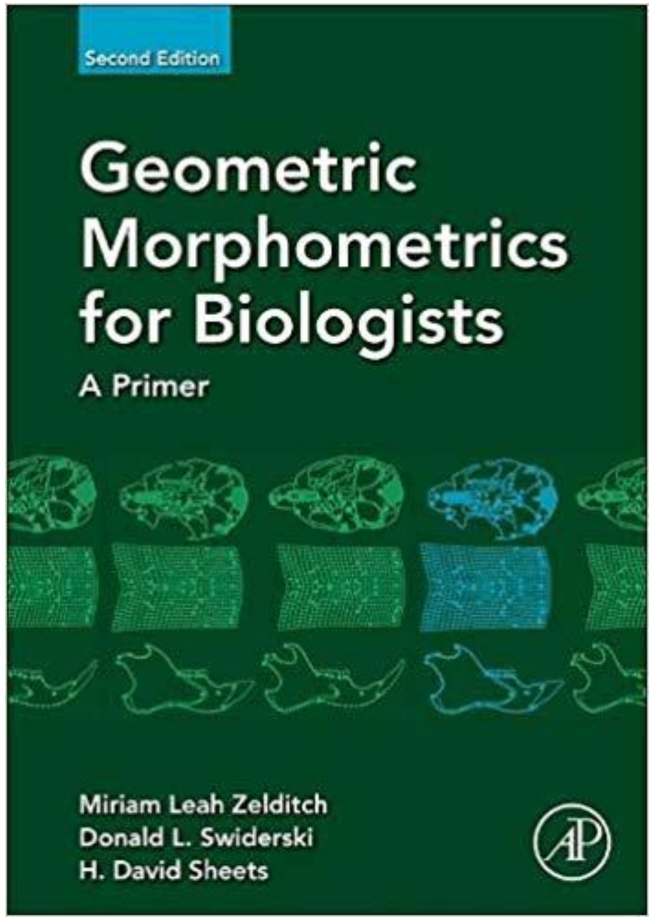
Some Definitions

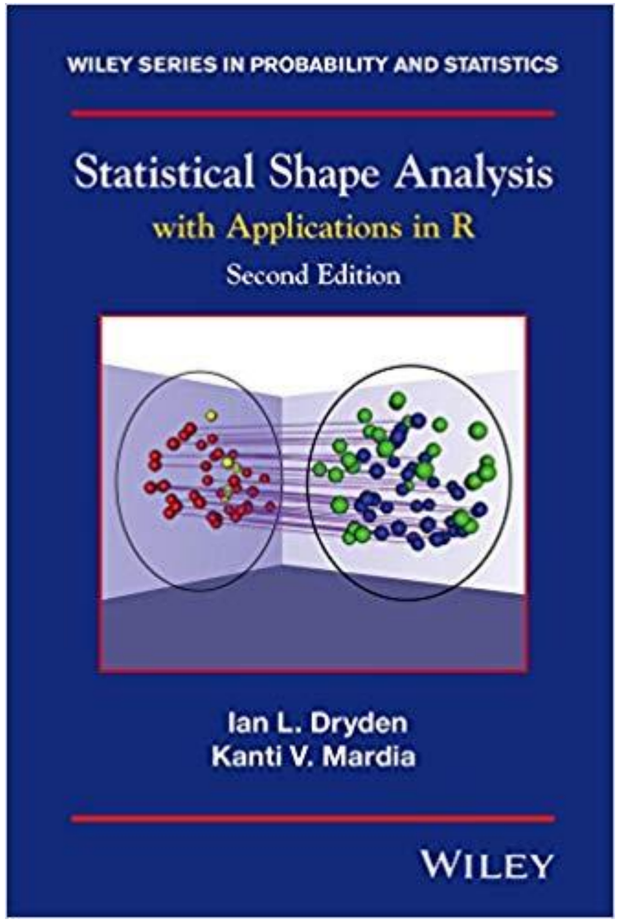
Procrustes coordinates: New set of coordinates after GPA
Consensus (mean) shape: The average of individual landmarks (red triangle)
Procrustes residuals: [Procrustes coordinate – consensus shape]; i.e. description of how each individual different from the consensus. (vector defined by the dashed line)
Procrustes distance: The squared sum of these differences, i.e., measure of how far the individual is from the mean (dashed lines).



https://life.bio.sunysb.edu/morph/glossary/gloss1.html
https://life.bio.sunysb.edu/morph/glossary/gloss2.html

Textbooks

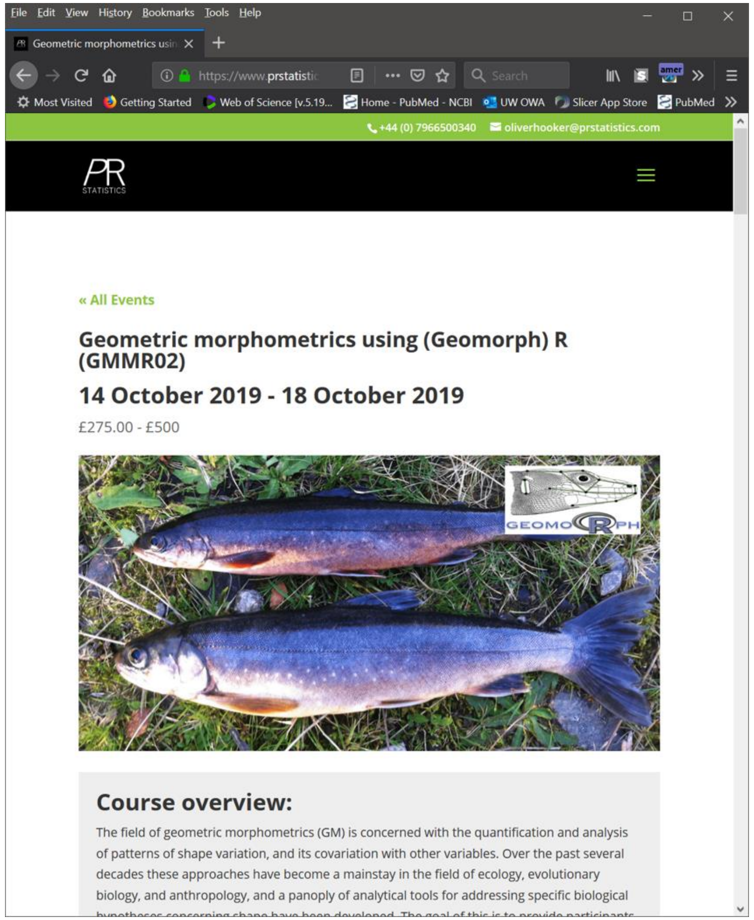




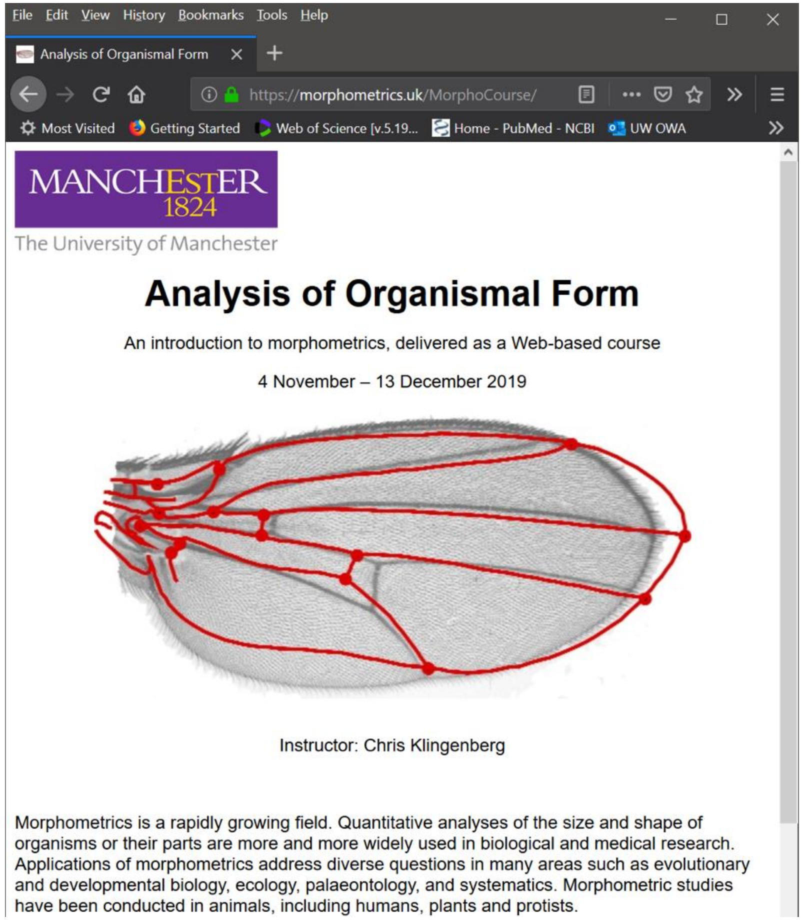
Also a short online paper https://geosci.uchicago.edu/~mwebster/Webster_and_Sheets_2010.pdf

A Practical Introduction to Landmark-based Geometric Morphometrics

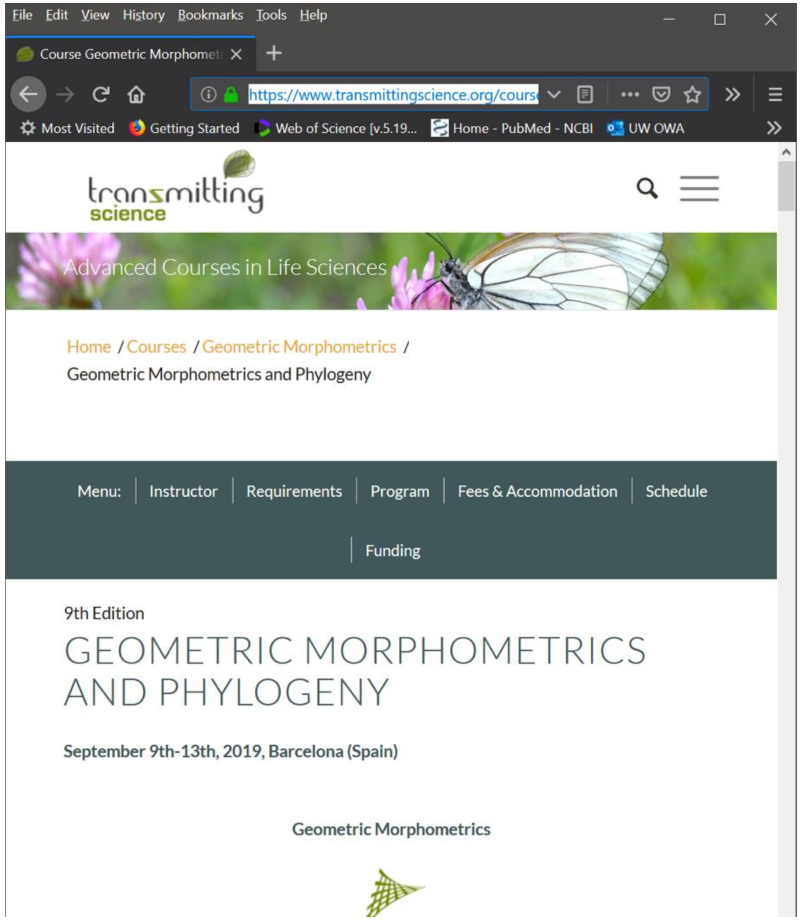
Geometric morphometrics specific short courses



https://www.prstatistics.com/course/geometric-morphometrics-using-r-gmmr02/ (Dean Adams)



https://morphometrics.uk/MorphoCourse/ (Chris Klingenberg)



https://www.transmittingscience.org/courses/geometric-morphometrics/geometric-morphometrics-phylogeny/

GPA Gotchas

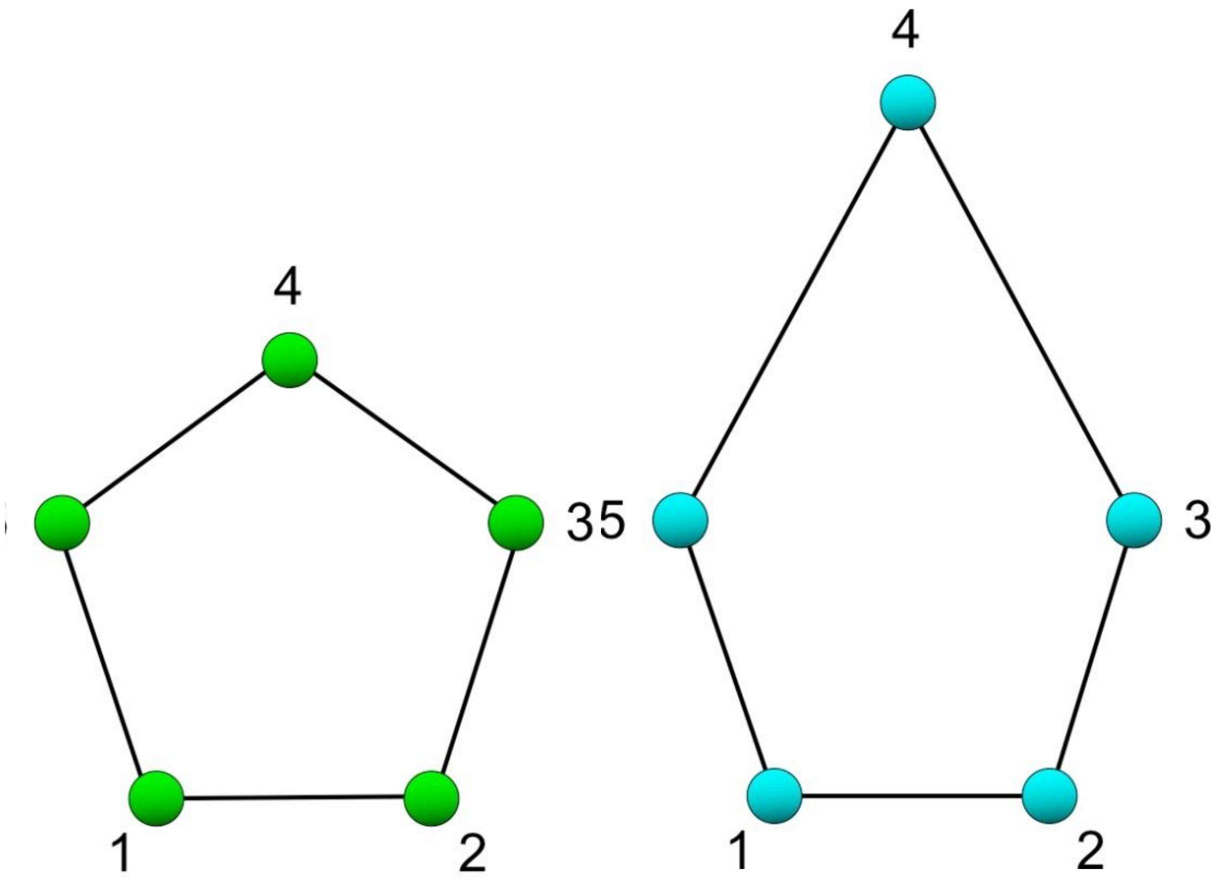
- In Generalized Least Squares (GLS) method of optimization variance will be distributed across landmarks. *

Warning about Procrustes

Procrustes-based analyses tell you about variation of the <u>entire</u> shape.

Procrustes analysis <u>does not</u> (easily) tell you how much variation occurs at <u>a particular landmark</u>.

With the removal of size and coordinate system, shape is being measured as displacement in each landmark <u>relative to</u> all the other landmarks.



Variations on a theme of Procrustes

Procrustes (Generalized Least Squares, GLS)

- minimizes sum-of-squares among homologous landmarks
- best method for minimizing difference in shape
- distributes differences equally among landmarks

Generalized Resistant Fit (GRF)

- uses median rather than mean for the fitting algorithms
- allows some landmarks to be more variable than others
- does not find smallest possible difference between shapes

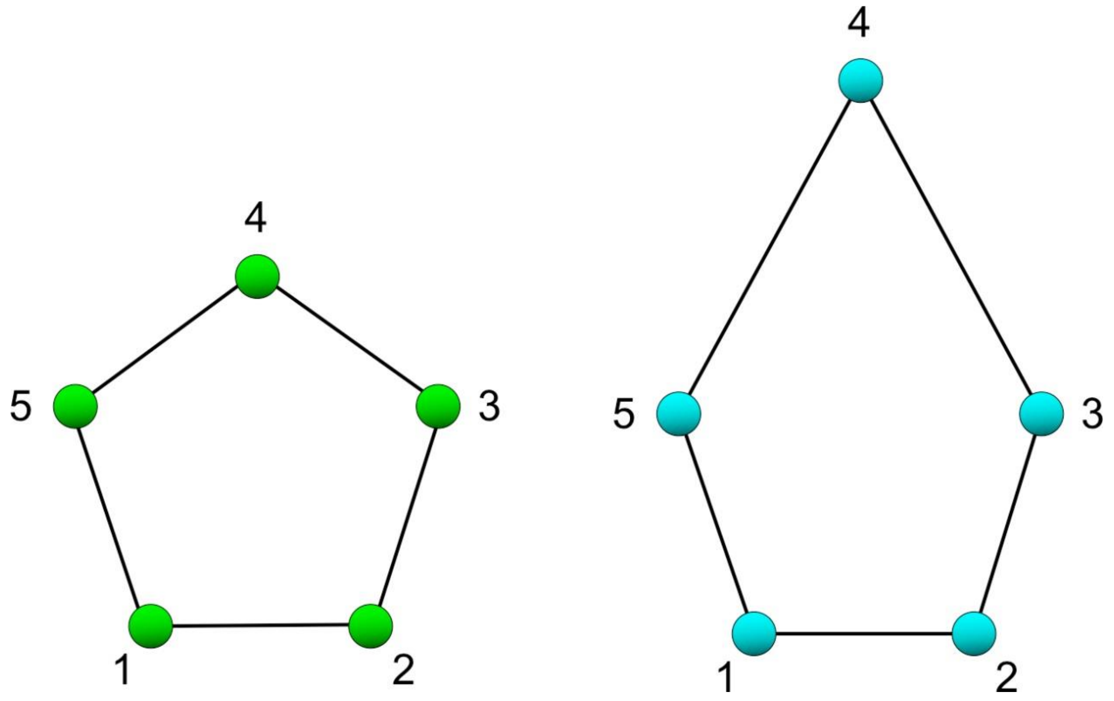
Bookstein Shape Coordinates

- uses two landmarks as baseline
- common in early "geometric morphometric" studies

GPA implementation of **SlicerMorph** (with the option of using Boas coordinates, which reintroduces the size during the PCA).

Euclidean Distance Matrix Analysis (EDMA)

- Can be useful in case/control study designs.
- Exploratory Method to identify significantly different differences between groups.
- All possible pairwise landmark distances are calculated. For N landmarks there are [Nx(N-1)]/2 pairwise distance.
- 5 LMs = 20 pairwise distances
- 50 LM = 1225 pairwise distances



'Pinocchio Effect'

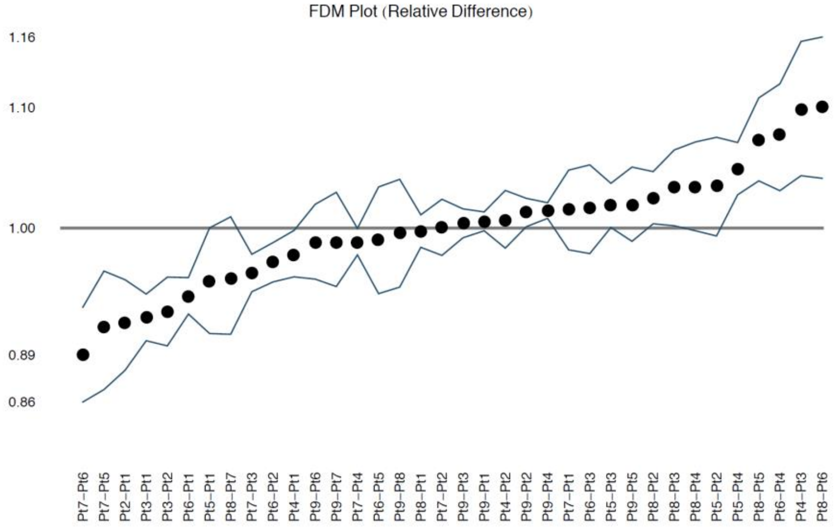
Lele S, Richtsmeier J. 1991. Euclidean Distance Matrix Analysis - a Coordinate-Free Approach for Comparing Biological Shapes Using Landmark Data. Am J Phys Anthropol 86:415–427.

Euclidean Distance Matrix Analysis

EDMA: example results

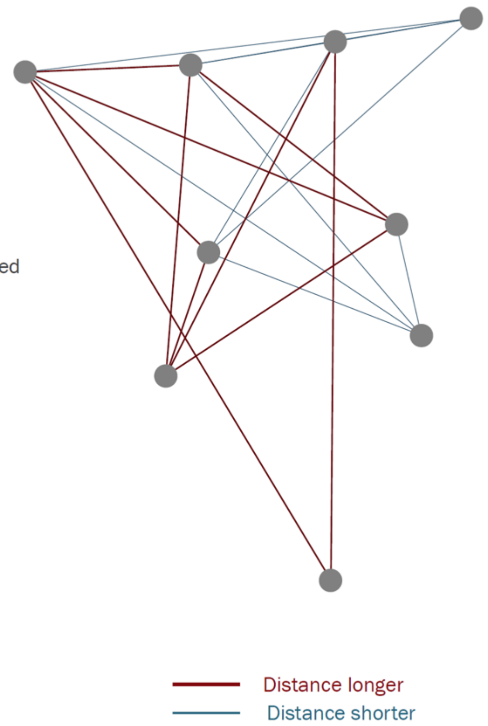
In this example the faces of left facing people and right facing people have been landmarked (landmarks at the corners of each eye, the edges of the nose, the corners of the mouth, and the chin). There were nine landmarks total, which gives 36 interlandmark distances.

The FDM plot shows the ratio of each interlandmark distance between the two samples (left facing over right facing), with those landmarks that got farther apart having ratios greater than 1.0, those that got closer having ratios less than 1.0. 95% confidence intervals were found with 1000 bootstrap replicates.



EDMA: example results

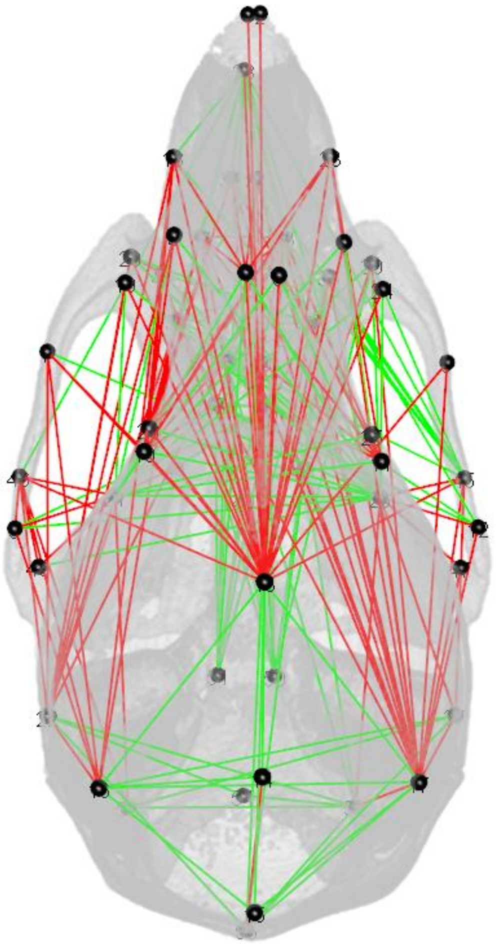
The consensus of the second sample is shown here with those interlandmark distances that got significantly longer in red and those that got significantly shorter in blue. For people who turned their heads from left to right, the nose and chin got significantly farther from the left eye and left corner of the mouth, whereas those same landmarks got closer to the right eye and the right corner of the mouth.



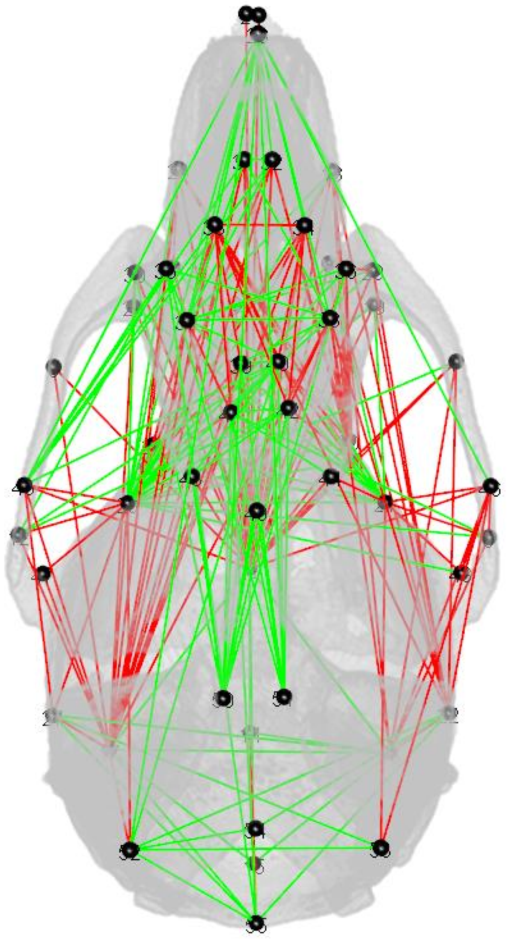
Effect of gestational (E0-E8) chronic alcohol exposure on skull development

Distances significantly smaller in EtOH group

Distances significantly larger in EtOH group



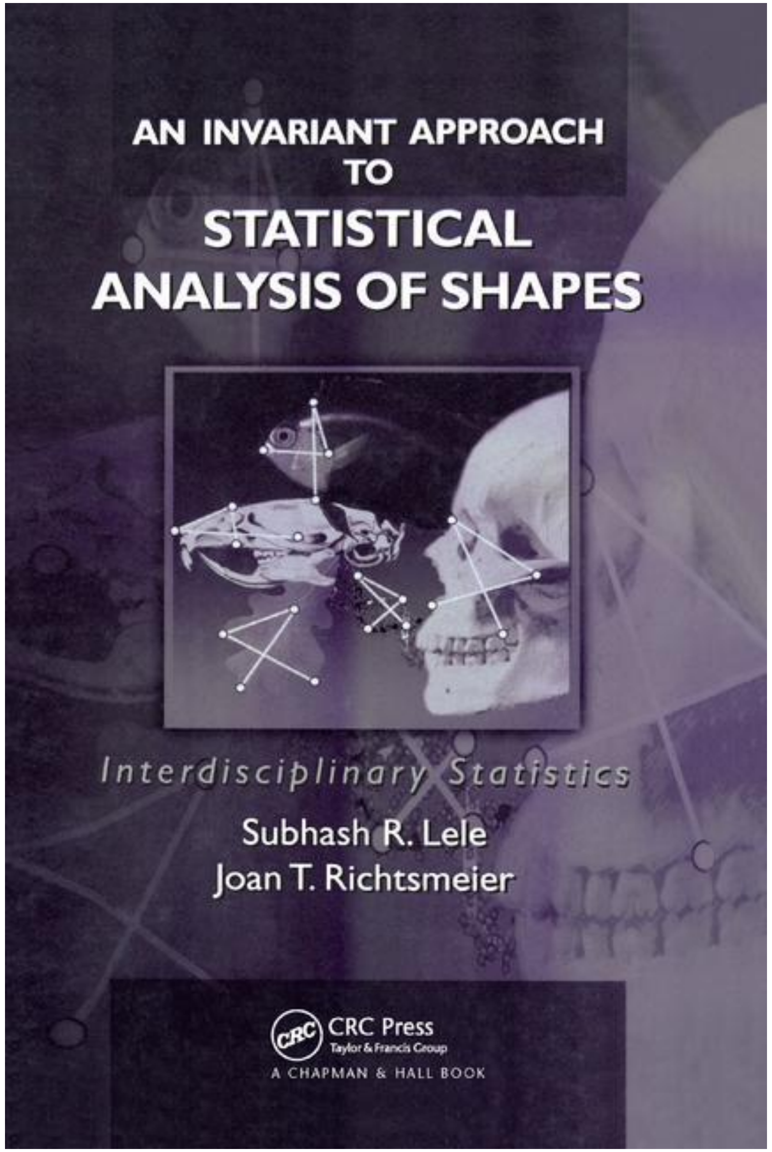
Superior / Dorsal view



Inferior / Ventral view

Student Project: EDMA is NOT available in SlicerMorph, but we have a prototype implementation that needs someone with time (and some python coding skills) to finesse and test.

EDMA References:



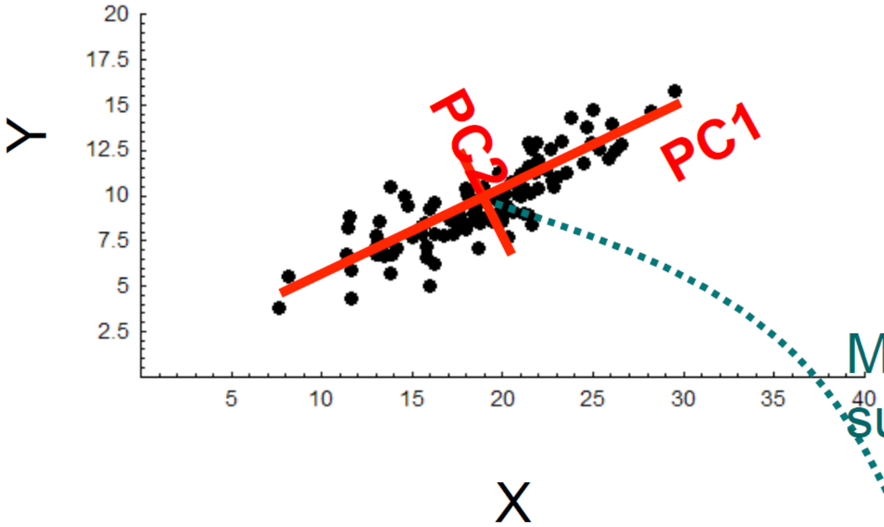
- Lele, Subhash, and Joan T. Richtsmeier. “Statistical Models in Morphometrics: Are They Realistic?” *Systematic Zoology* 39, no. 1 (March 1, 1990): 60–69. https://doi.org/10.2307/2992208.
- Lele, S., and Jt Richtsmeier. “Euclidean Distance Matrix Analysis - a Coordinate-Free Approach for Comparing Biological Shapes Using Landmark Data.” *American Journal of Physical Anthropology* 86, no. 3 (November 1991): 415–27. https://doi.org/10.1002/ajpa.1330860307.
- Lele, S., and T. M. Cole III. “A New Test for Shape Differences When Variance–Covariance Matrices Are Unequal.” *Journal of Human Evolution* 31, no. 3 (September 1996): 193–212. https://doi.org/10.1006/jhev.1996.0057.
- Lele, Subhash R, and Charles E McCulloch. “Invariance, Identifiability, and Morphometrics.” *Journal of the American Statistical Association* 97, no. 459 (September 1, 2002): 796–806. https://doi.org/10.1198/016214502388618609.

What does PCA do?

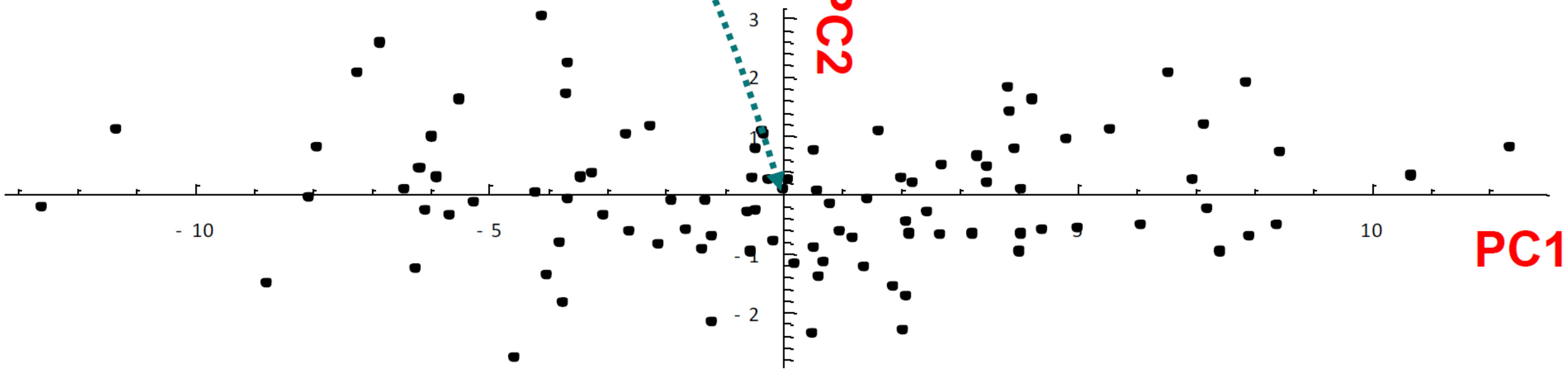
- 1. Rotates data to its major axes for easier visualization
- 2. Preserves original distances between data points (in other words, PCA does not **distort the variation data**, but **only if the covariance method** is used, which is standard in geometric morphometrics)
- 3. Removes correlations between variables to make further statistical analysis simpler
- 4. Can be a tool for ‘noise’ removal.

Principal components are a ‘rigid rotation’ of the original data

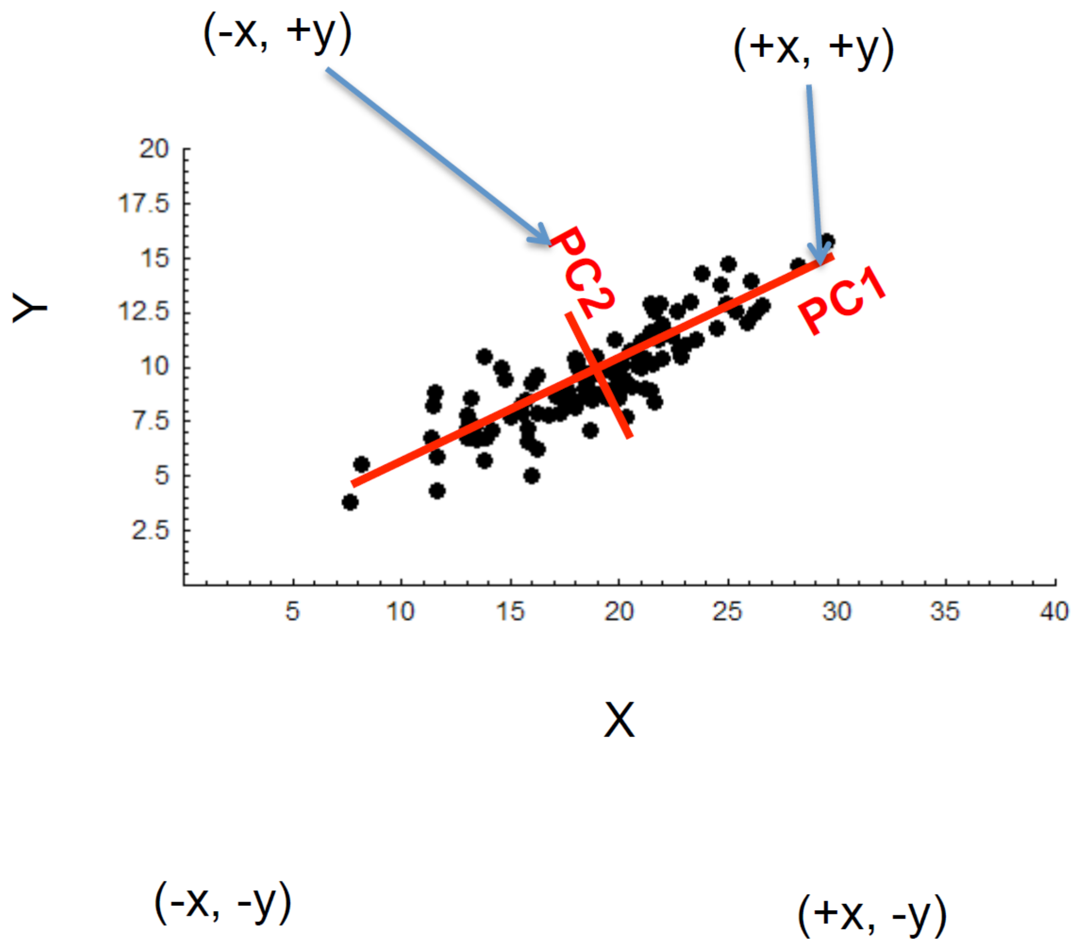
Data rotated so slope is 0.0



Mean changes to 0.0 (accomplished by subtracting mean from Procrustes points



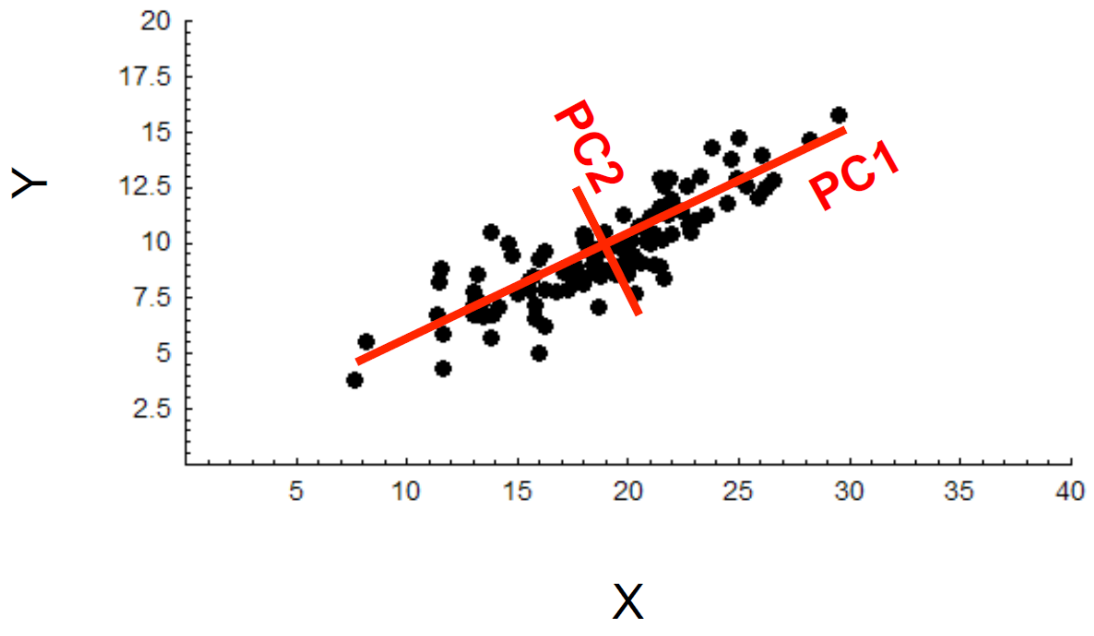
Eigenvector 'loadings' tell how each original variable contributes to the PC



Eigenvector Matrix

	PC1	PC2
X	0.89	-0.44
Y	0.44	0.89

Eigenvectors also describe how to transform data from original coordinate system to PCs and back

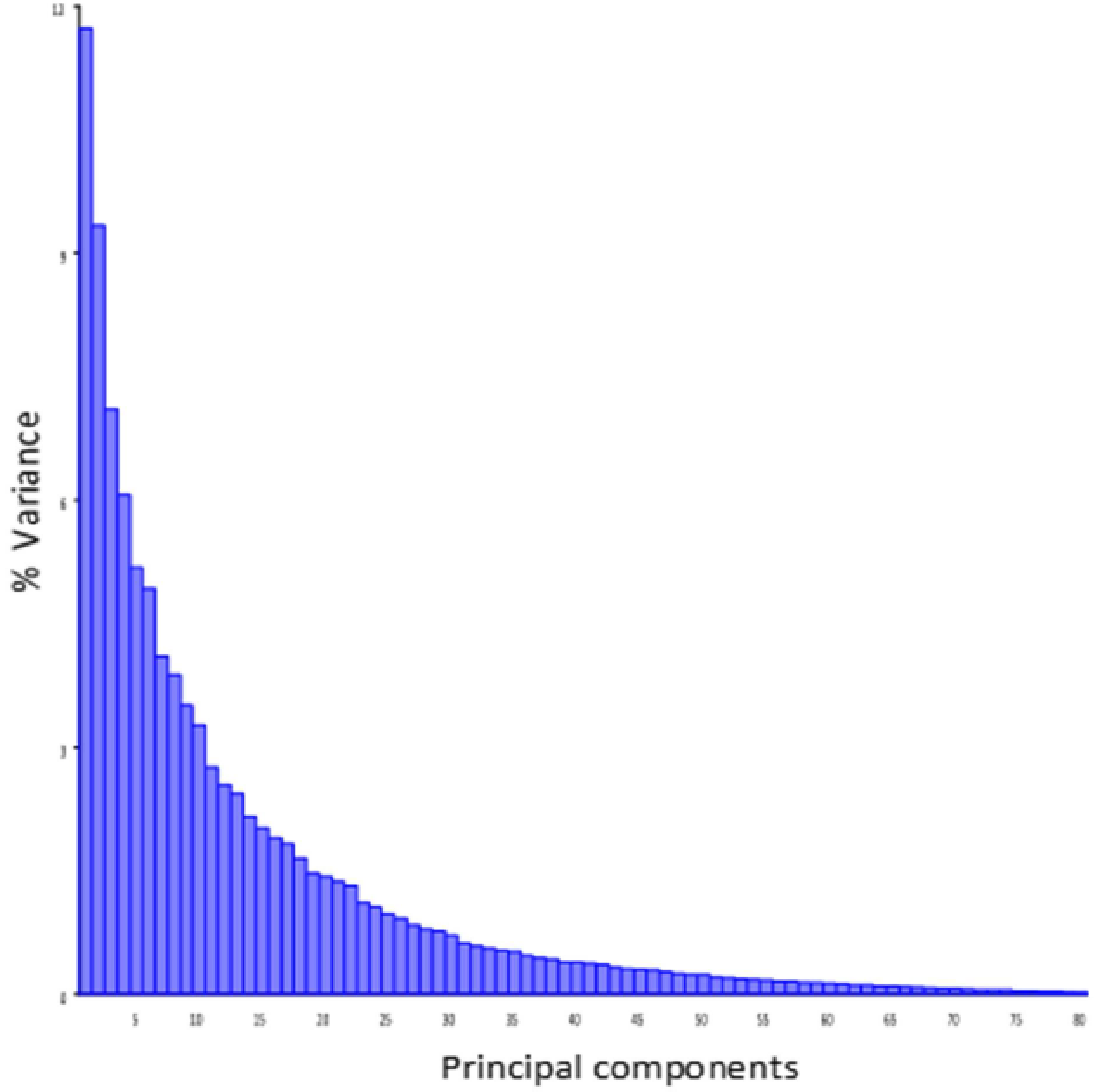


Eigenvector Matrix

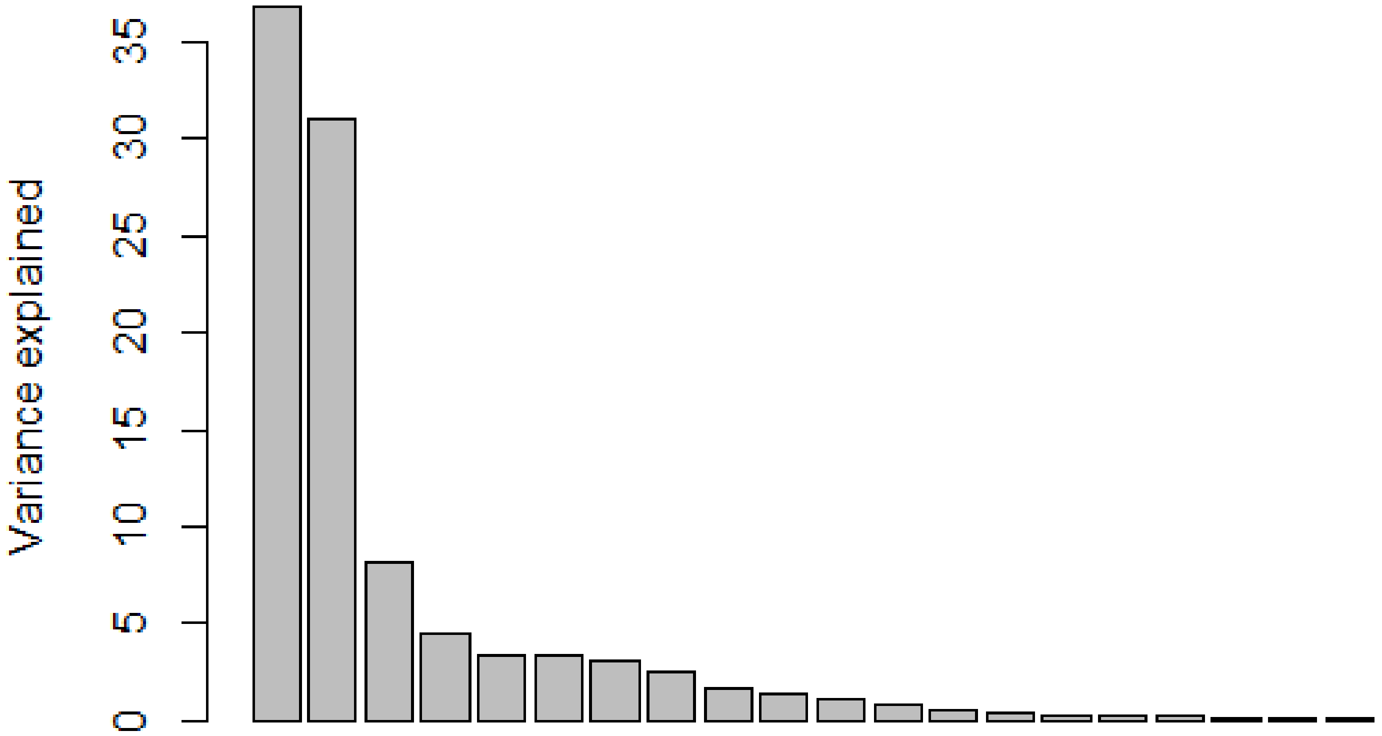
	PC1	PC2
X	0.89	-0.44
Y	0.44	0.89

(multiply PC1 X score by 0.89 and PC1 Y score by -0.44 and add back X,Y meanto get real X,Y)

Principle Components are ranked by variance



Mouse skulls from Maga et al. 2015



Plethodon Dataset from Geomorph

PCA is important in Geometric Morphometrics because....

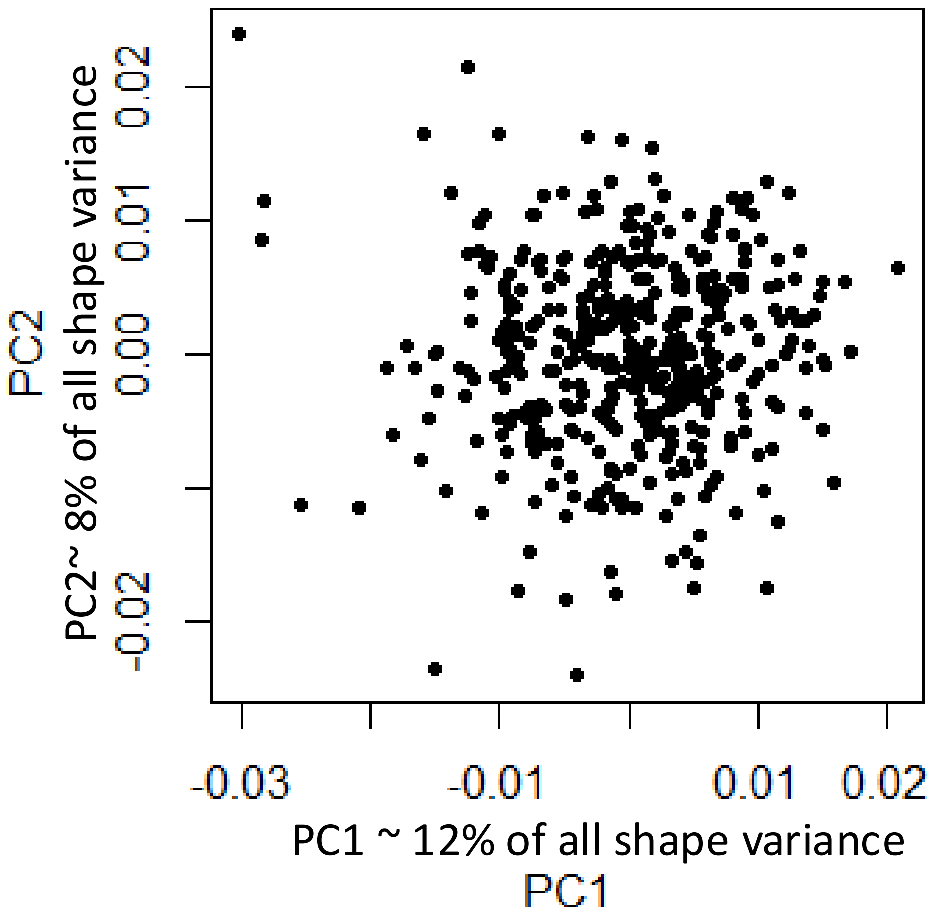
1. PCA scores are used as shape variables
2. Eigenvectors are convenient axes for shape space
3. Eigenvectors and their scores are uncorrelated as variables
4. Variance (eigenvalues) is partitioned across eigenvectors and scores in descending order
5. Scores can be safely used for all other statistical analyses, including tree building
6. Eigenvectors can be used to build shape models

PCA is important in Geometric Morphometrics because....

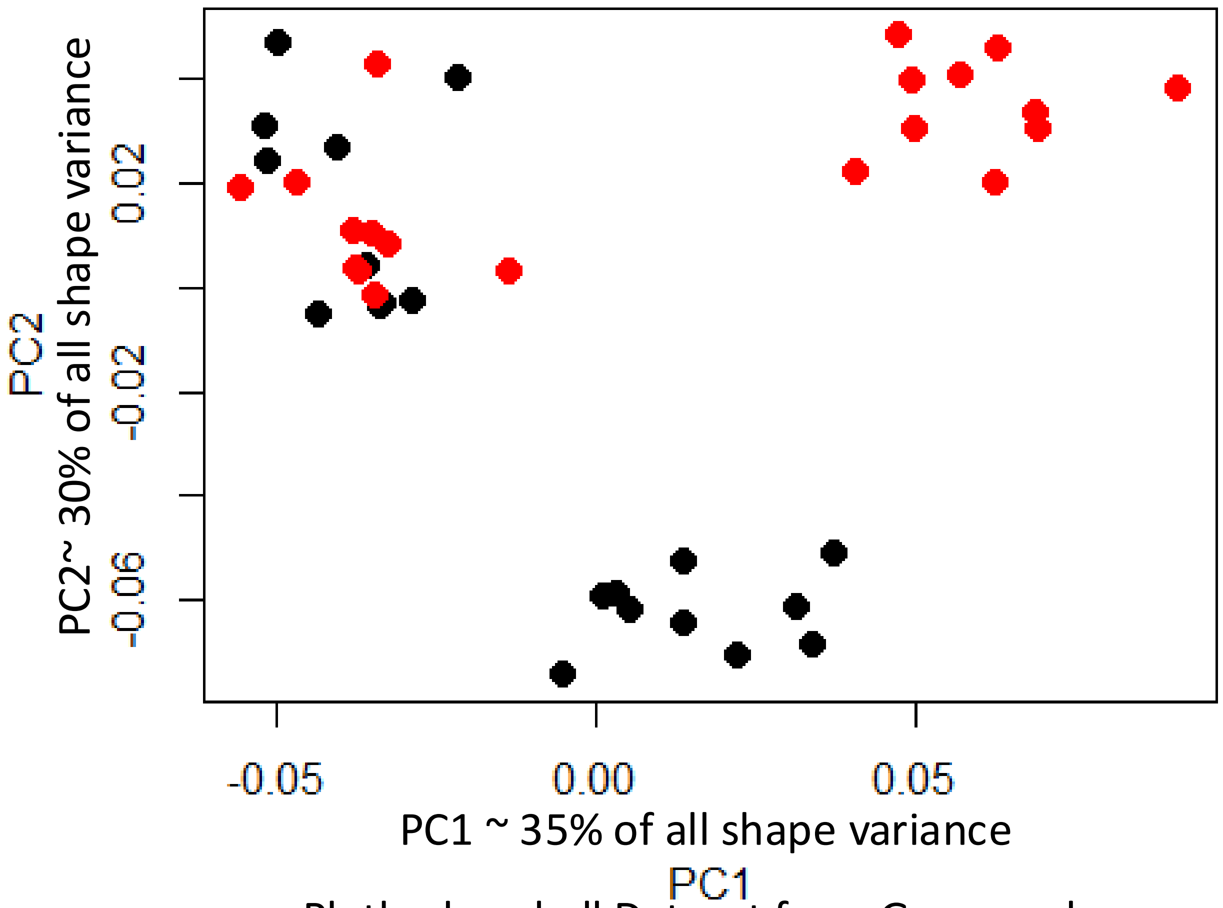
1. PCA scores are used as shape variables
2. Eigenvectors are convenient axes for shape space
3. Eigenvectors and their scores are uncorrelated as variables
4. Variance (eigenvalues) is partitioned across eigenvectors and scores in descending order
5. Scores can be safely used for all other statistical analyses, including tree building
6. Eigenvectors can be used to build shape models

PCs themselves have 0 (zero) inherent biological meaning. They are NOT traits.

…but some PCA results are more interpretable than others

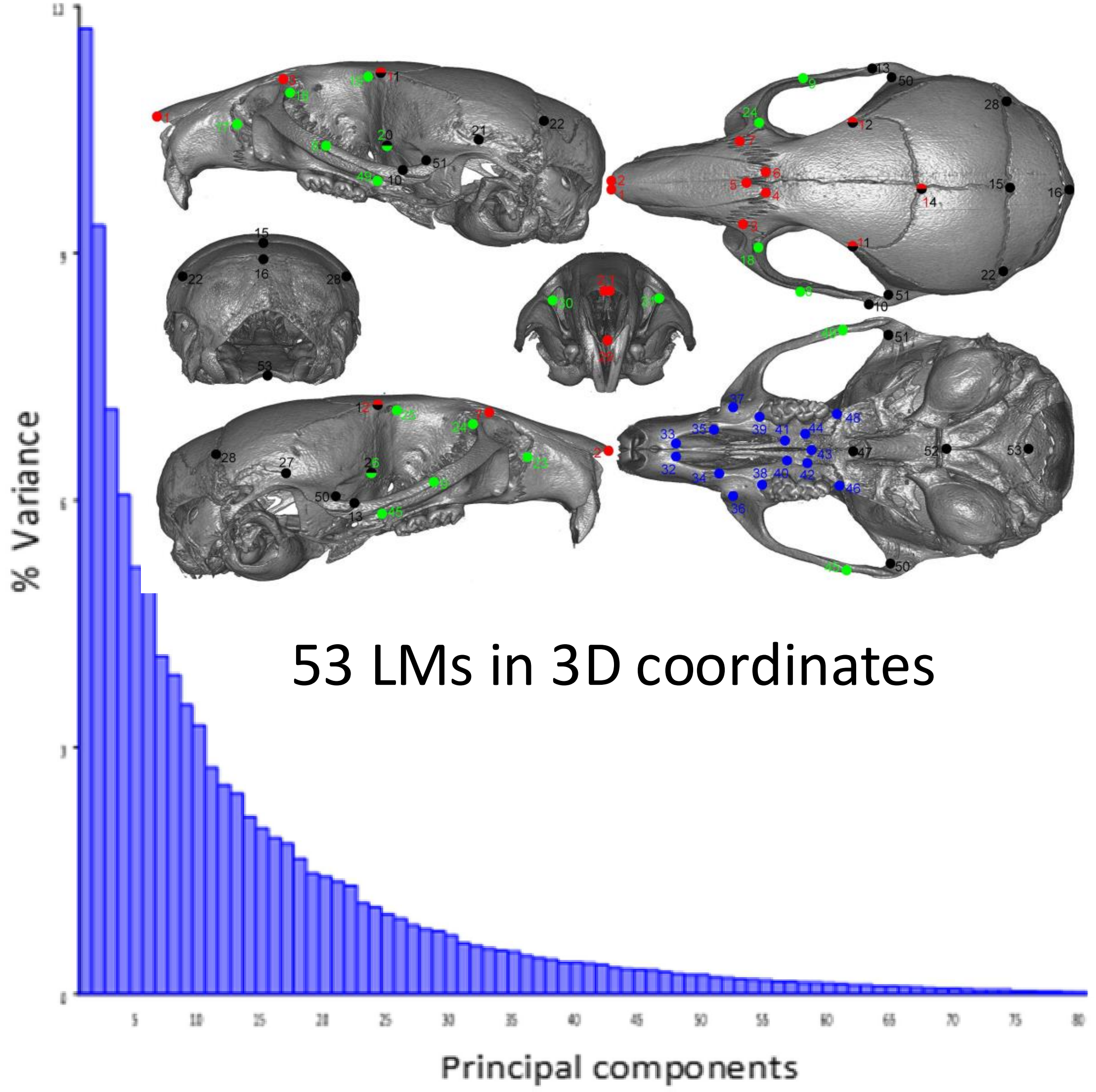


Mouse skulls from Maga et al. 2015. Frontiers in Physiology

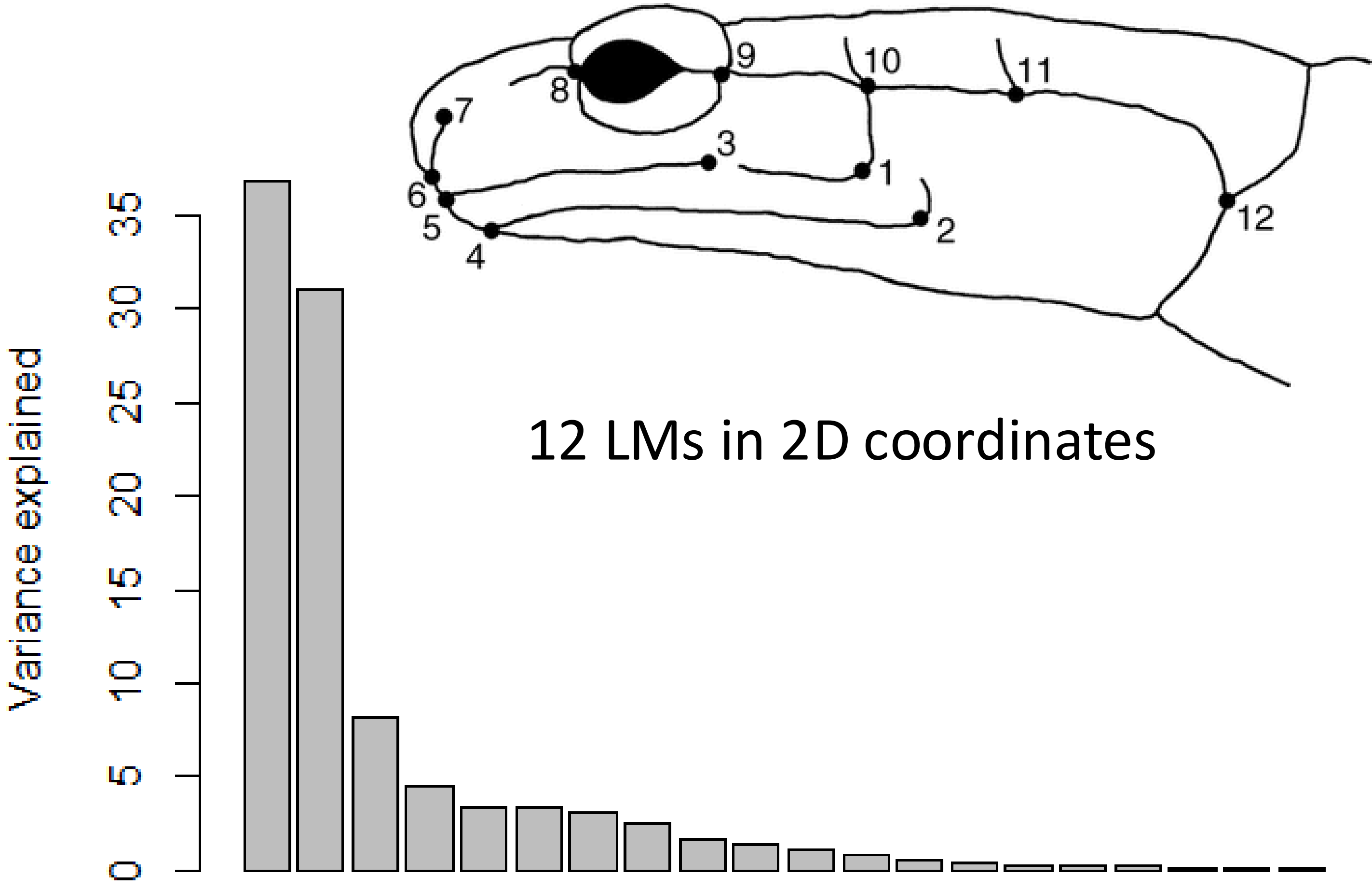


Plethodon skull Dataset from Geomorph
Adams 2004, *Ecology* 85, no. 10:2664–70.

How many landmarks to use?

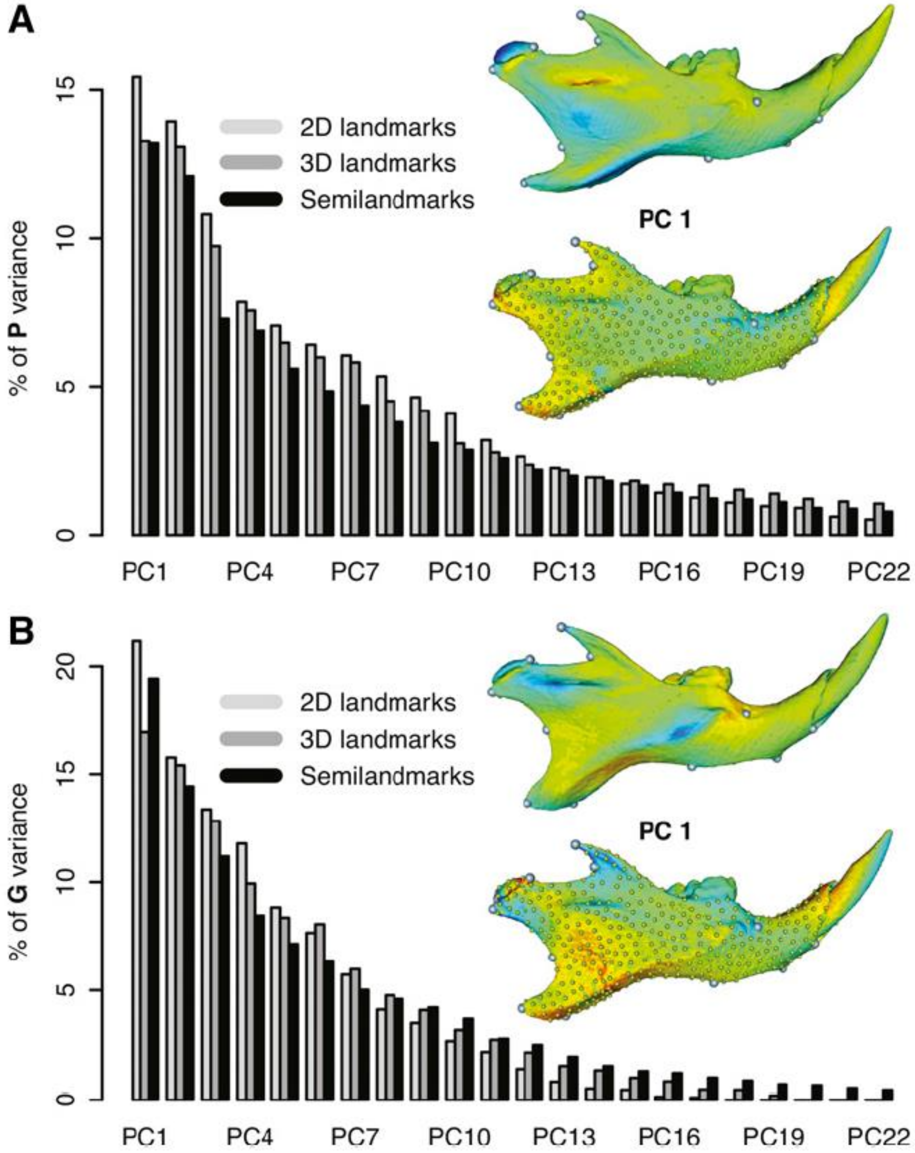


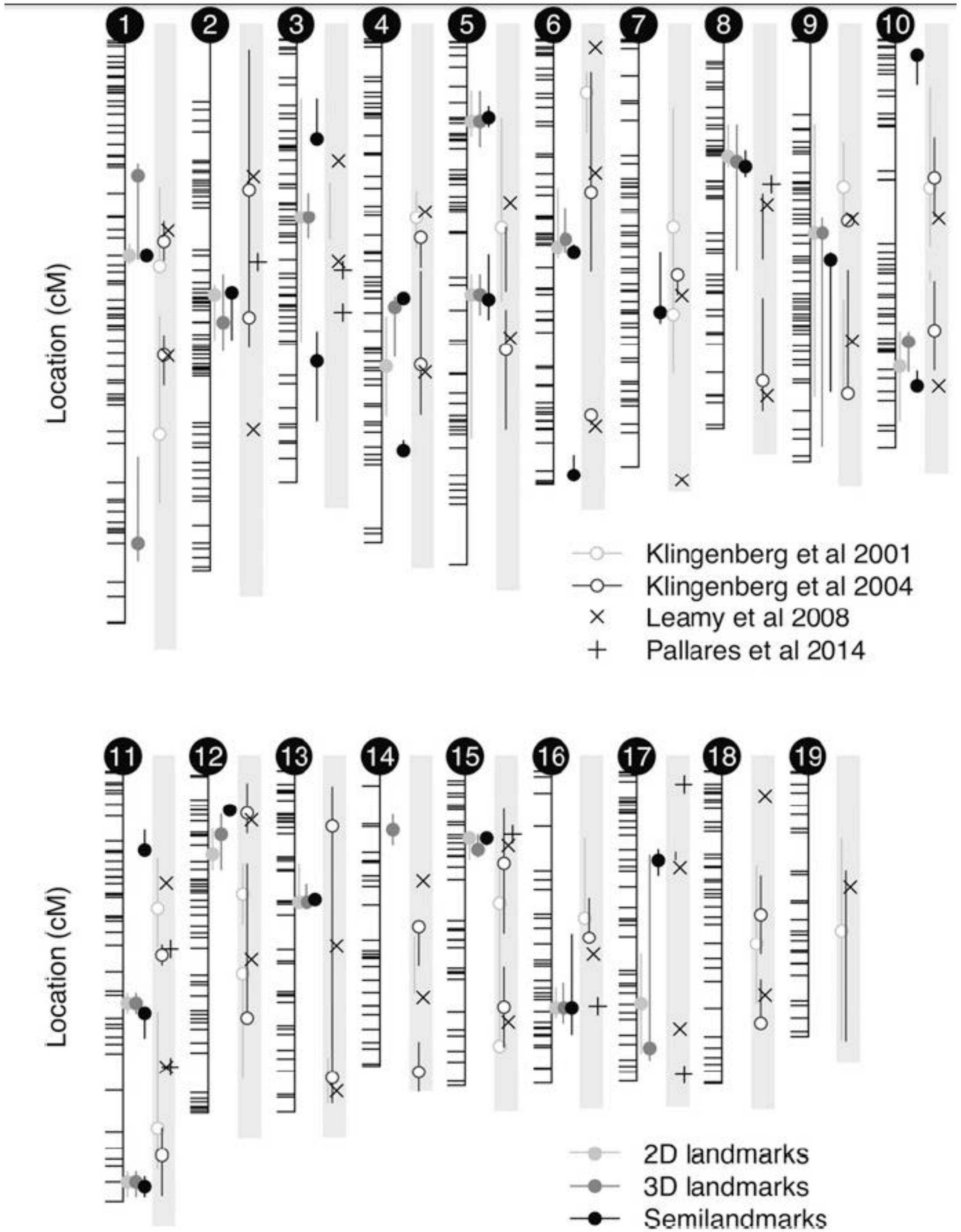
Mouse skulls from Maga et al. 2015



Plethodon skull Dataset from Geomorph
Adams 2004, *Ecology* 85, no. 10:2664–70.

2D vs 3D?





Navarro, Nicolas, and A. Murat Maga. "Does 3D Phenotyping Yield Substantial Insights in the Genetics of the Mouse Mandible Shape?" *G3: Genes, Genomes, Genetics* 6, no. 5 (May 1, 2016): 1153–63. https://doi.org/10.1534/g3.115.024372.

Curse of dimensionality

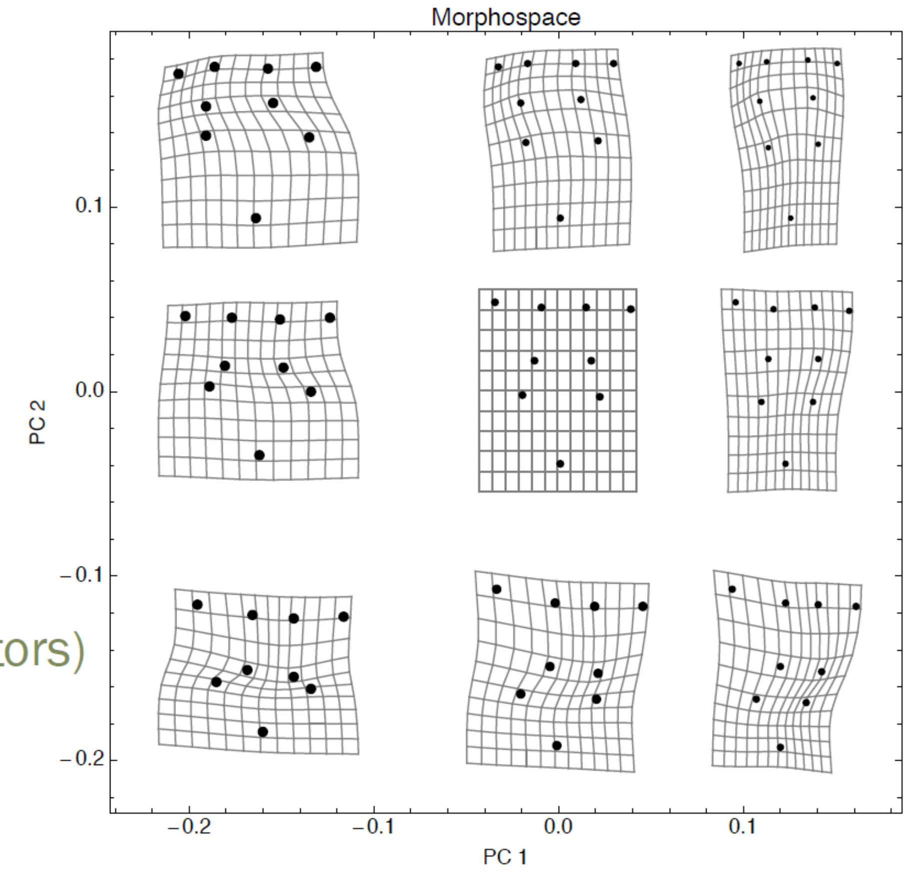
- **The curse of dimensionality:** When the dimensionality increases, the volume of the space increases so fast that the available data become sparse.

12 LMs in 2D = 20 dimensional space (24 variables - 4 DOF lost due to superimposition, scaling and rotation)
53 LMs in 3D = 152 dimensional space (159 variable - 7 DOF lost due to superimposition, scaling and rotation)

As a rule of thumb, you need at least sample size > your variable space
(in practice you may need sample size >>> your variable space, e.g., in mouse crosses)

Shape modelling…

How to construct models of shapes in morphospace

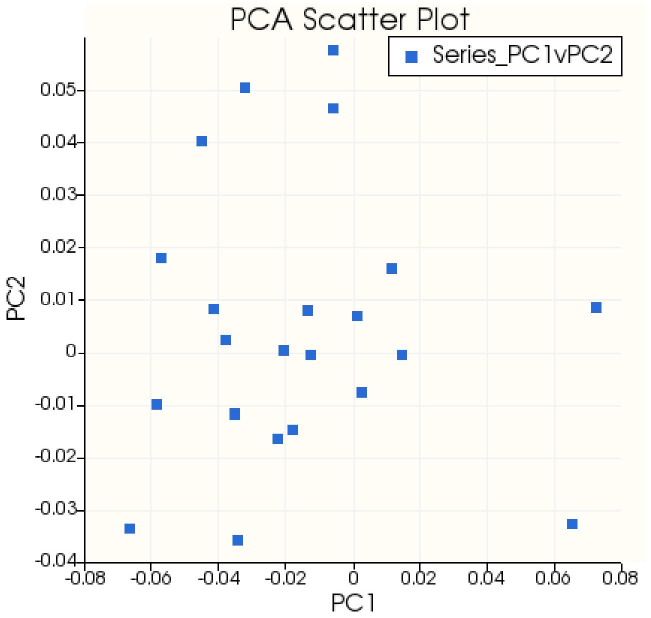


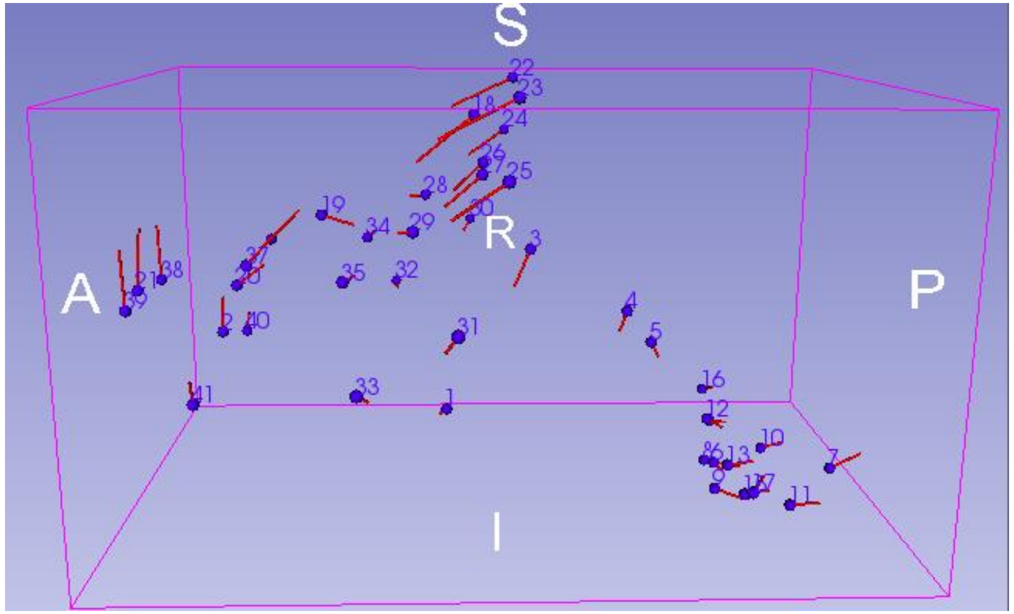
Ingredients:

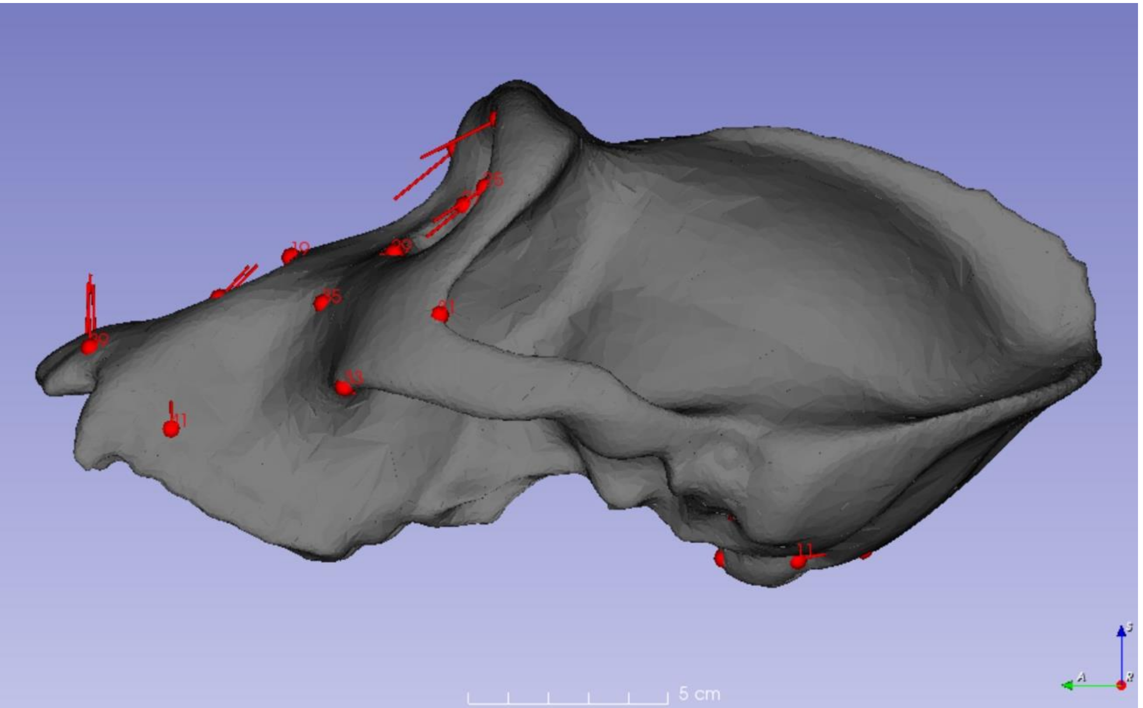
1. mean shape (consensus)
2. eigenvectors
3. the score (address) of the point to be modelled

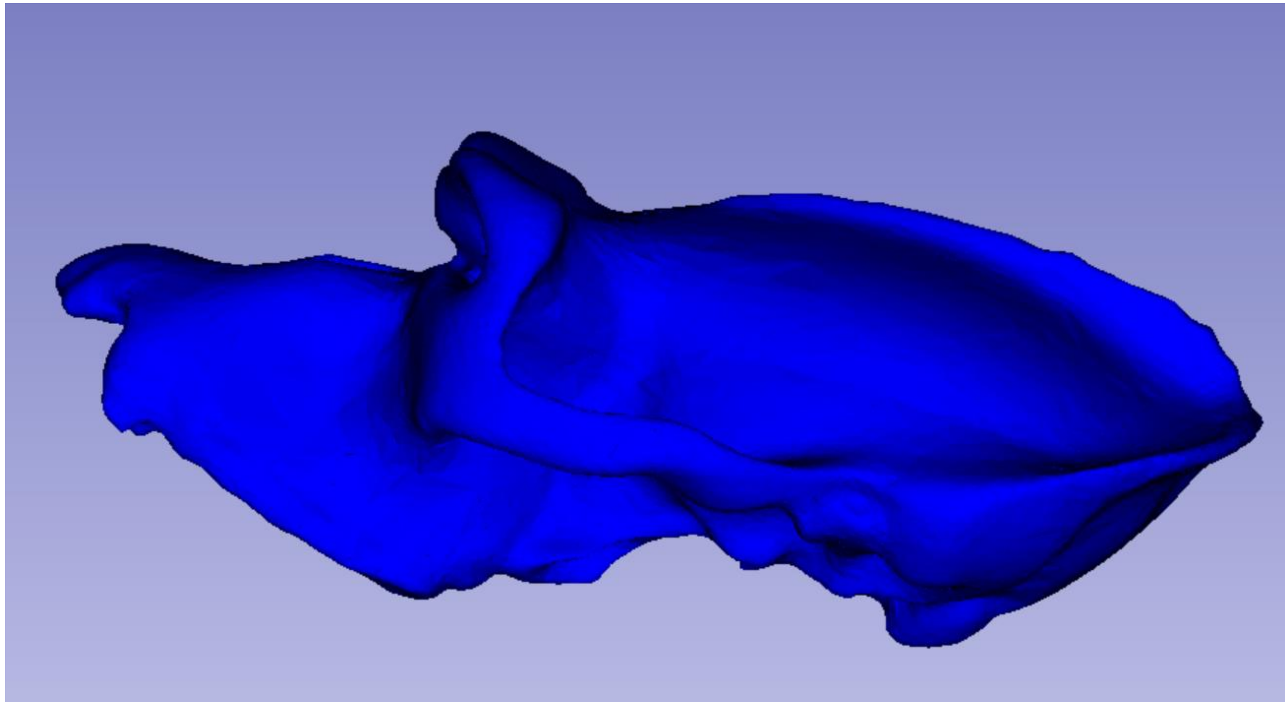
Model = consensus + ∑ (scores * eigenvectors)

Visualizing PCA results









Some other analyses on GM data

- Study of variability (Procrustes anova)
- Study of symmetry
- Study of allometry
- Study of modularity and integration
- Study of covariance through partial-least squares
- Study of phylogenetics

Procrustes ANOVA

Statistical assessment of the terms in the model using Procrustes distances among specimens. **Sum-of-squared Procrustes distances** are used as a measure of SS.

The observed SS are evaluated through permutation. In morphometrics this approach is known as a **Procrustes ANOVA**

Example from geomorph

```
#conduct GPA-alignment
Y.gpa <- gpagen(plethodon$land)
# construct a geomorph data frame
gdf <- geomorph.data.frame(Y.gpa, site = plethodon$site, species =
plethodon$species)
procD.lm(coords ~ species, data = gdf, iter = 999)
```

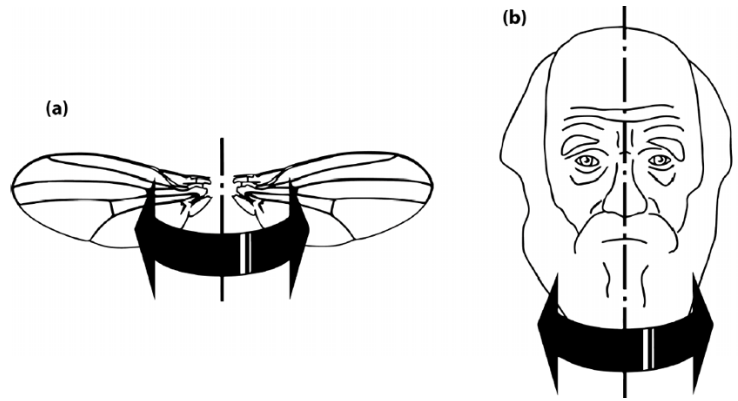
Output of the code

```
Analysis of Variance, using Residual Randomization
Permutation procedure: Randomization of null model residuals
Number of permutations: 1000
Estimation method: Ordinary Least Squares
Sums of Squares and Cross-products: Type I
Effect sizes (Z) based on SS distributions
          Df       SS        MS     Rsq      F      Z Pr(>SS)
species    1 0.029258 0.0292578 0.14856 6.6304 3.1637   0.001 **
Residuals 38 0.167682 0.0044127 0.85144
Total     39 0.196940
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Call: procD.lm(f1 = coords ~ species, iter = 999, data = gdf)
```

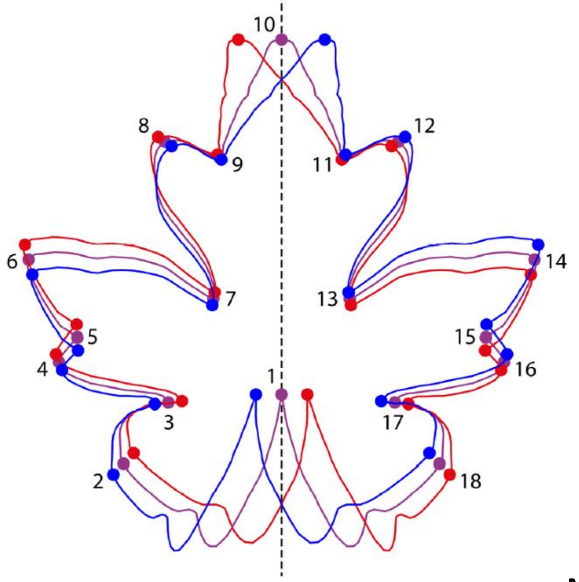
Goodall C. 1991. Procrustes Methods in the Statistical Analysis of Shape. Journal of the Royal Statistical Society Series B (Methodological) 53:285–339.

Analysis of symmetry with geometric morphometrics

Types of symmetry



Matching Object

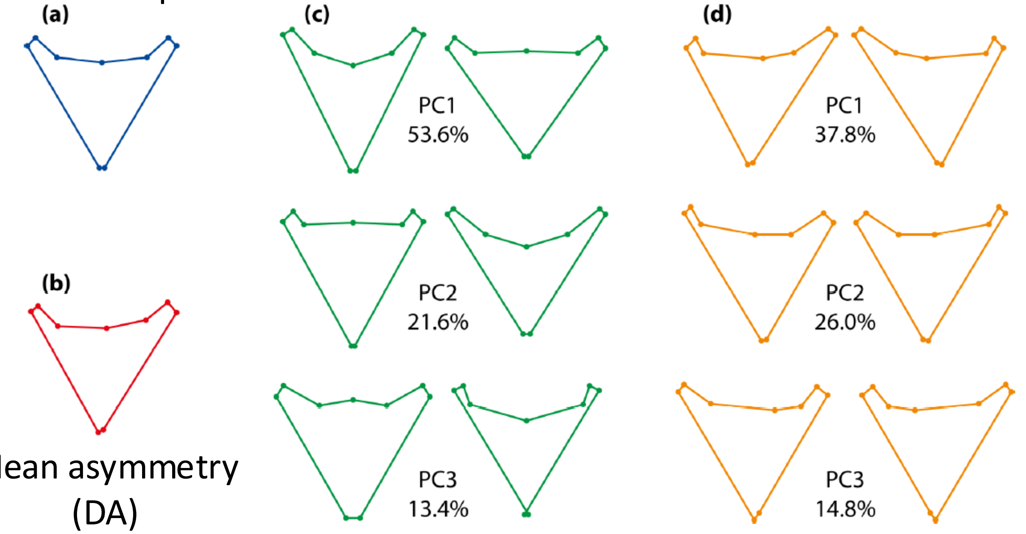


Original configuration
Reflected and relabed
Symmetrized

Symmetrized mean shape

PCA on Symmetric component

PCA on Asymmetric component



Mean asymmetry (DA)

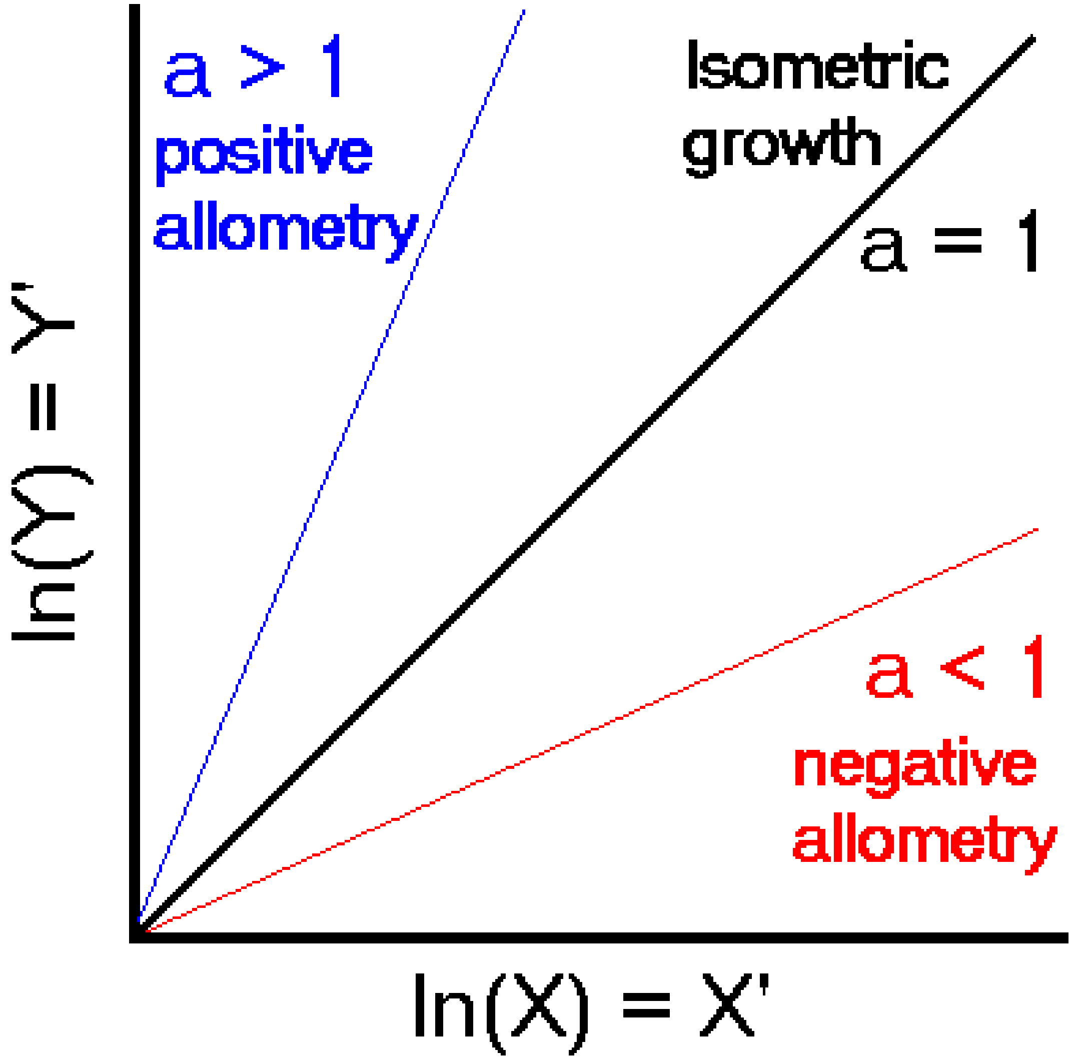
Directional asymmetry (DA) is the mean asymmetry in the population, and therefore can be estimated as the average of individual left–right shape differences over all the individual. Or equivalently, as the difference between the average of all left configurations and the average of all right configurations.

Fluctuating asymmetry (FA) is the variation of individual asymmetries around the average of directional asymmetry, and thus can be computed as each individual's left–right shape difference minus the overall average of the left–right shape differences.

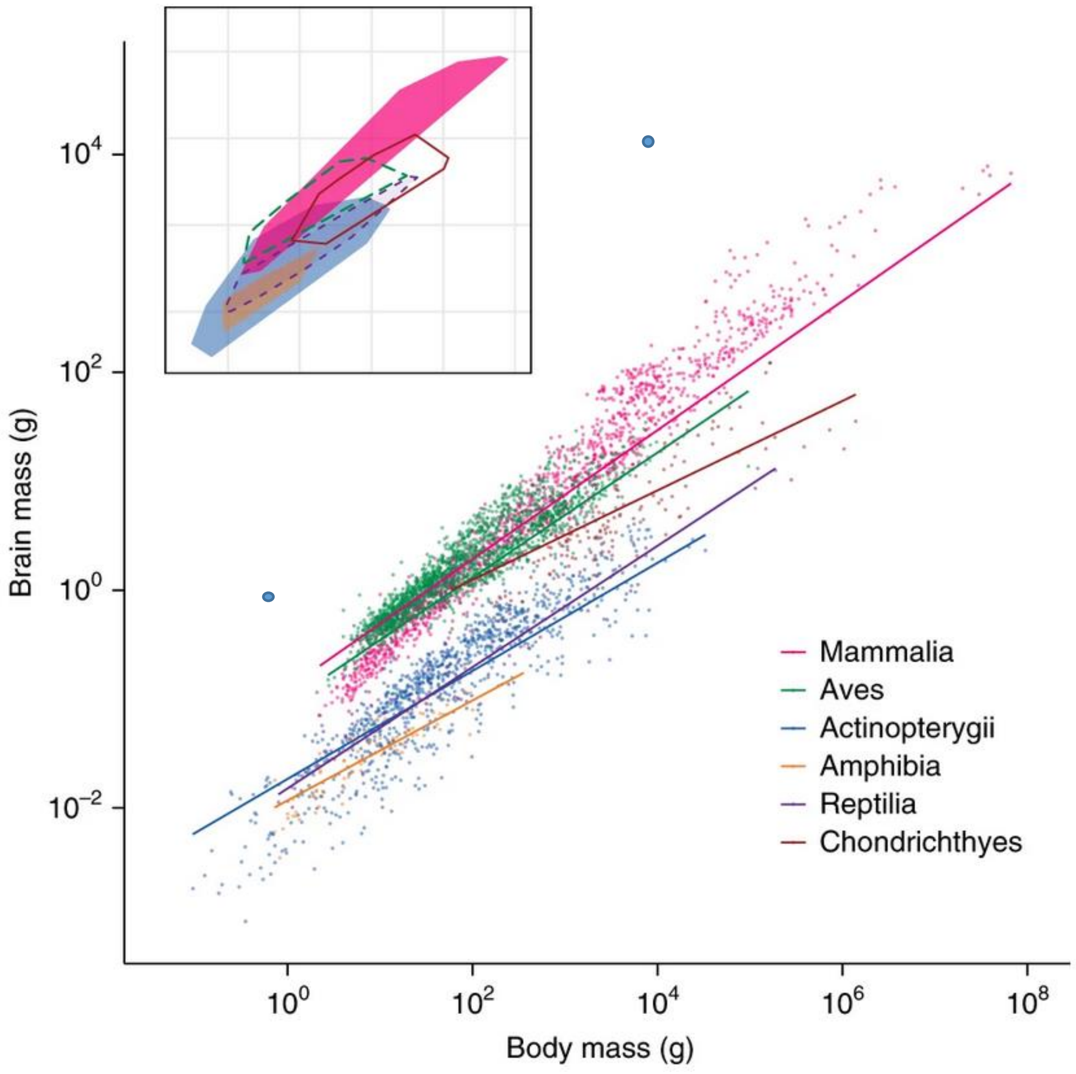
FA Score: Sqrt of sum of an individual's FA^2 (akin to Procrustes distance), is a metric used in studies of asymmetry and developmental instability.

- Klingenberg CP. 2015. Analyzing Fluctuating Asymmetry with Geometric Morphometrics: Concepts, Methods, and Applications. Symmetry 7:843–934.
- Klingenberg CP, Barluenga M, Meyer A. 2002. Shape Analysis of Symmetric Structures: Quantifying Variation Among Individuals and Asymmetry. Evolution 56:1909–1920.

Allometry



a > 1
positive
allometry
Isometric
growth
a = 1
a < 1
negative
allometry
ln(Y) = Y'
ln(X) = X'



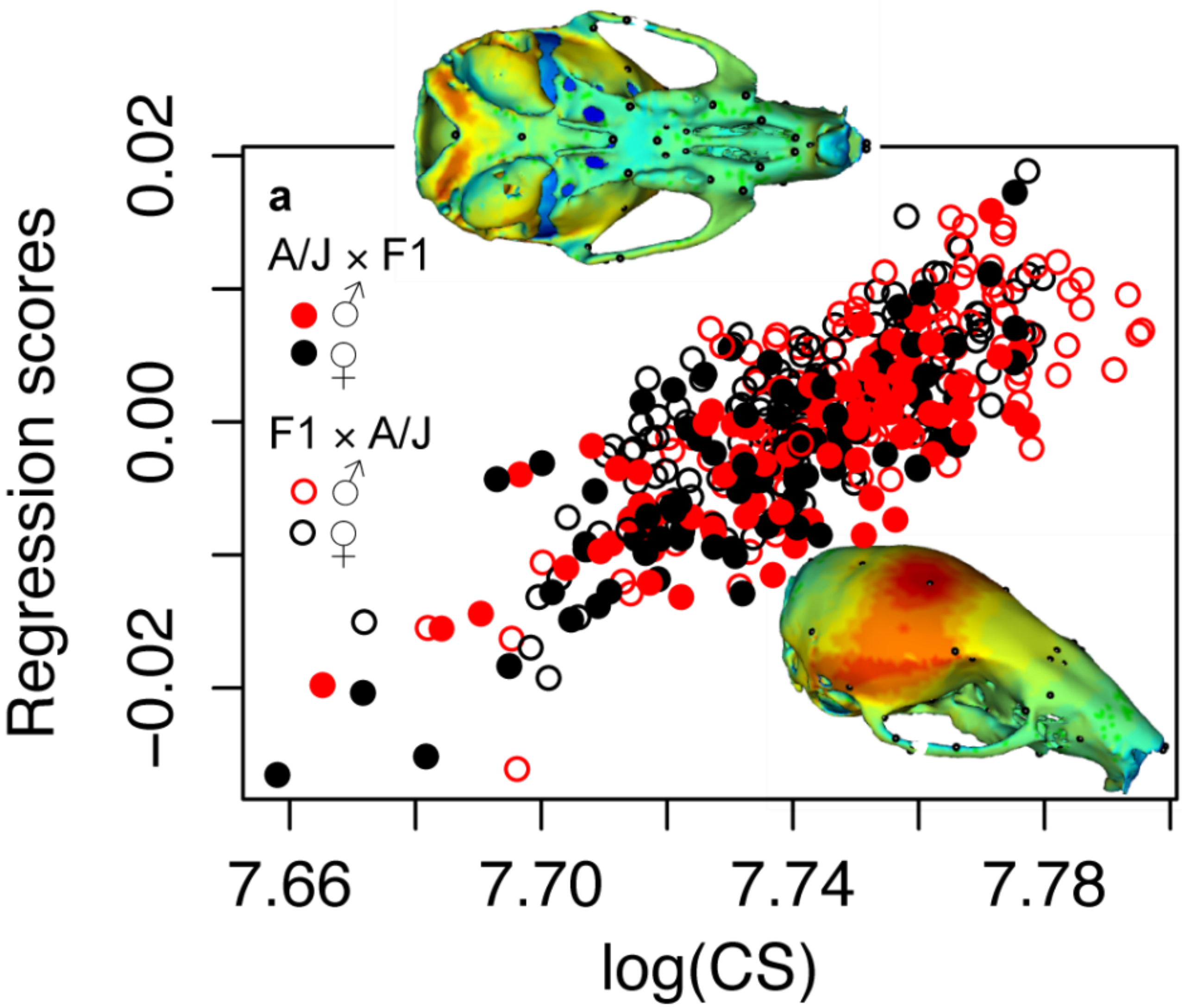
Brain mass (g)
Body mass (g)
10^{4}
10^{2}
10^{0}
10^{-2}
10^{0}
10^{2}
10^{4}
10^{6}
10^{8}
Mammalia
Aves
Actinopterygii
Amphibia
Reptilia
Chondrichthyes

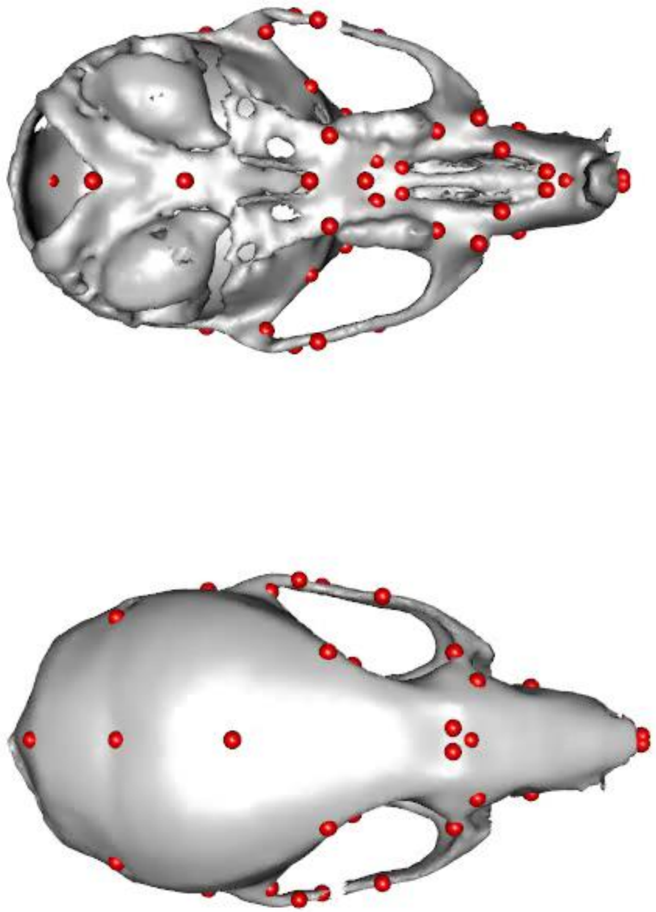
Study of allometry with geometric morphometrics

- For studying allometry with geometric morphometric data, the most straightforward method is to use a multivariate regression of the shape variables onto a measure of size (Monteiro 1999).
- The shape variables are the Procrustes aligned coordinates of the specimens and the shape measure **is centroid size** (or the logarithm of centroid size.)
- Test whether there is allometry by examining whether size and shape are correlated statistically. If so, the patterns of allometry can be characterized as the expected shape change per unit of increase in size.
- Compute regression scores by projecting the data points in shape space onto an axis in the direction of the regression vector (Drake and Klingenberg 2008). This is the shape variable that has the maximal covariation with centroid size and is therefore an optimal summary variable.

- Monteiro LR. 1999. Multivariate regression models and geometric morphometrics: the search for causal factors in the analysis of shape. Syst Biol 48:192–199.
- Drake AG, Klingenberg CP. 2008. The pace of morphological change: historical transformation of skull shape in St Bernard dogs. Proc Biol Sci 275:71–76.
- Klingenberg CP. 2016. Size, shape, and form: concepts of allometry in geometric morphometrics. Dev Genes Evol 226:113–137.

Effect of skull size on the skull shape



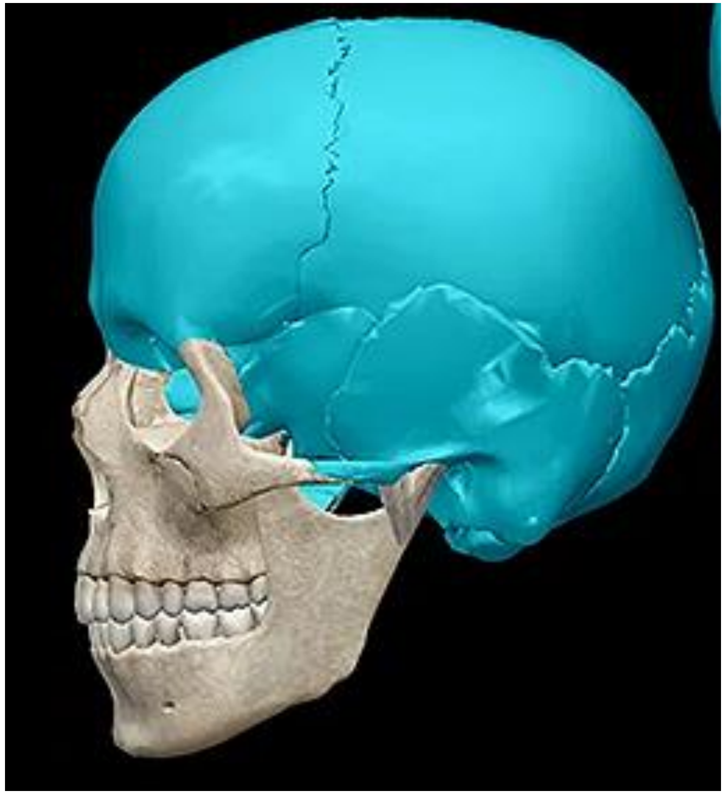


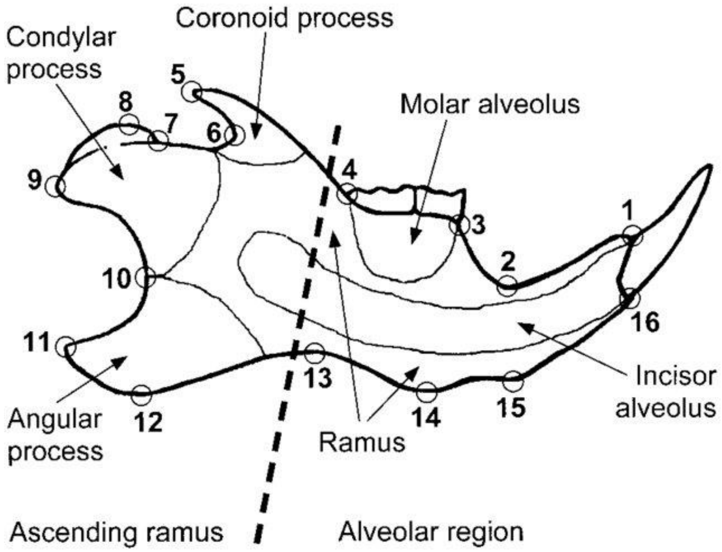
Maga AM, Navarro N, Cunningham ML, Cox TC. 2015. Quantitative trait loci affecting the 3D skull shape and size in mouse and prioritization of candidate genes in-silico. Frontiers in Physiology | Craniofacial Biology 6:92.

Modularity and integration

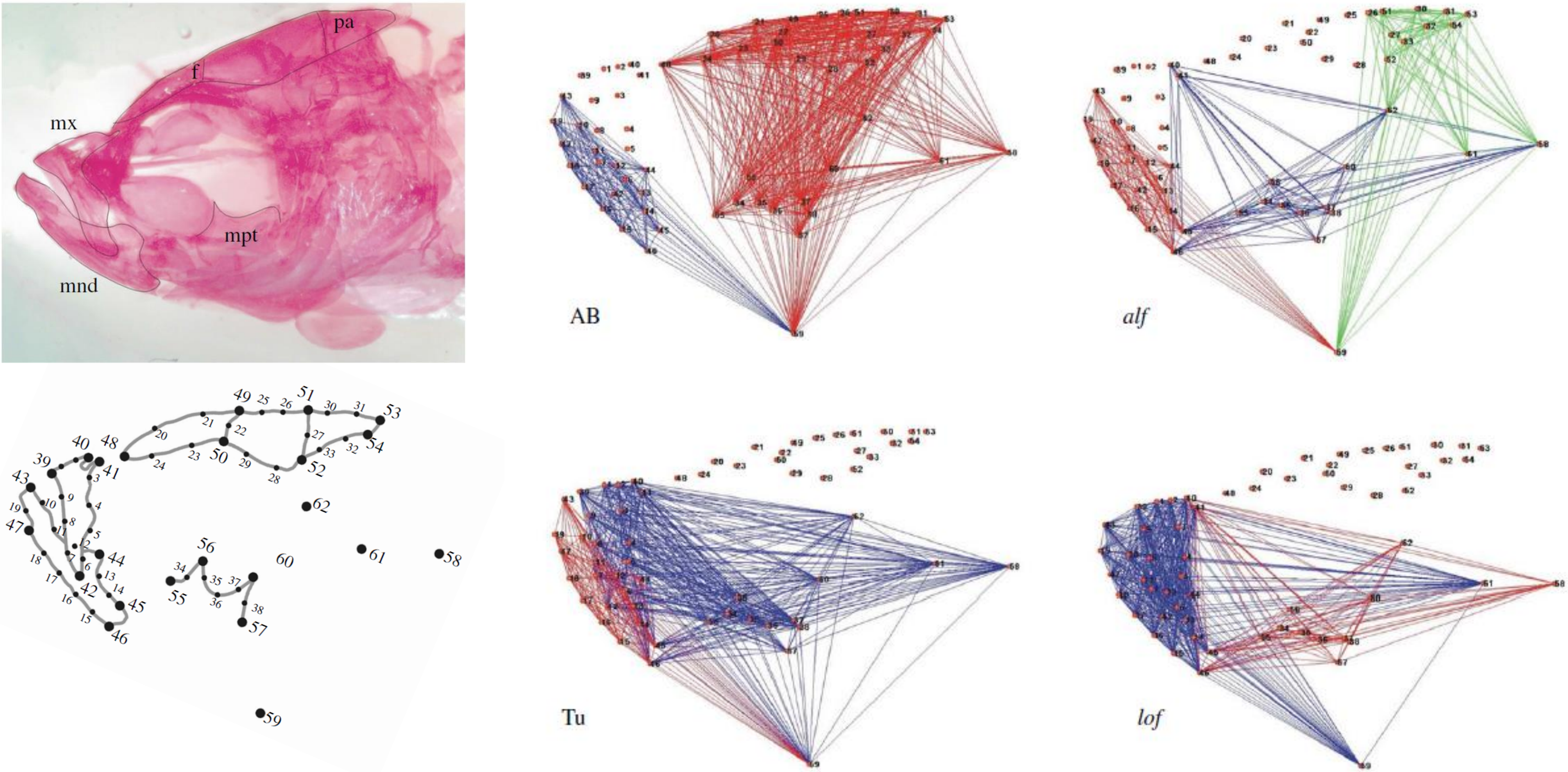
- A system is modular, if it can be divided into multiple sets of strongly interacting parts that are relatively autonomous with respect to each other.

- Typical examples:
  - splanchnocranium (face) vs neurocranium
  - Ascending ramus vs body of mandible





Modularity in zebrafish skull



Parsons Kevin J., Son Young H., Crespel Amelie, Thambithurai Davide, Killen Shaun, Harris Matthew P., and Albertson R. Craig. “Conserved but Flexible Modularity in the Zebrafish Skull: Implications for Craniofacial Evolvability.” *Proceedings of the Royal Society B: Biological Sciences* 285, no. 1877 (April 25, 2018): 20172671. https://doi.org/10.1098/rspb.2017.2671.

What can you do with SlicerMorph GPA module?

- **GPA in SlicerMorph will provide**
  - Procrustes distances
  - Aligned Procrustes coordinates
  - Centroid sizes
  - 2D/3D plot of mean shapes
  - Visualization of individual landmark variances

  (**not for analysis, but for QC!!!**)
  - PCA and PC scores
  - 2D/3D eigenvectors (lollipop plots)
  - Visualization along PCA axes
  - Deformation of a reference shape model along PC axes using TPS
  - 3D animations of PCA results

- **SlicerMorph won't currently do:**
  - Procrustes anova
  - Multivariate regression of shape coordinates on CS
  - Symmetry analysis
  - Modularity/Integration
  - Essentially any sort of statistics based on analysis of coordinate data
  - For these, import the output from GPA into R and do all these analyses in geomorph/Morpho/shapes packages. (Exercise for day 4)

Some key literature on landmark based statistical shape analysis

- Adams DC, Rohlf FJ, Slice DE. 2013. A field comes of age: geometric morphometrics in the 21st century. Hystrix 24:7–14.
- Adams, Dean C. "Evaluating Modularity in Morphometric Data: Challenges with the RV Coefficient and a New Test Measure." *Methods in Ecology and Evolution*, January 1, 2016, n/a-n/a. https://doi.org/10.1111/2041-210X.12511.
- Bookstein FL. 1989. Principal warps: thin-plate splines and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence 11:567–585.
- Bookstein FL. 1997. Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. Medical Image Analysis 1:225–243.
- Drake AG, Klingenberg CP. 2008. The pace of morphological change: historical transformation of skull shape in St Bernard dogs. Proc Biol Sci 275:71–76.
- Klingenberg CP, Barluenga M, Meyer A. 2002. Shape Analysis of Symmetric Structures: Quantifying Variation Among Individuals and Asymmetry. Evolution 56:1909–1920.
- Klingenberg CP. 1998. Heterochrony and allometry: the analysis of evolutionary change in ontogeny. Biological Reviews 73:79–123.
- Klingenberg CP. 2009. Morphometric integration and modularity in configurations of landmarks: tools for evaluating a priori hypotheses. Evol Dev 11:405–421.
- Klingenberg CP. 2010. Evolution and development of shape: integrating quantitative approaches. Nat Rev Genet 11:623–635.
- Klingenberg CP. 2016. Size, shape, and form: concepts of allometry in geometric morphometrics. Dev Genes Evol 226:113–137.
- Lele S, Richtsmeier J. 1991. Euclidean Distance Matrix Analysis - a Coordinate-Free Approach for Comparing Biological Shapes Using Landmark Data. Am J Phys Anthropol 86:415–427.
- Lele S, Richtsmeier JT. 1995. Euclidean distance matrix analysis: confidence intervals for form and growth differences. Am J Phys Anthropol 98:73–86.
- Rohlf FJ, Corti M. 2000. Use of two-block partial least-squares to study covariation in shape. Syst Biol 49:740–753.
- Rohlf FJ, Slice D. 1990. Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks. Systematic Zoology 39:40–59.
- Rohlf FJ. 2000. On the use of shape spaces to compare morphometric methods. Hystrix-Italian Journal of Mammalogy 11:9–25.