

Procrustes and PCA

the details

P. David Polly

Department of Earth and Atmospheric Sciences

Adjunct in Biology and Anthropology

Indiana University

Bloomington, Indiana 47405 USA

pdpolly@indiana.edu



Outline

Behind the scenes of Procrustes analysis

- Translation, scaling, and rotation

Behind the scenes of Principal Components Analysis (PCA)

- PCA as a data rotation
- Covariance matrices as rotation matrices
- Eigenvectors, eigenvalues, and scores

Properties of PCA scores (shape variables)

- Equivalence of PCA scores and Procrustes residuals
- Orthogonality
- Descending order of variance
- “Significance” of PC axes

Curved spaces and tangent spaces

- Why is morphospace curved?
- What is tangent space?
- Does it matter?

Procrustes superimposition

also known as...

Procrustes analysis

Procrustes fitting

Generalized Procrustes Analysis (GPA)

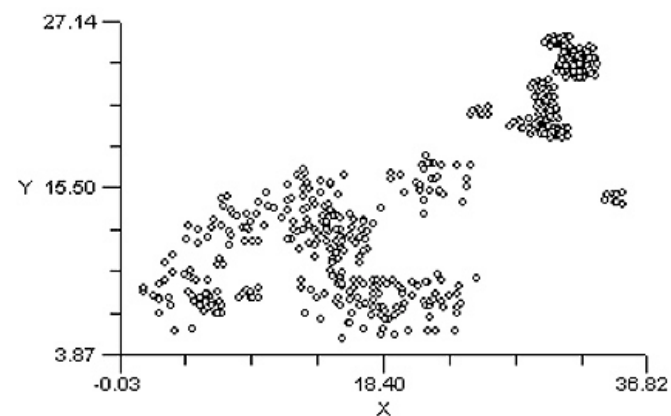
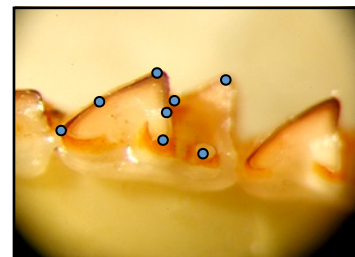
Generalized least squares (GLS)

Least squares fitting

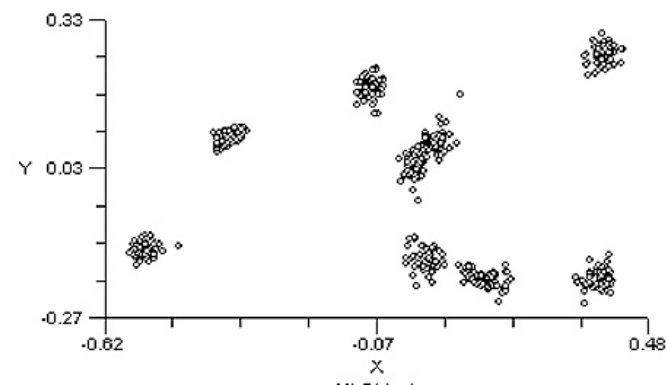
- Centers all shapes at the origin (0,0,0)
- Usually scales all shapes to the same size (usually “unit size” or size = 1.0)
- Rotates each shape around the origin until the sum of squared distances among them is minimized (similar to least-squares fit of a regression line)
- Ensures that the differences in shape are minimized

Procrustes superimposition

- Translates, scales, and rotates landmarks to a common coordinate system
- Removes degrees of freedom in the process
 - $2 + 1 + 1 = 4$ for 2D data
 - $3 + 1 + 3 = 7$ for 3D data
- Creates statistical challenges
 - collinearity
 - “singular” covariance matrix
 - “curved” shape space



Before

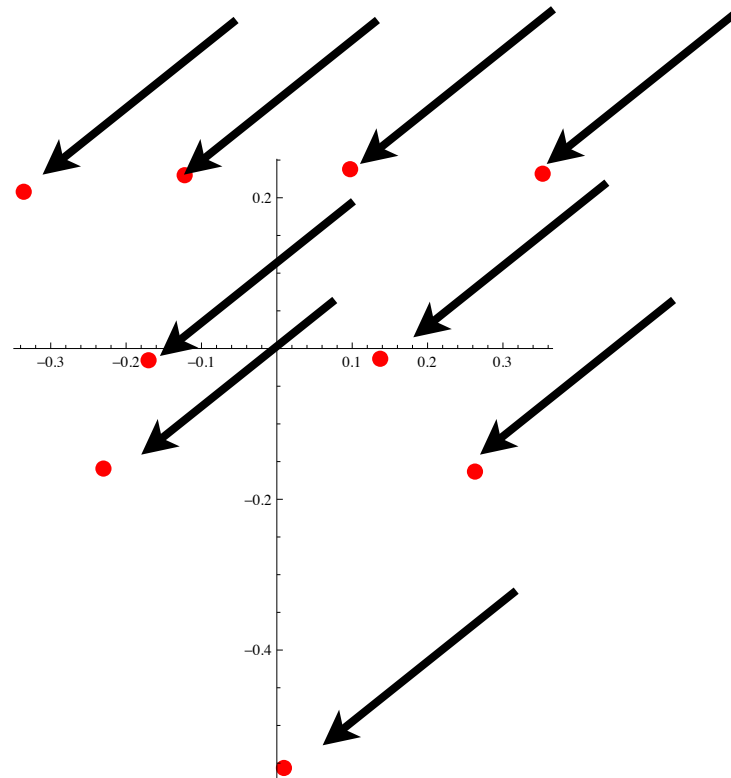


After

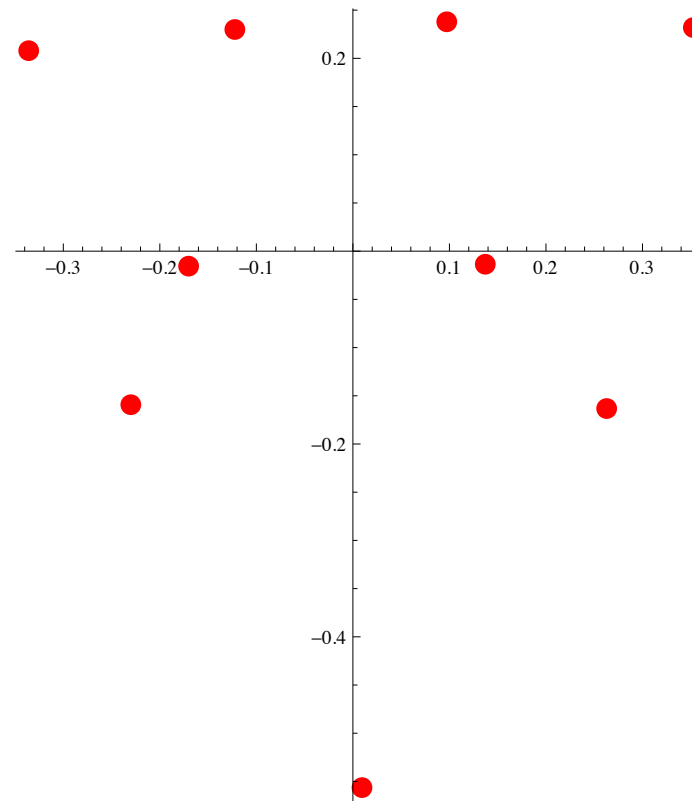
Procrustes: Translation, Rotation, Rescaling

Thought problem...

How do you translate one of your original landmark shapes to the point {0,0} on a graph?



How do you scale landmarks that have been centered?



How do you rotate landmarks?

rotated shape = landmarks.RotationMatrix

$$R = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

theta is the angle in radians
(2 pi radians in 360 degrees)

The true goal of all morphometric analyses...

...to measure morphological similarity and difference

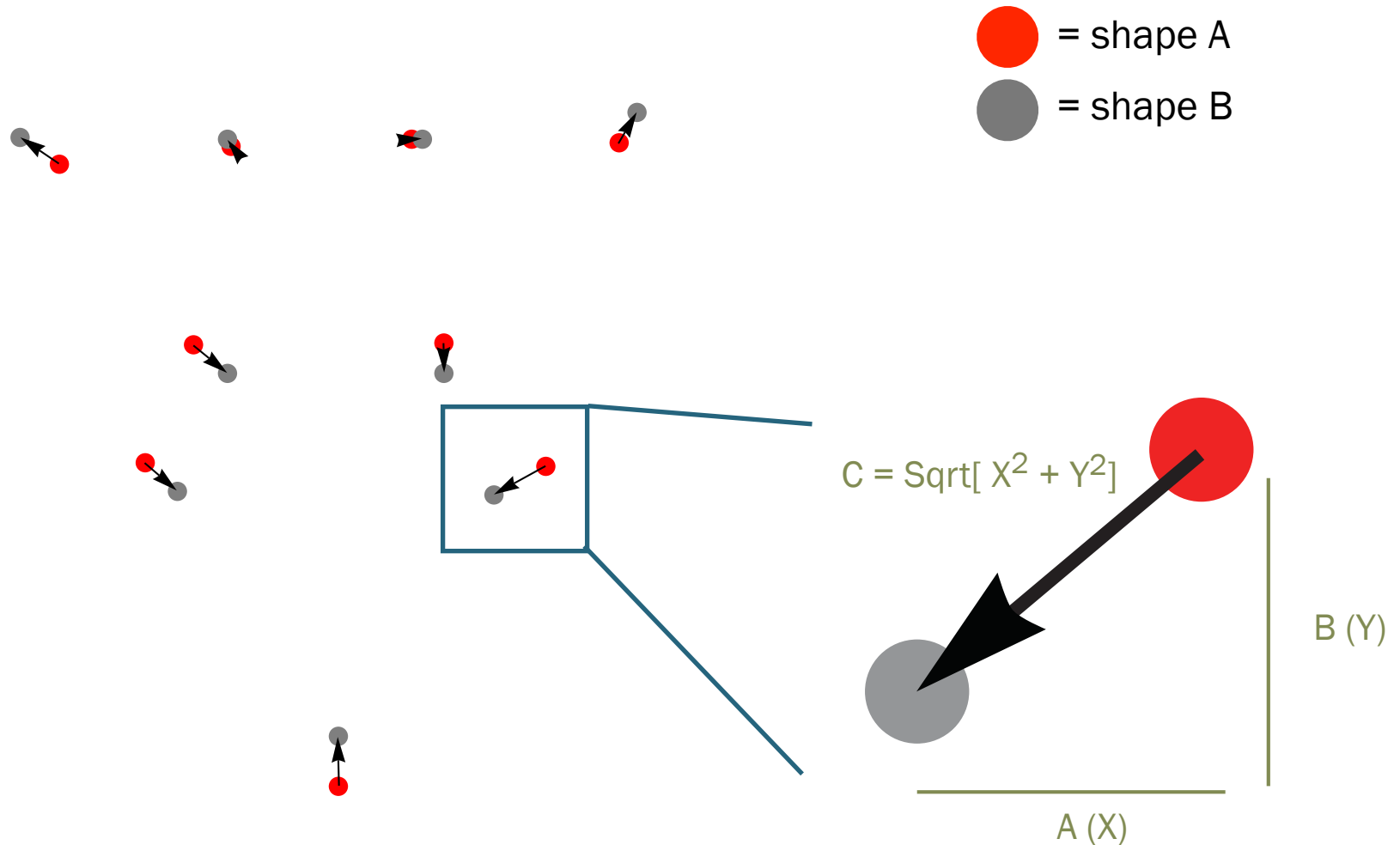
Morphometric distances are the main measure of difference

Measured as the difference between objects (which might be specimens or means of species, or whatever) on all the variables being used

In GMM, the main measure of difference is the Procrustes distance, the distance between shapes after they have been superimposed

Procrustes distance

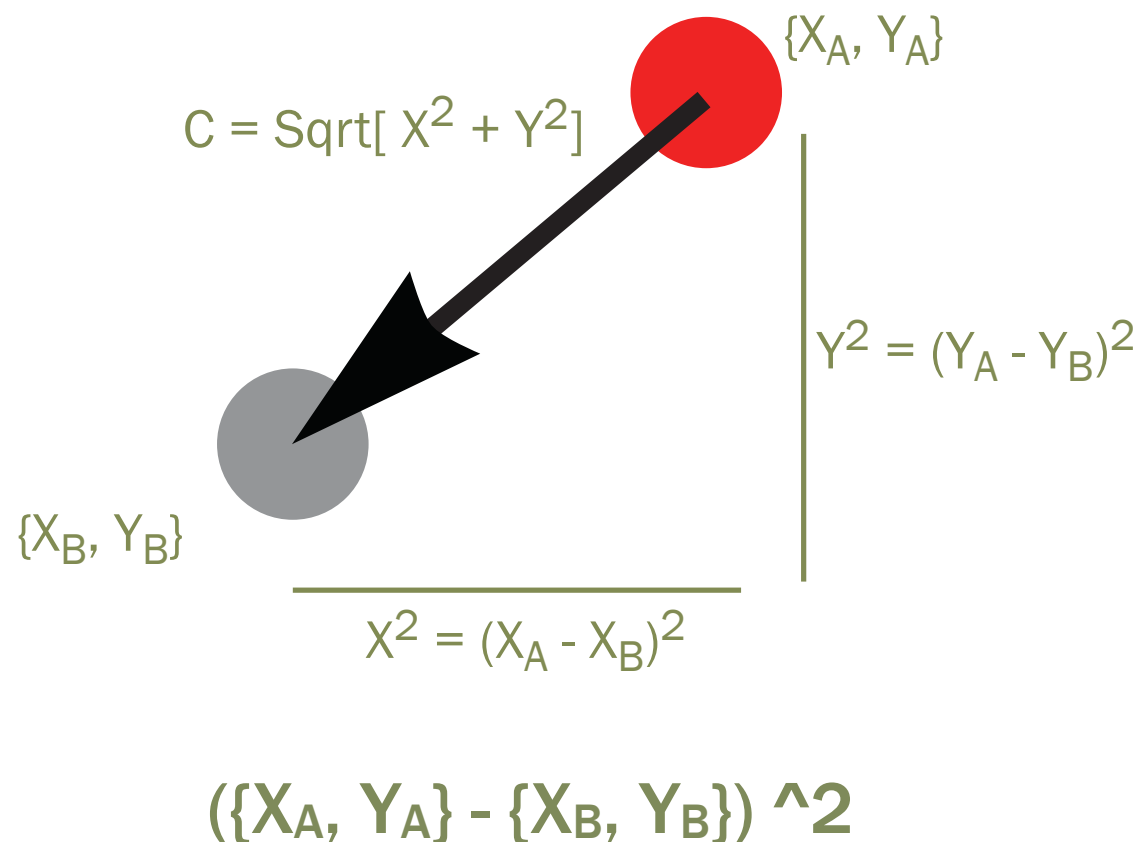
Sum of the distances between corresponding landmarks of two shapes.



Procrustes distance in Mathematica

How would you calculate the Procrustes distance between the mean shape (consensus) and any one of the objects?

1. How to calculate X^2 ? Y^2 ?



Warning:

Procrustes analysis tells you about variation of the entire shape in your sample.

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

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

Sometimes it is interesting to know whether part of a structure is more variable than another part.

Procrustes analysis is not the way to answer such a question...

416

P. DAVID POLLY

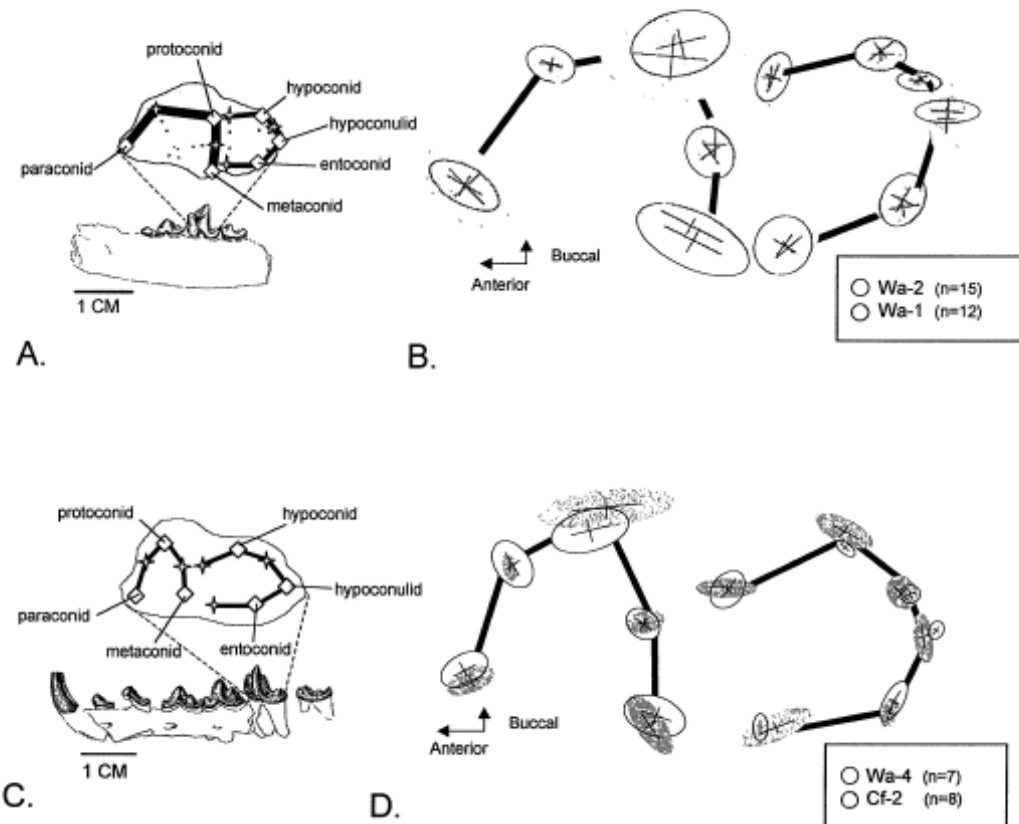
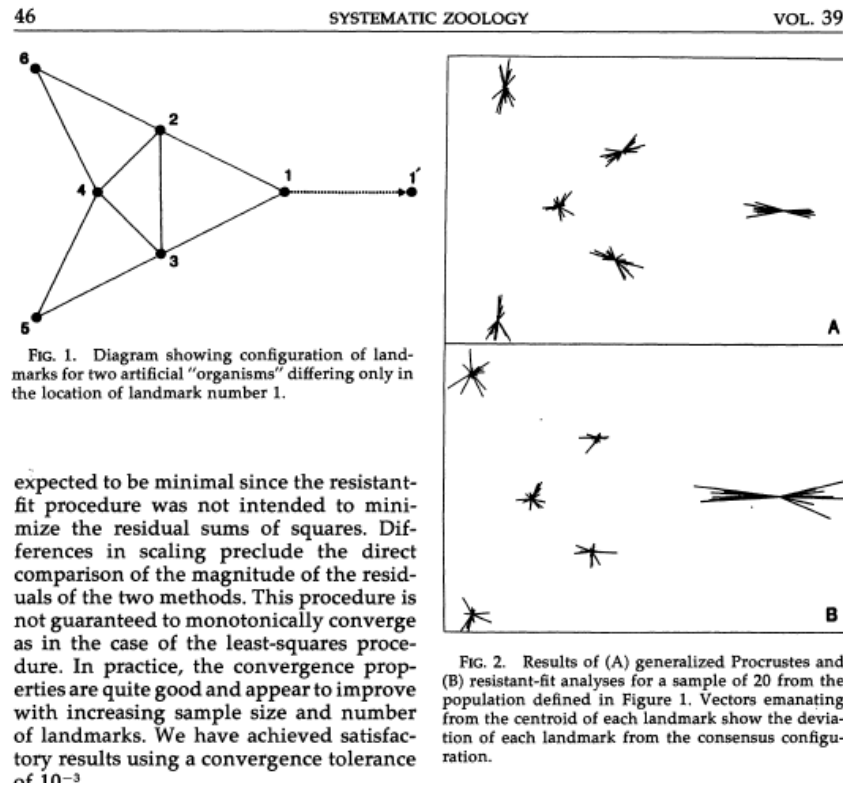


FIGURE 5. Variation and shape change in M_1 . A, Lingual view of a *Viverravus* lower jaw fragment and an enlarged occlusal view inset showing the M_1 landmarks used in this study. These include six cusps (labeled and represented by diamonds) and five intercuspal notches (unlabeled and represented by stars). B, Variation and shape change in the *Viverravus* cusp landmarks. The heavy lines connect landmarks as in A. The ellipses represent the variation present in each landmark (Wa-1, gray; Wa-2, colorless). Major axes cross at the landmark centroid in each ellipse. The difference in centroid position represents evolutionary change between the two samples. C, *Didymictis* M_1 landmarks used in this study. Landmarks are the same as in *Viverravus*. D, Variation and shape change in *Didymictis* M_1 landmarks.

Polly, P.D. 1998. Variability, selection, and constraints: development and evolution in viverravid (Carnivora, Mammalia) molar morphology. *Paleobiology*, 24: 409-429.

The 'Pinocchio Effect' is an example:
variance that occurs in only one landmark is redistributed to
other landmarks during superimposition

Simulated shapes where
only one landmark differs
(like Pinocchio's nose)

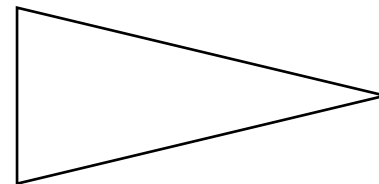
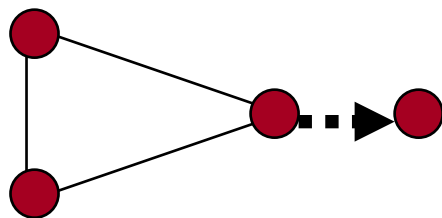


Procrustes
superposition
redistributes variation
to all landmarks

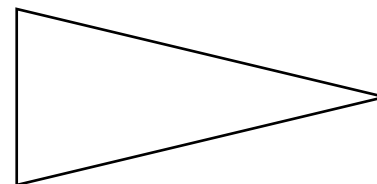
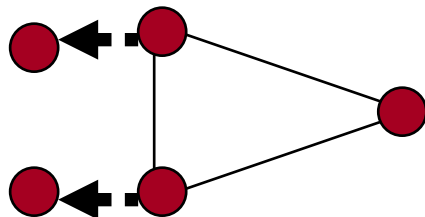
Other methods, like
GRF, concentrate
variation in subsets of
landmarks

GMM provides information about the relative displacements of all landmarks

Moving one point to elongate a triangle...



Is geometrically equal to moving the other two points in the opposite direction...



GMM cannot distinguish between the two. Think carefully when you make biological or evolutionary interpretations.

Variants on 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)

- allows some landmarks to be more variable than others (removes the Pinocchio effect)
- does not find smallest possible difference between shapes

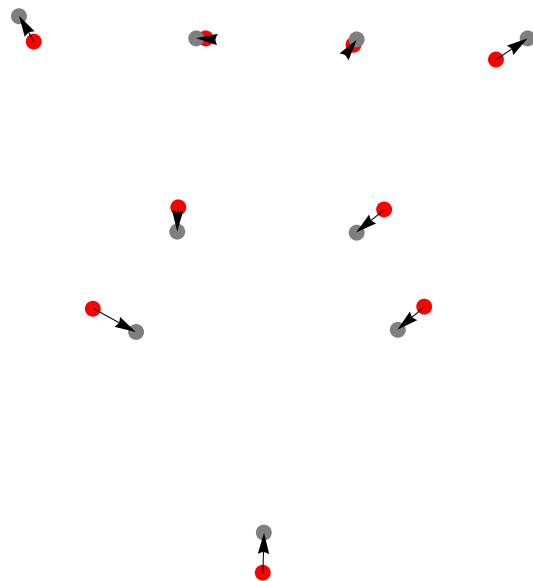
Bookstein Shape Coordinates

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

Euclidean Distance Matrix Analysis (EDMA)

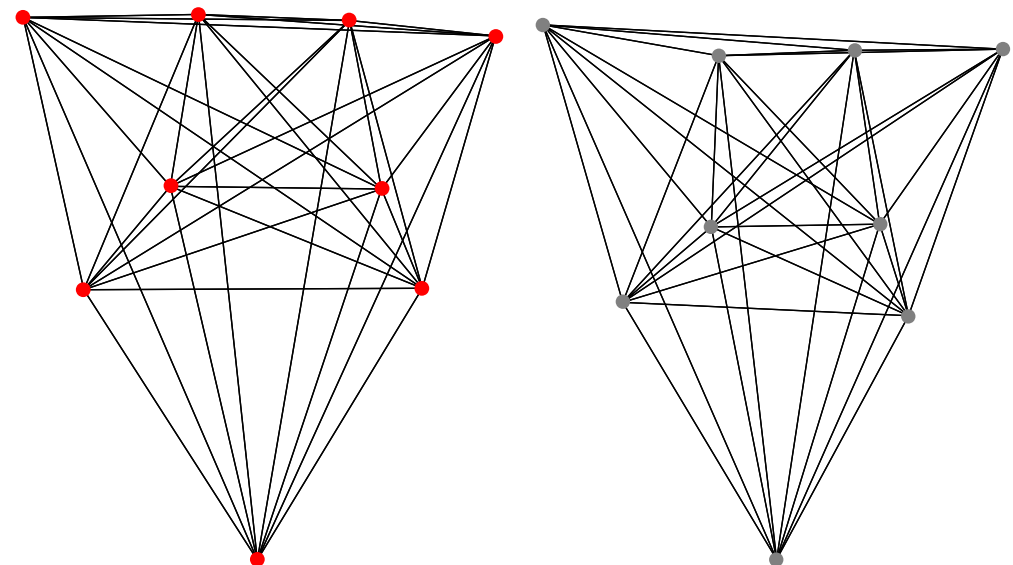
- Method for localizing variation in particular landmarks
- Based on matrix of distances between landmarks

Procrustes approach:
summed distance between shapes



faces 1 & 2

EDMA approach:
matrix of distances between landmarks
compared between shapes



face 1

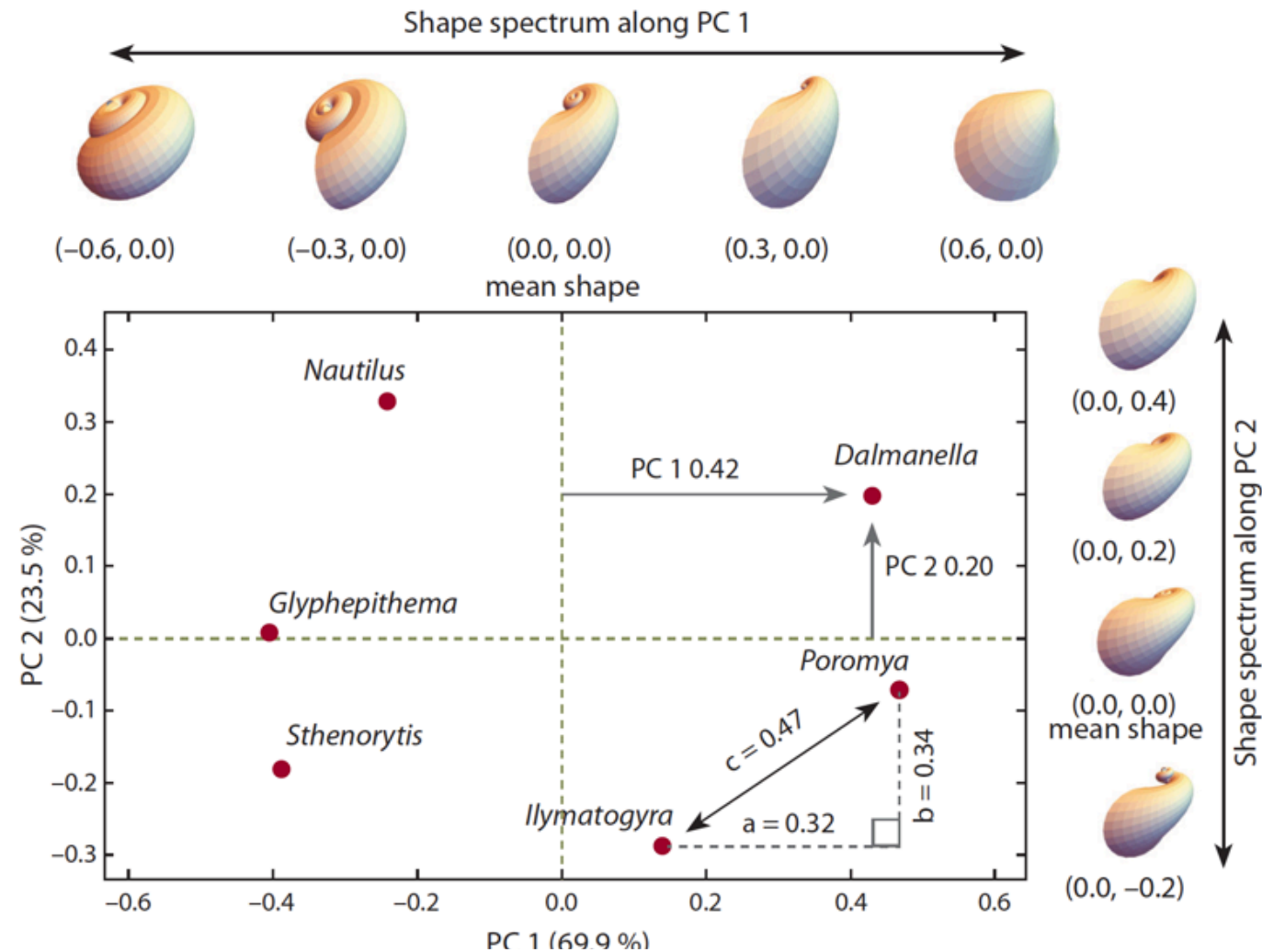
face 2

Intro to EDMA literature: Richtsmeier, J.T., Burke Deleon, V. and Lele, S.R., 2002. The promise of geometric morphometrics. *American Journal of Physical Anthropology*, 119: 63-91.

What does PCA do?

1. Rotates data to its major axes for better 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

Anatomy of morphospace



Important points: the “meaning” of PCA

1. Principal components analysis finds the axes of greatest variation in a data set
2. PCA removes correlations from the data
3. Principal components scores are “shape variables” that are the basis for further analysis
4. But PCA is nothing more than a rotation of the data!

Behind the scenes in PCA of landmarks

Procrustes

This aligns shapes and minimizes differences between them to ensure that only real shape differences are measured.

1. **Subtract mean (consensus) from each shape to produce “residuals”**
This centers the PC axes on the mean (consensus) shape.
2. **Calculate covariance matrix of residuals**
Estimates variance and covariance among the original variables
3. **Calculate eigenvalues and eigenvectors of covariance matrix**
Finds the major axes of the data and the variation along them.
4. **Multiply residuals times eigenvectors to produce scores**
Rotates the original data onto the major axes and gives the coordinates for their new position.

Output of PCA

Eigenvalues

variance on each PC axis

(In Mathematica: *Eigenvalues[CM]*)

Eigenvectors

loading of each original variable on each PC axis

(In Mathematica: *Eigenvectors[CM]*)

Scores (=shape variables)

location of each data point on each PC axis

(In Mathematica: *PrincipalComponents[resids]*)

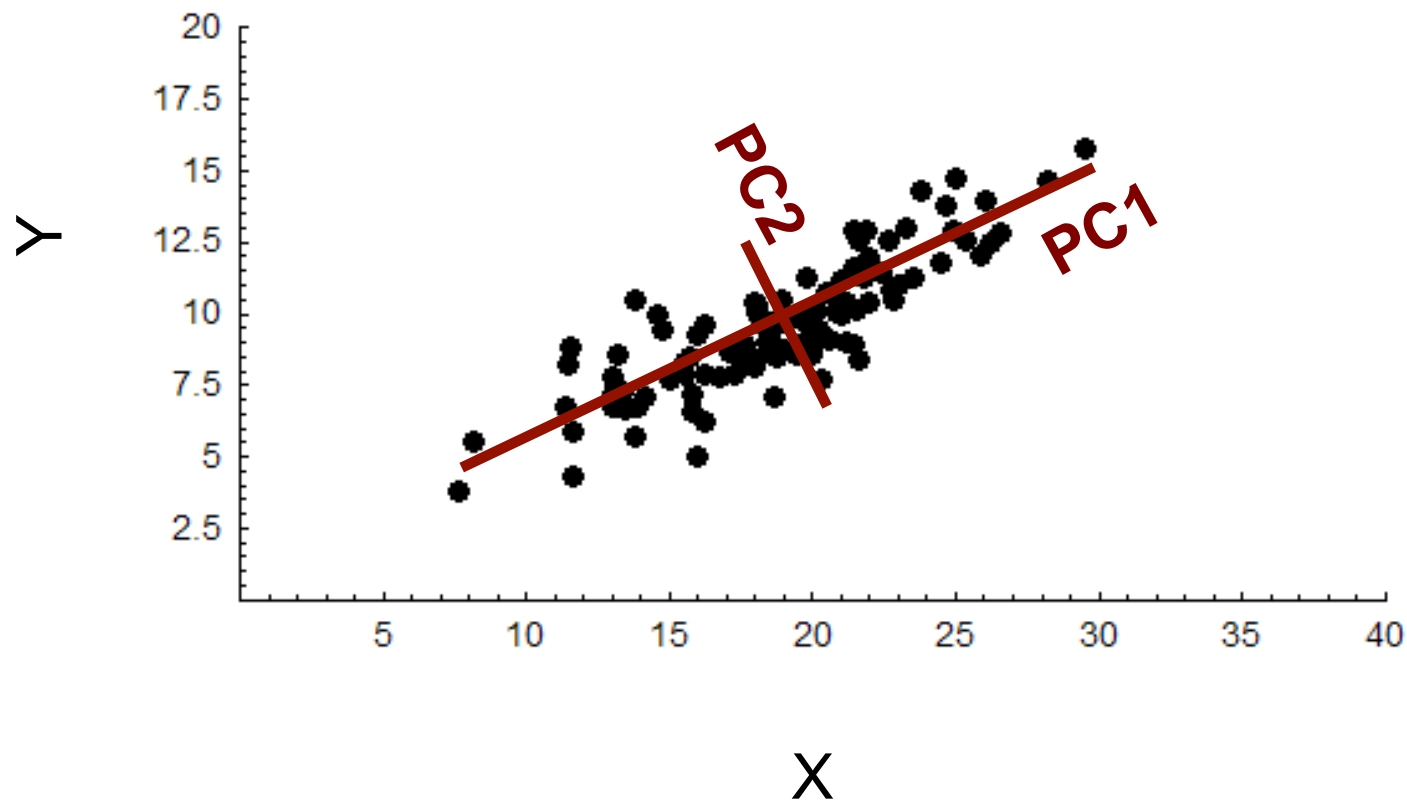
resids are the residuals of the Procrustes coordinates

CM is the covariance matrix of the residuals

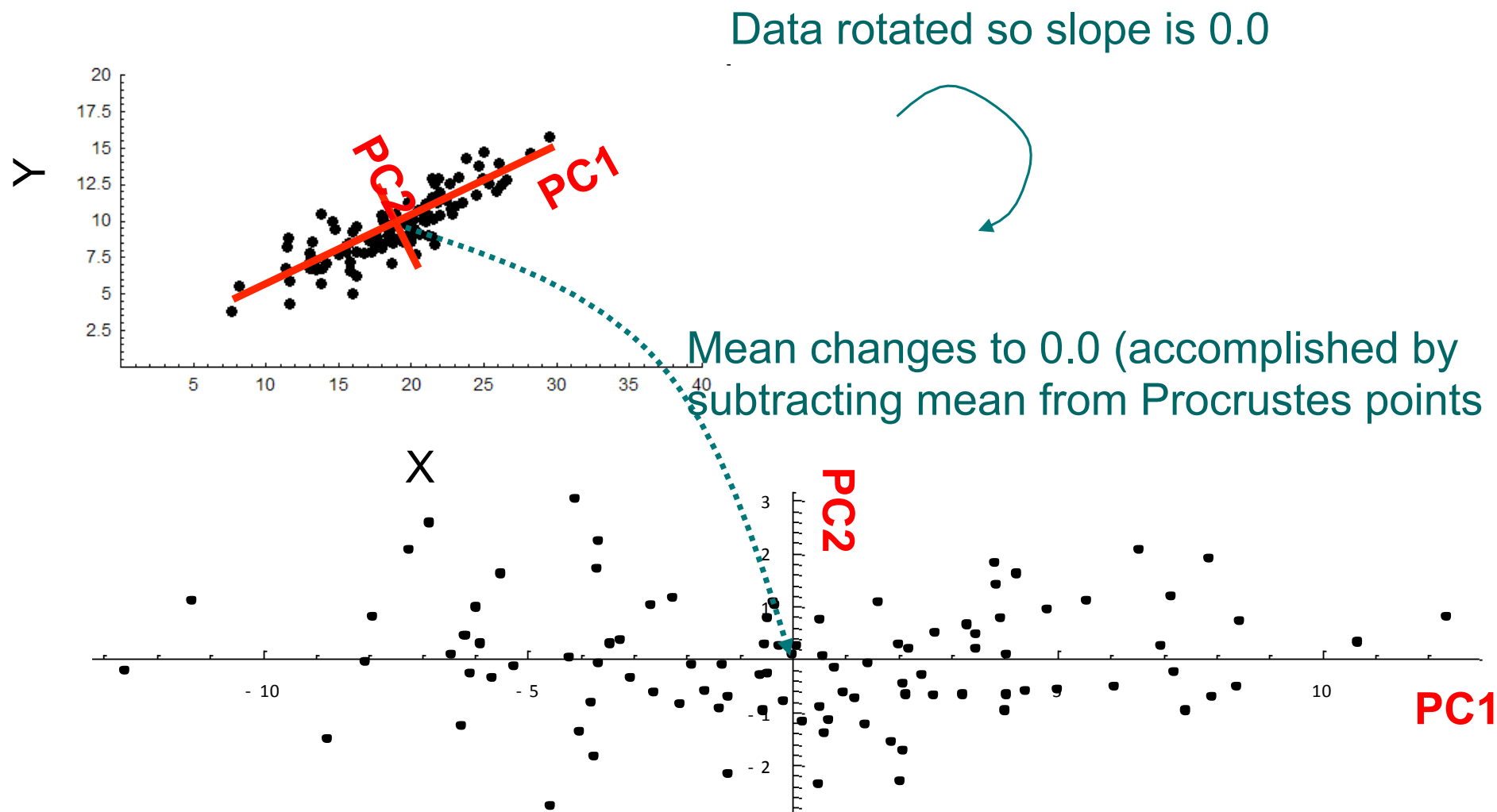
Use this method to get eigenvalues and eigenvectors

{vects, vals, z} = SingularValueDecomposition[CM]

The principal components (PCs) of a data set are its major axes



Principal components are a 'rigid rotation' of the original data

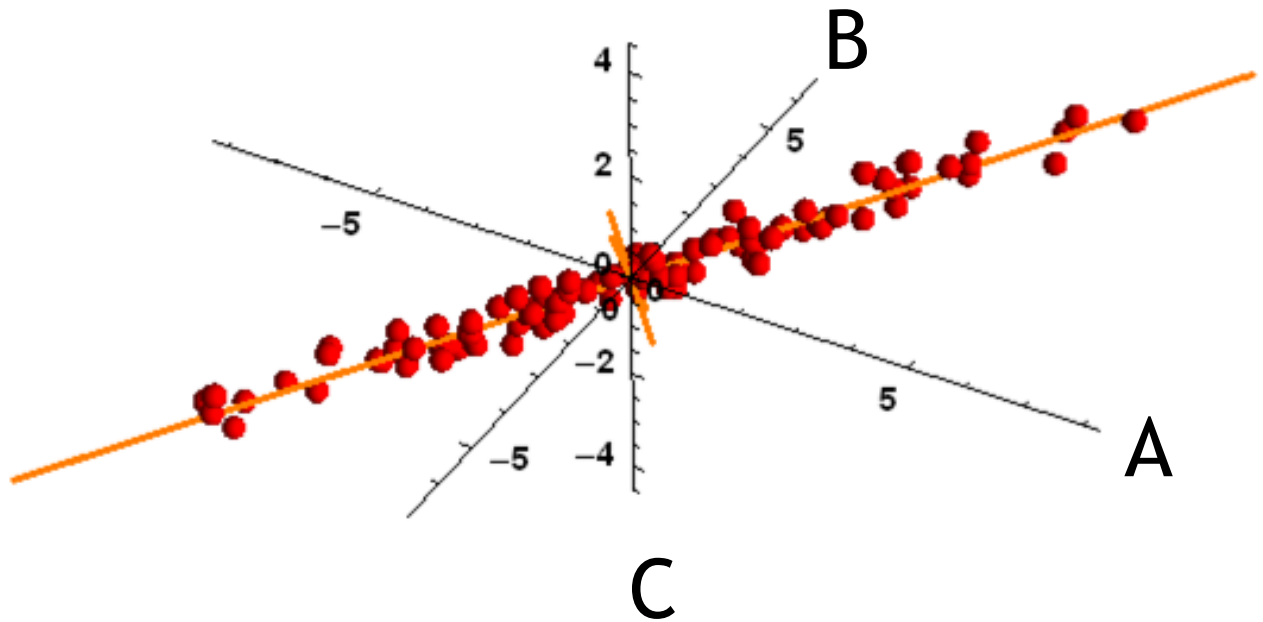


Note that variance increases along horizontal axis, but decreases along vertical axis.

PCA is based on the covariance matrix

Diagonal elements are variances, off-diagonal are covariances (slopes)

	A	B	C
A	6.56	4.69	2.59
B	4.69	4.21	1.38
C	2.59	1.38	1.36



Eigenvalues

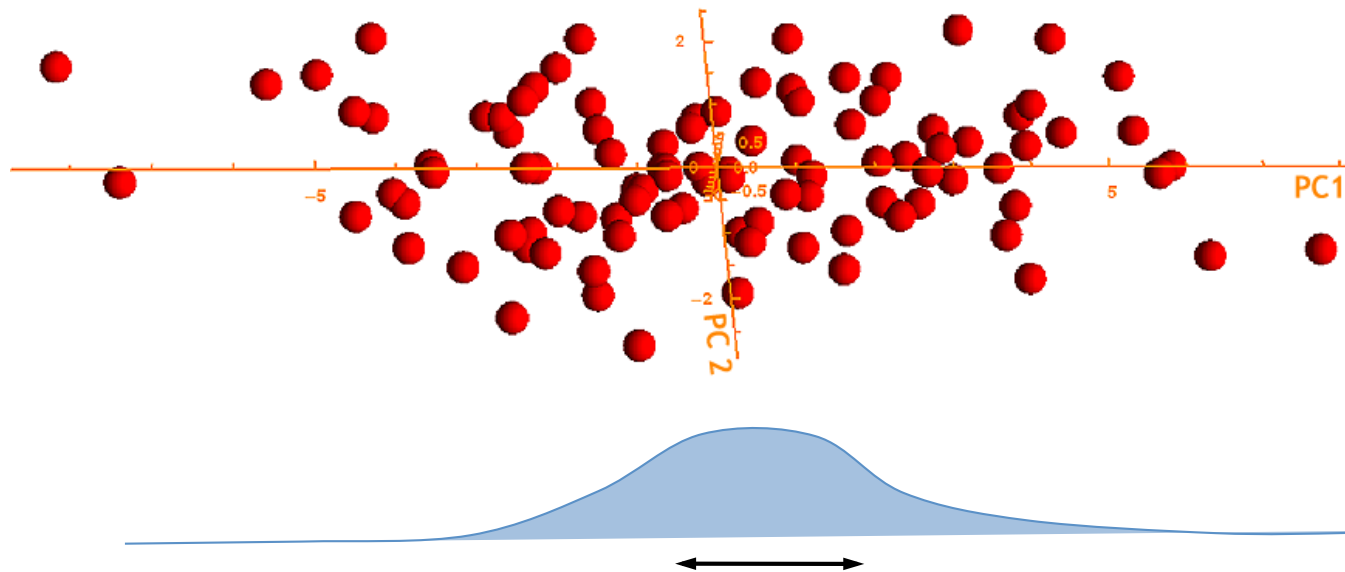
Variance of data along each PC axis

PC 1 = 11.08

PC 2 = 1.01

PC 3 = 0.04

	PC1	PC2	PC3
PC1	11.08	0	0
PC2	0	1.01	0
PC3	0	0	0.04



Important point: the meaning of eigenvalues

Between 95% and 99% of data lie within 2.0 SDs of the mean

1. If you know the variance, you know the standard deviation is its square-root;
2. You know that nearly all the data have a range of $4 * SD$;
3. If the mean is 0.0, then nearly all the data lie between $-2 * SD$ and $+2 * SD$;
4. The eigenvalues (or singular values) of a PC are variances, therefore the range of data on that PC can be calculated from them.

Important point: the meaning of eigenvalues (cont.)

Total variance of morphometric data set is the total amount of shape variation, which can be calculated three ways:

1. Summing squared distances between landmark points and the consensus (sample mean) for all the objects and dividing by $(n-1)$;
2. Summing the eigenvalues that are returned by the PCA;
3. Summing squared PC scores (have a mean of zero so no subtraction is required) and dividing by $(n-1)$;

If these three calculations don't give the same number, something is wrong

Useful variants on Eigenvalues

Eigenvalues

PC 1 = 11.08

PC 2 = 1.01

PC 3 = 0.04

12.13

Percent explained

$100 * 11.08 / 12.13 = 91.3$

$100 * 1.01 / 12.13 = 8.3$

$100 * 0.04 / 12.13 = 0.3$

100.0

Standard Deviation

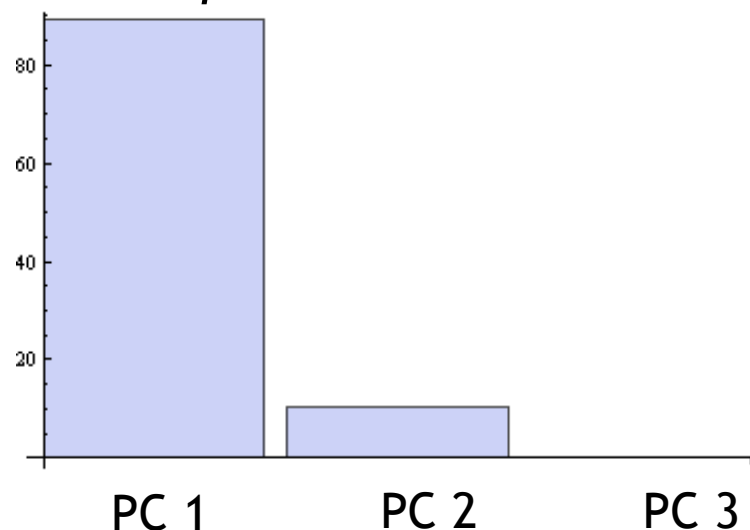
(procGPA reports this)

$11.08^{0.5} = 3.33$

$1.01^{0.5} = 1.00$

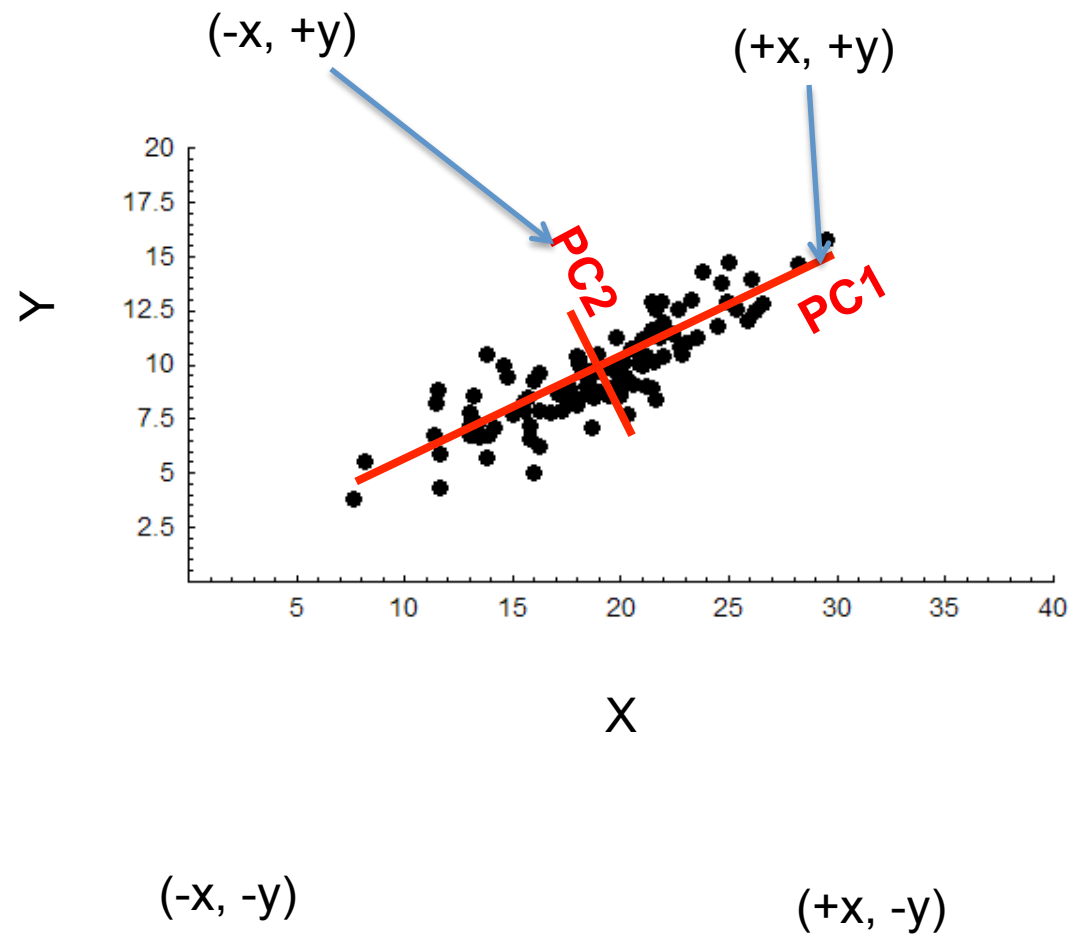
$0.04^{0.5} = 0.20$

Scree plot



BarChart[Eigenvalues[CM]

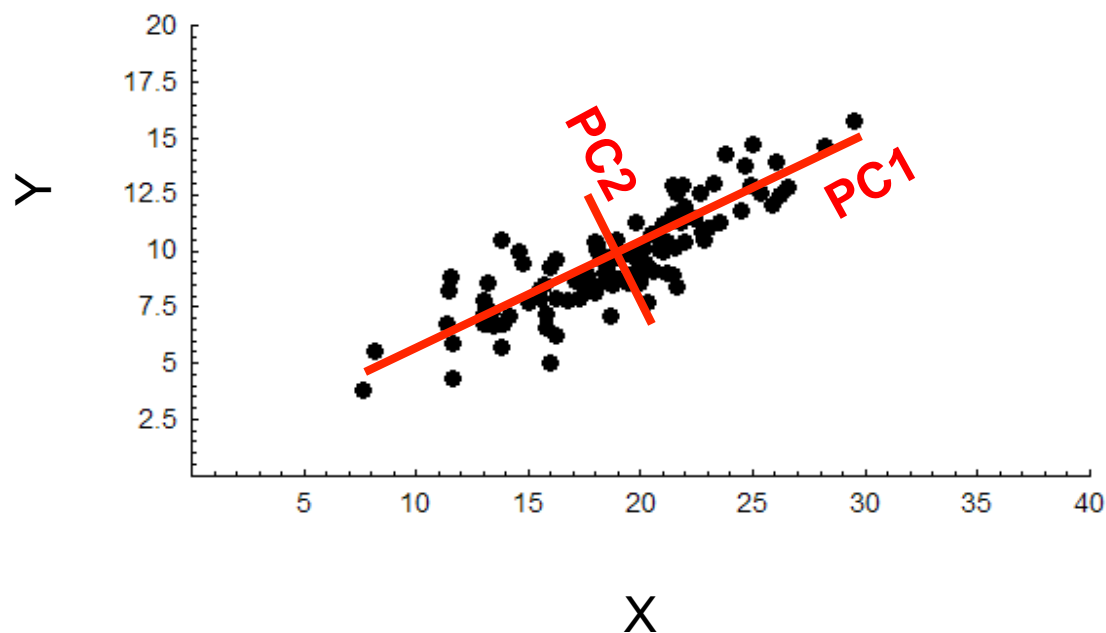
Eigenvector 'loadings' tell how each original variable contributes to the PC



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

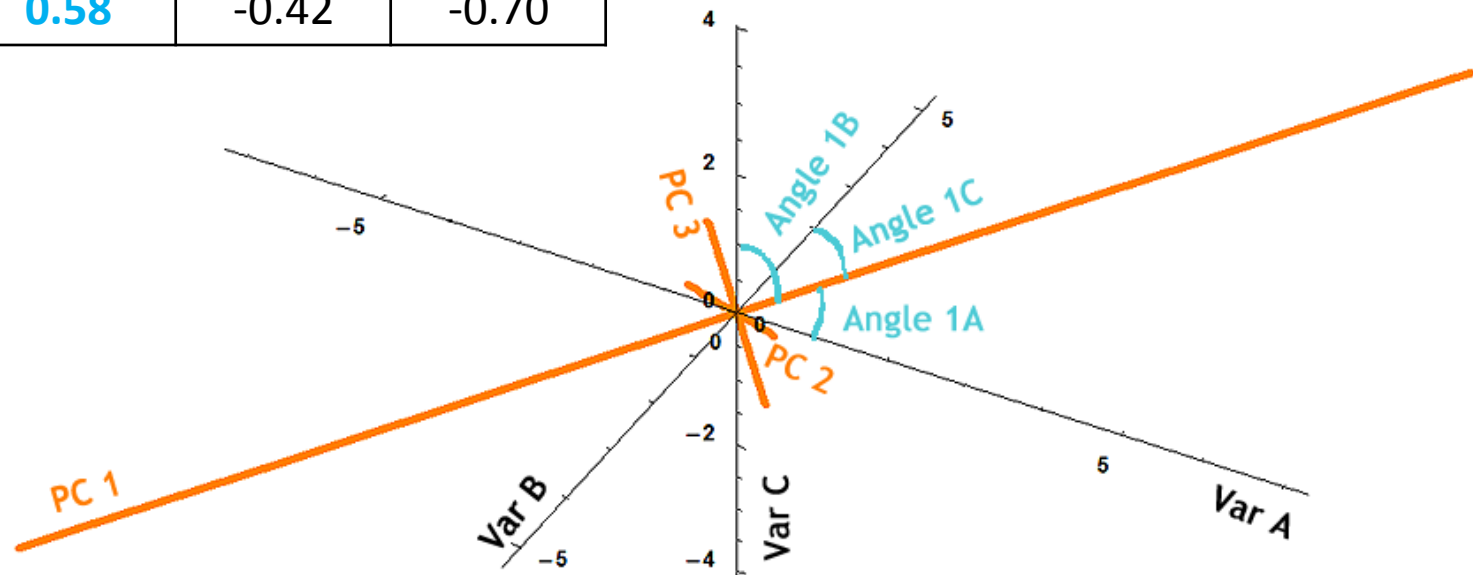
(multiply PC1 X score by 0.89 and PC1 Y score by -0.44 and add back X, Y meant to get real X, Y)

Eigenvectors: definition 1

Angles between PC and original variables

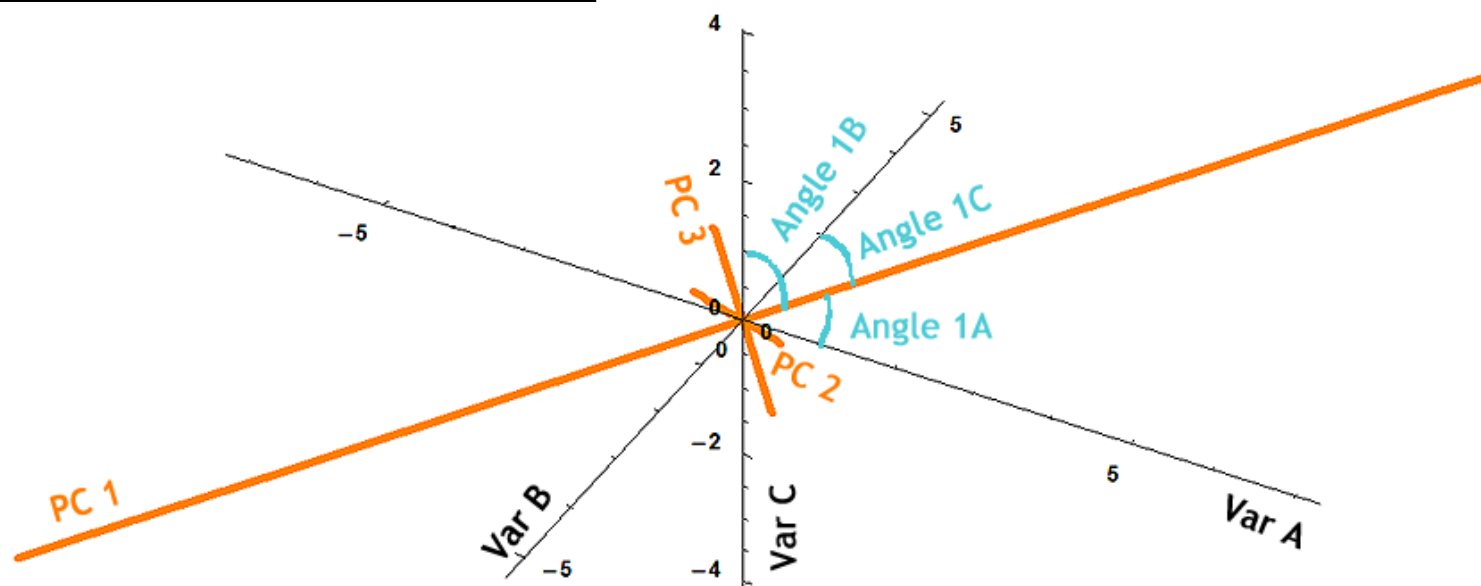
(the eigen vector matrix is a rotation matrix in radians)

	PC1	PC2	PC3
Var A	-0.76	-0.58	-0.29
Var B	0.28	-0.69	0.66
Var C	0.58	-0.42	-0.70



Same Eigenvectors converted from radians to degrees

	PC1	PC2	PC3
Var A	-43.8	-33.1	-16.4
Var B	16.2	-39.9	37.7
Var C	33.2	-24.2	-39.9

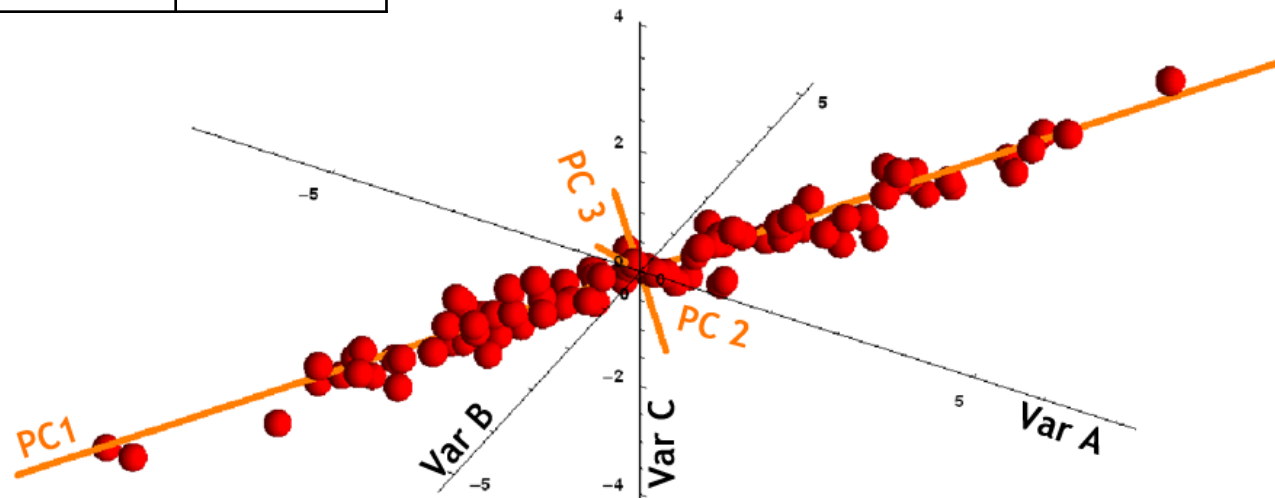


Eigenvectors: definition 2

Loading (or importance) of each variable to the PC.

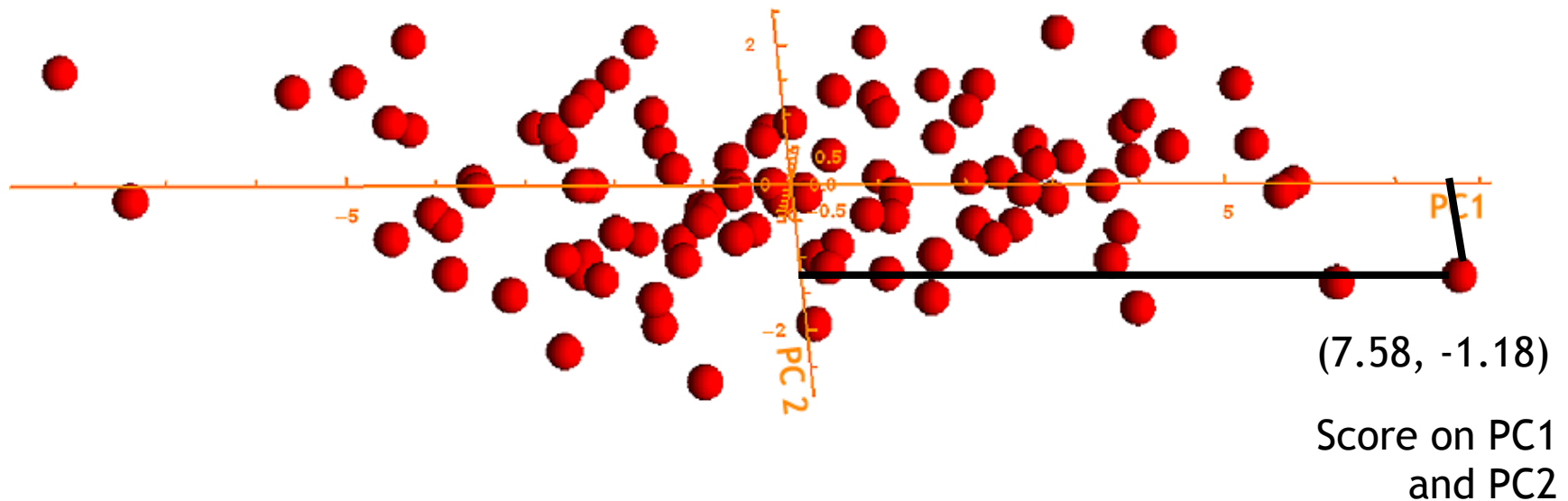
The larger the absolute value, the more important the variable.

	PC1	PC2	PC3
Var A	-0.76	-0.58	-0.29
Var B	0.28	-0.69	0.66
Var C	0.58	-0.42	-0.70

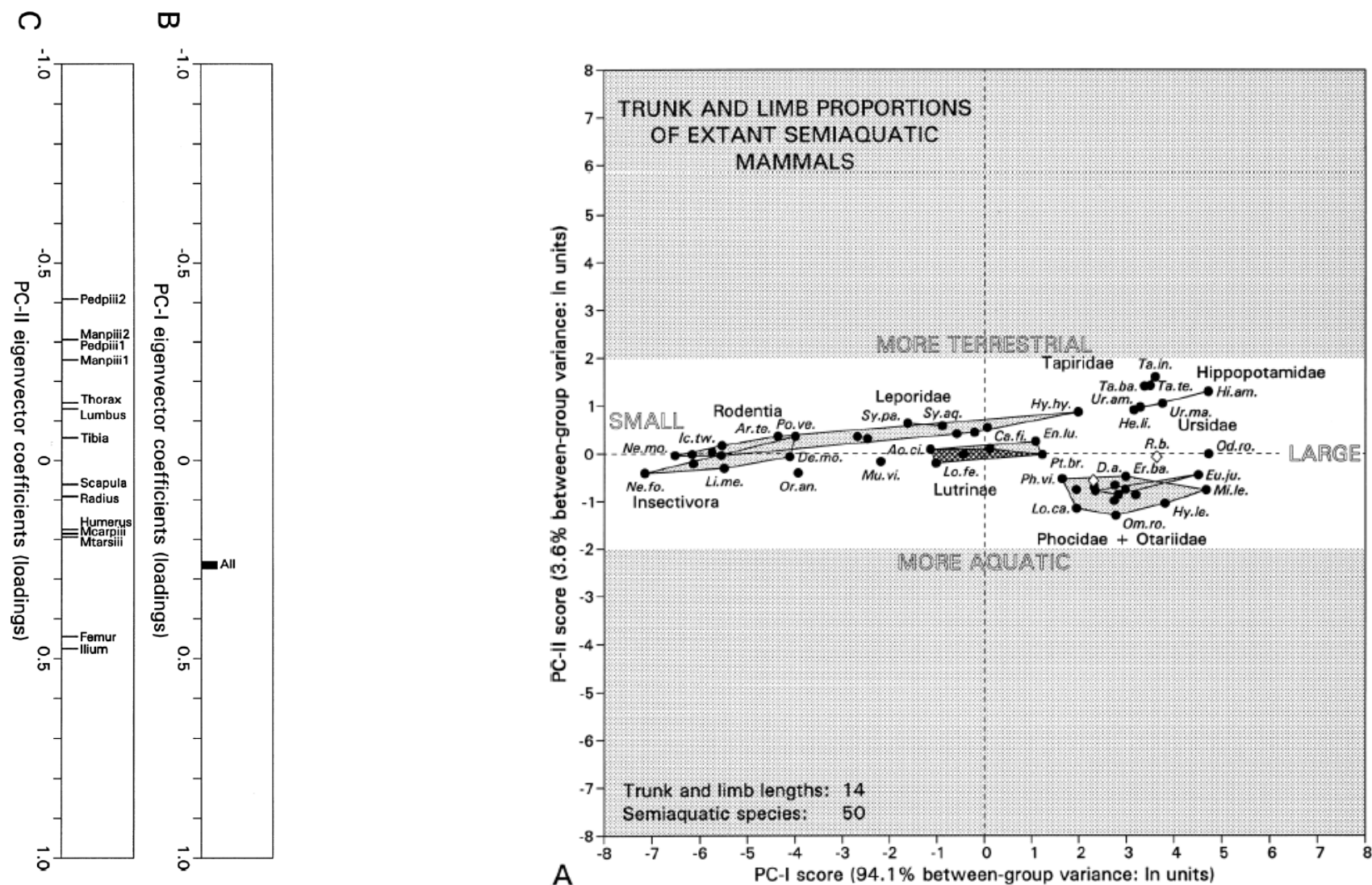


Scores

The coordinates of each data point on the PC axes.
These numbers can be thought of as new variables, or shape variables.



PCA plots have lots of meaning



Gingerich, P.D. 2003. Land-to-sea transition in early whales: evolution of Eocene archaeoceti (Cetacea) in relation to skeletal proportions and locomotion of living semiaquatic mammals. *Paleobiology*, 29: 429-454.

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 vs Relative Warps vs Partial Warps

Relative warps = Principal components

Relative warps/Principal components organize shape variation so that the greatest amount is explained on PC1, second greatest on PC2, etc. Also PC1 is uncorrelated with PC2 is uncorrelated with PC3, etc.

Partial warps (can safely be ignored)

Partial Warps measure the “scale” of shape variation over the entire object down to a small part of the object. NOT principal components (even though the plots look alike). Partial warp 1 explains variation in ALL the landmarks, Partial warp 2 explains variation in part of the landmarks, Partial warp 3 in a smaller number, etc. Partial Warp 1 MAY be correlated with Partial warp 2, etc.

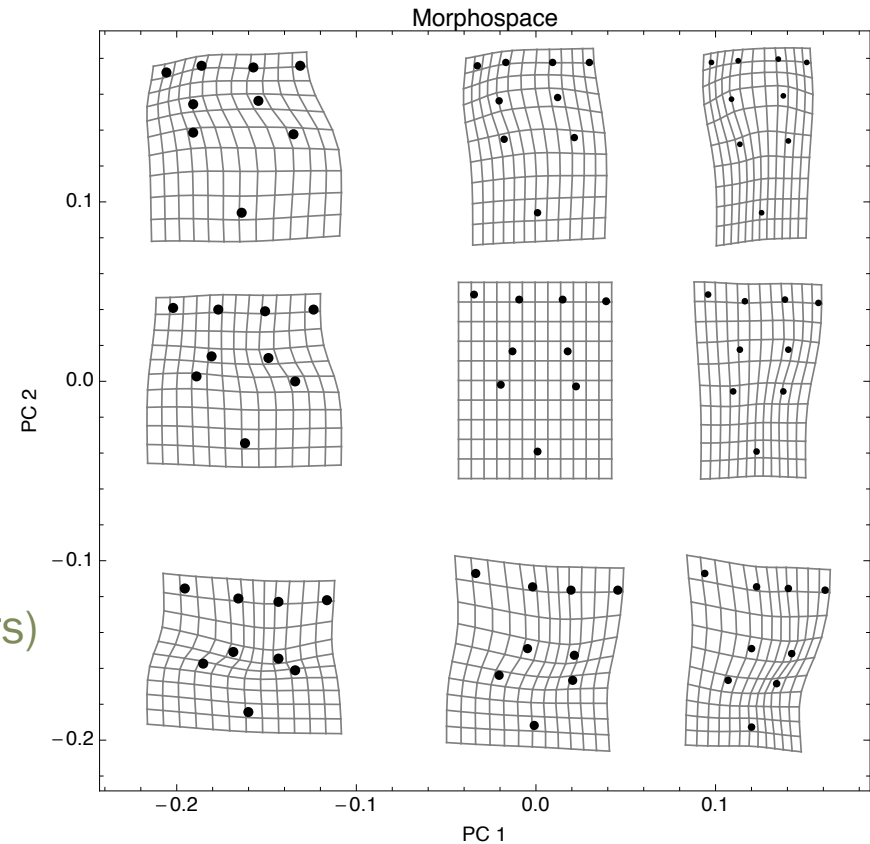
Shape modelling...

How to construct models of shapes in morphospace

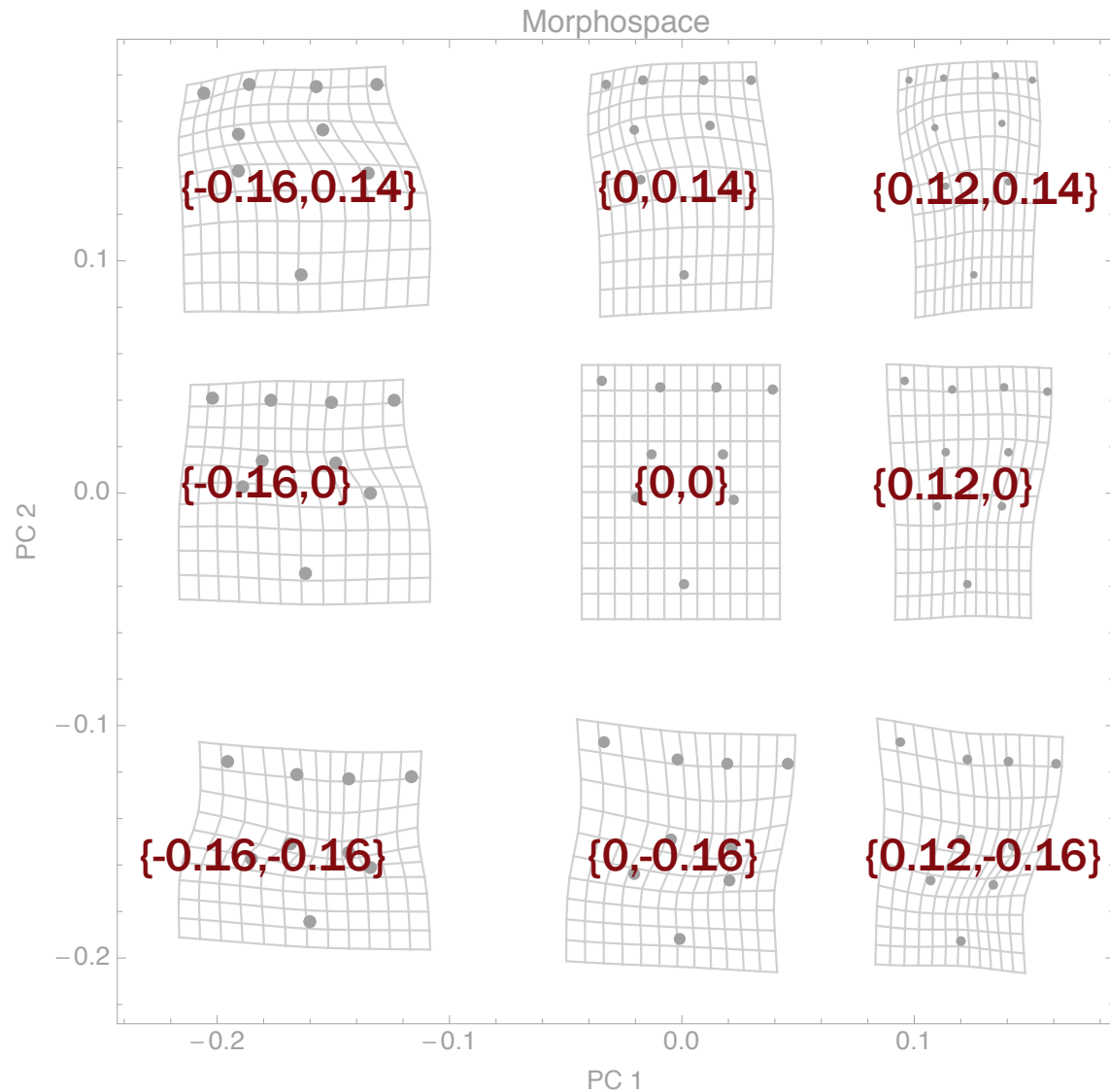
Ingredients:

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

$$\text{Model} = \text{consensus} + \sum (\text{scores} * \text{eigenvectors})$$



Shape modelling: note scores of the models



Shape models

Basic equation for 2D shape space:

$$\text{model} = \text{score PC1} * \text{vector PC1} + \text{score PC2} * \text{vector PC2} + \text{consensus}$$

Model at center of the space:

$$\begin{aligned}\text{model} &= 0 * \text{vector PC1} + 0 * \text{vector PC2} + \text{consensus} \\ &= \text{consensus}\end{aligned}$$

Model at center right:

$$\begin{aligned}\text{model} &= 0.12 * \text{vector PC1} + 0 * \text{vector PC2} + \text{consensus} \\ &= 0.12 * \text{vector PC1} + \text{consensus}\end{aligned}$$

Model at upper right:

$$\text{model} = 0.12 * \text{vector PC1} + 0.14 * \text{vector PC2} + \text{consensus}$$