Geometric morphometrics

Outline analyses

Manuel F. G. Weinkauf

Univerzita Karlova, Prague, Czech Republic

26–27 August 2022

Section 1

Introduction

Describe the true overall shape

- We use the x- and y-coordinates (semi-landmarks) of the outline of the structure
- The outline is then re-described using complex mathematics to make it comparable with other outlines

Describe the true overall shape

- We use the x- and y-coordinates (semi-landmarks) of the outline of the structure
- The outline is then re-described using complex mathematics to make it comparable with other outlines
- **Advantage:** Great for structures that otherwise lack comparable or homologue features

Describe the true overall shape

- We use the x- and y-coordinates (semi-landmarks) of the outline of the structure
- The outline is then re-described using complex mathematics to make it comparable with other outlines
- **Advantage:** Great for structures that otherwise lack comparable or homologue features
- **Disadvantage:** Dismisses all information other then silhouette shape

We extract the outline of a structure

- Ideal for structures with little internal characteristics

Carcharadontosaurus saharicus

http://www.fossilmall.com

We extract the outline of a structure

- Ideal for structures with little internal characteristics
- Identify the object of interest in the image

Carcharadontosaurus saharicus

http://www.fossilmall.com

We extract the outline of a structure

- Ideal for structures with little internal characteristics
- Identify the object of interest in the image
- Extraxt x- and y coordinates along outline
- The first outline point is mostly a **well defined homologue structure**

Carcharadontosaurus saharicus

http://www.fossilmall.com

That was easy . . . now it gets complicated

Problem 1: Unequal number of outline points

Specimens vary in size and image resolution

We have a number of x- and y-coordinates, but their number differs between specimens → **we have now way to compare them**

Hanafi Idris et al. (2008) J. Biol. Sci. 8: 882–8

That was easy . . . now it gets complicated

Problem 2: Non-comparable outline points

Specimens vary in size and shape

Even if we force the same number of points for all specimens, their **individual positions are not comparable** because the specimens are not identical

Hanafi Idris et al. (2008) J. Biol. Sci. 8: 882–8

That was easy . . . now it gets complicated

Problem 2: Non-comparable outline points

Specimens vary in size and shape

Even if we force the same number of points for all specimens, their **individual positions are not comparable** because the specimens are not identical

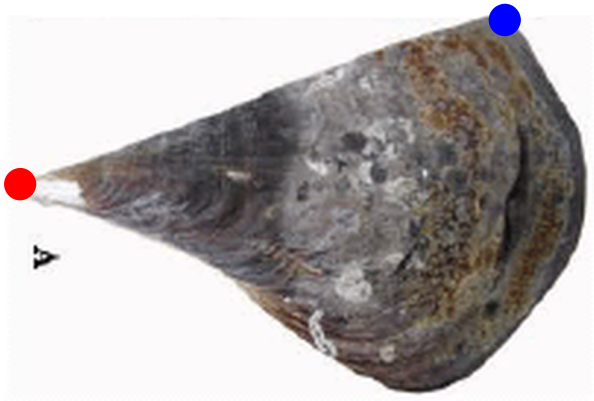
Hanafi Idris et al. (2008) J. Biol. Sci. 8: 882–8

That was easy . . . now it gets complicated

Problem 2: Non-comparable outline points

Specimens vary in size and shape

Even if we force the same number of points for all specimens, their **individual positions are not comparable** because the specimens are not identical



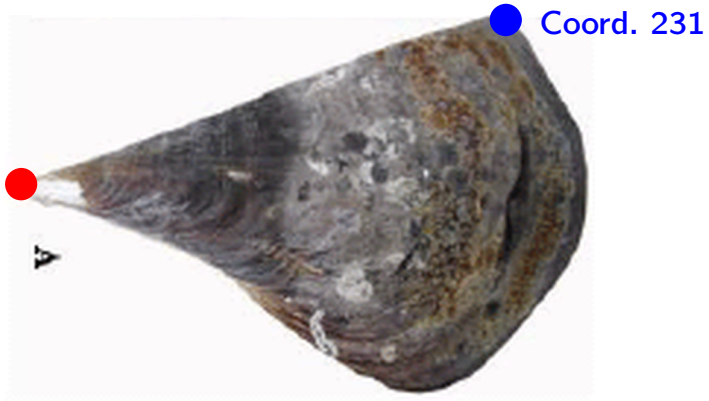
Hanafi Idris et al. (2008) J. Biol. Sci. 8: 882–8

That was easy . . . now it gets complicated

Problem 2: Non-comparable outline points

Specimens vary in size and shape

Even if we force the same number of points for all specimens, their **individual positions are not comparable** because the specimens are not identical



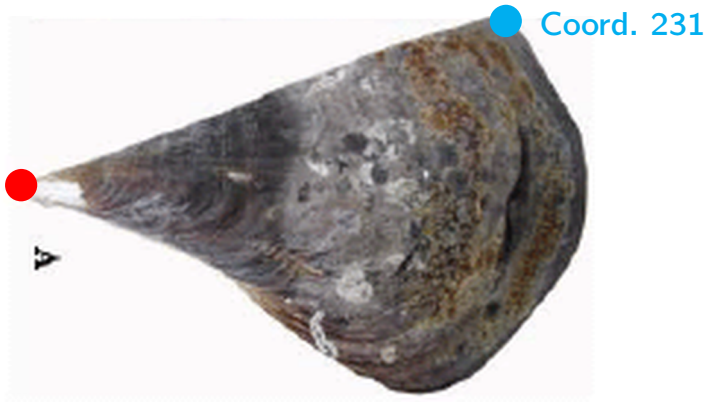
Hanafi Idris et al. (2008) J. Biol. Sci. 8: 882–8

That was easy . . . now it gets complicated

Problem 2: Non-comparable outline points

Specimens vary in size and shape

Even if we force the same number of points for all specimens, their **individual positions are not comparable** because the specimens are not identical



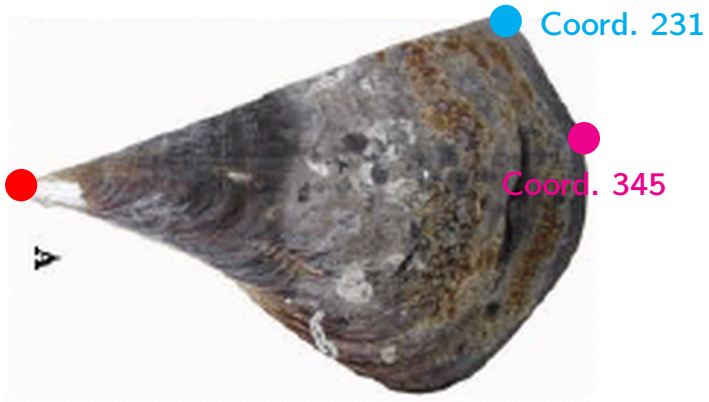
Hanafi Idris et al. (2008) J. Biol. Sci. 8: 882–8

That was easy . . . now it gets complicated

Problem 2: Non-comparable outline points

Specimens vary in size and shape

Even if we force the same number of points for all specimens, their **individual positions are not comparable** because the specimens are not identical



Hanafi Idris et al. (2008) J. Biol. Sci. 8: 882–8

That was easy . . . now it gets complicated

How do we solve this predicament?

- We have some problems . . .
  1. We have a number of x- and y-coordinates, but their number differs between specimens $\rightarrow$ we have now way to compare them
  2. Even if we force the same number of points for all specimens, their individual positions are not similar because the specimens are not identical

That was easy . . . now it gets complicated

How do we solve this predicament?

- We have some problems . . .
  1. We have a number of x- and y-coordinates, but their number differs between specimens $\rightarrow$ we have now way to compare them
  2. Even if we force the same number of points for all specimens, their individual positions are not similar because the specimens are not identical
- . . . and a potential solution
  1. We re-describe the outlines in a way that makes them comparable $\Rightarrow$ **the same number-position should describe the same morphological characteristic in all specimens**

Section 2

Fourier transformation et al.

The brilliant idea of a French gentleman

- Complex waveforms are difficult to describe

Jean-Baptiste Joseph Fourier

https://en.wikipedia.org

The brilliant idea of a French gentleman

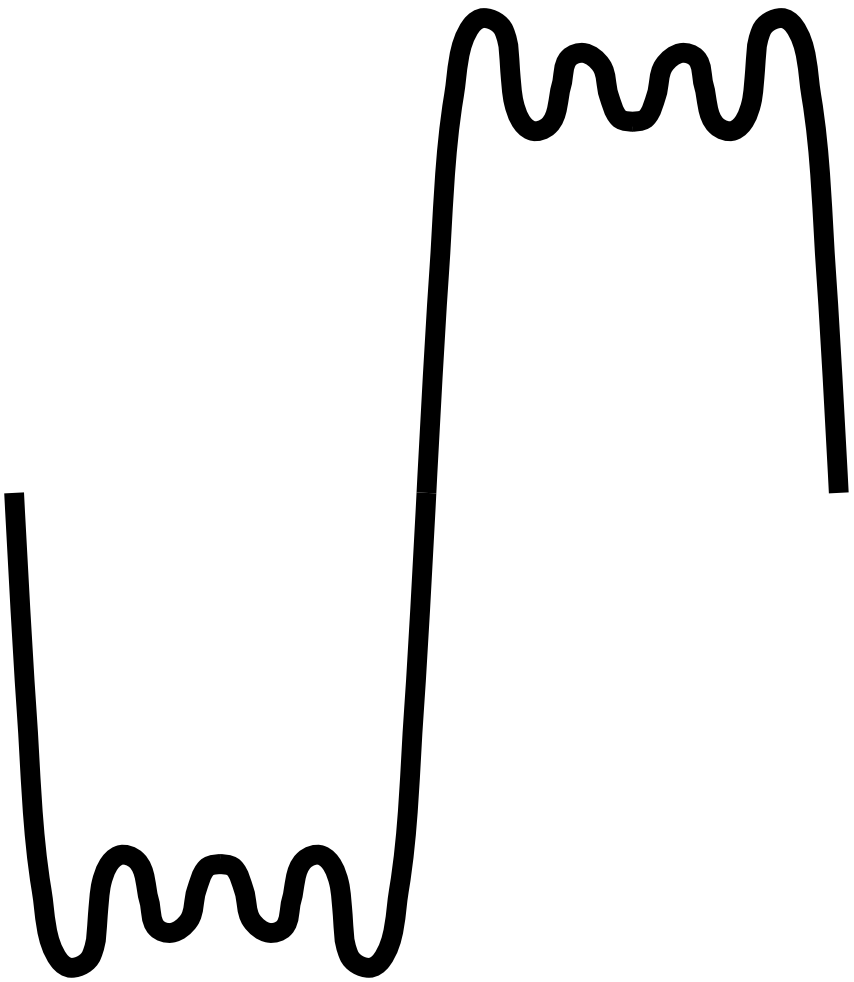
- Complex waveforms are difficult to describe
- What if we describe them as a sum of simpler waveforms?

Jean-Baptiste Joseph Fourier

https://en.wikipedia.org

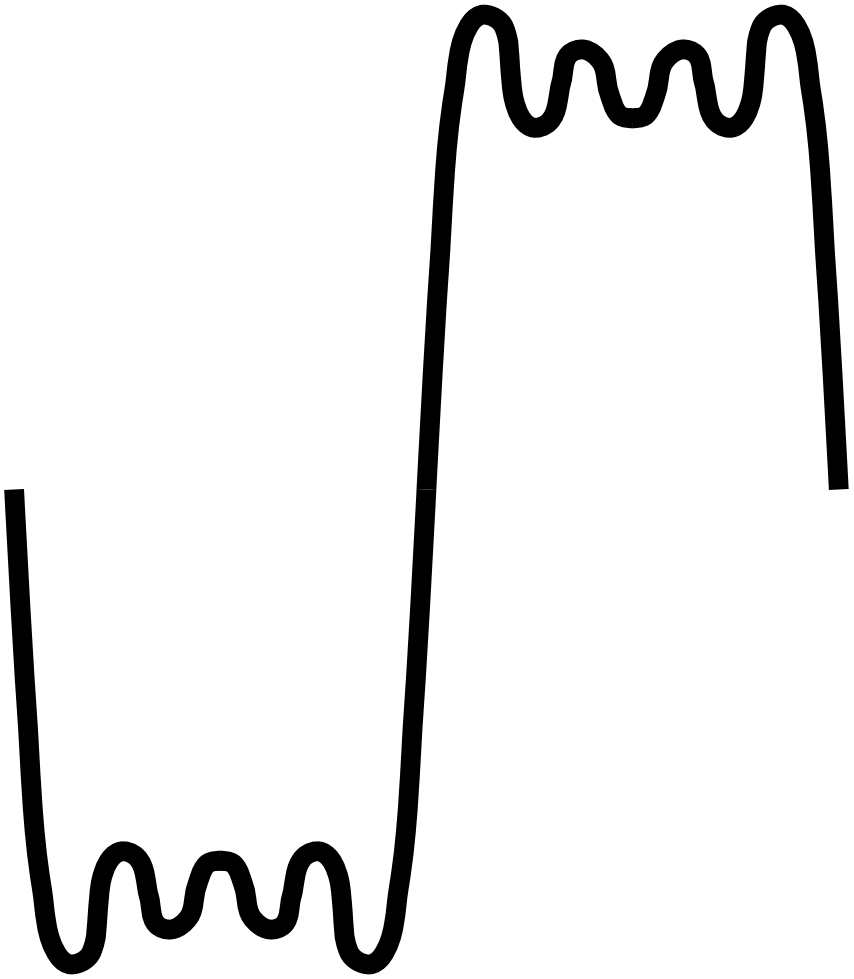
The idea of Fourier analyses

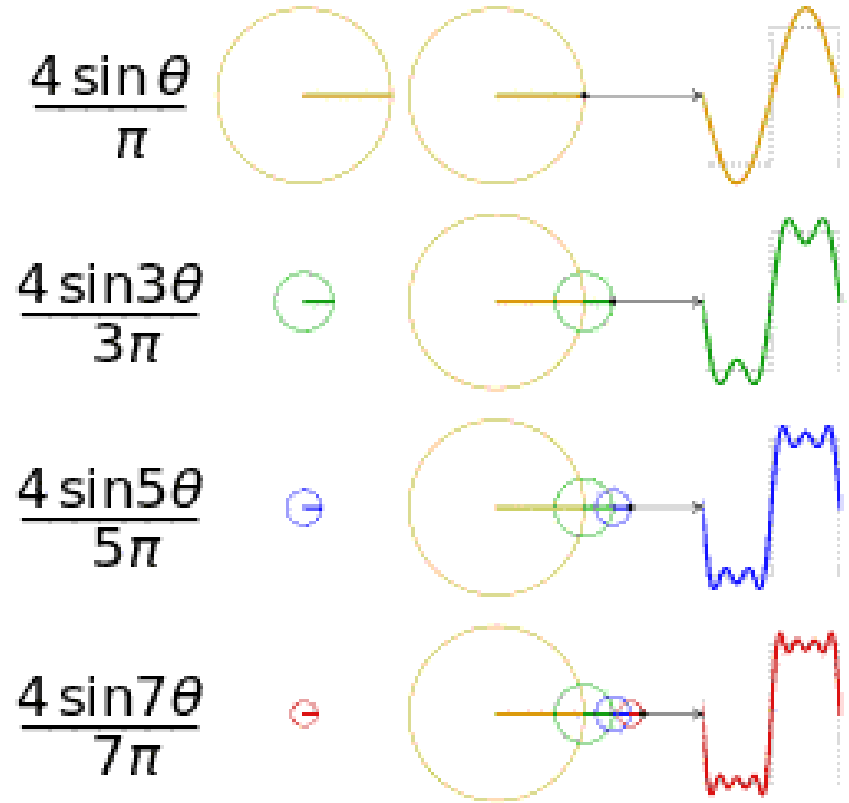
You have a complex form . . .



The idea of Fourier analyses

You have a complex form . . . and describe it as a sum sin- and cos-functions

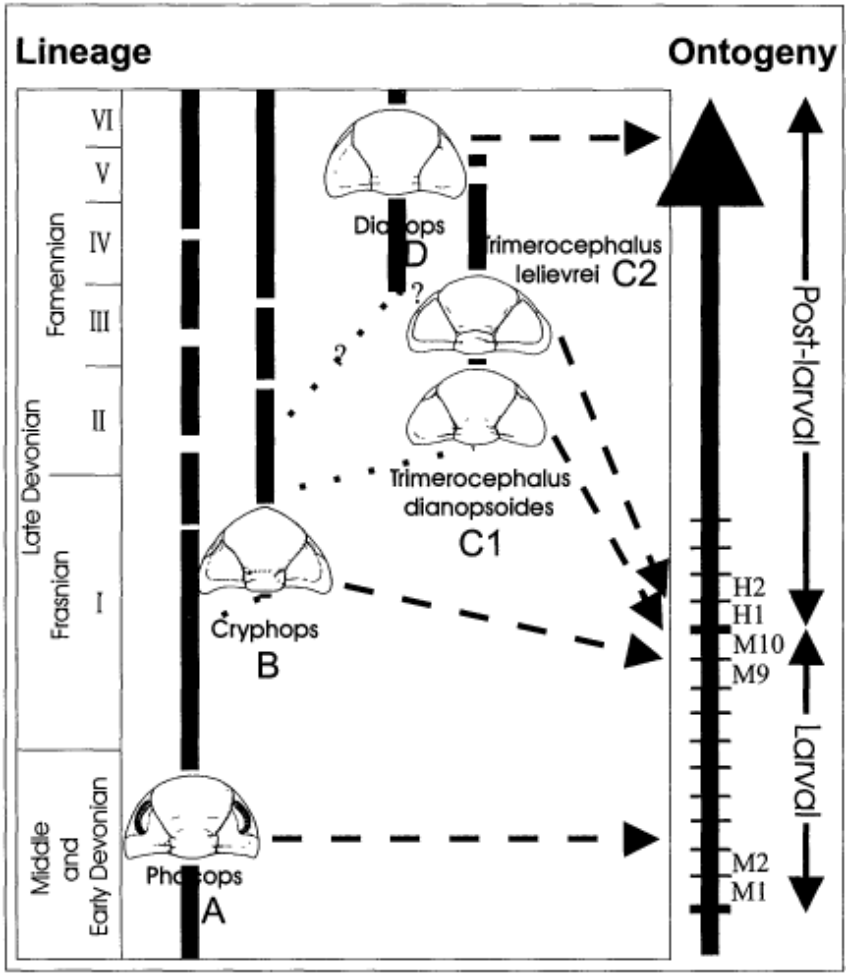


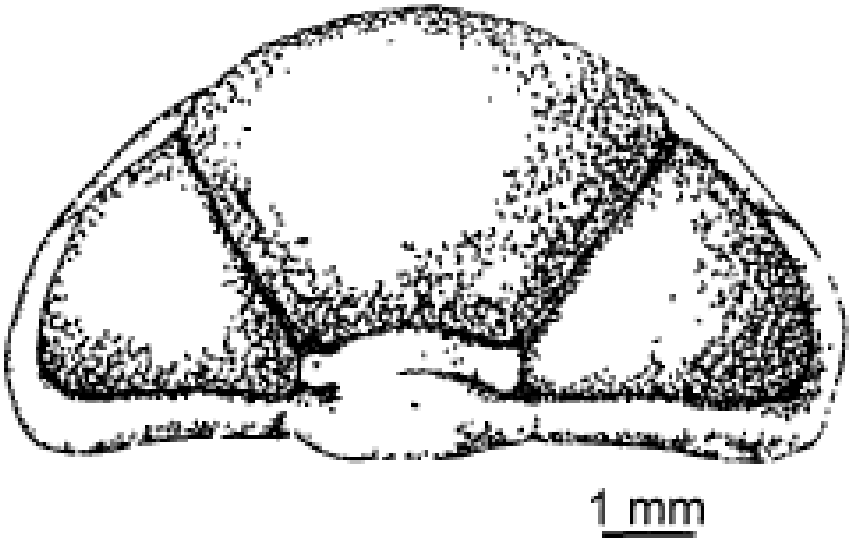


https://en.wikipedia.org

An example: Trilobite cephalons

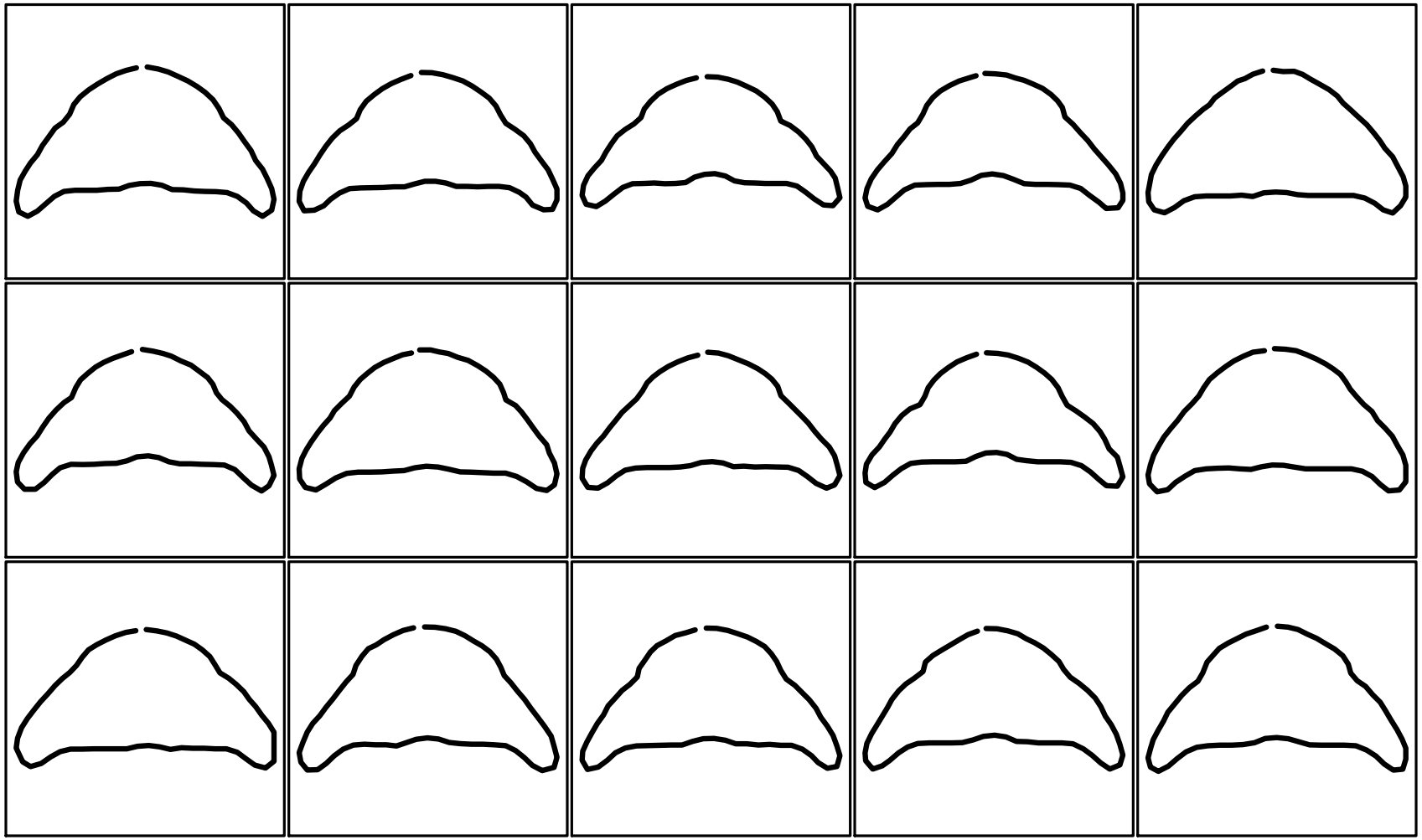
Crônier et al. (1998) *Paleobiology* 24 (3): 359–70





An example: Trilobite cephalons

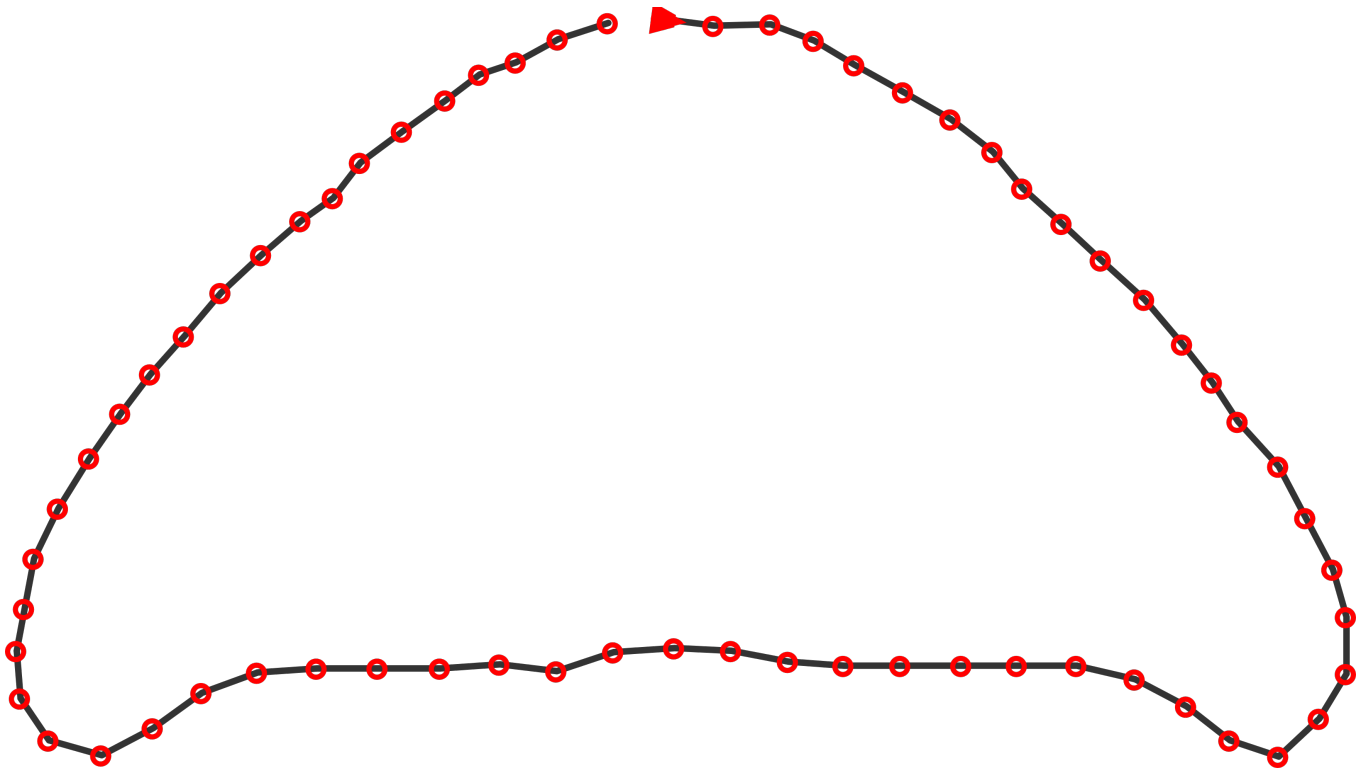
Outlines of the cephalon of 50 specimens of *Trimerocephalus lelievrei*



The original Fourier shape analysis

Step 1

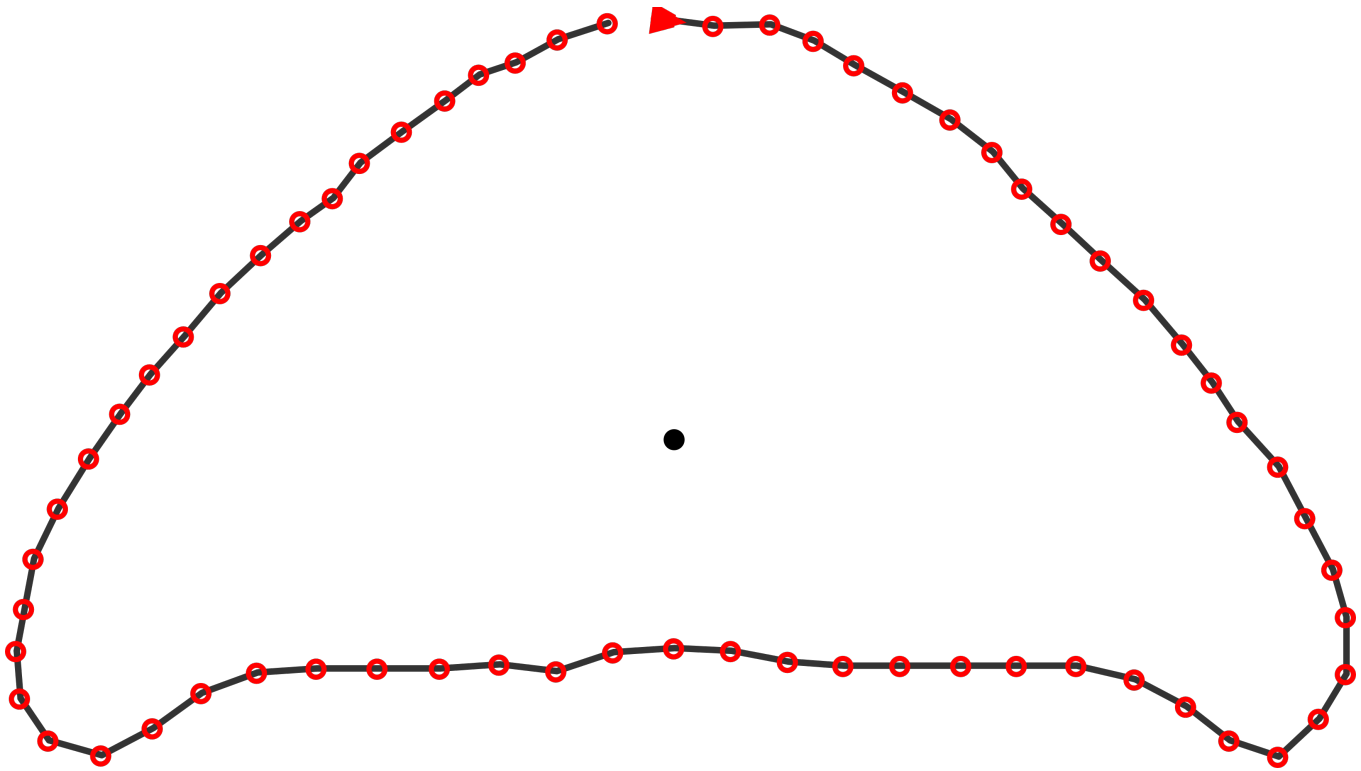
Resample outline for fixed number of equidistant points



The original Fourier shape analysis

Step 2

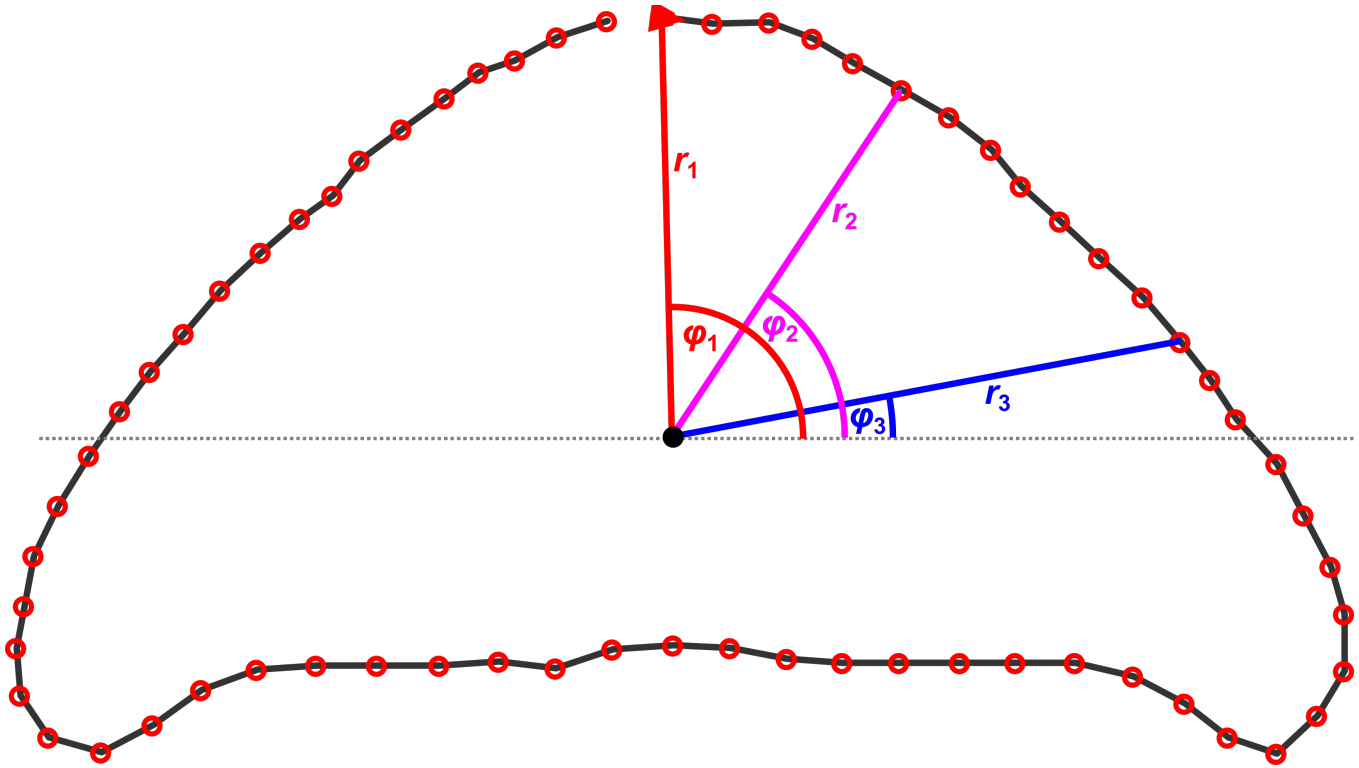
Find the centre of the shape



The original Fourier shape analysis

Step 3

Draw vectors from centre to each point



Fourier re-description of the shape

1. Each radius you draw gives you an angle ϕ from the horizontal and a length r between the centre point and the outline point

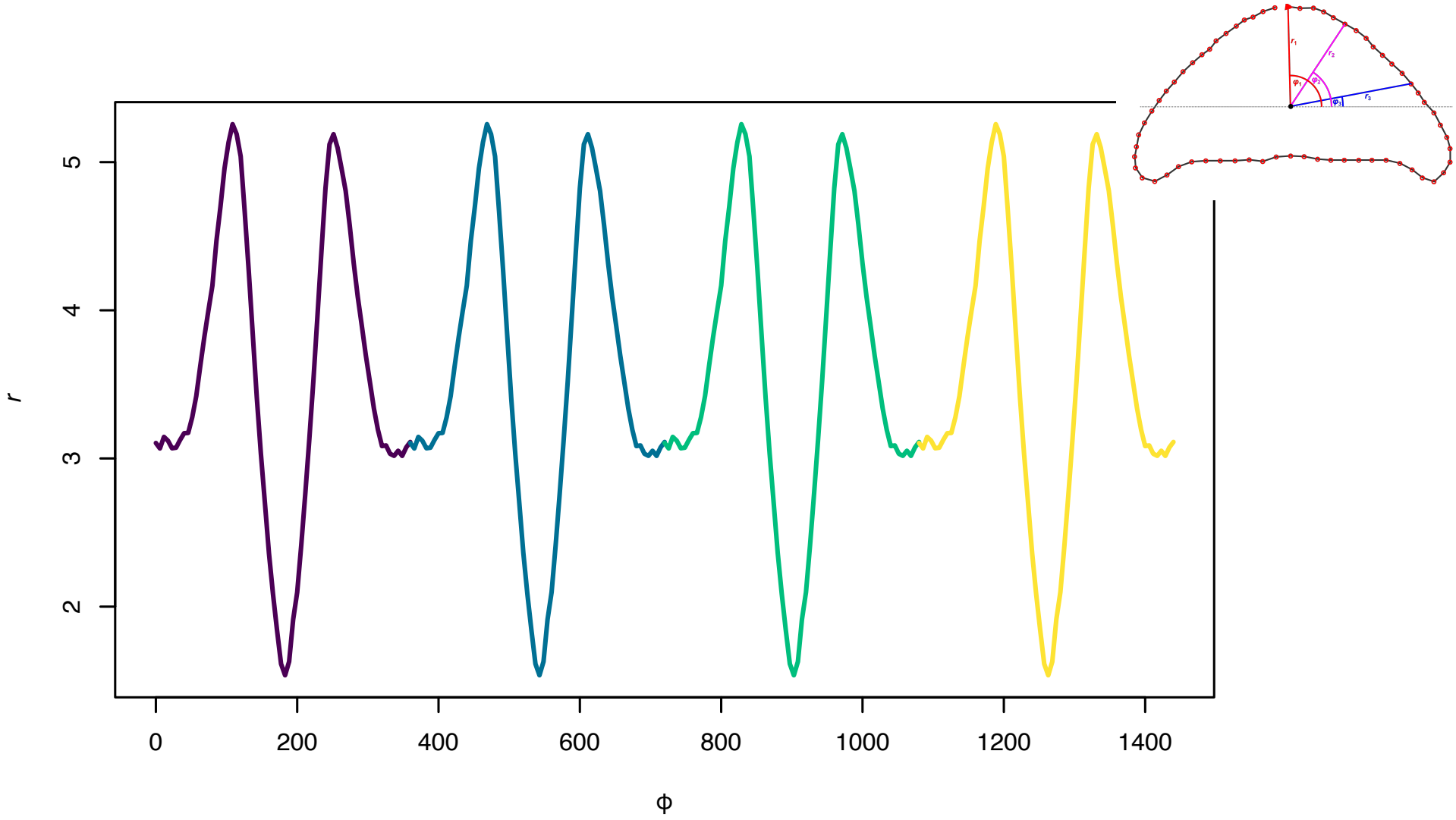
Fourier re-description of the shape

1. Each radius you draw gives you an angle ϕ from the horizontal and a length r between the centre point and the outline point
2. The radius will depend on the angle, because it is the distance between centre and outline of the shape at that angle $\Rightarrow$ r is a function of ϕ: $r(\phi)$

Fourier re-description of the shape

1. Each radius you draw gives you an angle ϕ from the horizontal and a length r between the centre point and the outline point
2. The radius will depend on the angle, because it is the distance between centre and outline of the shape at that angle $\Rightarrow$ r is a function of ϕ: $r(\phi)$
3. When you reach the last point in the outline and go a step further, you end up where you began $\Rightarrow$ the function $r(\phi)$ is periodic and can be described by Fourier functions

Fourier re-description of the shape

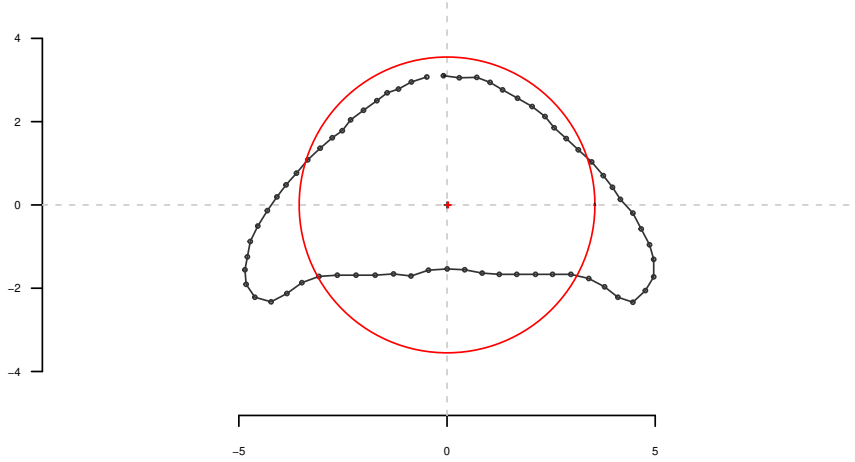


Fourier re-description of the shape

- We use increasingly many functions (called 'harmonics') added together to increasingly better describe the shape

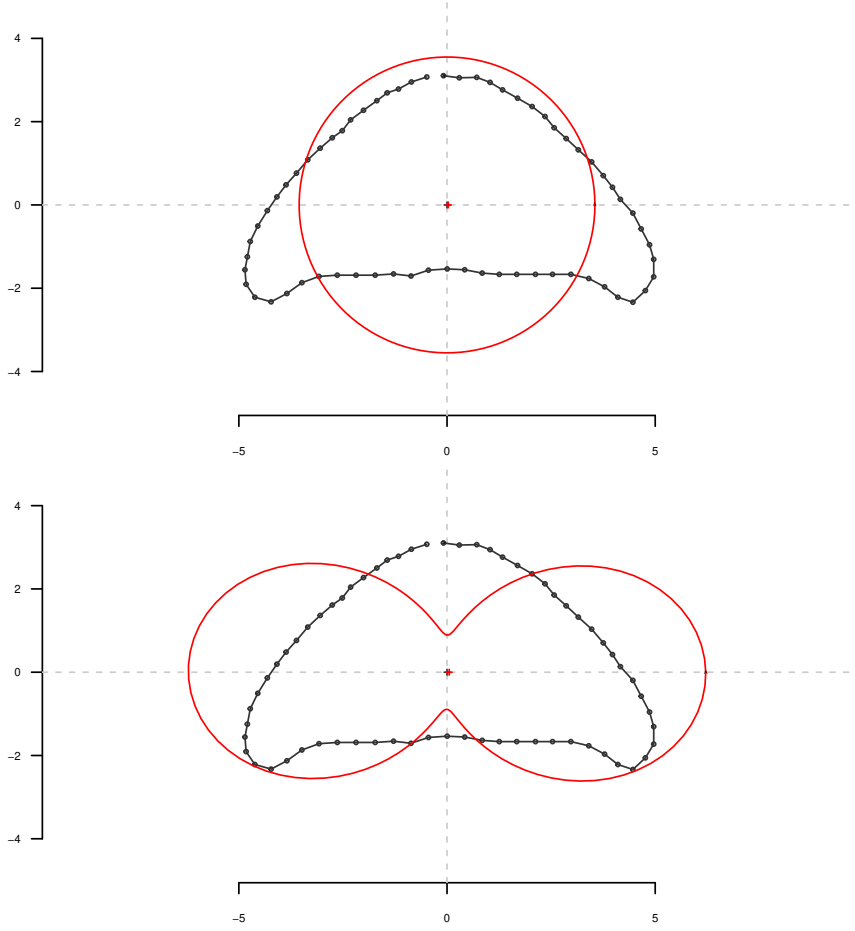
Fourier re-description of the shape

- We use increasingly many functions (called 'harmonics') added together to increasingly better describe the shape
  1. One harmonic can only describe a circle

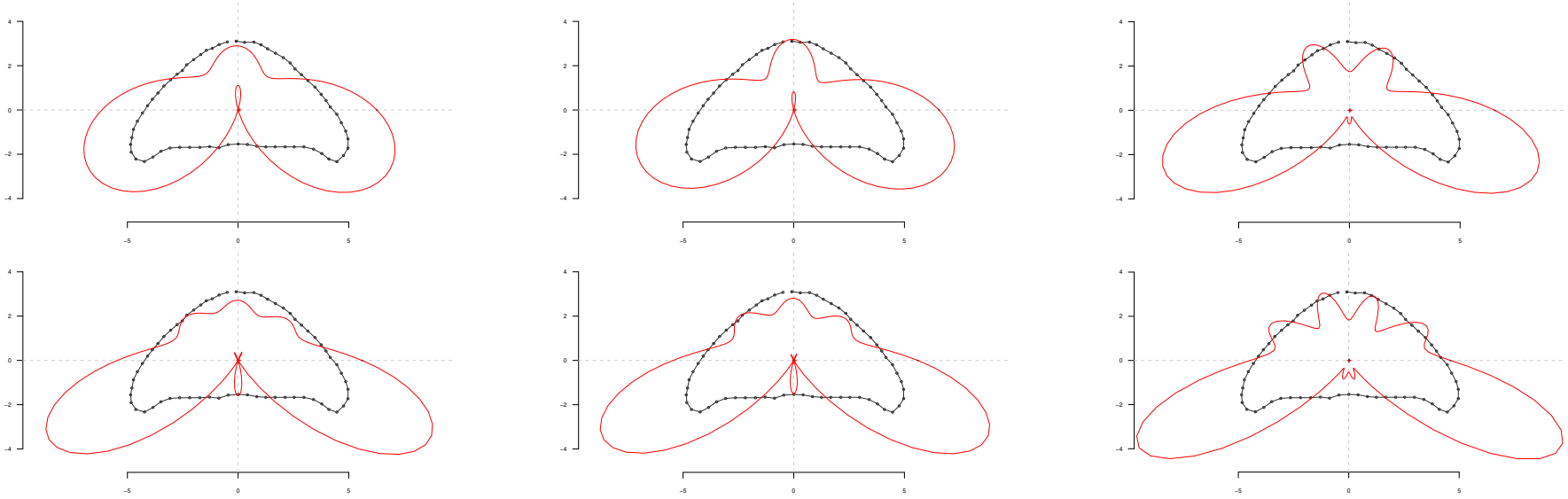


Fourier re-description of the shape

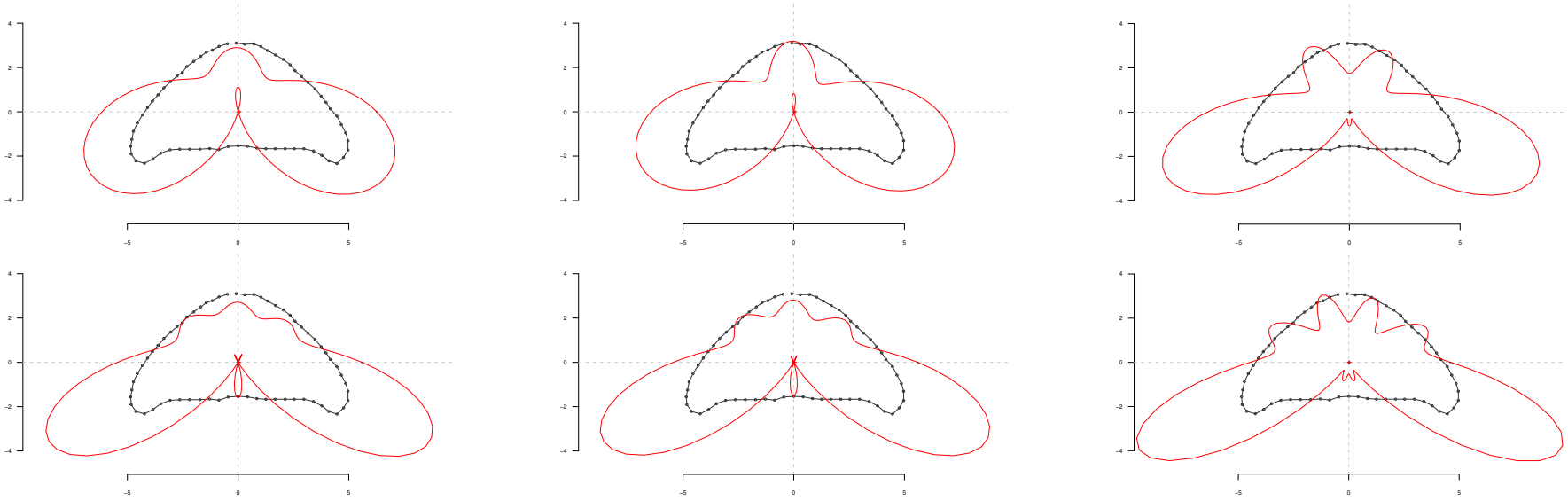
- We use increasingly many functions (called 'harmonics') added together to increasingly better describe the shape
  1. One harmonic can only describe a circle
  2. Two harmonics can describe an ovoid shape



The more harmonics, the better, but . . .

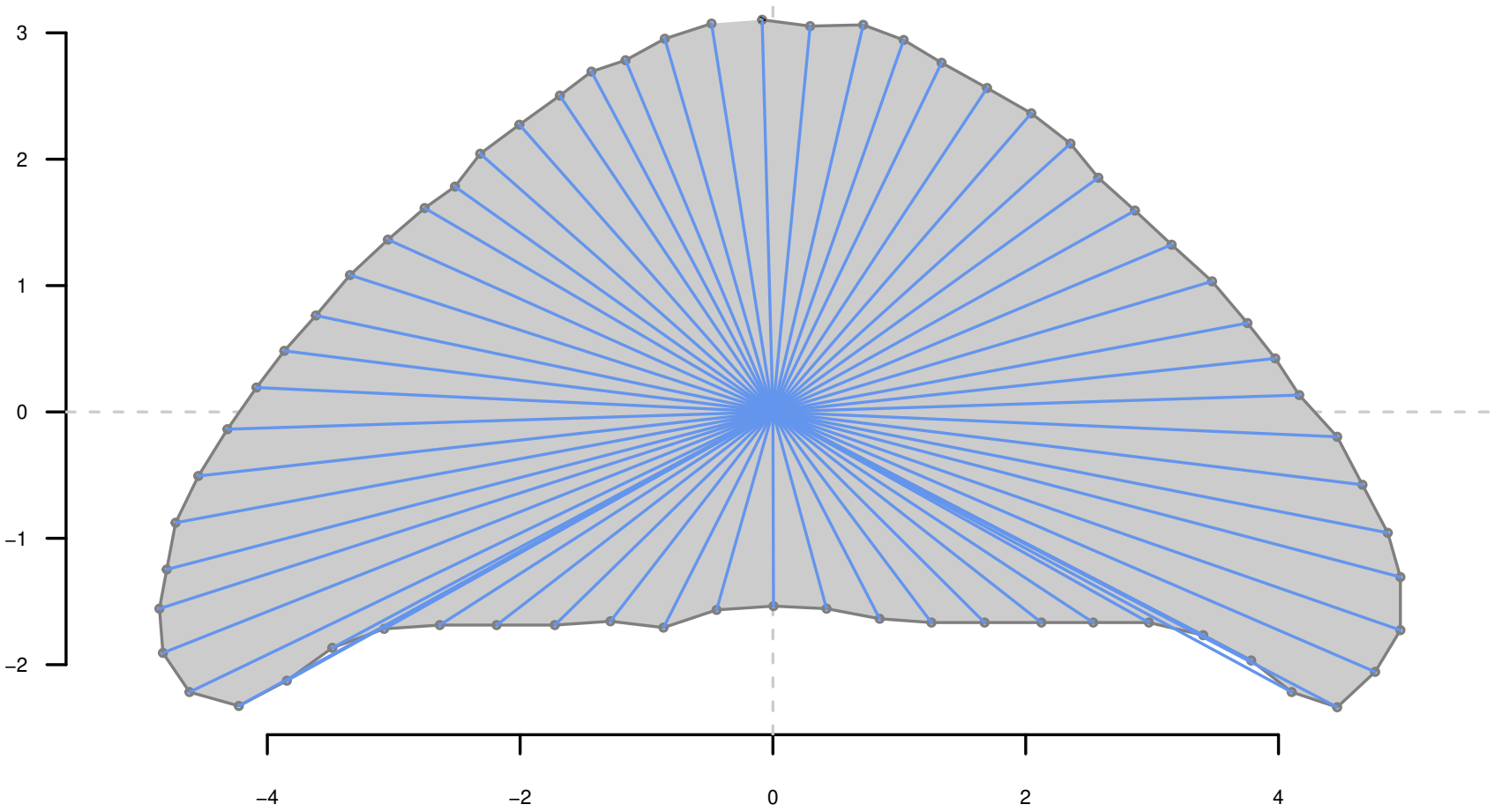


The more harmonics, the better, but . . .

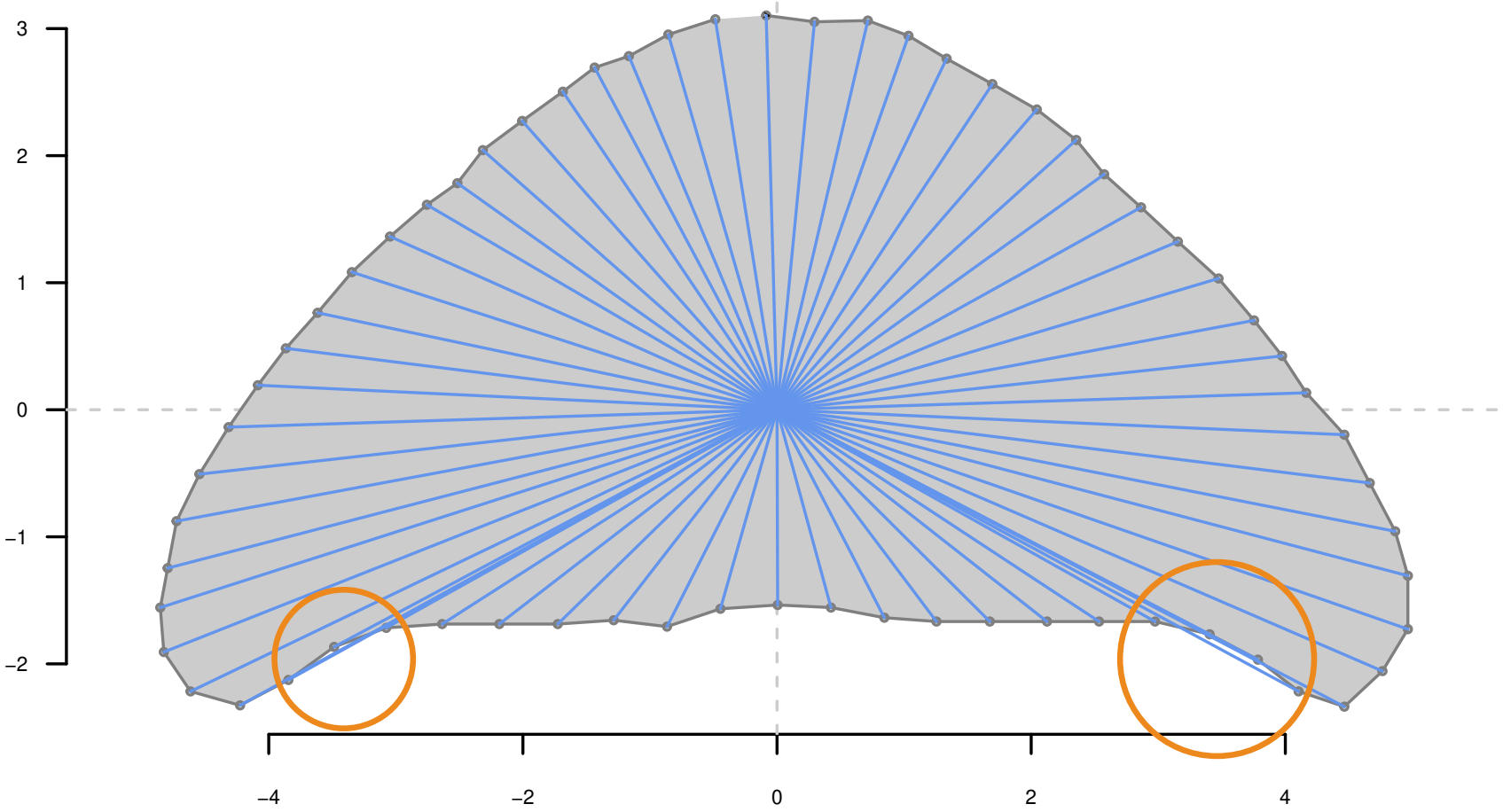


Problem 1: This simple approach is not very good at describing even only fairly complex shapes

The other problem with radii Fourier analysis



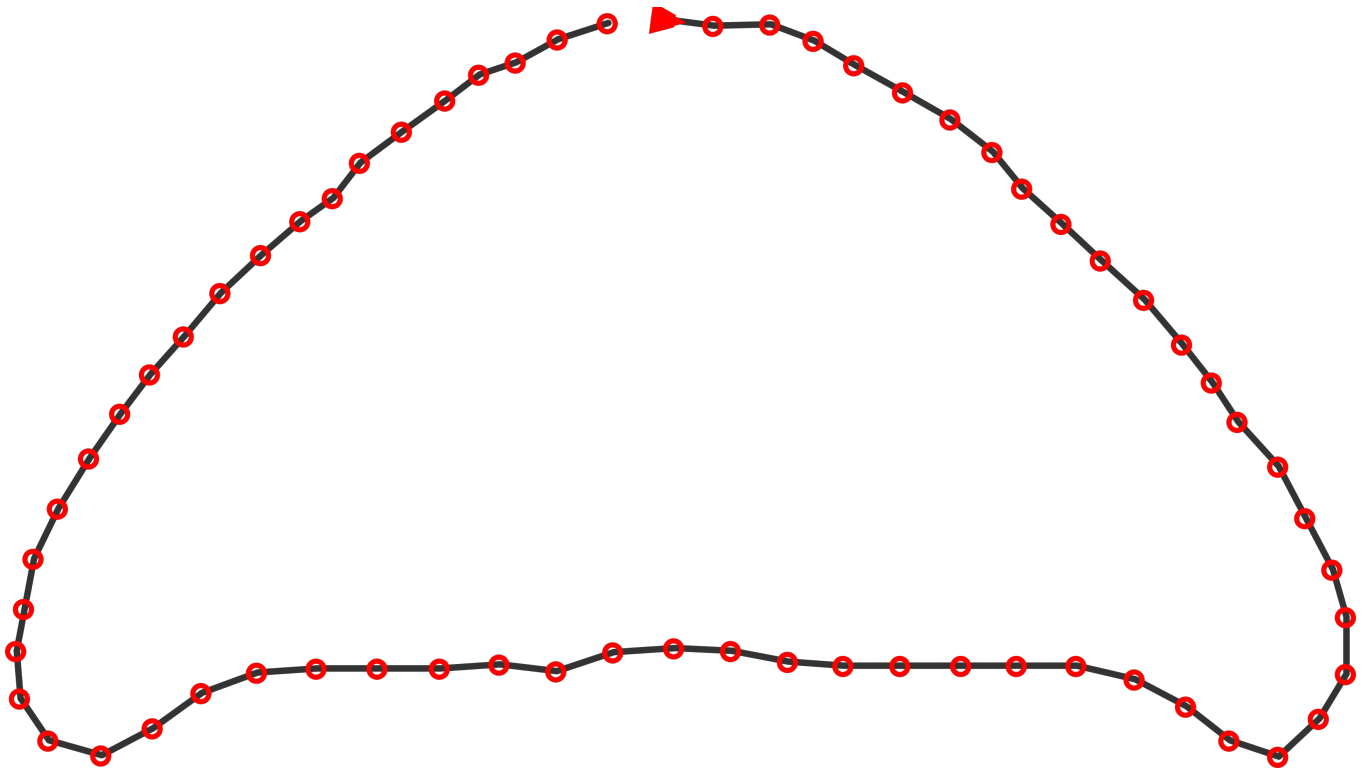
The other problem with radii Fourier analysis



Elliptic Fourier analysis

Step 1

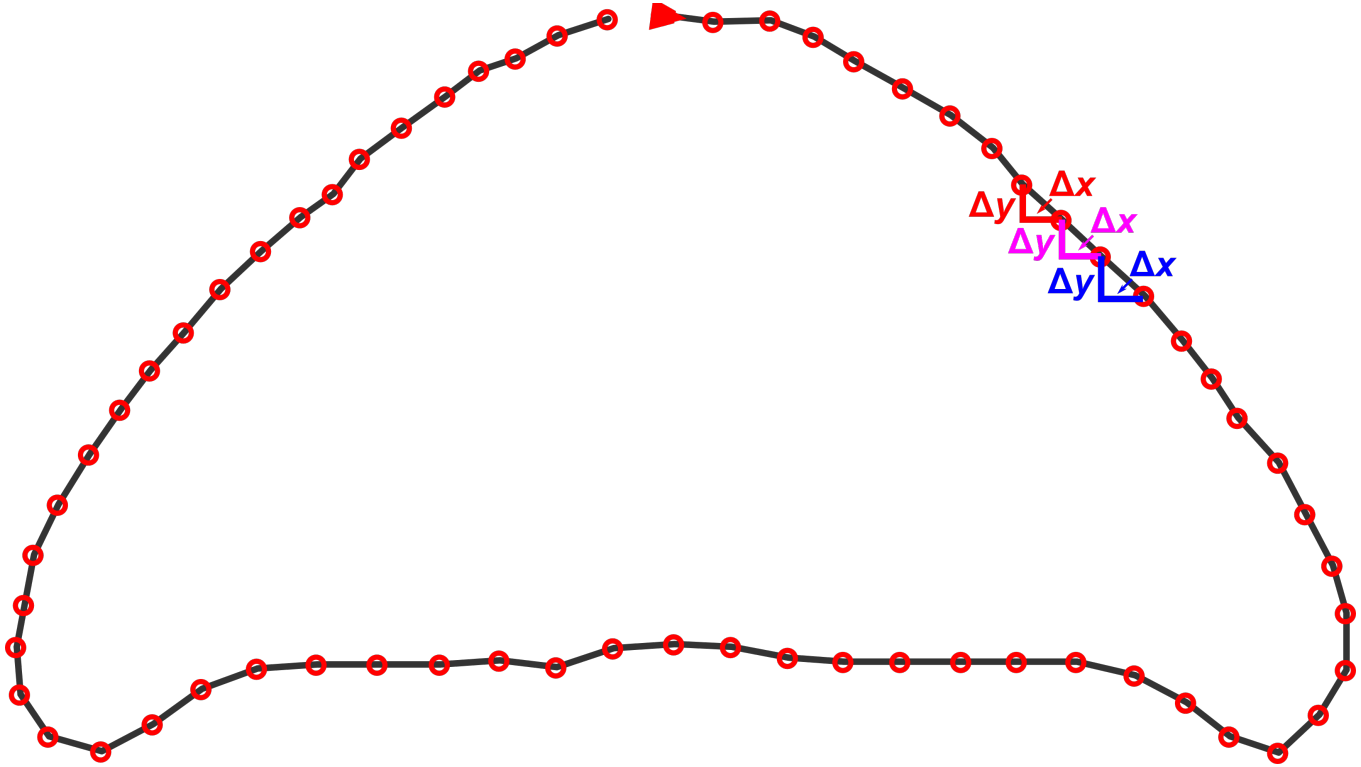
Resample outline for fixed number of equidistant points



Elliptic Fourier analysis

Step 2

Gather step size (Δx, Δy) from point to point



Getting the harmonics out of these data

- The Fourier decomposition of these $\Delta x/\Delta y$ is even more complicated than with radii variation Fourier analysis
- In the end, you get a set of harmonics, that satisfies important conditions

Getting the harmonics out of these data

- The Fourier decomposition of these $\Delta x/\Delta y$ is even more complicated than with radii variation Fourier analysis
- In the end, you get a set of harmonics, that satisfies important conditions
  1. All specimens are described by the same number of harmonics

Getting the harmonics out of these data

- The Fourier decomposition of these $\Delta x/\Delta y$ is even more complicated than with radii variation Fourier analysis
- In the end, you get a set of harmonics, that satisfies important conditions
  1. All specimens are described by the same number of harmonics
  2. The harmonic components are now independent of the original position of the outline points

Getting the harmonics out of these data

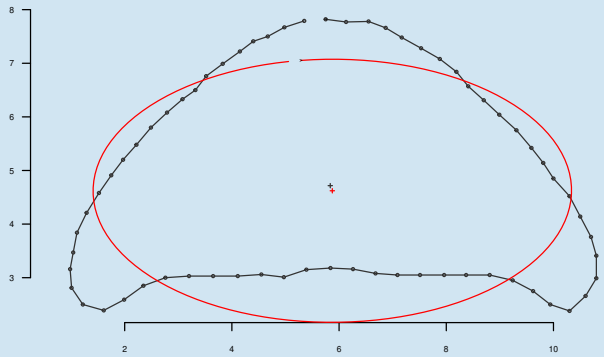
- The Fourier decomposition of these $\Delta x/\Delta y$ is even more complicated than with radii variation Fourier analysis
- In the end, you get a set of harmonics, that satisfies important conditions
  1. All specimens are described by the same number of harmonics
  2. The harmonic components are now independent of the original position of the outline points
  3. Corresponding harmonics are homologous between specimens

Getting the harmonics out of these data

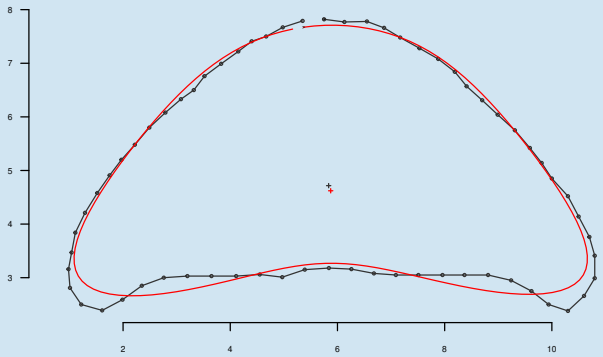
- The Fourier decomposition of these $\Delta x/\Delta y$ is even more complicated than with radii variation Fourier analysis
- In the end, you get a set of harmonics, that satisfies important conditions
  1. All specimens are described by the same number of harmonics
  2. The harmonic components are now independent of the original position of the outline points
  3. Corresponding harmonics are homologous between specimens
- You still have many data points, indeed four times as many as original outline points, but these **data points are now comparable with each other**

Much better performance via radii variation method

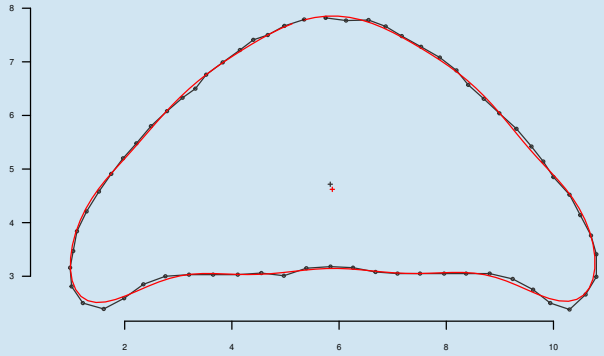
1 Harmonic



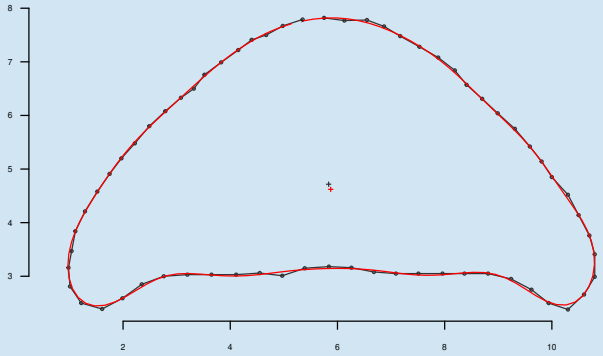
4 Harmonics



8 Harmonics

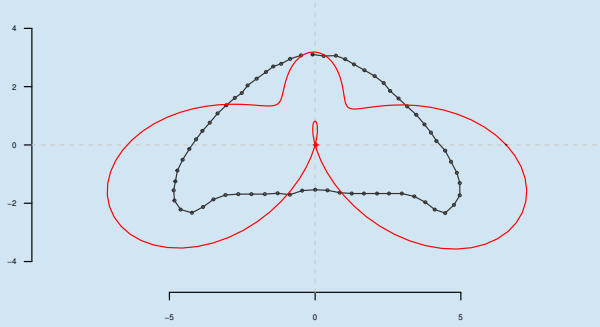


12 Harmonics

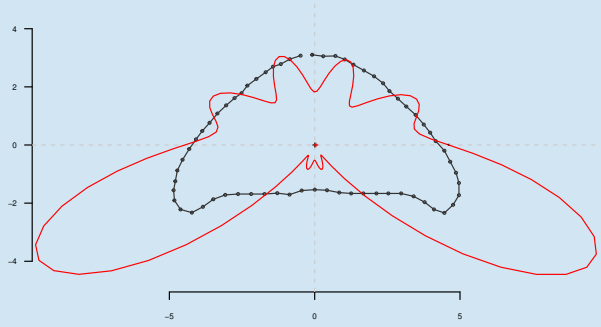


Much better performance via radii variation method

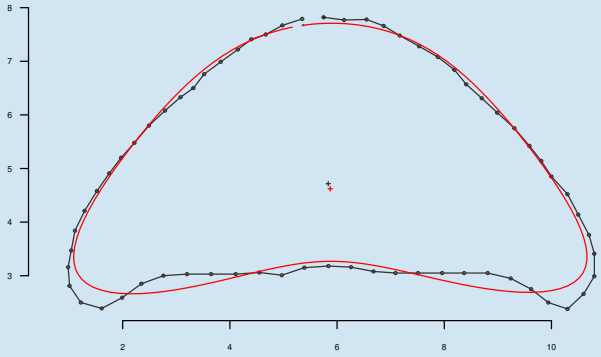
4 Harmonic, RVFA



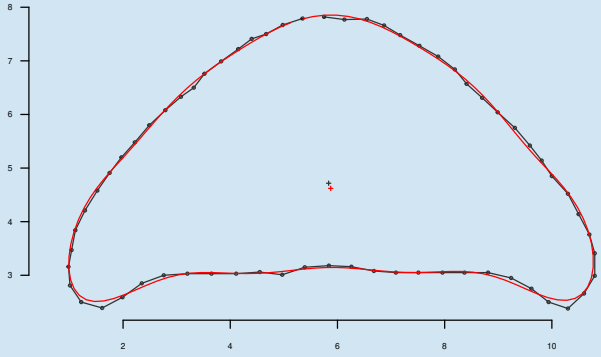
8 Harmonics, RVFA



4 Harmonics, EFA



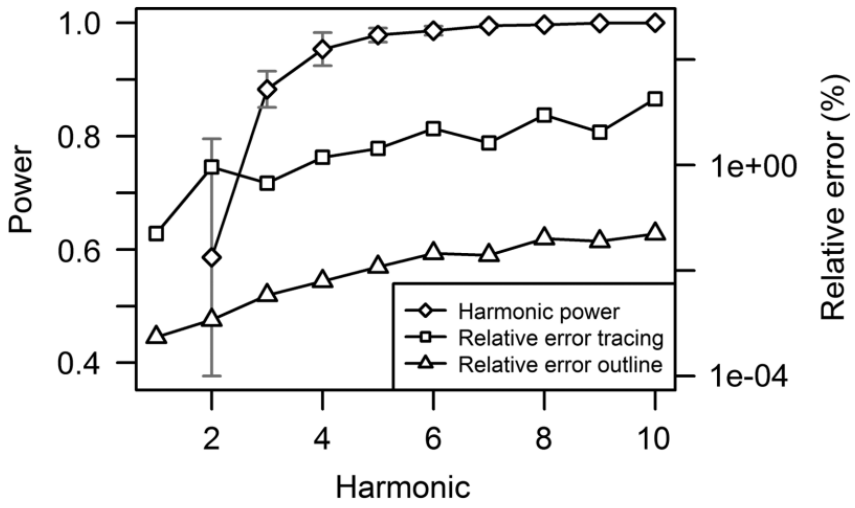
8 Harmonics, EFA



Enough is enough

Finding the optimal number of harmonics

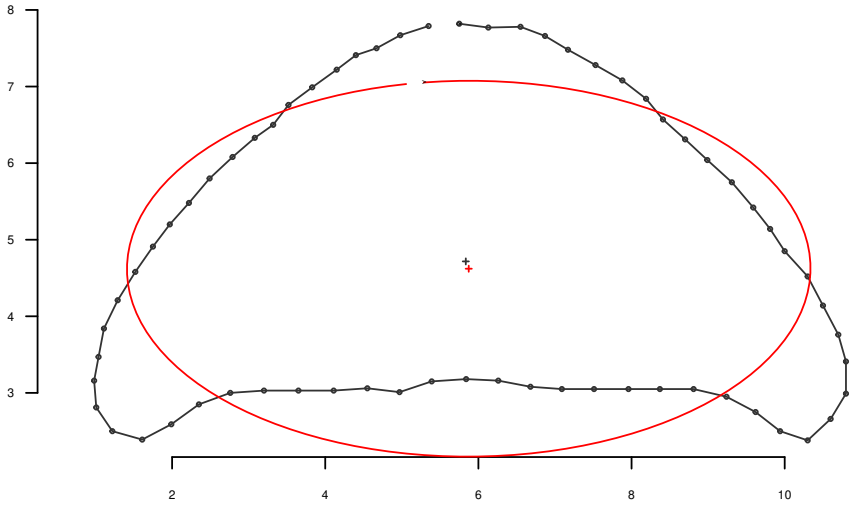
- Cumulative descriptive power of harmonics can be calculated for each specimen
- Optimal number of harmonics can be chosen at level where average or minimum power of harmonic is above a defined threshold across all specimens
- Threshold often chosen as 99 %, but can be different



Hoffmann et al. (2017) Paleobiology 43 (2): 304–20

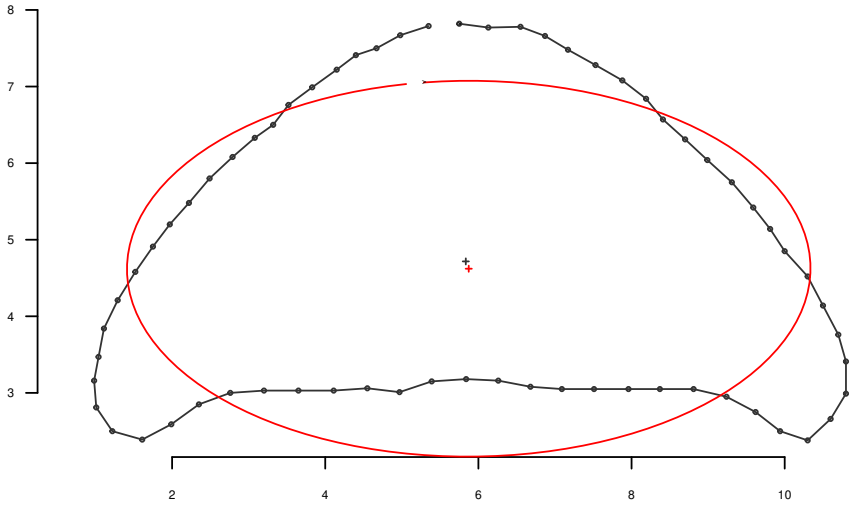
Normalized EFA

- Usually, outlines are rotated during analysis to eliminate arbitrary differences due to object orientation in the image
  1. Based on a well-defined baseline (requires two reliable landmarks)



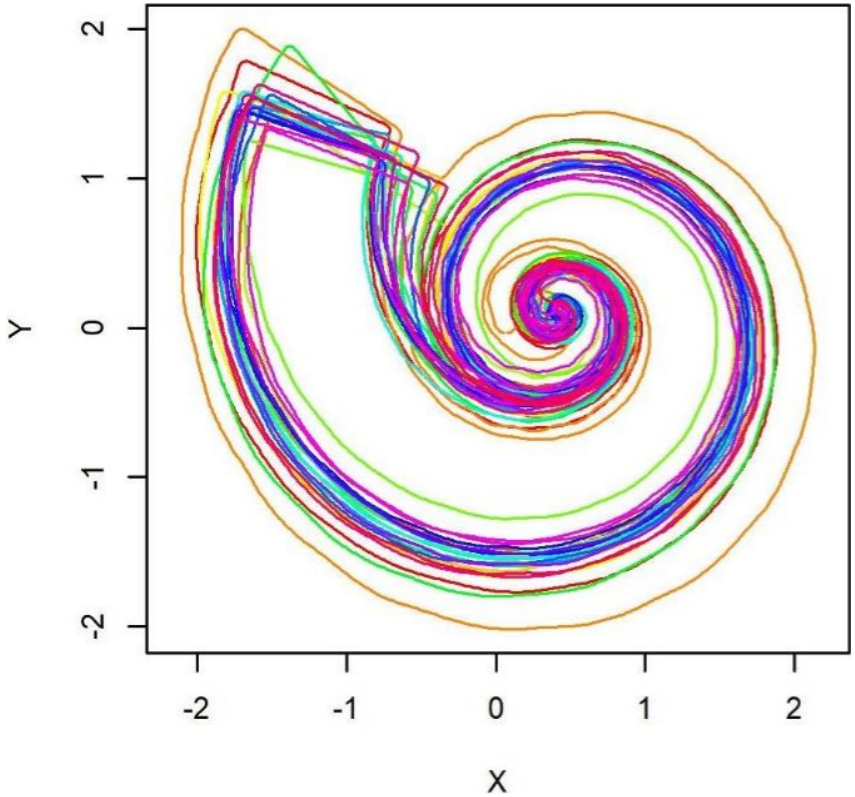
Normalized EFA

- Usually, outlines are rotated during analysis to eliminate arbitrary differences due to object orientation in the image
  1. Based on a well-defined baseline (requires two reliable landmarks)
  2. Based on the longest axis of the first harmonic



Normalized EFA

- Usually, outlines are rotated during analysis to eliminate arbitrary differences due to object orientation in the image
  1. Based on a well-defined baseline (requires two reliable landmarks)
  2. Based on the longest axis of the first harmonic
- Normalization ensures that only relevant shape information is used for analyses



Hoffmann et al. (2021) J. Molluscan Stud. 87: eyab001

Other outline procedures

- There exist other methods of outline analyses as well, for instance
  1. Zahn–Roskies Fourier analysis
  2. Fast Fourier transform analysis
  3. Eigenshape analysis
- All methods have advantages and disadvantages, but elliptic Fourier analysis is mostly a good choice

R

Example of outline data preparation

For a look at outline data preparation in R, we move on to exercise № 2

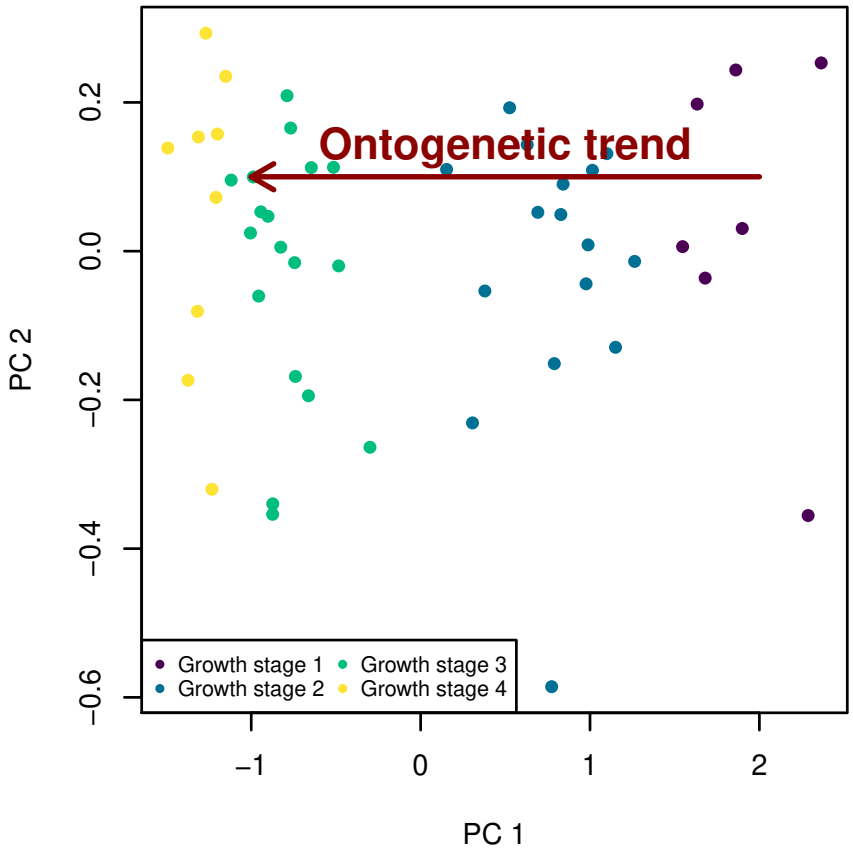
Open the exercise sheet for instructions and code examples

Section 3

Outline data analysis

We now have a multivariate dataset

- We can analyse the data in the same way as we analyse other multivariate data
  1. PCA to visualize differences
  2. Cluster analyses to define groups
  3. LDA/CVA to distinguish *a priori* defined groups
  4. MANOVA to test for between-group differences



Principal component analysis (PCA)

- PCA is an ideal method to display multivariate data like the many harmonics that result from any outline analysis

Principal component analysis (PCA)

- PCA is an ideal method to display multivariate data like the many harmonics that result from any outline analysis
- A PCA of harmonics is mostly performed on the variance–covariance matrix, as the harmonics are all measured on the same scale and have the same variance

Principal component analysis (PCA)

- PCA is an ideal method to display multivariate data like the many harmonics that result from any outline analysis
- A PCA of harmonics is mostly performed on the variance–covariance matrix, as the harmonics are all measured on the same scale and have the same variance
- The PCA allows to see the two or three most important components of the morphological description to see differences and/or clusters of points

Principal component analysis (PCA)

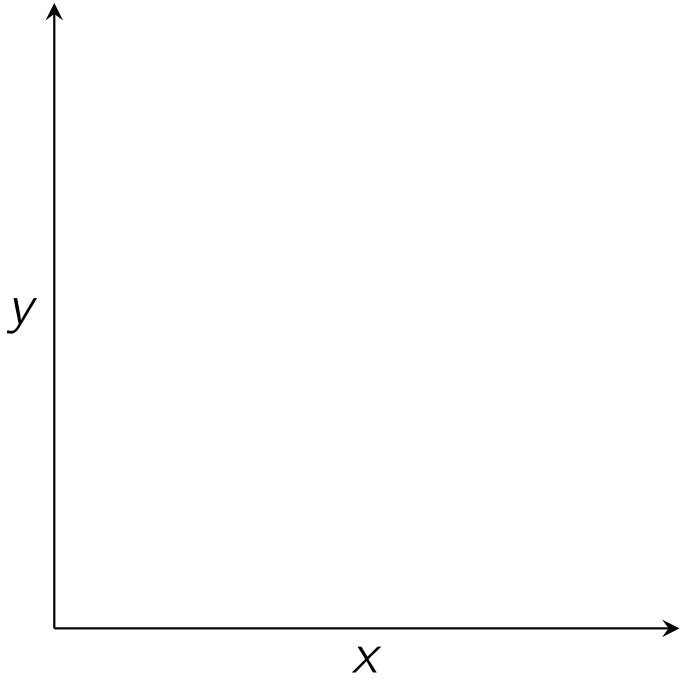
- PCA is an ideal method to display multivariate data like the many harmonics that result from any outline analysis
- A PCA of harmonics is mostly performed on the variance–covariance matrix, as the harmonics are all measured on the same scale and have the same variance
- The PCA allows to see the two or three most important components of the morphological description to see differences and/or clusters of points
- The PCA shows the **morphospace** of the analysed organisms

Principal component analysis (PCA)

- PCA is an ideal method to display multivariate data like the many harmonics that result from any outline analysis
- A PCA of harmonics is mostly performed on the variance–covariance matrix, as the harmonics are all measured on the same scale and have the same variance
- The PCA allows to see the two or three most important components of the morphological description to see differences and/or clusters of points
- The PCA shows the **morphospace** of the analysed organisms
- PCA is **only data visualization** for data mining, **not statistics**

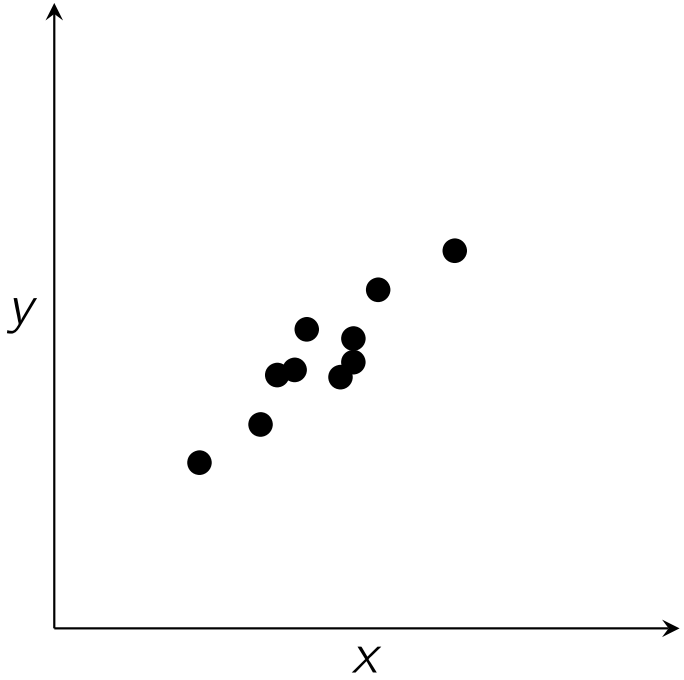
Principal component analysis

Example in 2 D



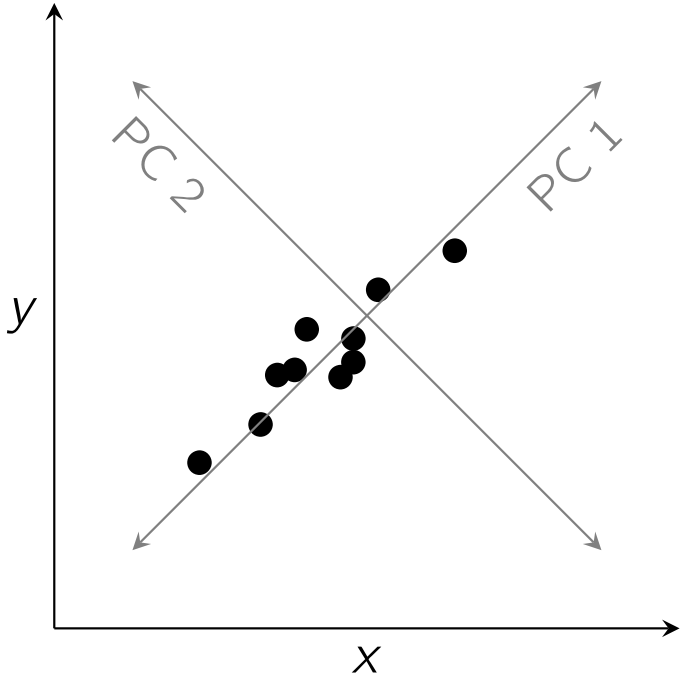
Principal component analysis

Example in 2 D



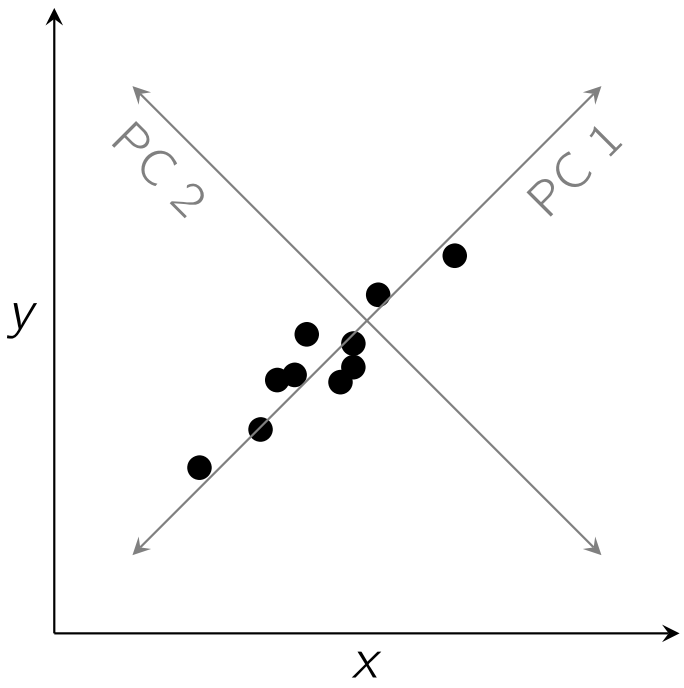
Principal component analysis

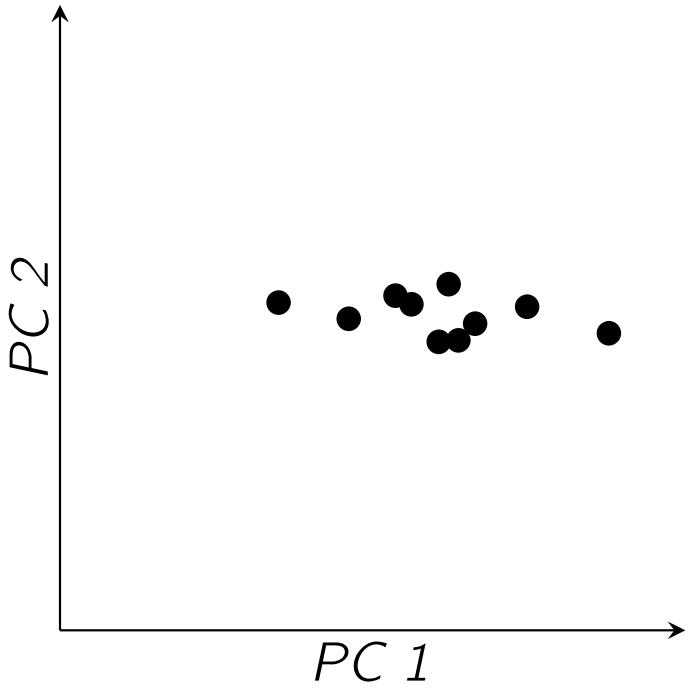
Example in 2 D



Principal component analysis

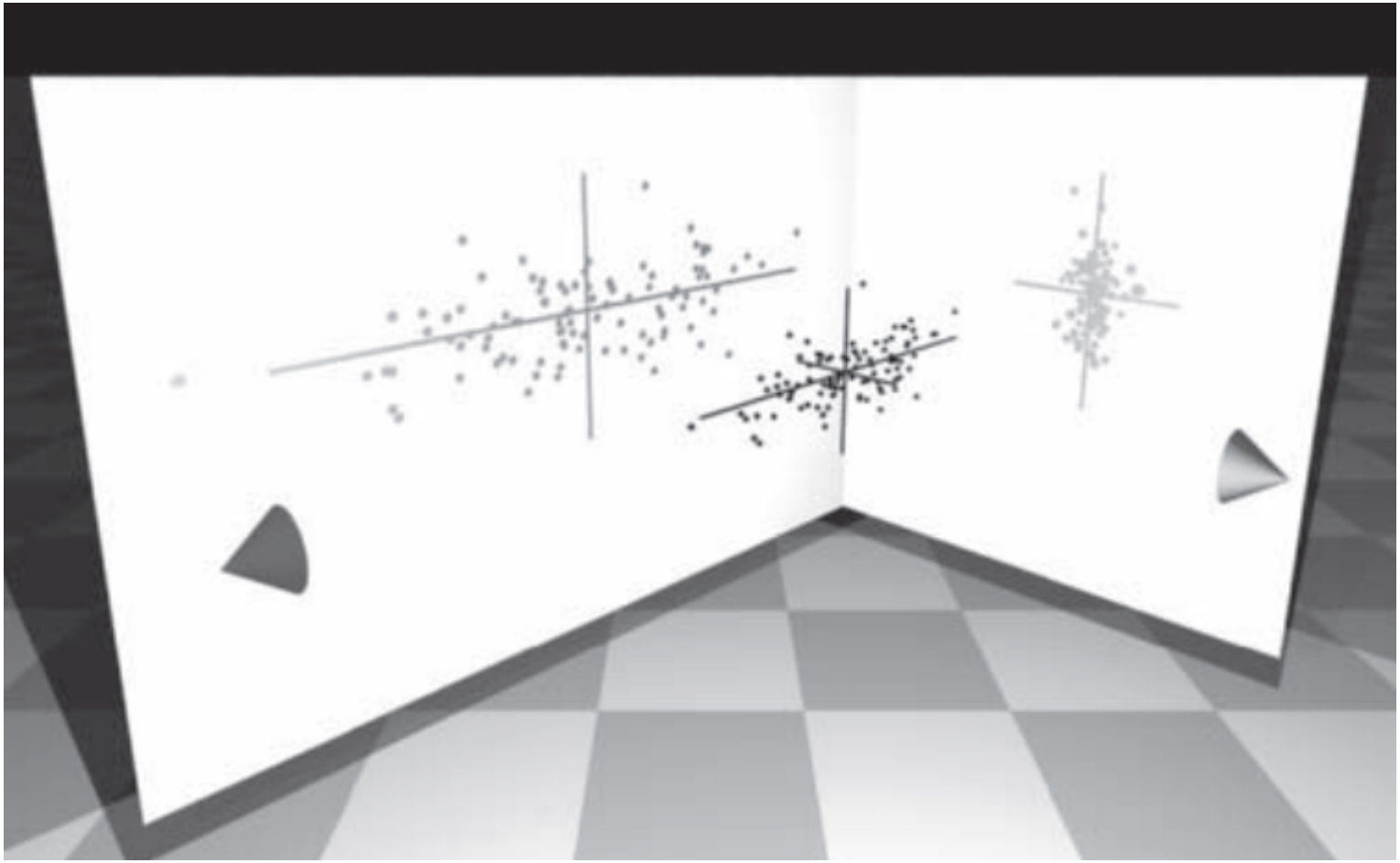
Example in 2 D





Principal component analysis

Example in 3 D



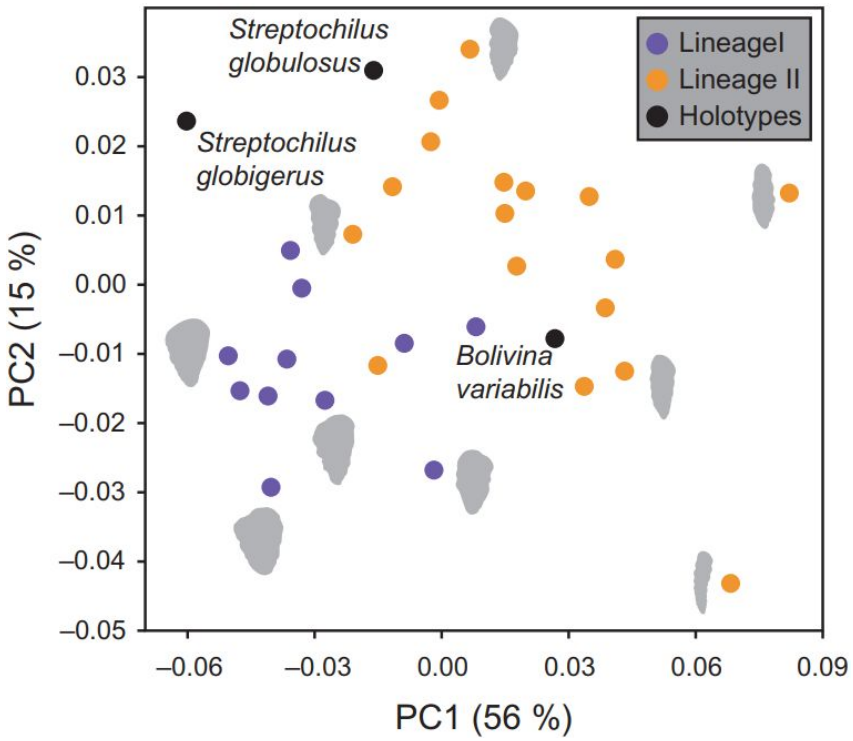
Hammer and Harper (2006) Palaeontological Data Analysis. (Blackwell Publishing: Malden, Oxford, Carlton)

Main output of a PCA

- **Scores**: The coordinates for each specimen in the entire PCA space (i.e. one value per PC)—corresponds to position of this specimen within the morphospace. **Scores can be used as variables for follow-up analyses**
- **Loadings**: The contribution of each parameter to each PC axis. In outline analysis, this is rarely interesting, but it can be used to check the dominance of Harmonic 1
- **Standard deviation of each PC axis**: This is helpful to estimate the degree of dispersion of data points along this axis

Interpretation of PCA

- PCA can be performed on the data without any prior knowledge of grouping
- If *a priori* groups are known, they can be indicated in the PCA plot but do not affect the ordination solution
- PCA can be investigated for visible group-clusters or trends, which may be further interpreted



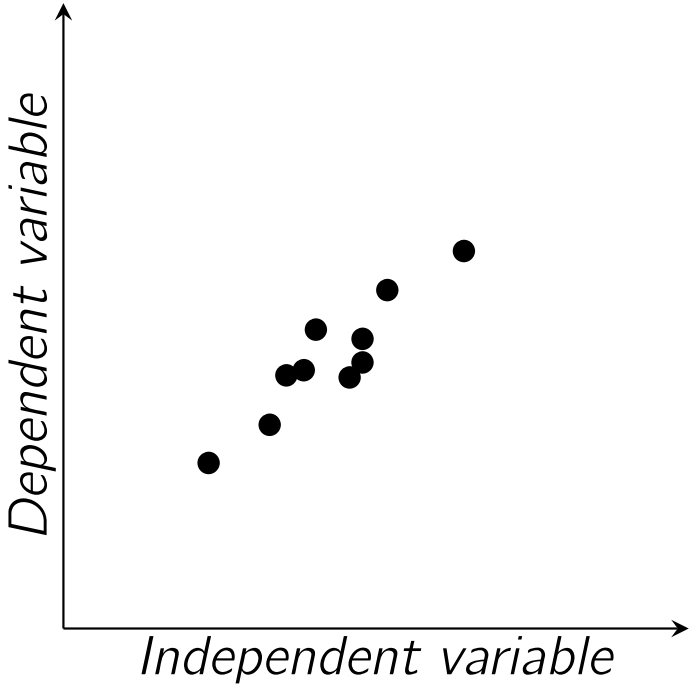
Kucera et al. (2017) J. Plankton. Res. 39 (3): 436–49

PC scores as analytical variable

- The scores of a morphometric PCA can also be used as variable for further analyses
- Normally, the scores of specimens along PC 1 (the axis that explains the greatest amount of variation) are used
- If you have good reason to believe that PC 1 is dominated by some underlying factor (e.g. environmental parameters), you can also use higher PCs for this kind of analysis
- Follow-up analyses can involve practically every type of standard analysis, but regression analyses are probably most common

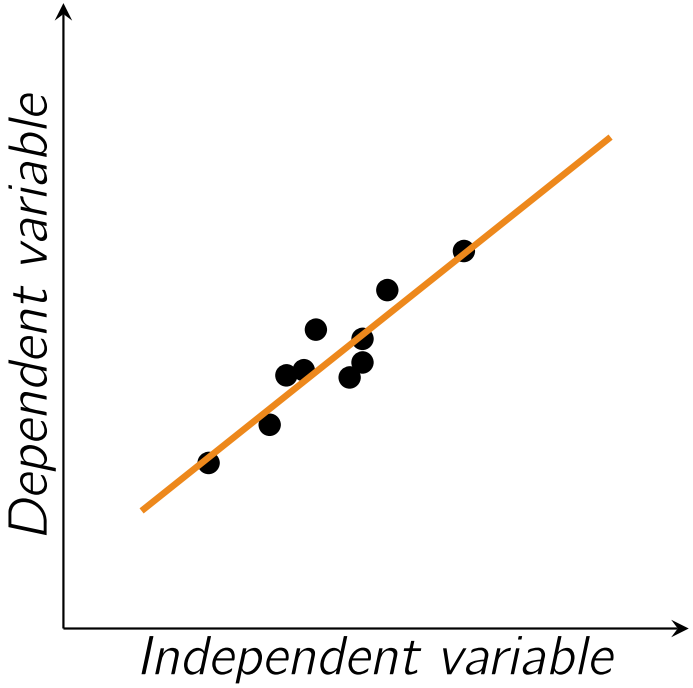
Regression analyses

- Regression checks for causal relationship between independent (x) and dependent (y) variable



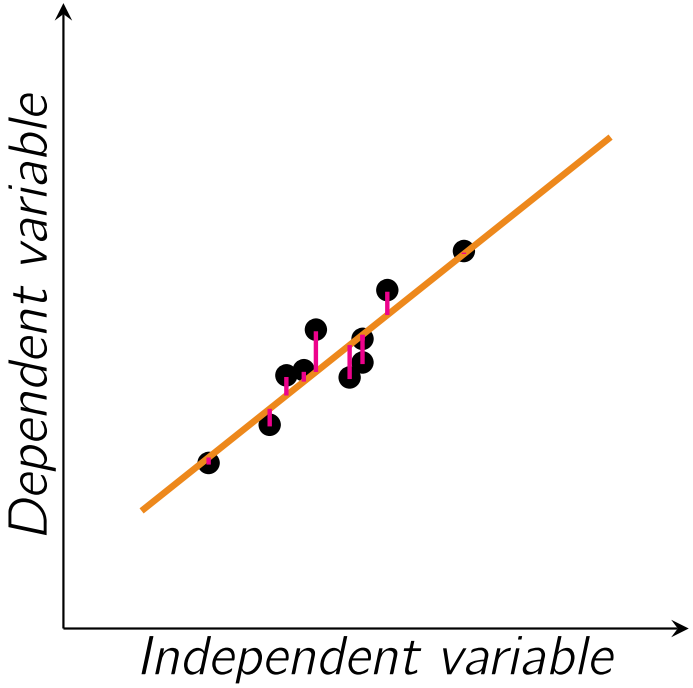
Regression analyses

- Regression checks for causal relationship between independent (x) and dependent (y) variable
- A predictive equation is fitted to the data



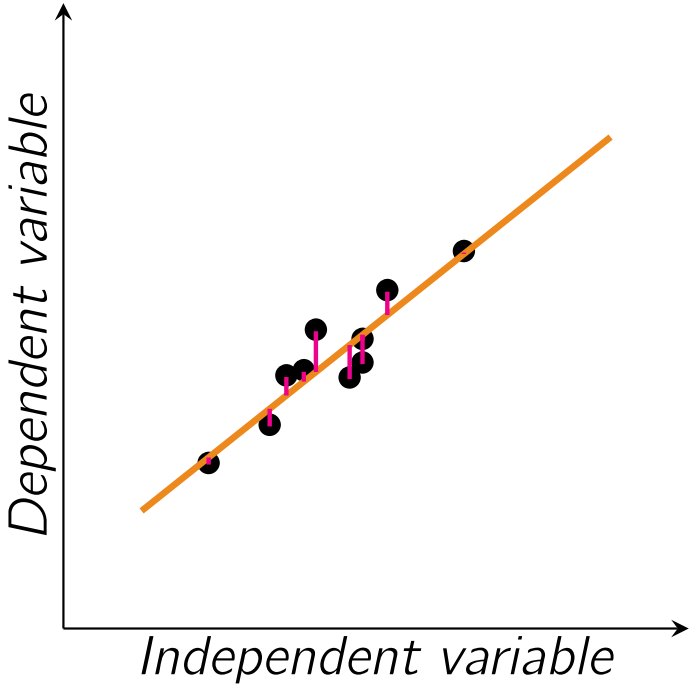
Regression analyses

- Regression checks for causal relationship between independent (x) and dependent (y) variable
- A predictive equation is fitted to the data with overall distance to all data points minimal

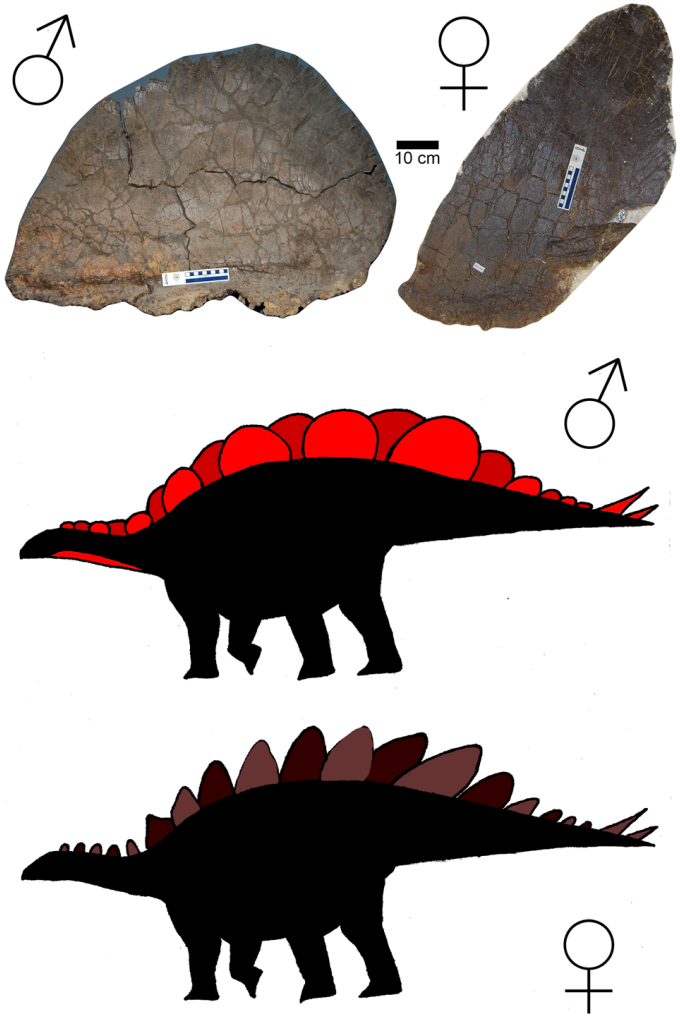


Regression analyses

- Regression checks for causal relationship between independent (x) and dependent (y) variable
- A predictive equation is fitted to the data with overall distance to all data points minimal
  - p-value: Probabilistic support of the null-hypothesis H_0
  - R^2-value (coefficient of determination): Fraction of variation explained
  - Correlation coefficient: Measure of effect size



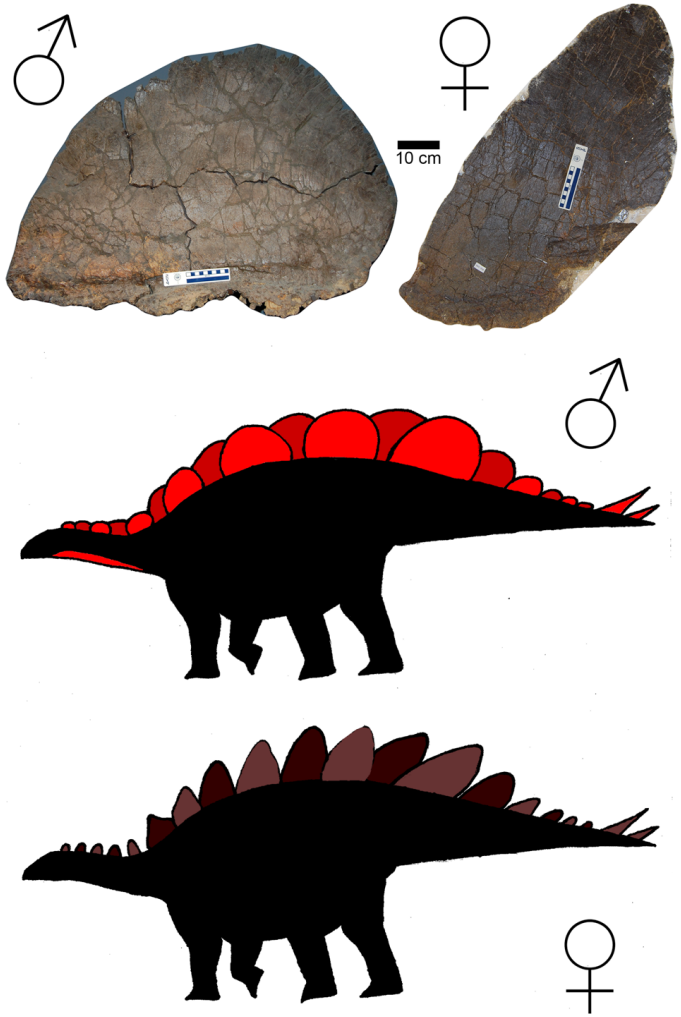
Excursion: The probabilistic *p*-value



http://viralportal.net

1. You start working on *Stegosaurus* and observe, that there are males and females to be found

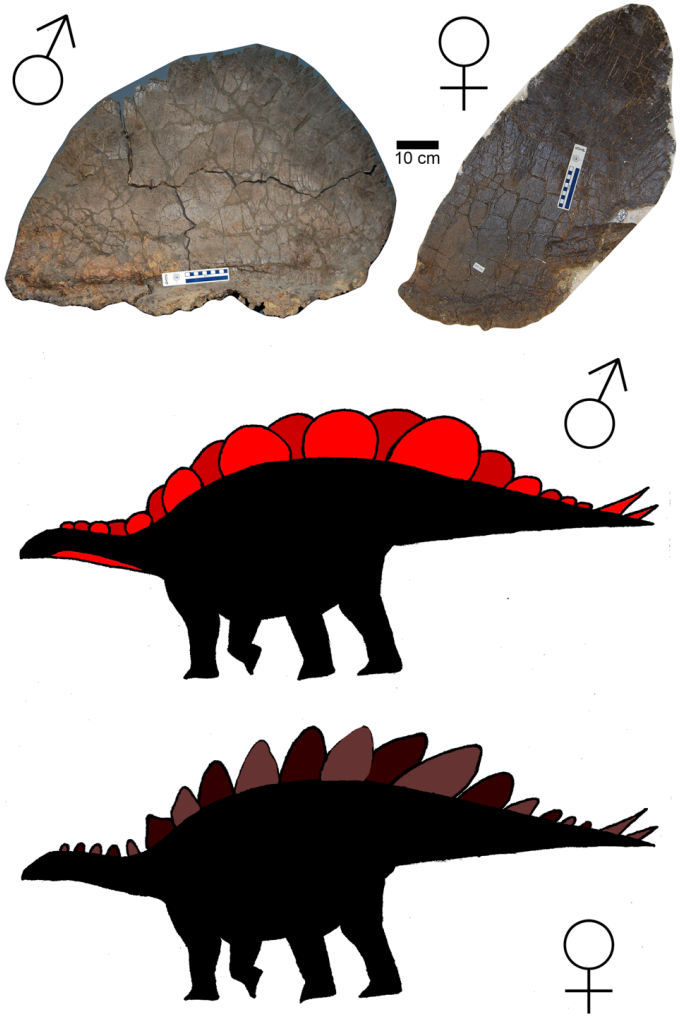
Excursion: The probabilistic *p*-value



http://viralportal.net

1. You start working on *Stegosaurus* and observe, that there are males and females to be found
2. You have the impression you find more females than males

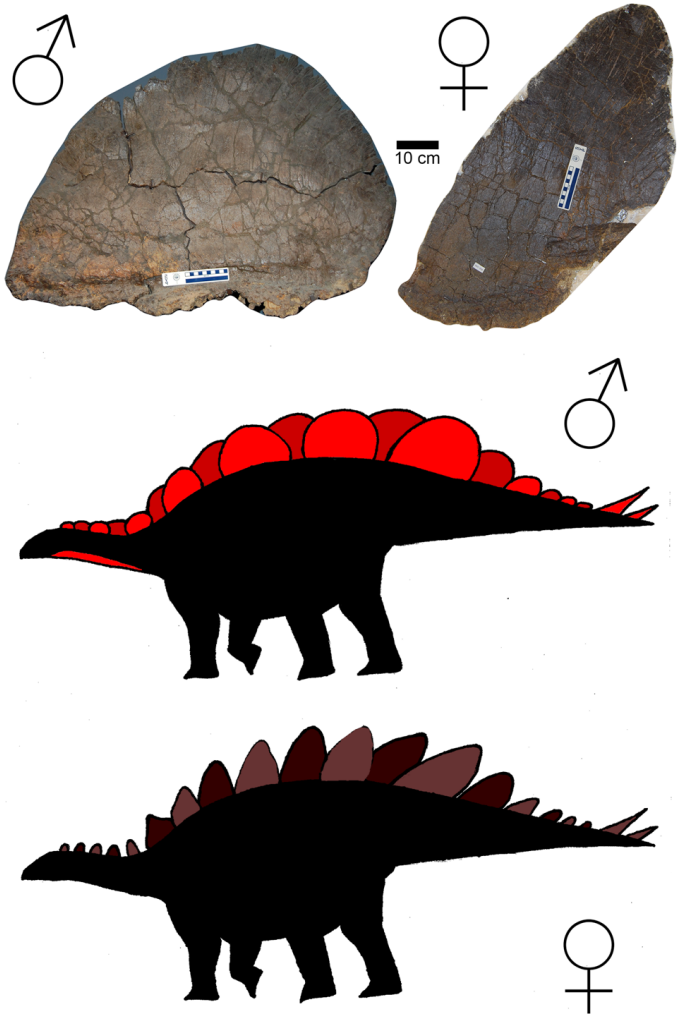
Excursion: The probabilistic *p*-value



http://viralportal.net

1. You start working on *Stegosaurus* and observe, that there are males and females to be found
2. You have the impression you find more females than males
3. You develop a hypothesis: The sex-ratio in *Stegosaurus* is not 1:1

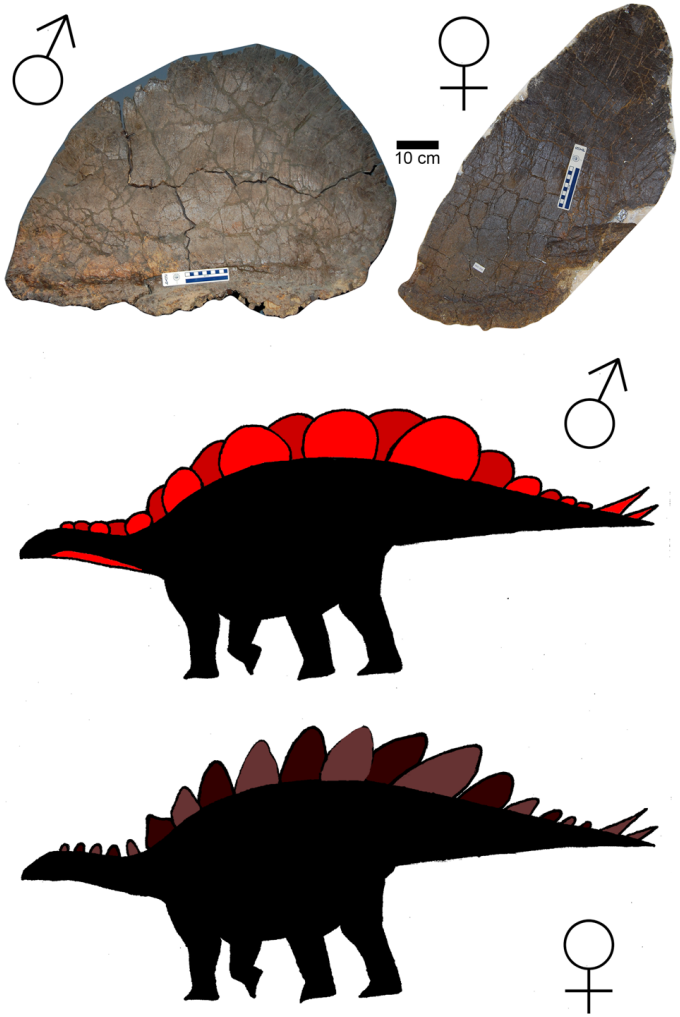
Excursion: The probabilistic *p*-value



http://viralportal.net

1. You start working on *Stegosaurus* and observe, that there are males and females to be found
2. You have the impression you find more females than males
3. You develop a hypothesis: The sex-ratio in *Stegosaurus* is not 1:1
4. You postulate the null-hypothesis: **H_0 = The sex-ratio in *Stegosaurus* is equal**

Excursion: The probabilistic *p*-value



http://viralportal.net

1. You start working on *Stegosaurus* and observe, that there are males and females to be found
2. You have the impression you find more females than males
3. You develop a hypothesis: The sex-ratio in *Stegosaurus* is not 1:1
4. You postulate the null-hypothesis: **H_0 = The sex-ratio in Stegosaurus is equal**
5. You try to **disprove** H_0

Excursion: The probabilistic *p*-value

The process of hypothesis evaluation

- You go in the field and calculate a statistic: For instance, you look at 10 *Stegosaurus* and count how many are male and female

Excursion: The probabilistic *p*-value

The process of hypothesis evaluation

- You go in the field and calculate a statistic: For instance, you look at 10 *Stegosaurus* and count how many are male and female
- You then calculate the probability to make this observation **under the assumption that H_0 is correct**

Excursion: The probabilistic *p*-value

The process of hypothesis evaluation

- You go in the field and calculate a statistic: For instance, you look at 10 *Stegosaurus* and count how many are male and female
- You then calculate the probability to make this observation **under the assumption that H_0 is correct**
- If this probability is lower than a predefined threshold, you assume that H_0 is incorrect and accept your original hypothesis

Excursion: The probabilistic *p*-value

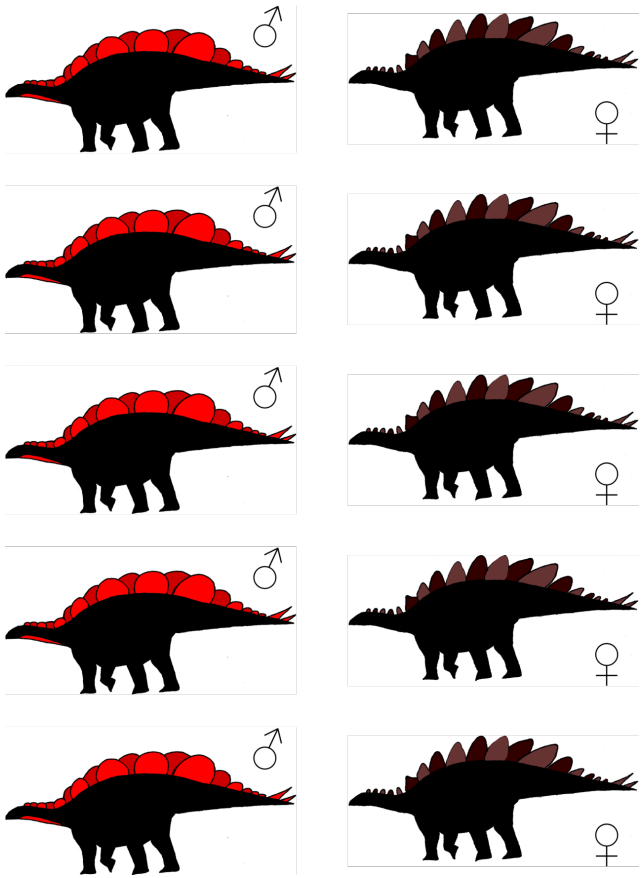
The process of hypothesis evaluation

- You go in the field and calculate a statistic: For instance, you look at 10 *Stegosaurus* and count how many are male and female
- You then calculate the probability to make this observation **under the assumption that H_0 is correct**
- If this probability is lower than a predefined threshold, you assume that H_0 is incorrect and accept your original hypothesis
- In natural sciences, this threshold is normally $\alpha = 0.05$

Excursion: The probabilistic *p*-value

The sex-distribution in *Stegosaurus*

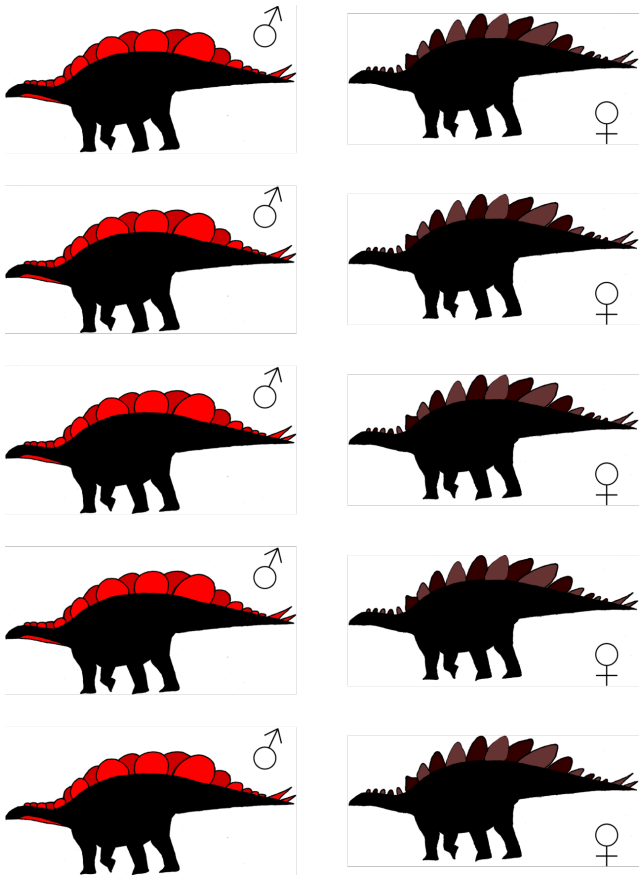
Assumption of H_0



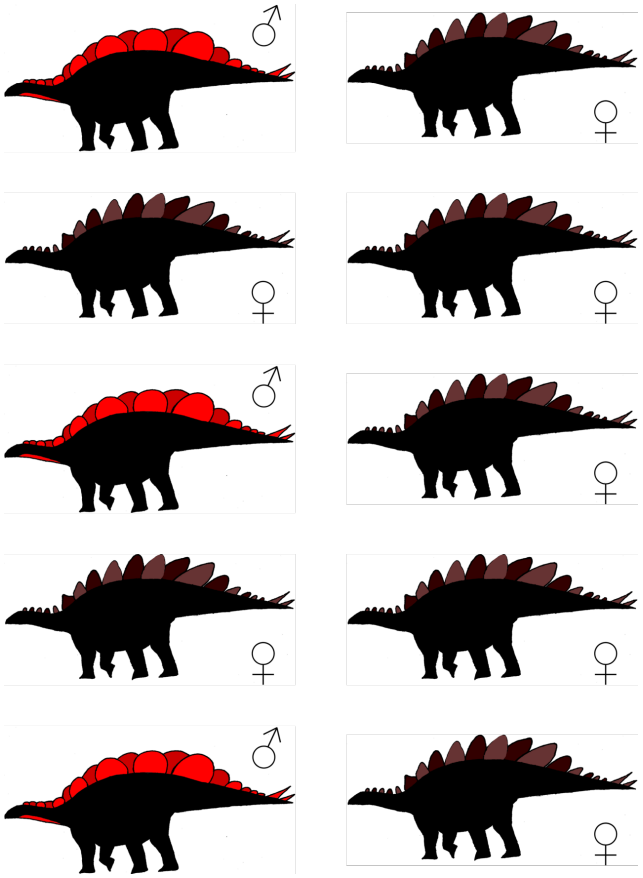
Excursion: The probabilistic *p*-value

The sex-distribution in *Stegosaurus*

Assumption of H_0



Observation



Excursion: The probabilistic p-value

How do we interpret that?

The binomial probability is defined as

$$p(x) = \binom{n}{x} \times p_0^x \times (1 - p_0)^{n-x}$$

Excursion: The probabilistic p-value

How do we interpret that?

The binomial probability is defined as

$$p(x) = \binom{n}{x} \times p_0^x \times (1 - p_0)^{n-x}$$

In our example: $x = 3$ male *Stegosaurus*

$$p(3) = \frac{10!}{3! \times (10-3)!} \times 0.5^3 \times 0.5^{10-3}$$

$$p(3) = \frac{362880}{2 \times 720} \times 0.125 \times 0.0078125$$

$$p(3) = 0.2460938$$

Excursion: The probabilistic p-value

How do we interpret that?

The binomial probability is defined as

$$p(x) = \binom{n}{x} \times p_0^x \times (1 - p_0)^{n-x}$$

In our example: $x = 3$ male *Stegosaurus*

$$p(3) = \frac{10!}{3! \times (10-3)!} \times 0.5^3 \times 0.5^{10-3}$$

$$p(3) = \frac{362880}{2 \times 720} \times 0.125 \times 0.0078125$$

$$p(3) = 0.2460938$$

If H_0 was true, we would observe a result like this in $c.25\,\%$ of all cases

Excursion: The probabilistic *p*-value

How do we interpret that?

The binomial probability is defined as

$$p(x) = \binom{n}{x} \times p_0^x \times (1 - p_0)^{n-x}$$

In our example: $x = 3$ male *Stegosaurus*

$$p(3) = \frac{10!}{3! \times (10-3)!} \times 0.5^3 \times 0.5^{10-3}$$

$$p(3) = \frac{362880}{2 \times 720} \times 0.125 \times 0.0078125$$

$$p(3) = 0.2460938$$

If H_0 was true, we would observe a result like this in *c*.25 % of all cases

We observe that $p > 0.05$ and therefore accept H_0

Excursion: The probabilistic p-value

Beware of the limitations

- The p-value gives you a probability, no clear answer: If you choose $\alpha = 0.05$ you will falsely reject the null-hypothesis in 5 % of all cases (Type I error)

Excursion: The probabilistic p-value

Beware of the limitations

- The p-value gives you a probability, no clear answer: If you choose $\alpha = 0.05$ you will falsely reject the null-hypothesis in 5 % of all cases (Type I error)
- There is also a chance to falsely accept the null-hypothesis although it is false (Type II error)—this depends on the power of your test

Excursion: The probabilistic p-value

Beware of the limitations

- The p-value gives you a probability, no clear answer: If you choose $\alpha = 0.05$ you will falsely reject the null-hypothesis in 5 % of all cases (Type I error)
- There is also a chance to falsely accept the null-hypothesis although it is false (Type II error)—this depends on the power of your test
- The equations will become much more complex for other data distributions and more complex tests: Do not calculate by hand, that is what software is for

Excursion: The probabilistic p-value

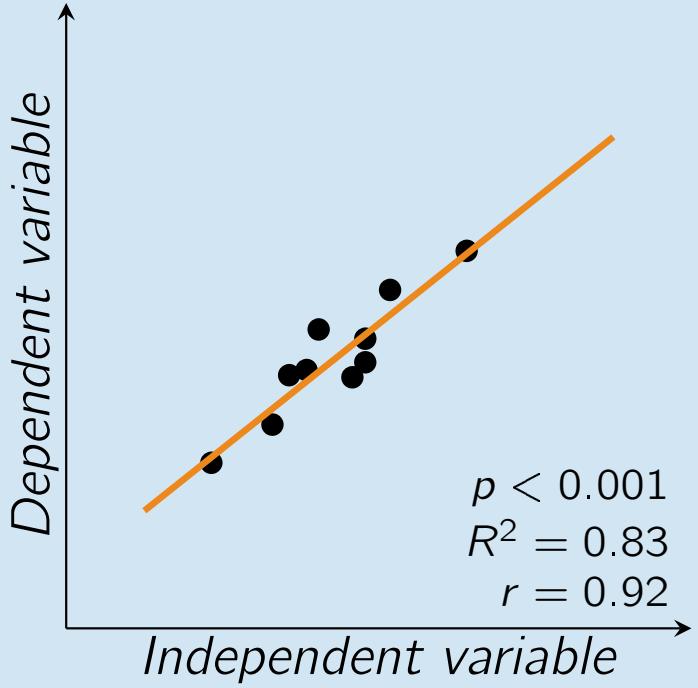
Beware of the limitations

- The p-value gives you a probability, no clear answer: If you choose $\alpha = 0.05$ you will falsely reject the null-hypothesis in 5 % of all cases (Type I error)
- There is also a chance to falsely accept the null-hypothesis although it is false (Type II error)—this depends on the power of your test
- The equations will become much more complex for other data distributions and more complex tests: Do not calculate by hand, that is what software is for
- Besides frequentist statistics there exist maximum likelihood statistics and Bayesian statistics: They have advantages and disadvantages of their own

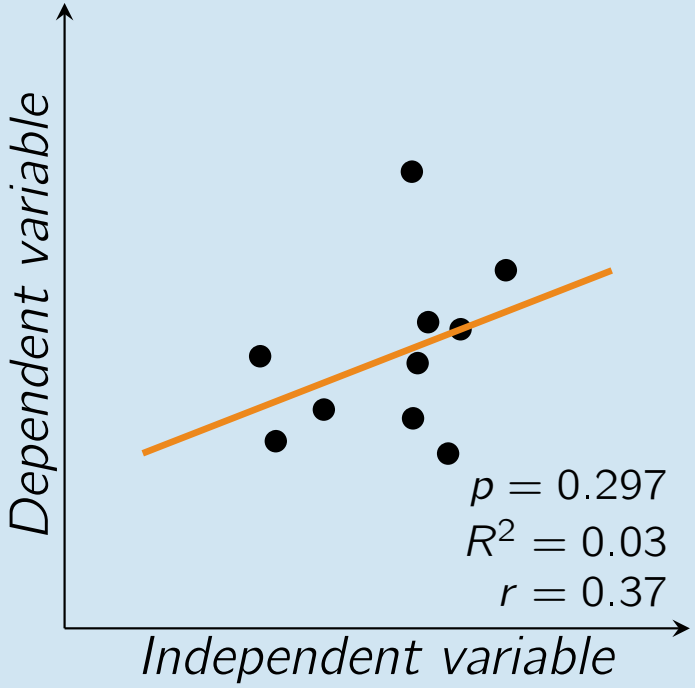
Regression analyses

How to interpret regression results

Low spread, significant



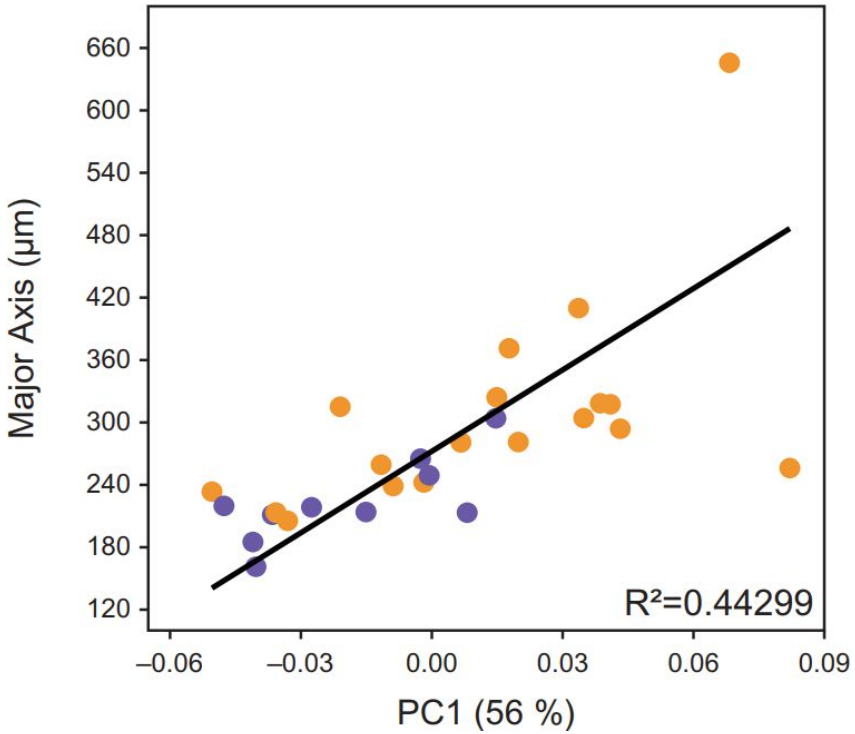
Large spread, insignificant



Regression analyses in morphometrics

Checking for ontogenetic trends

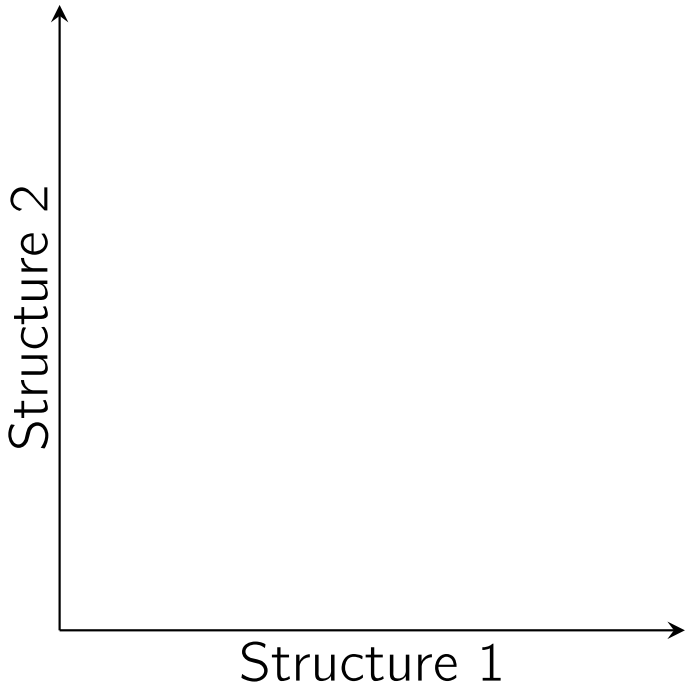
- Investigate the relationship of morphology with other parameters
  1. With environmental factors: **Ecophenotypes**
  2. With size parameter: **Ontogenetic trends**
  3. With geological sample age: **Anagenetic trends**
  4. With any other continuous (and technically ordinal) variable



Kucera et al. (2017) J. Plankton. Res. 39 (3): 436–49

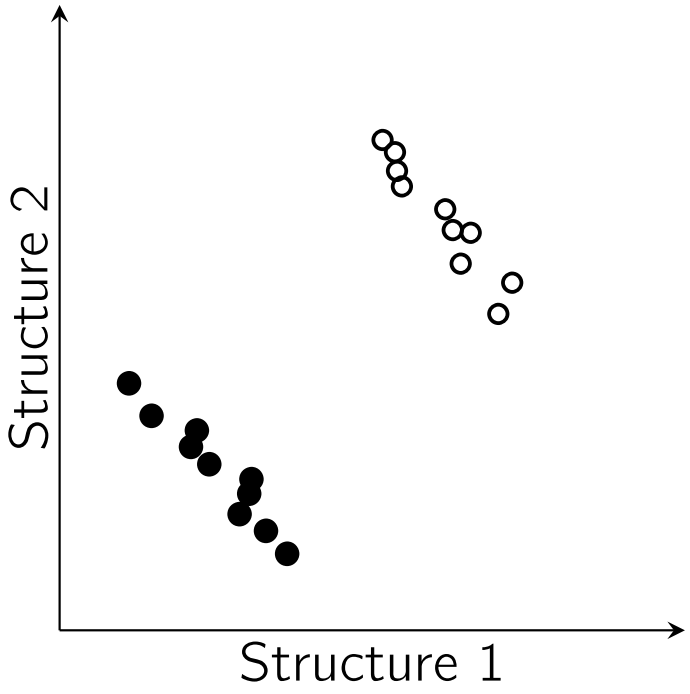
Linear discriminant analysis

Example in 2 D



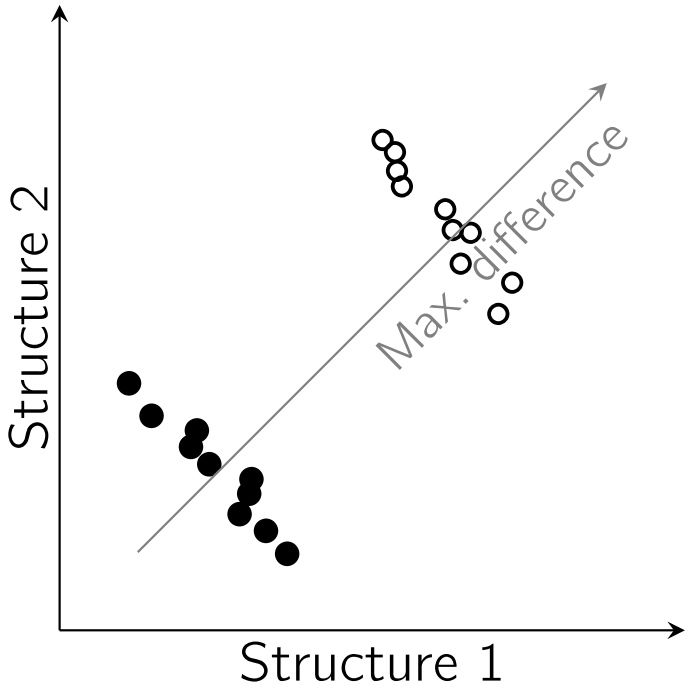
Linear discriminant analysis

Example in 2 D



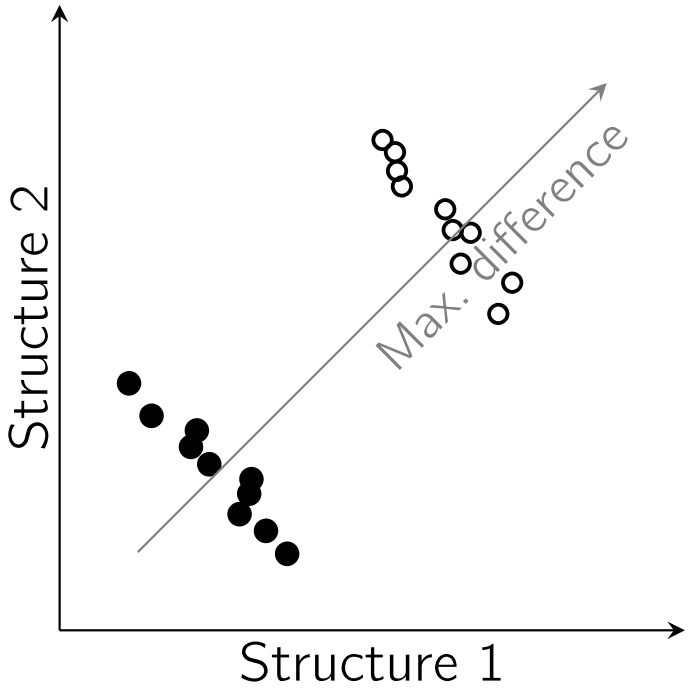
Linear discriminant analysis

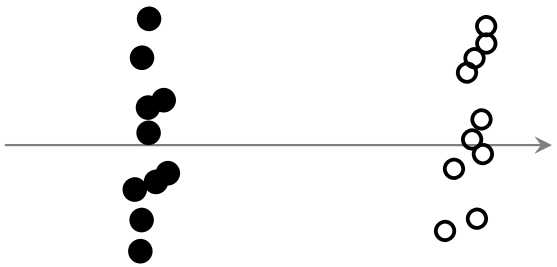
Example in 2 D



Linear discriminant analysis

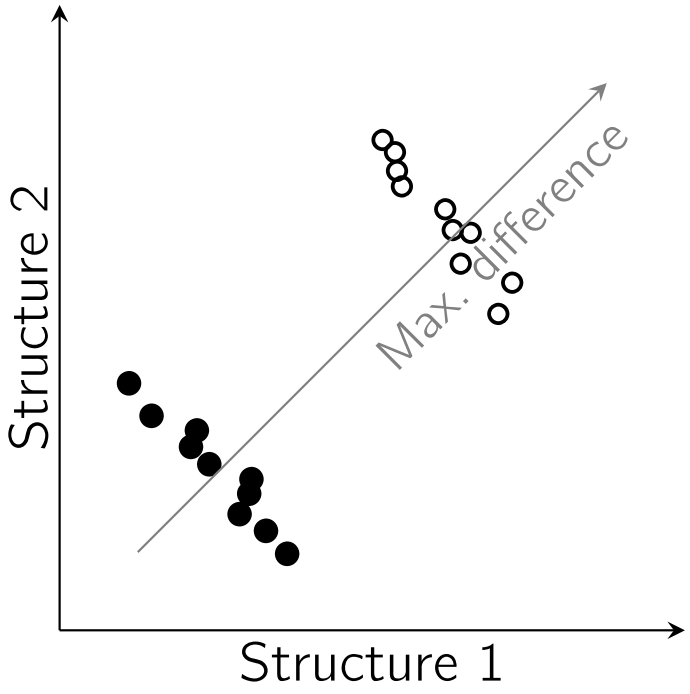
Example in 2 D

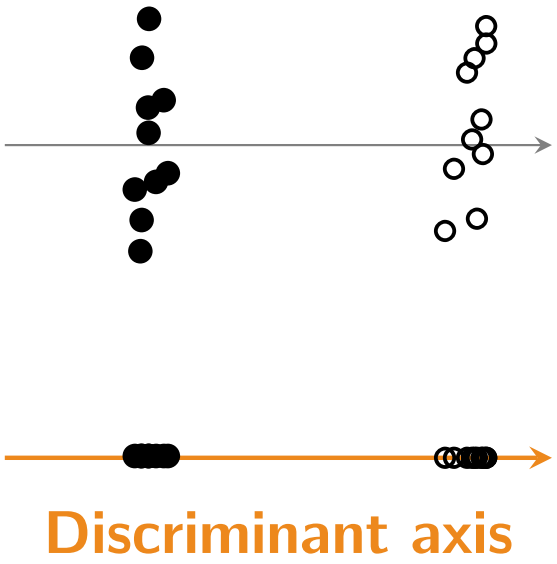




Linear discriminant analysis

Example in 2 D





Linear discriminant analysis

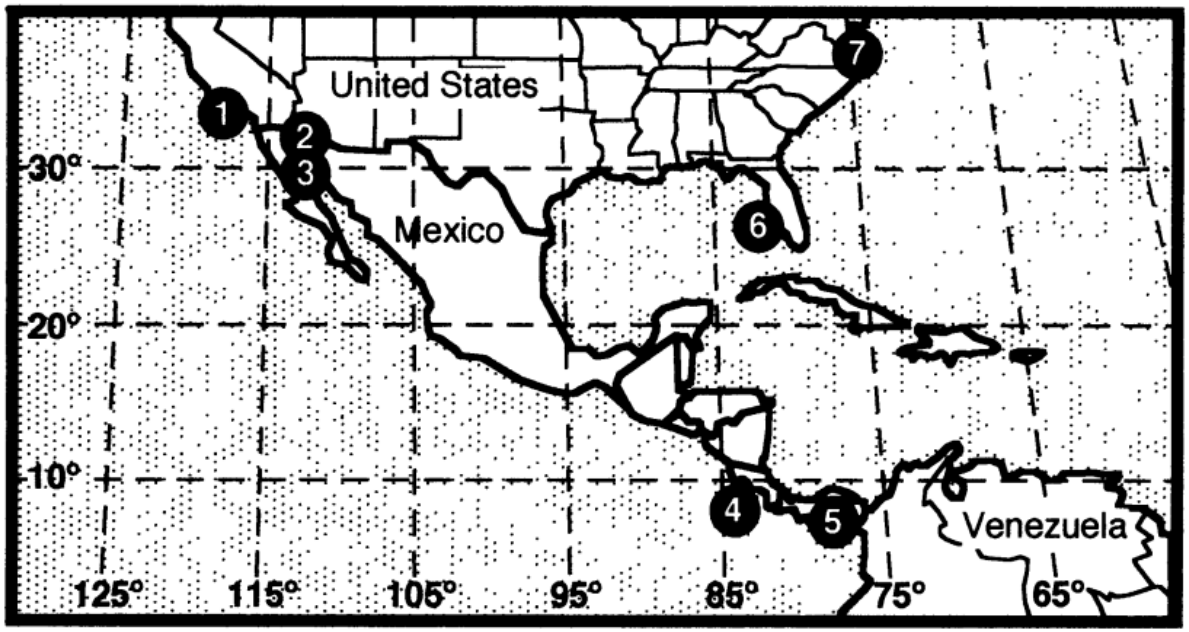
Example: Brachiopods of the genus *Glottidia*

http://paleopolis.rediris.es

Linear discriminant analysis

Example: Brachiopods of the genus *Glottidia*

http://paleopolis.rediris.es

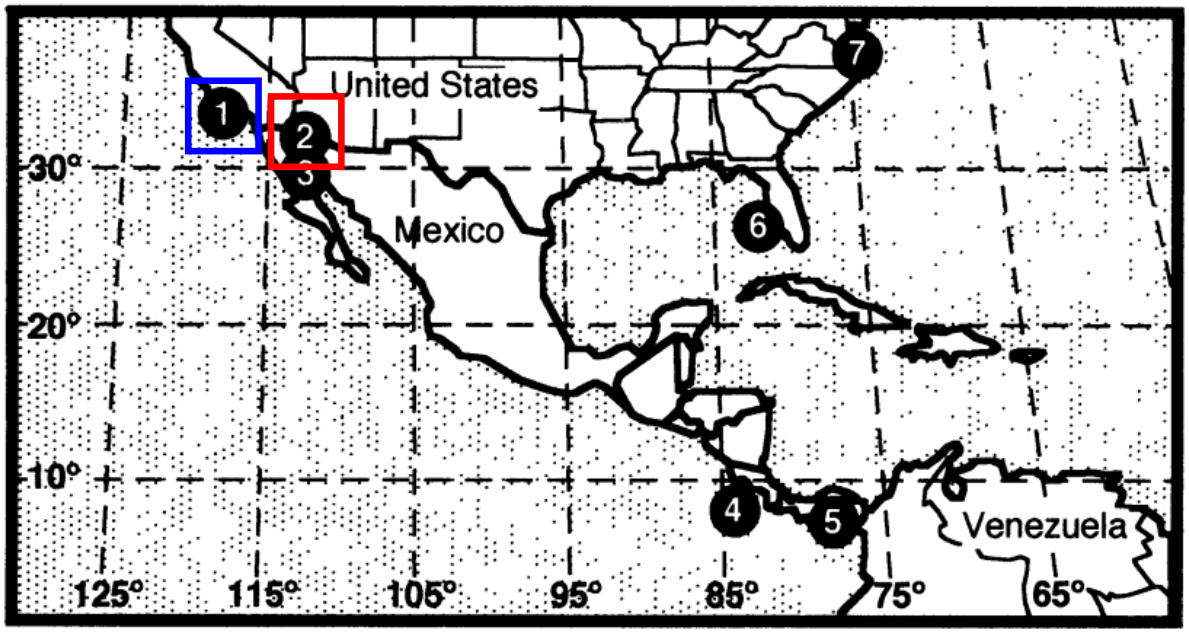


Kowalewski et al. (1997) Paleobiology 23 (4): 444–69

Linear discriminant analysis

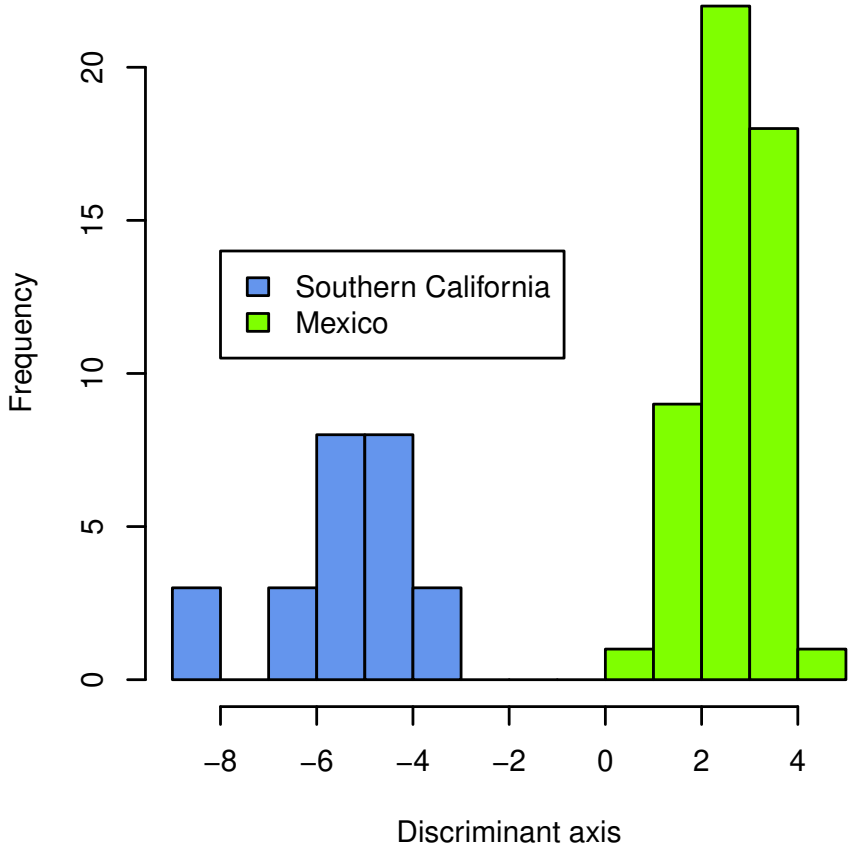
Example: Brachiopods of the genus *Glottidia*

http://paleopolis.rediris.es



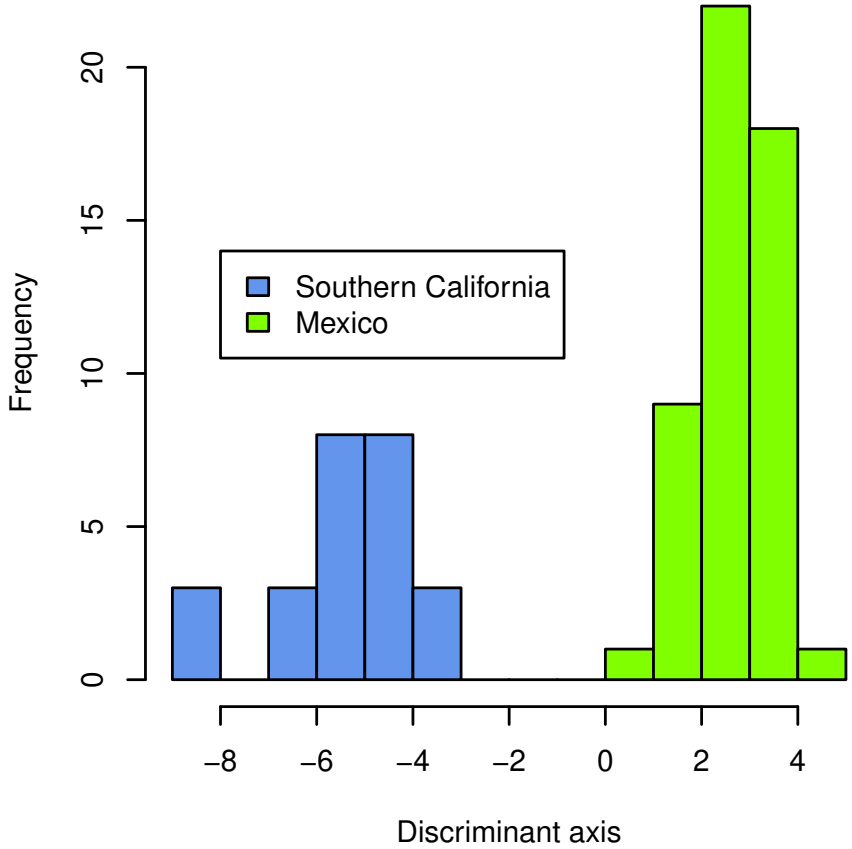
Kowalewski et al. (1997) Paleobiology 23 (4): 444–69

LDA: Compare morphology between two groups



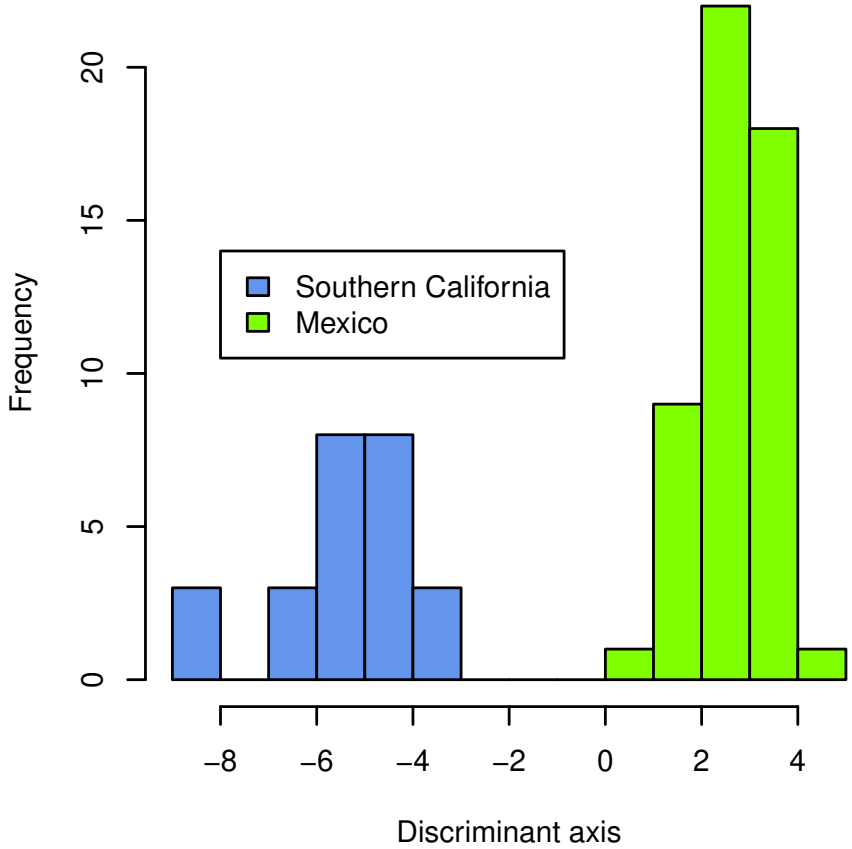
- The LDA returns:
  1. A graphic result of the distribution of specimens of both groups along the discriminant axis → a clear difference in positions of the groups is a good sign

LDA: Compare morphology between two groups



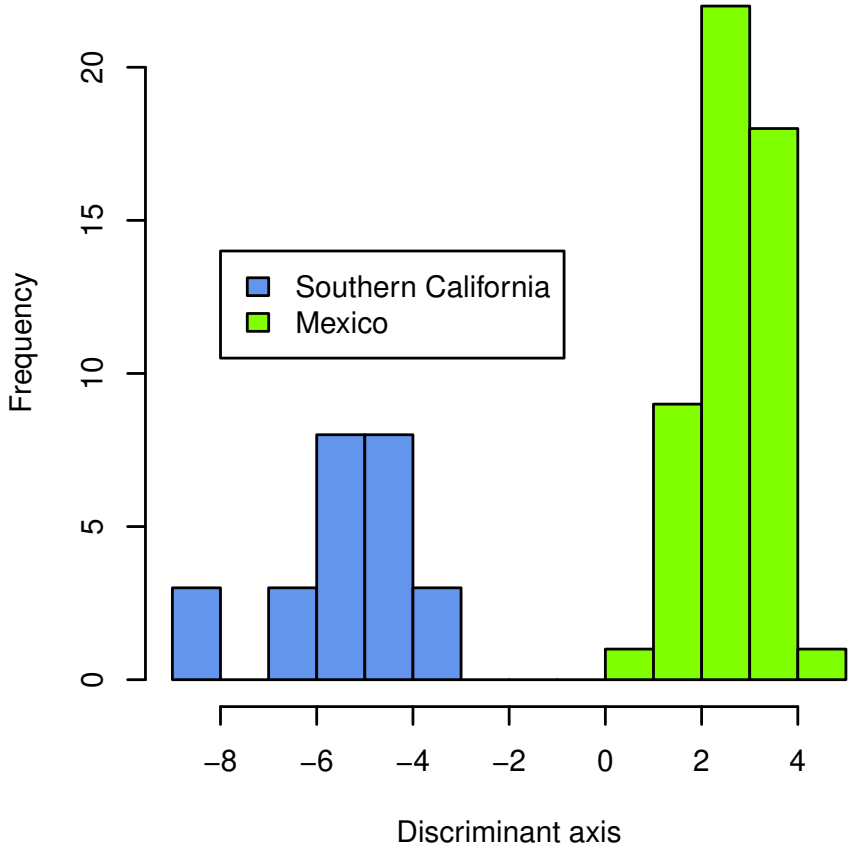
- The LDA returns:
  1. A graphic result of the distribution of specimens of both groups along the discriminant axis → a clear difference in positions of the groups is a good sign
  2. The *a posteriori* allocations of specimens to groups to validate the quality of the LDA solution

LDA: Compare morphology between two groups



- For a statistical evaluation of group differences, we use **M**ultivariate **AN**alysis **O**f **VA**riances (MANOVA)

LDA: Compare morphology between two groups

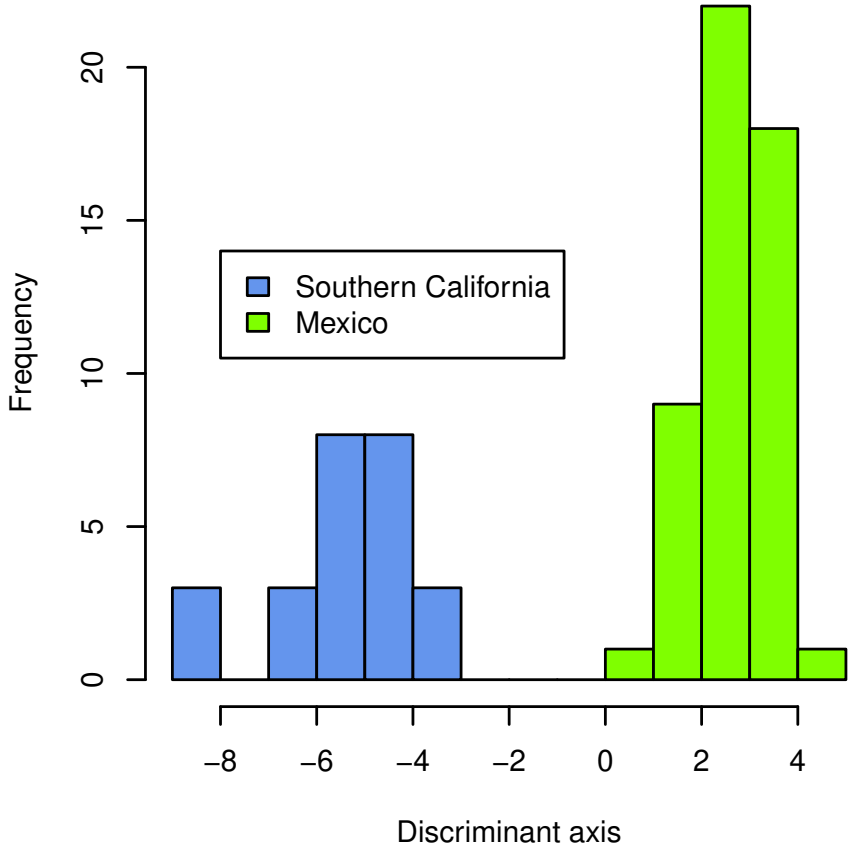


- For a statistical evaluation of group differences, we use **M**ultivariate **AN**alysis **O**f **VA**riances (MANOVA)

$$\text{Pillai} = 0.938$$

$$p < 0.001$$

LDA: Compare morphology between two groups



- For a statistical evaluation of group differences, we use **M**ultivariate **AN**alysis **O**f **VA**riances (MANOVA)

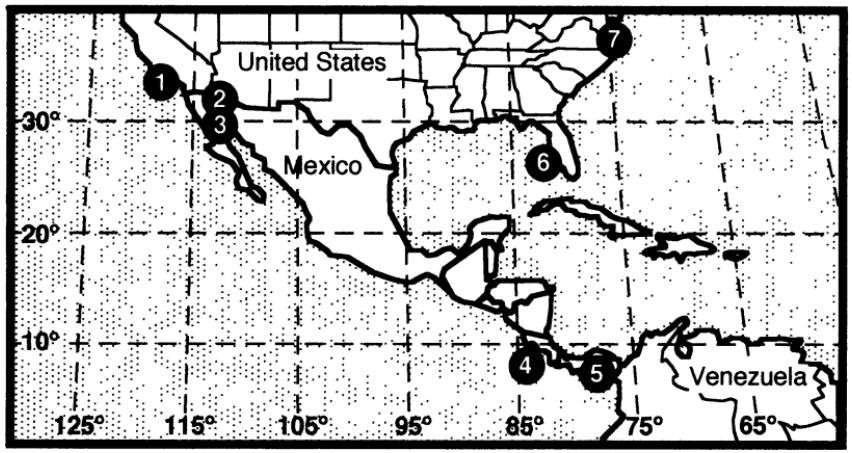
$$\text{Pillai} = 0.938$$

$$p < 0.001$$

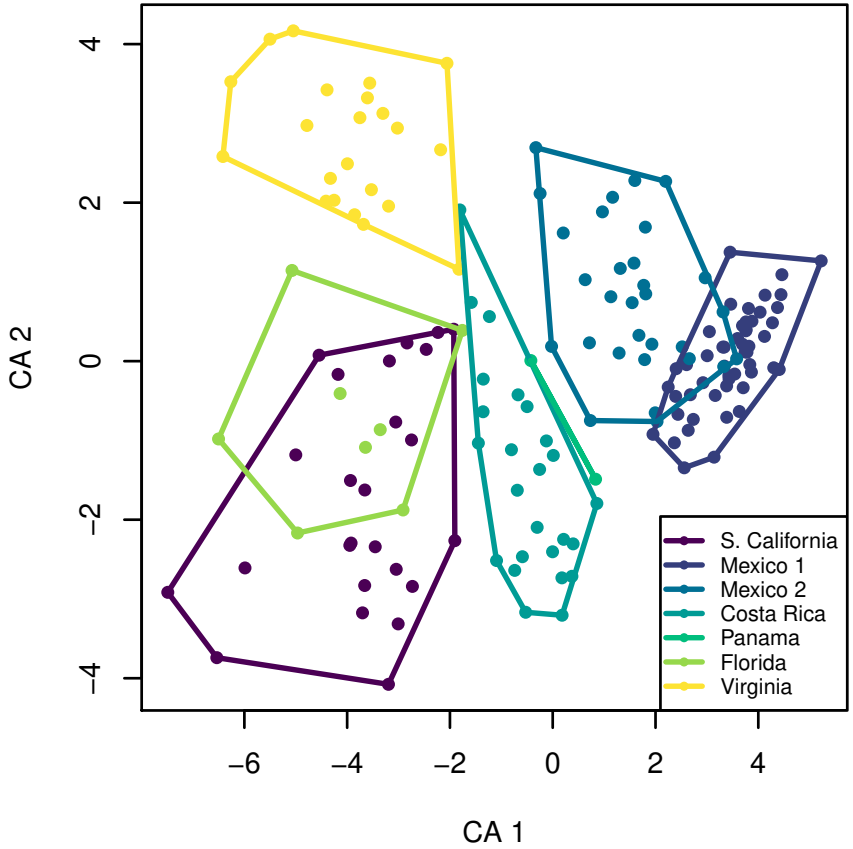
The brachipods from southern California and Mexico differ significantly in their morphology

Canonical variates analysis

The extension of LDA for more than two groups



Kowalewski et al. (1997) Paleobiology 23 (4): 444–69



Of sharpshooters and misconceptions

- LDA/CVA and MANOVA can only be used when the groups were defined **a priori** and **independently** of the morphological measures

Of sharpshooters and misconceptions

- LDA/CVA and MANOVA can only be used when the groups were defined **a priori** and **independently** of the morphological measures
- DO: Test for instance if groups from different localities/times or species defined based on genetics differ morphologically

Of sharpshooters and misconceptions

- LDA/CVA and MANOVA can only be used when the groups were defined **a priori** and **independently** of the morphological measures
- DO: Test for instance if groups from different localities/times or species defined based on genetics differ morphologically
- **DO NOT:** Test if groups differ morphologically that you defined **based on the same morphological data** (e.g. via PCA)

R

Example of outline data analysis

For a look at outline data analysis in R, we move on to exercise № 3

Open the exercise sheet for instructions and code examples