The Data Science Revolution in Biomechanics:

Traditional statistical tests *vs* modern machine learning methods in the study of lizard locomotion

Kathleen L. Foster, Ball State University
Alessandro M. Selvitella, Purdue University Fort Wayne

BALL STATE UNIVERSITY

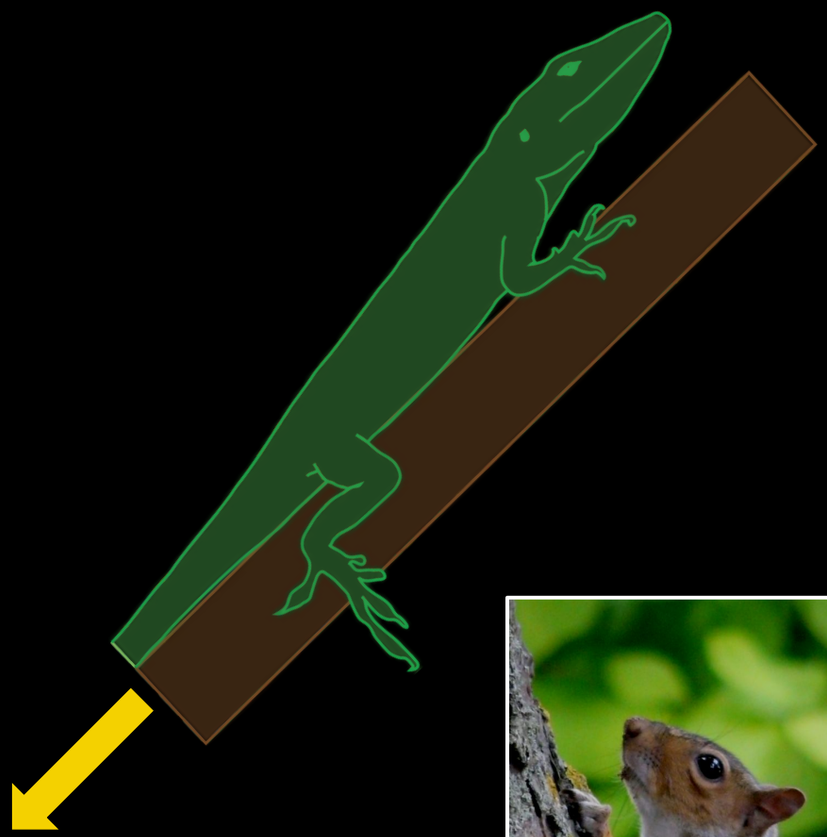
PURDUE

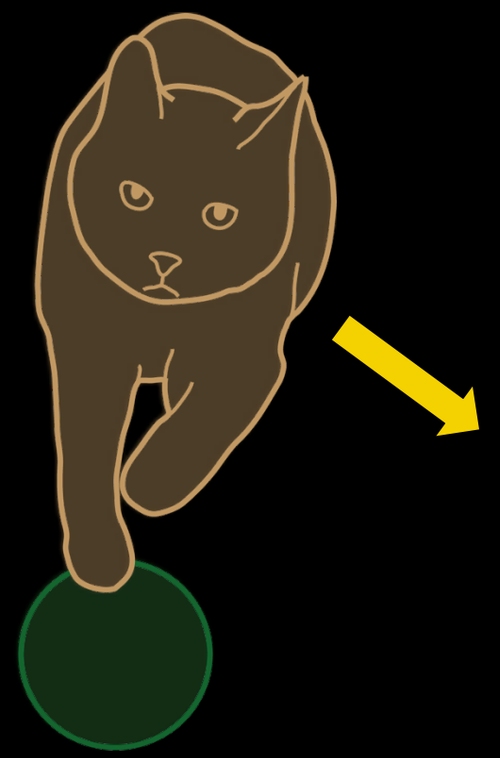
The environment is complex

Animals do all sorts of wild and wacky things

The challenges posed by arboreality

The challenges posed by arboreality





Morphological adaptation meets functional demand

primates

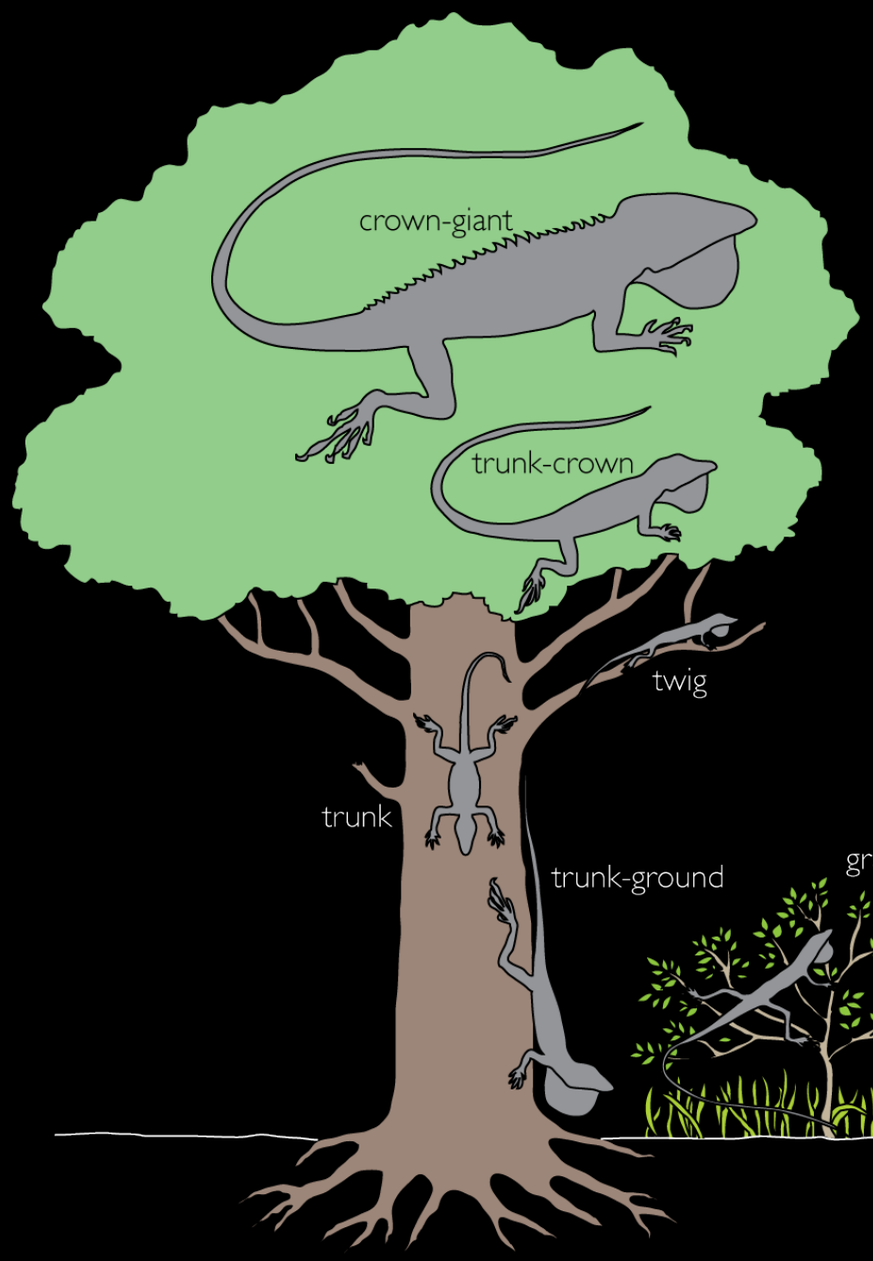
rodents

lizards

marsupials

kinkajous

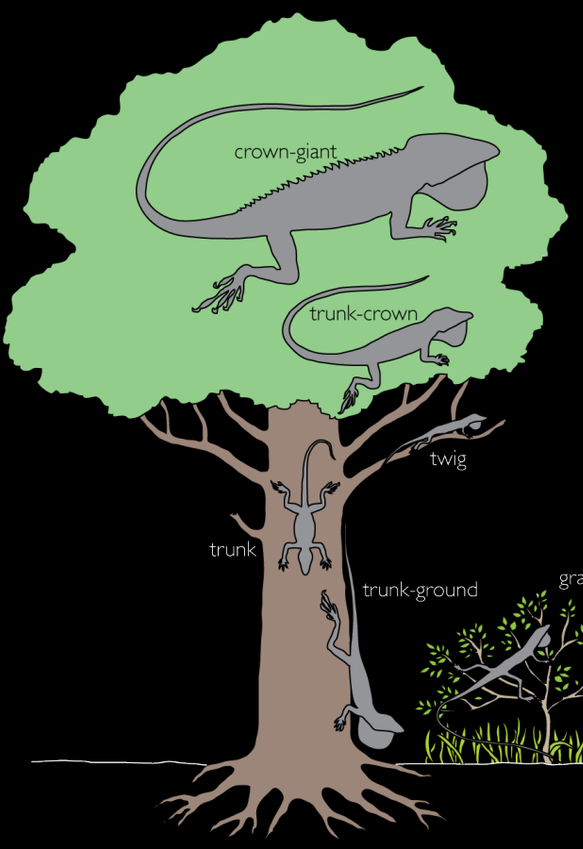
Morphological adaptation meets functional demand



Anolis ecomorphs

Losos (2009) *Lizards in an evolutionary tree*

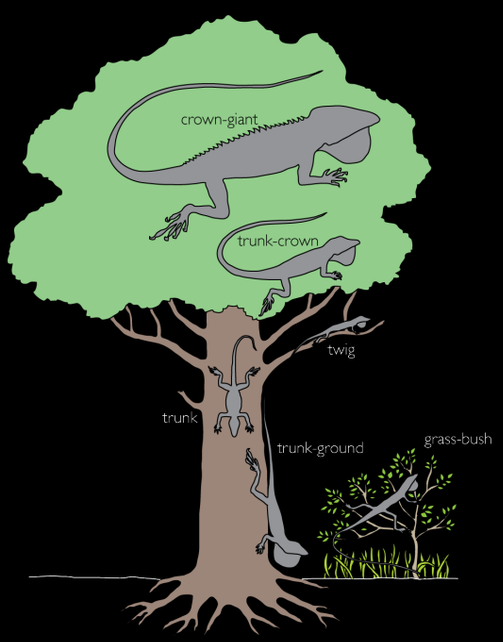
Morphological adaptation meets functional demand



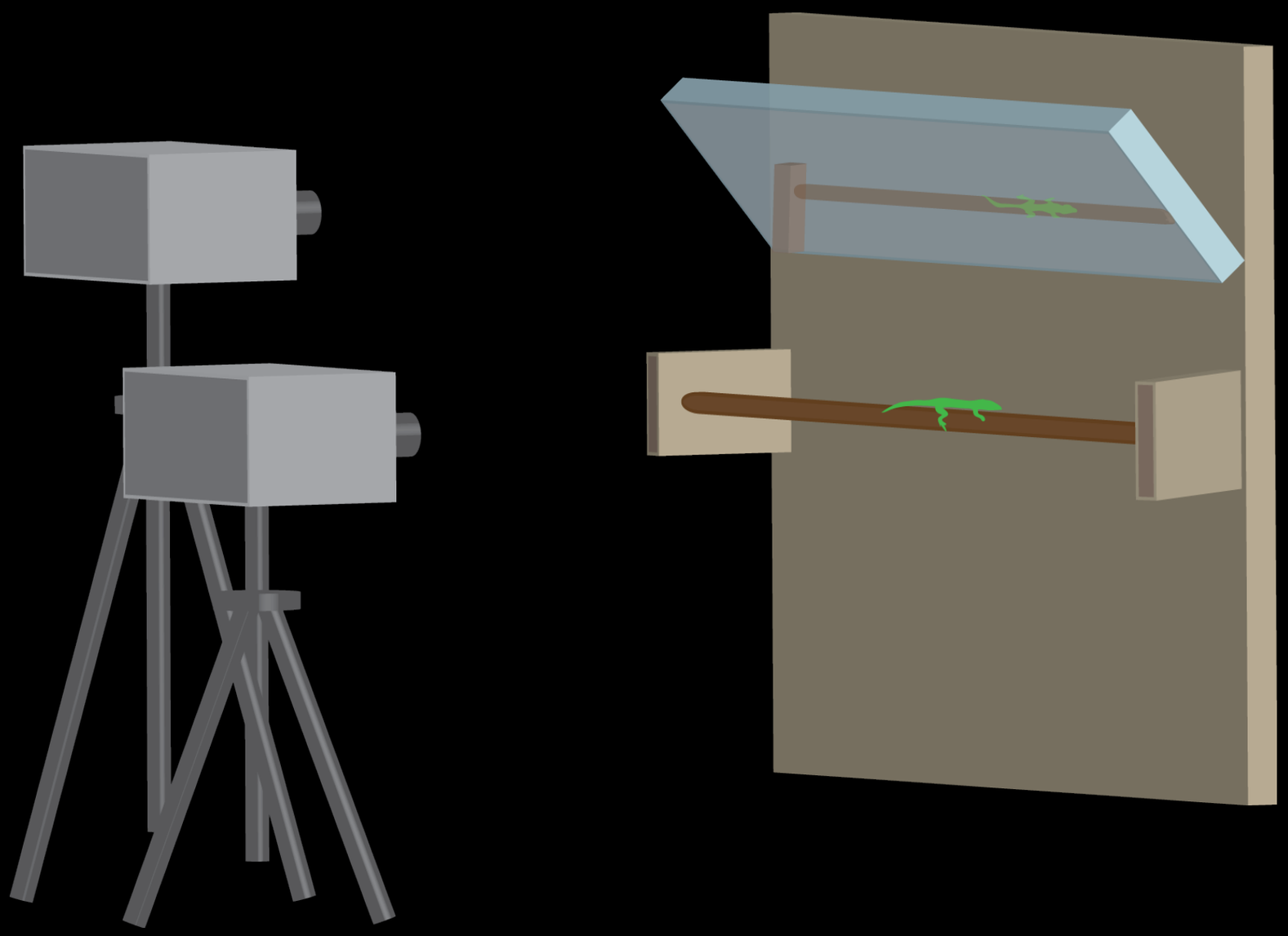
How does the environment impact locomotor behavior of *Anolis* lizards?

How does the environment impact locomotor behavior of *Anolis* lizards?

- How do the **movements** of the forelimbs and hind limbs change on different arboreal surfaces?
- How does the **coordination** of forelimb and hind limb joints change on different arboreal surfaces?

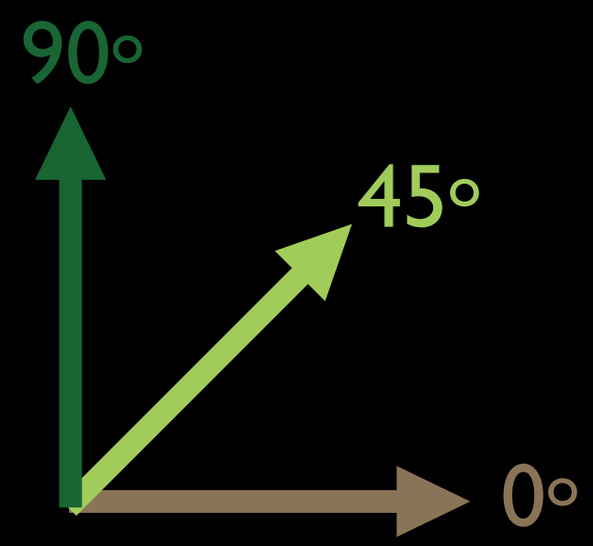


Experimental setup



narrow (1.3cm)

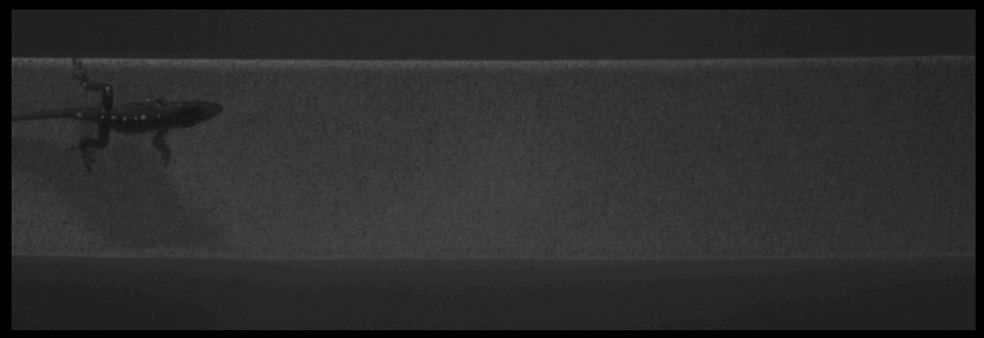
broad

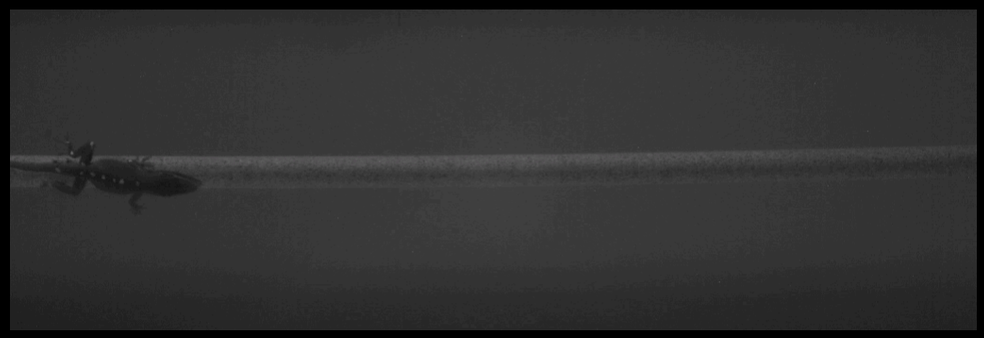


Foster and Higham (2012) *J. Exp. Biol.*

Broad, flat perch

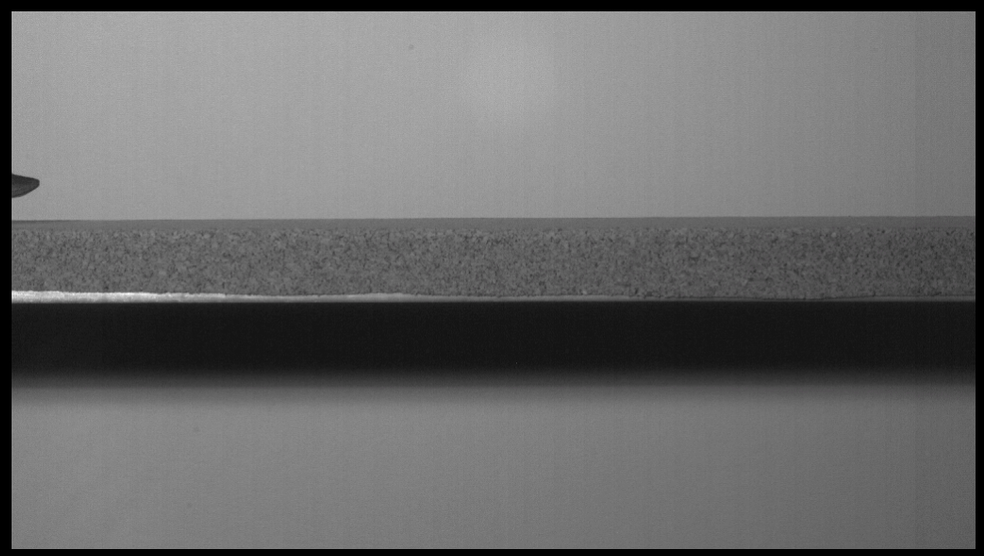
Narrow perch

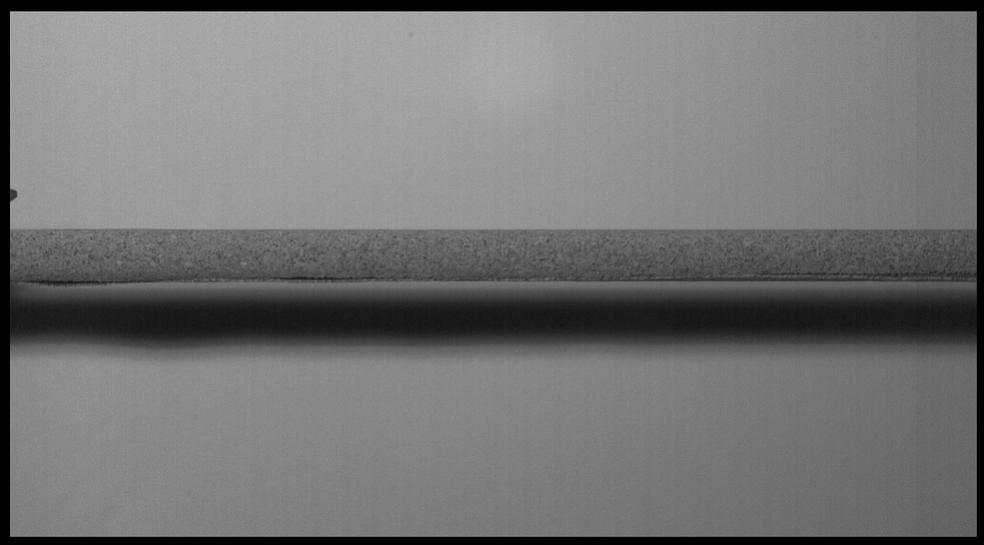




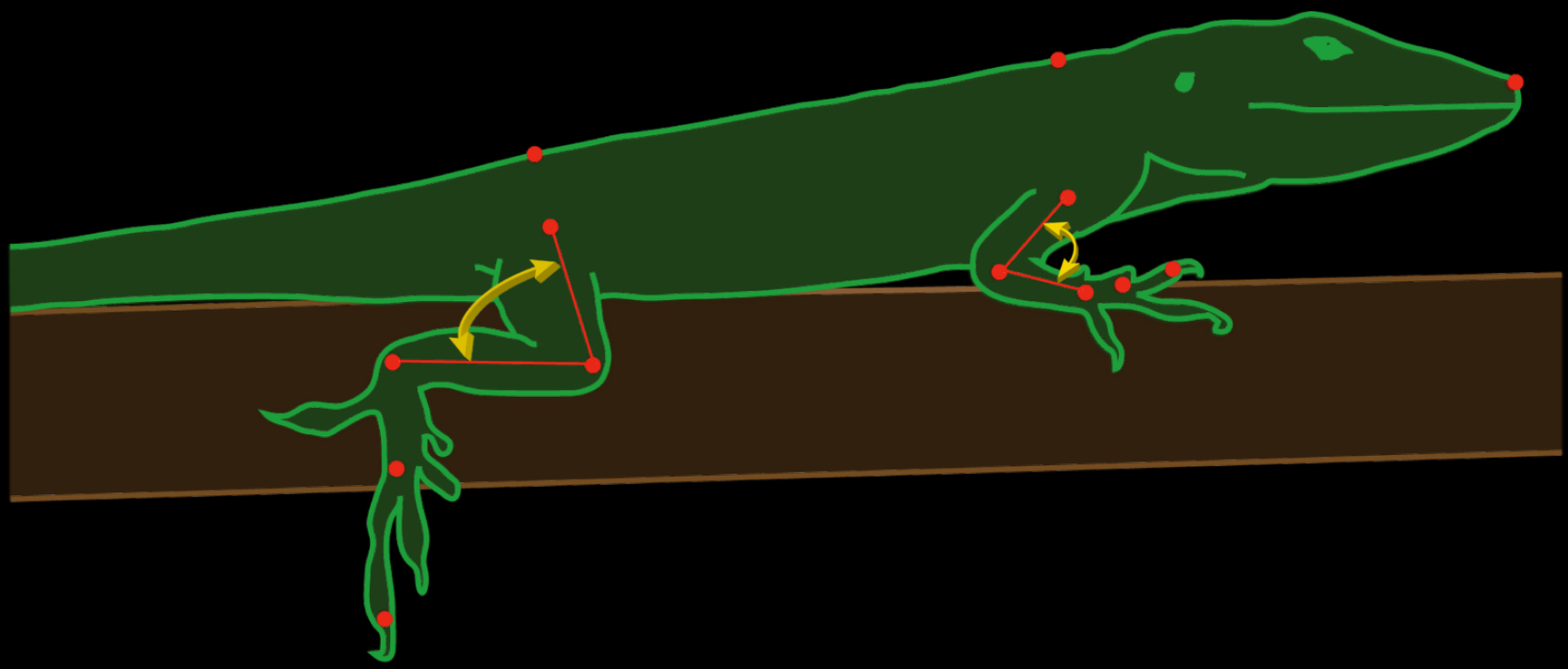
Top view

Side view





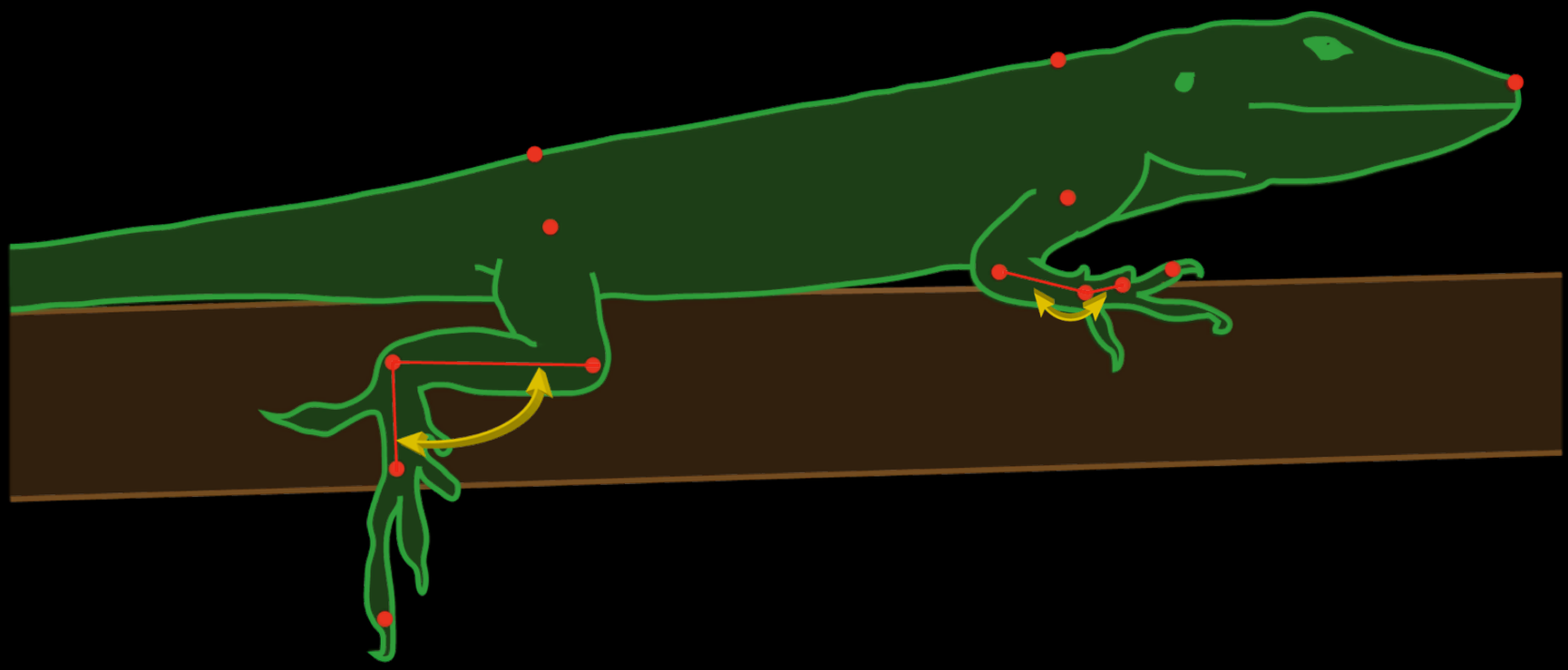
3D digitized points and angles



Elbow and knee angle

Foster and Higham (2012) *J. Exp. Biol.*

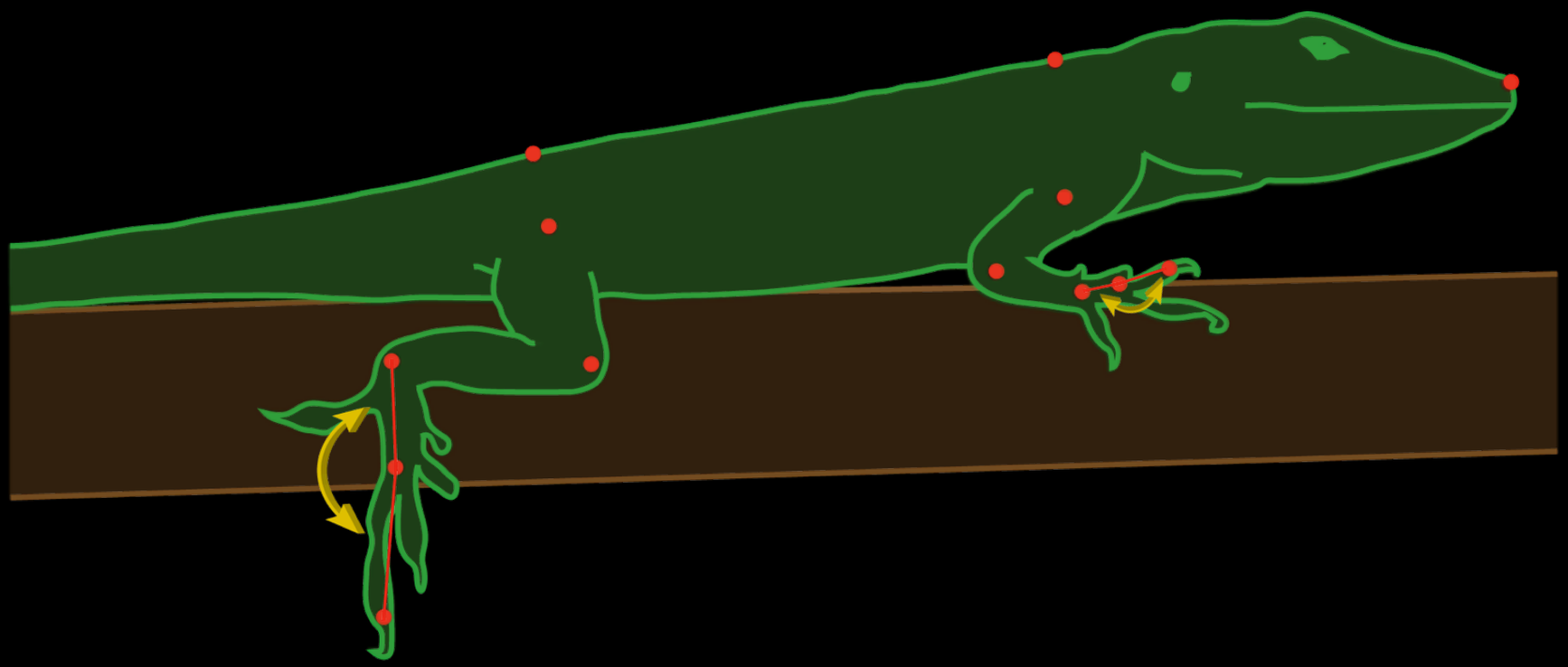
3D digitized points and angles



Wrist and ankle angle

Foster and Higham (2012) *J. Exp. Biol.*

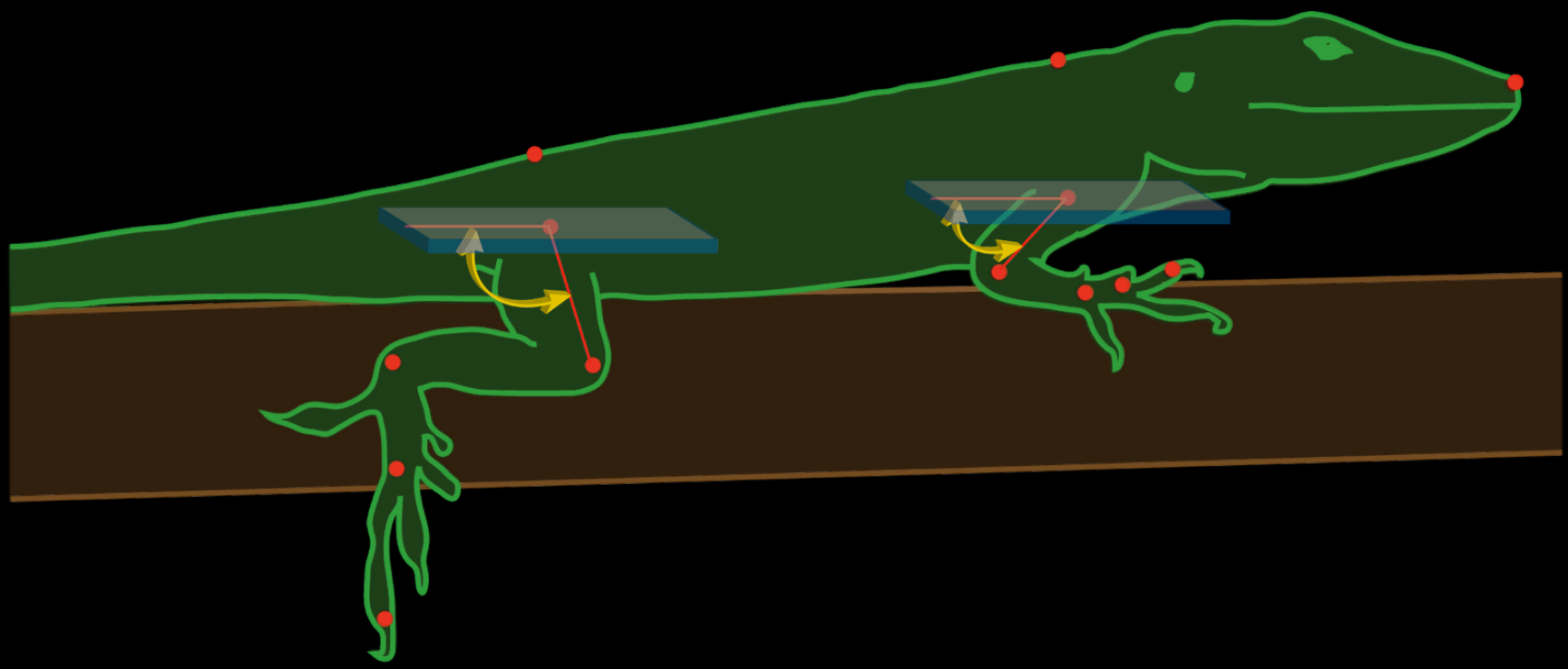
3D digitized points and angles



Fore- and hind toe angle

Foster and Higham (2012) *J. Exp. Biol.*

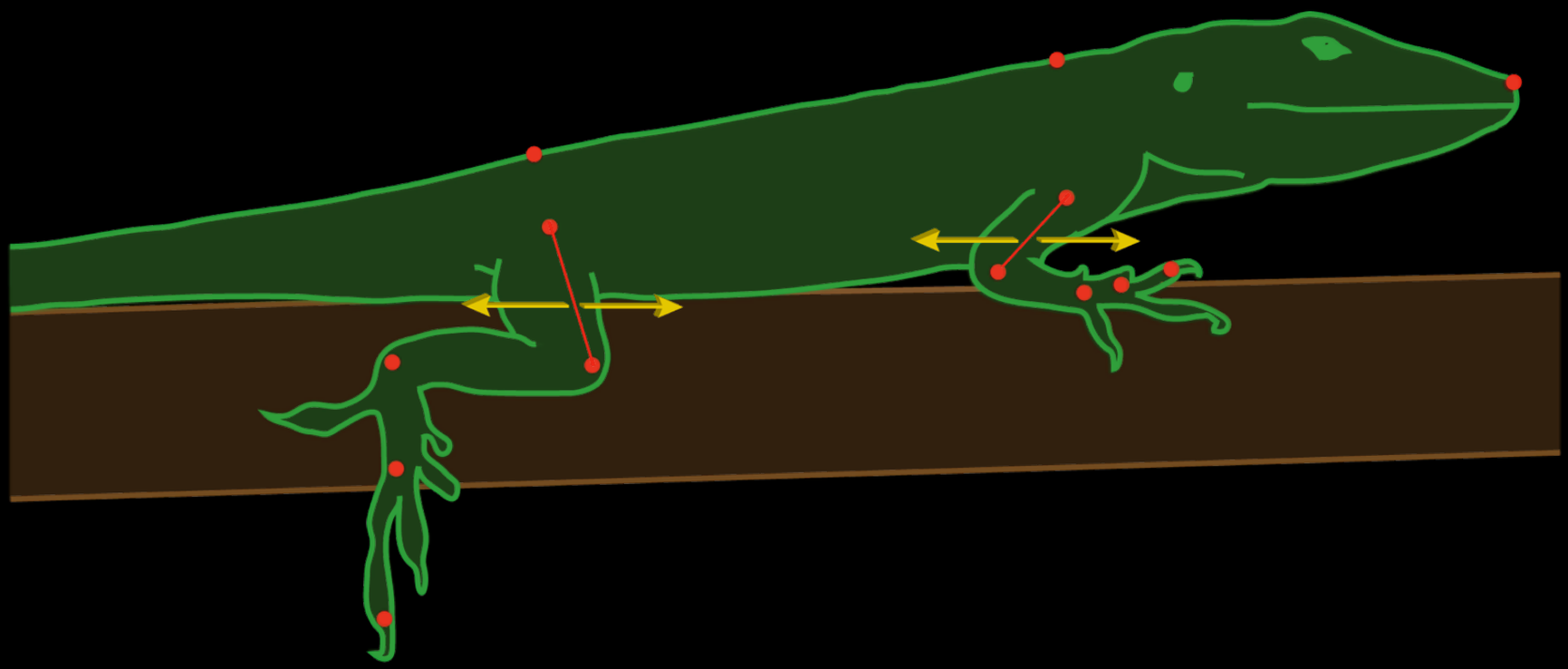
3D digitized points and angles



Humerus and femur depression

Foster and Higham (2012) *J. Exp. Biol.*

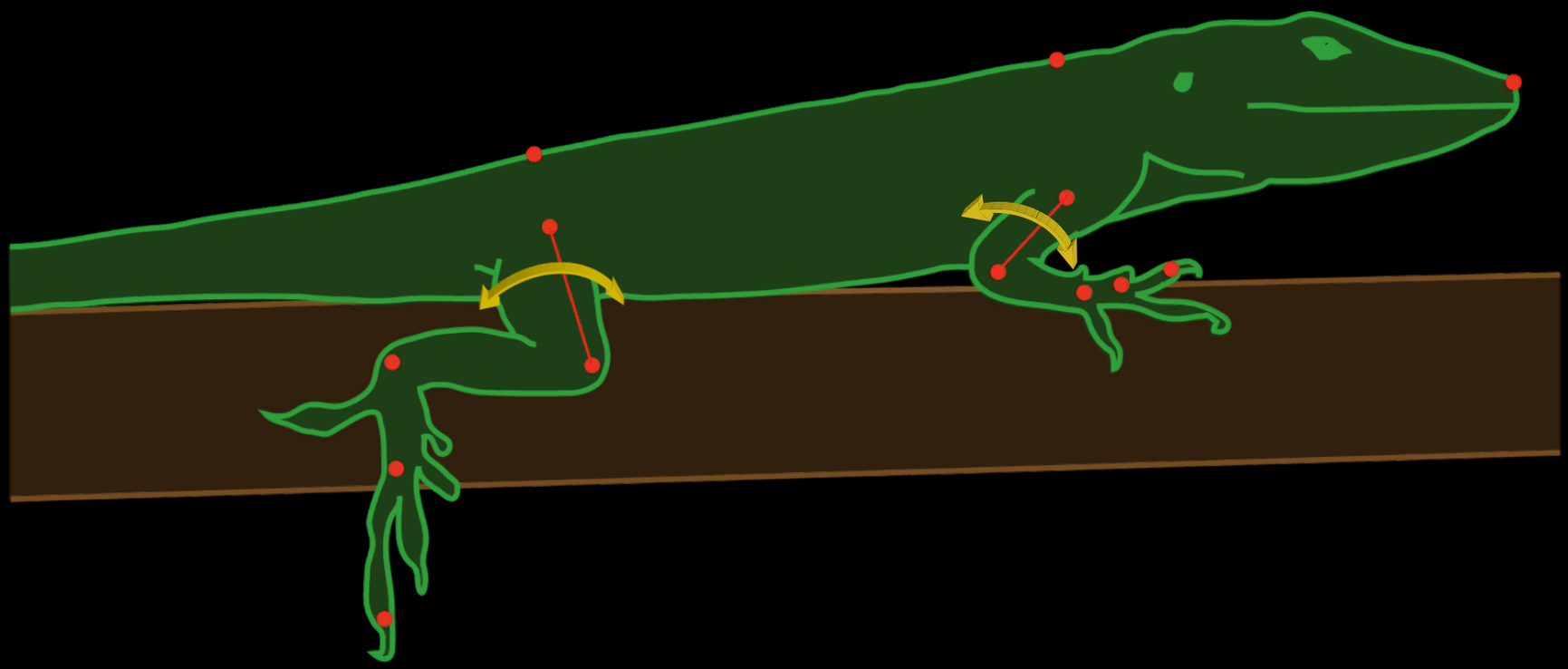
3D digitized points and angles



Humerus and femur protraction and retraction

Foster and Higham (2012) *J. Exp. Biol.*

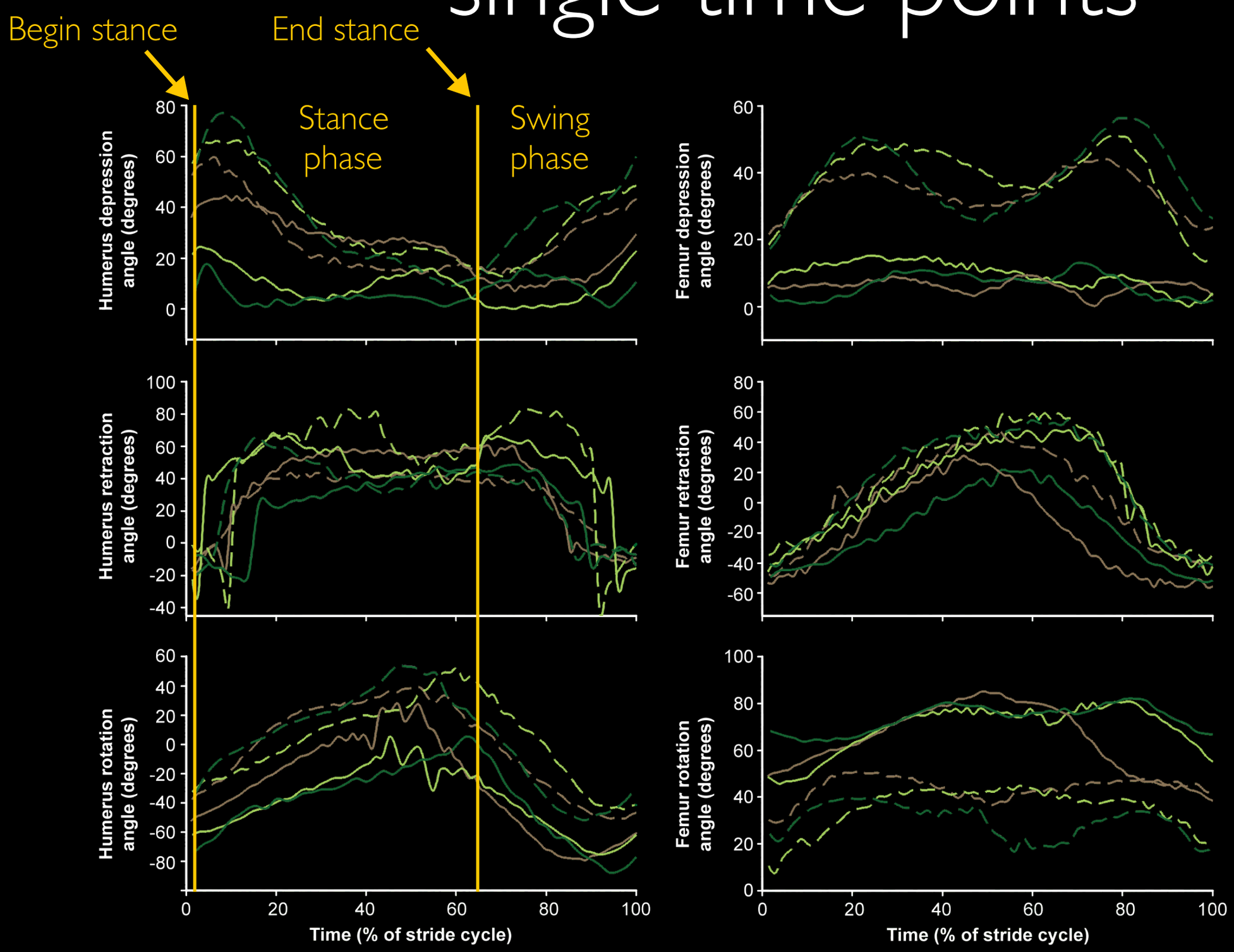
3D digitized points and angles

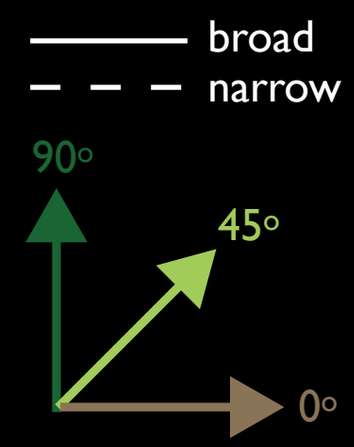


Humerus and femur long-axis rotation

Foster and Higham (2012) *J. Exp. Biol.*

Joint angle - time series & single time points





Foster and Higham (2012) *J. Exp. Biol.*

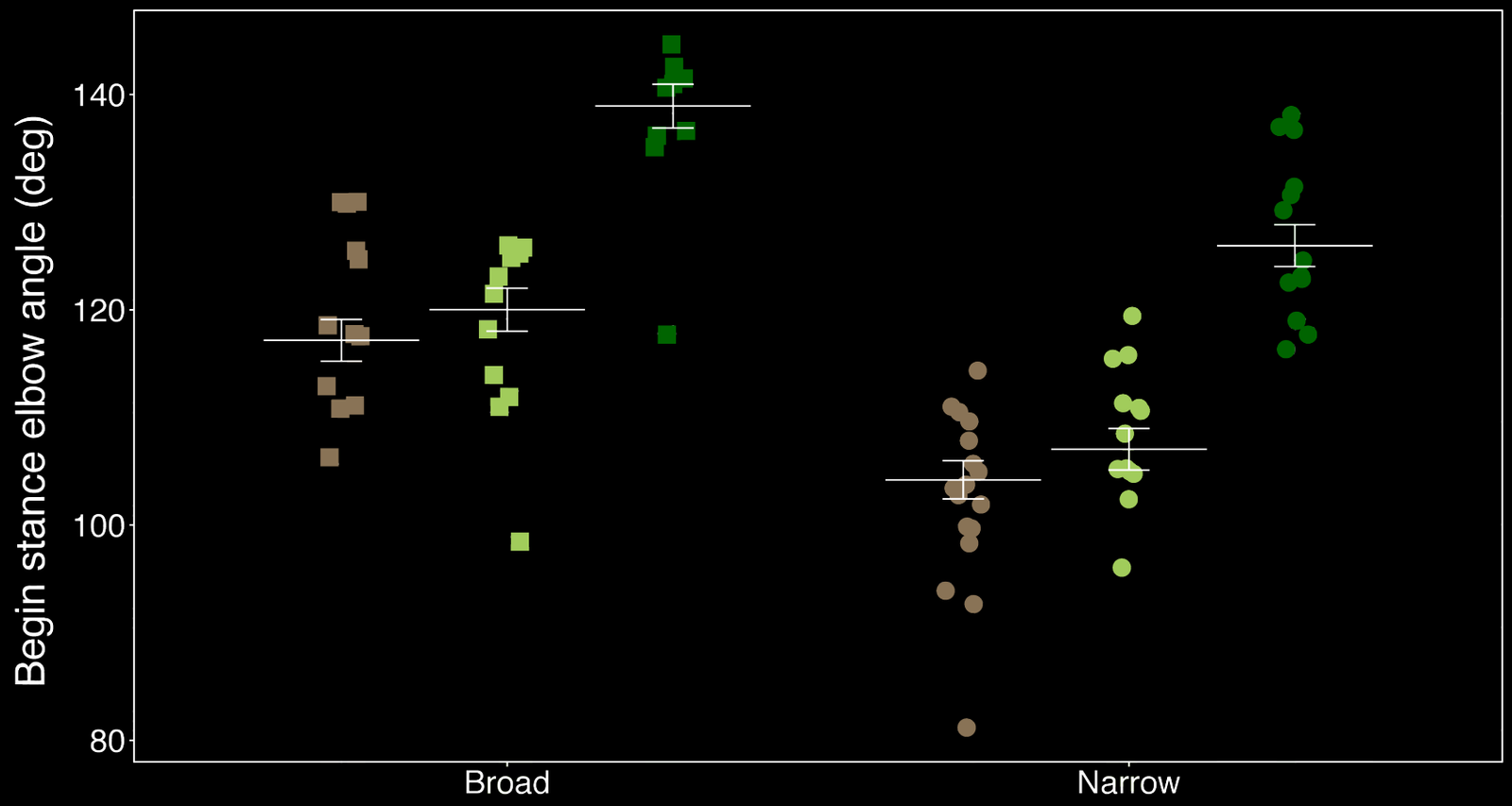
How should we analyze this dataset?

	Univariate	Multivariate	Time series
Visualization	✓	✓	✓
Hypothesis test	✓	✓	
Classification		✓	

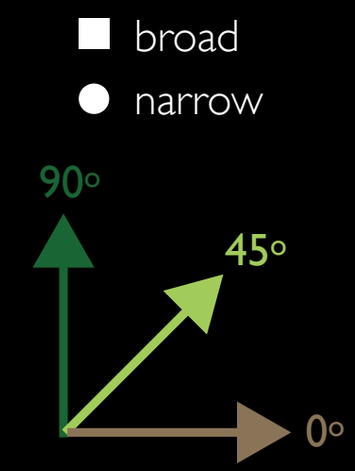
How should we analyze this dataset?

- **Univariate methods**
  - ANOVA to test for differences between substrates **in each variable (time point)** - visualized using box-type plots
- **Multivariate methods**
  - MANOVA to test for differences between substrates **when considering all variables together**
  - Canonical Discriminant Analysis (CDA) for visualization of maximal group separation and to determine how **all variables together** contribute to group separation
  - Linear Discriminant Analysis (LDA) to classify observations (strides) according to group (environment) using the **combined variables**
- **Time series analyses**
  - Dynamic Time Warping to visualize distances between the **time series of all joint angles**

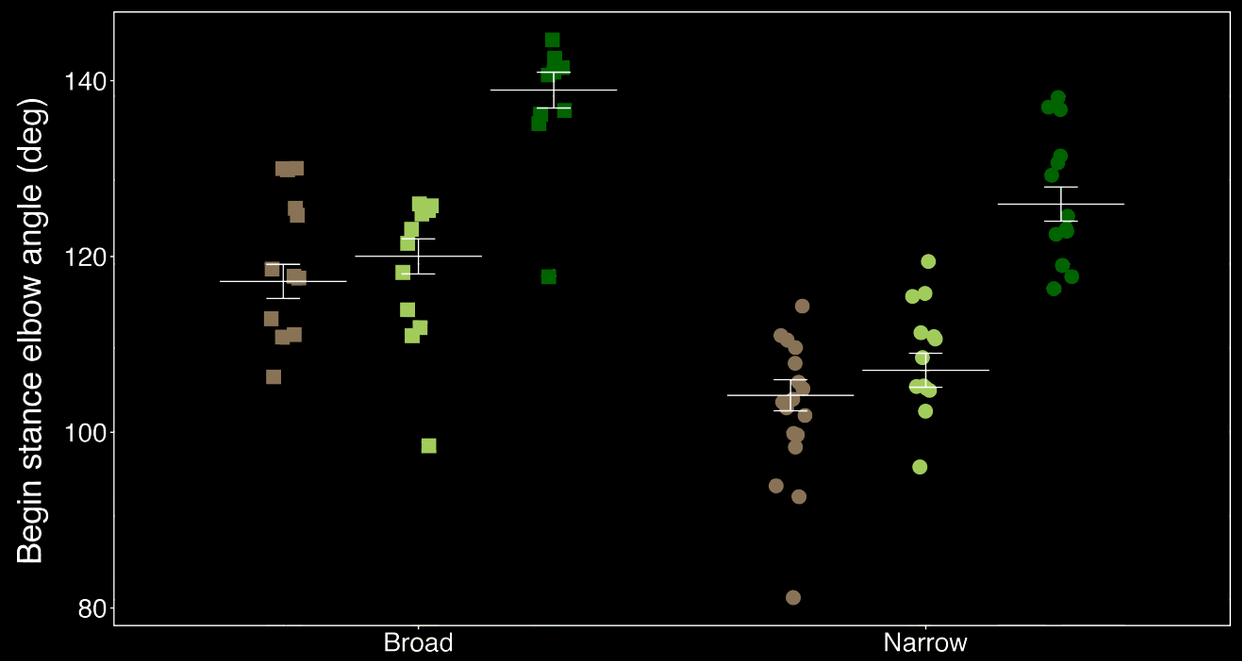
Two-way mixed model ANOVA



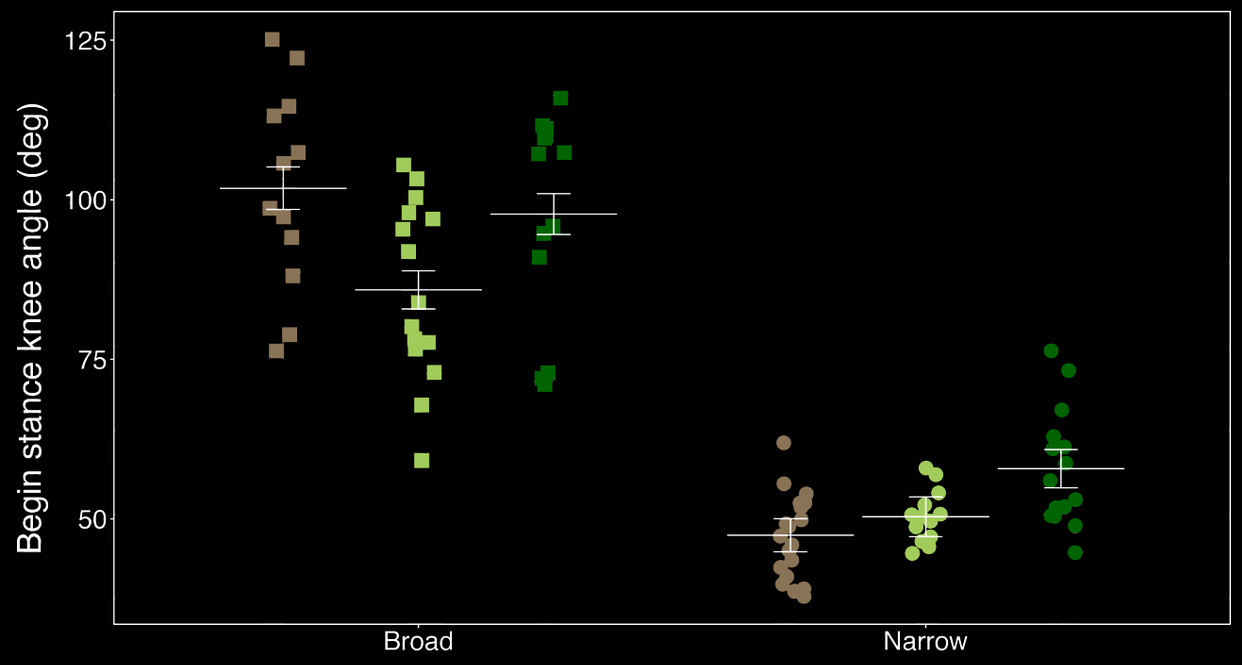
Perch diameter: $P < 0.0001$
Incline: $P < 0.0001$
Interaction: P = N.S.



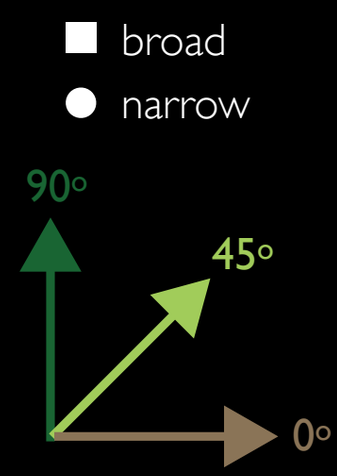
Two-way mixed model ANOVA



Perch diameter: P < 0.0001
Incline: P < 0.0001
Interaction: P = N.S.

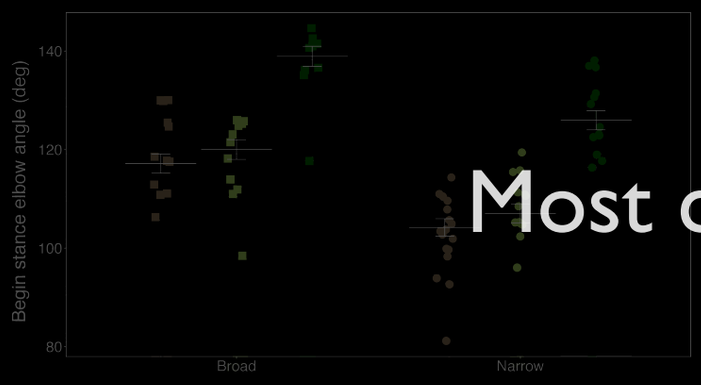


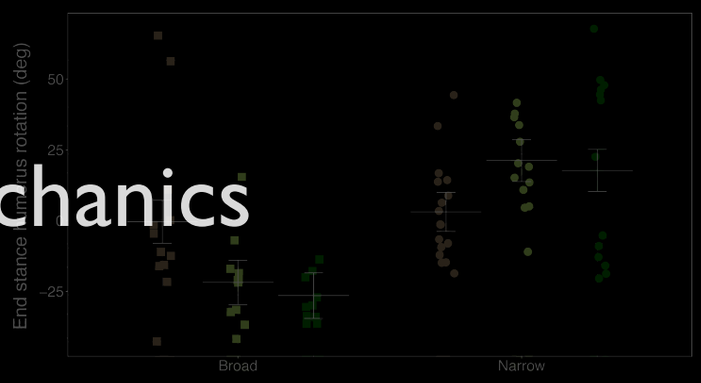
Perch diameter: P < 0.0001
Incline: P = 0.0032
Interaction: P = 0.0115

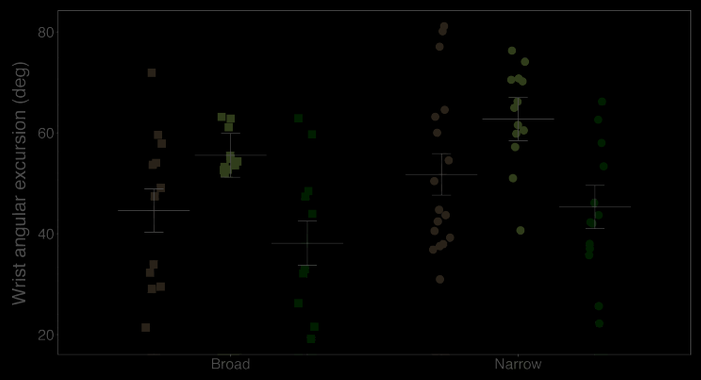


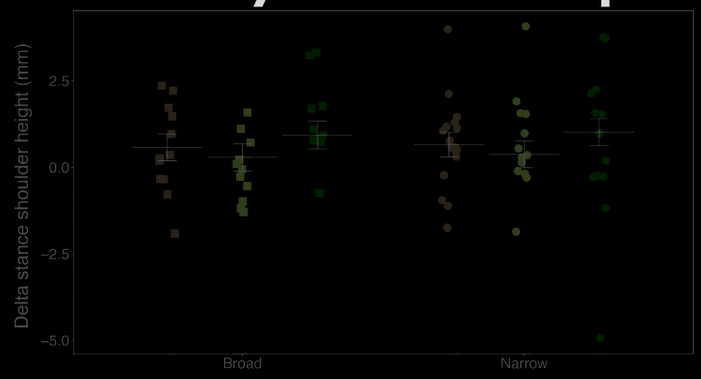
Two-way mixed model ANOVA

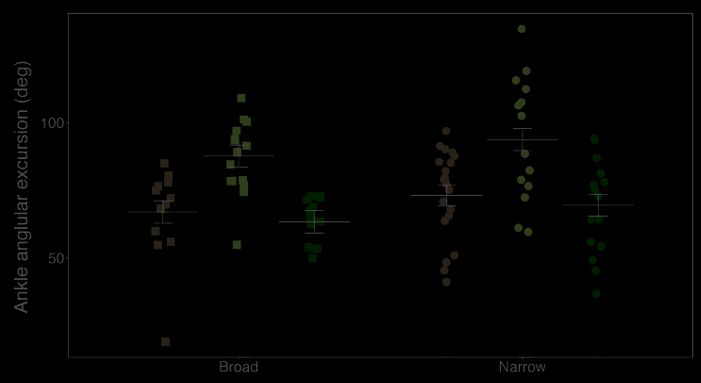
Most common strategy in biomechanics
AND easy to interpret

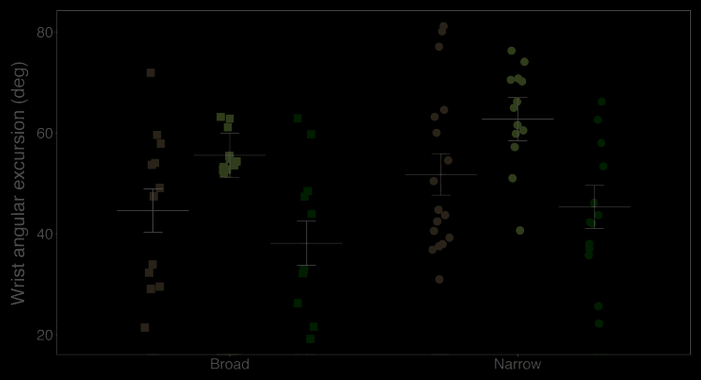


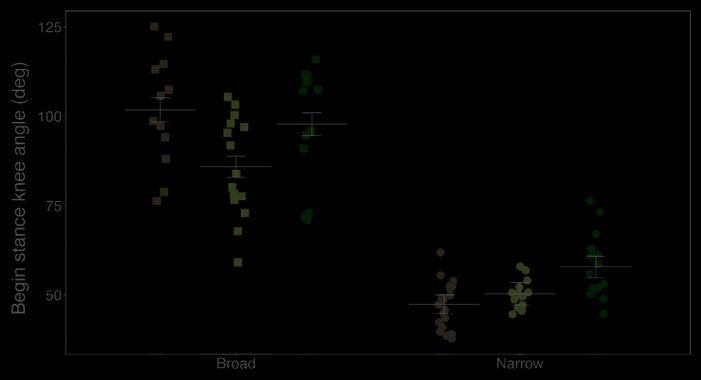


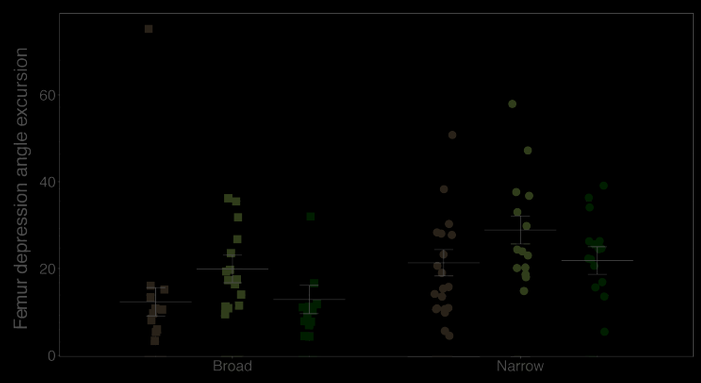


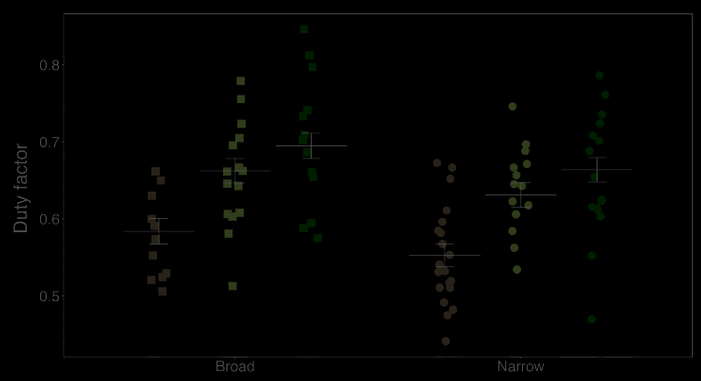




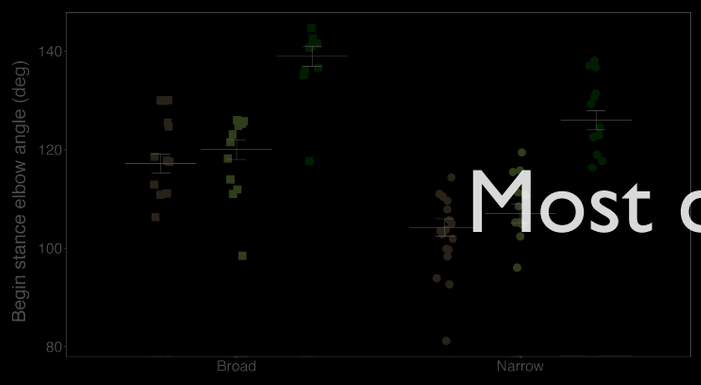


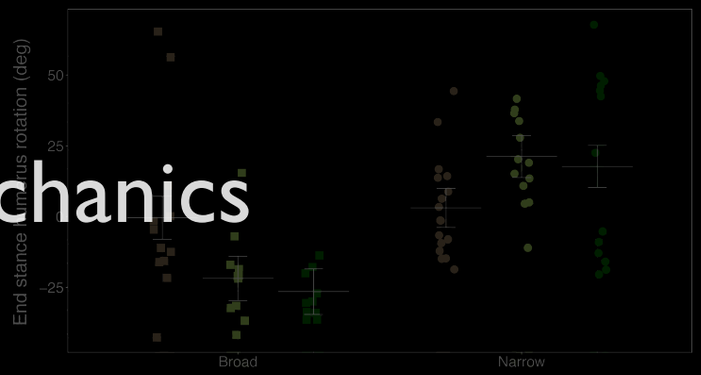






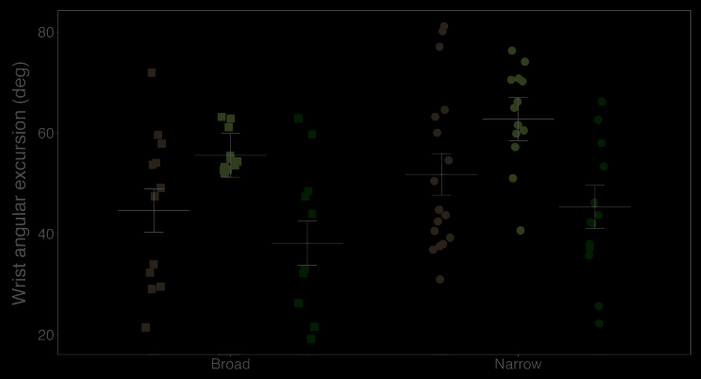
Two-way mixed model ANOVA

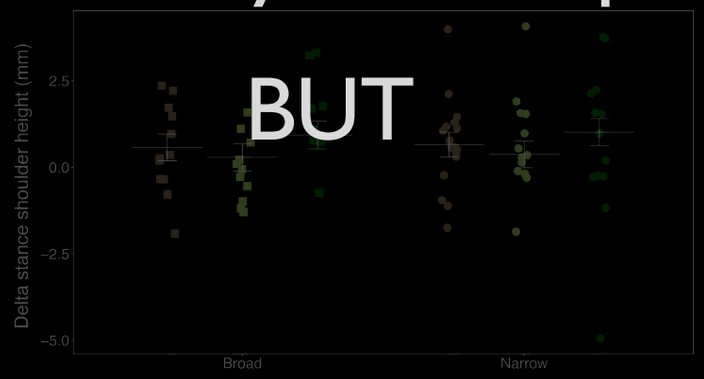


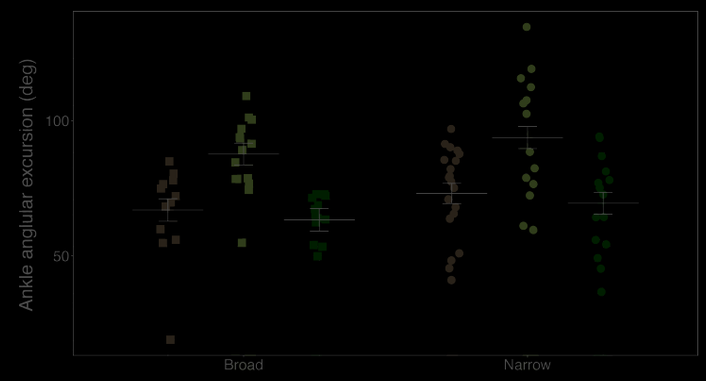


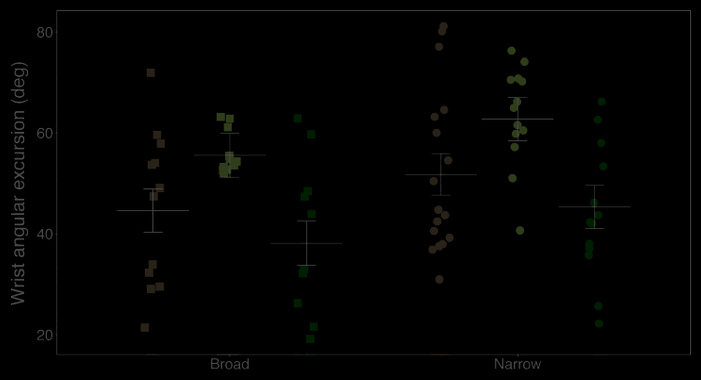
Most common strategy in biomechanics
AND easy to interpret

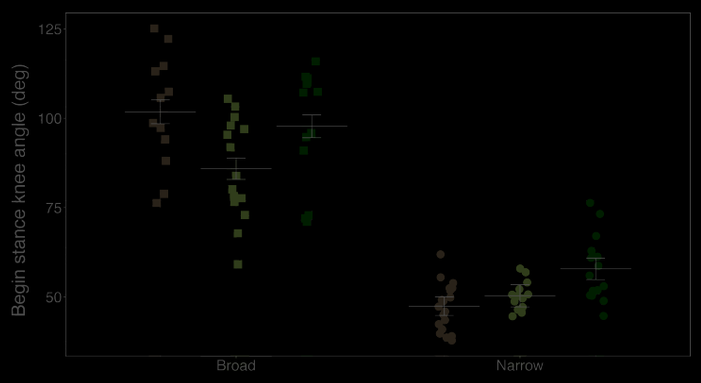
BUT

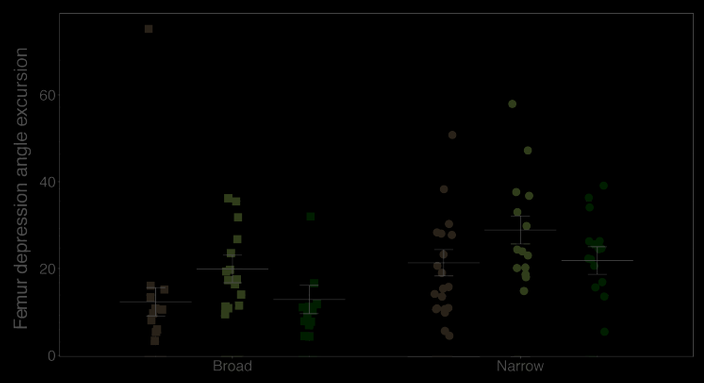


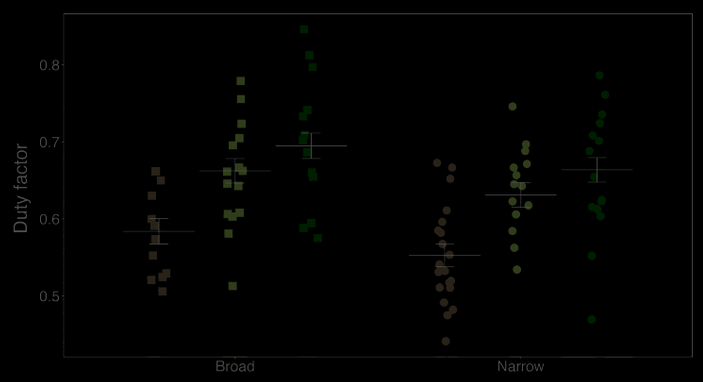












Two-way mixed model ANOVA

Most common strategy in biomechanics
AND easy to interpret

BUT

- There are 72 individual variables - which are important?
- Need to do some sort of correction for multiple tests since some results must be significant by random chance (e.g. Bonferroni, Benjamini-Hochberg)
- Does not take into account correlations between variables or time

Two-way mixed model ANOVA

	% variables significantly affected	
	Perch diameter	Incline
Forelimb	42	64
Hind limb	69	69

- Significant changes to at least one variable at nearly every joint in response to changes in incline or perch diameter (88 significant results in total)
- Generally, decreased height above surface, greater limb flexion, slower running speed, and greater duty factor on steeper/ narrow surfaces ⟶ crouching for stability

Multivariate Analysis

- MANOVA
  - to **test** for differences between substrates **when considering all variables together**
- Canonical Discriminant Analysis (Classical perspective)
  - to visualize maximal group separation and to determine **how all variables together** contribute to group separation
- Linear Discriminant Analysis (Modern perspective)
  - to classify strides according to environment based on the **combined variables**

MANOVA

- Multivariate ANOVA to test the hypothesis that the **vectors** of means for two or more groups are different
- Tested by calculating the ratio of the multivariate between-group sum of squares to within-group sum of squares → F-distribution

↓

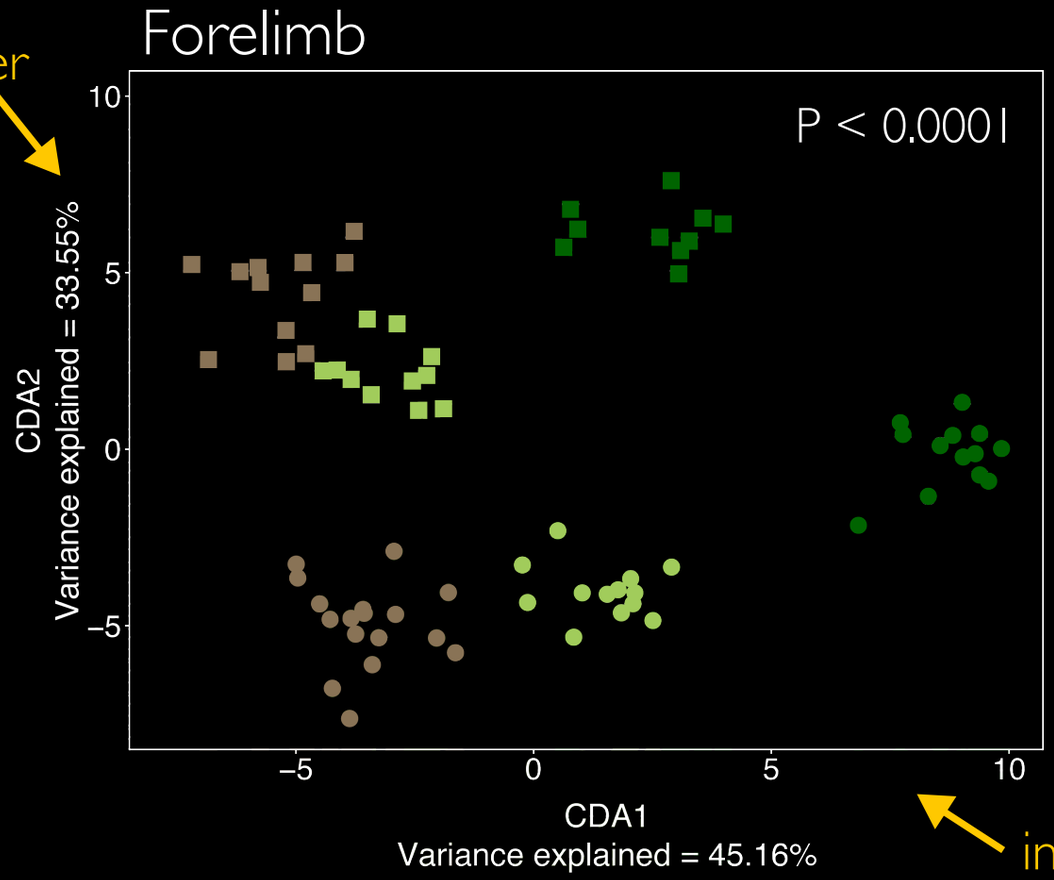
Canonical Discriminant Analysis

Canonical Discriminant Analysis

- Goal: visualization of group separation & determine **how** the groups are separated
- Eigenanalysis based on variance/covariance matrices between and within user-defined groups
- Defines "canonical axes" as linear combinations of variables that maximize the between-group variation relative to within-group variation
- Correlation of variables against CDA axes can be used to interpret the source of the variation contributing to group separation

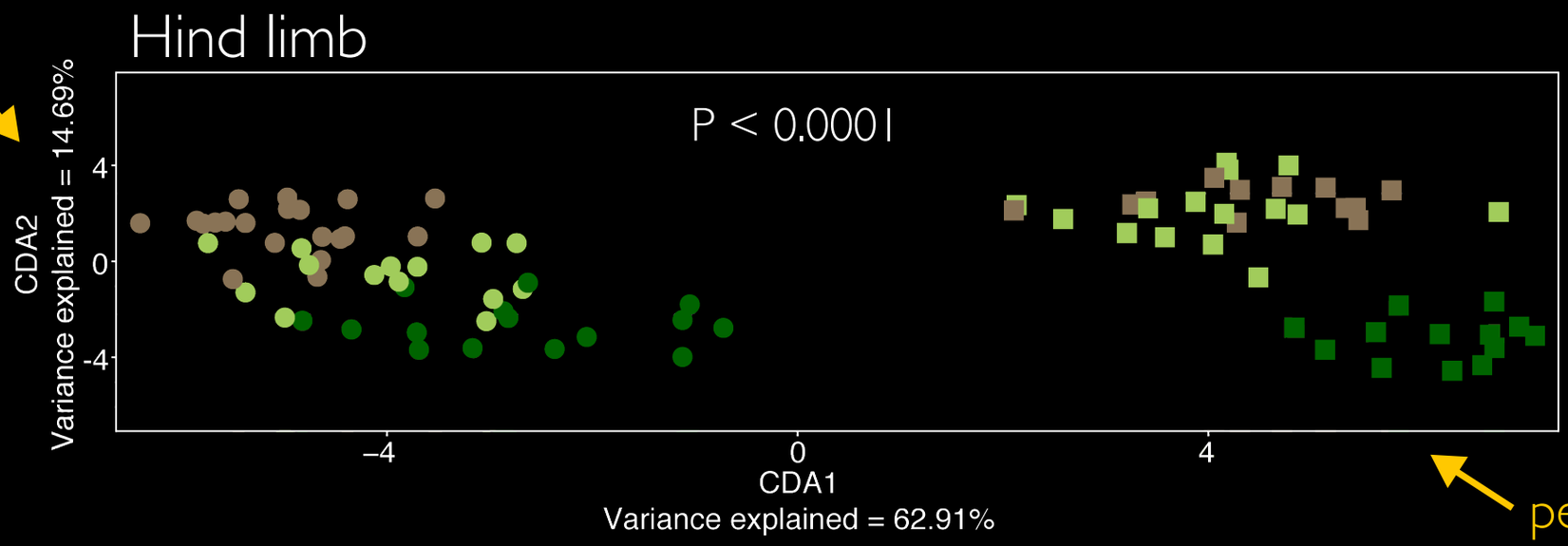
Canonical Discriminant Analysis

perch diameter

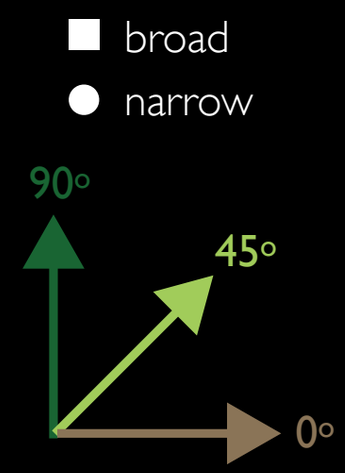


incline

incline



perch diameter

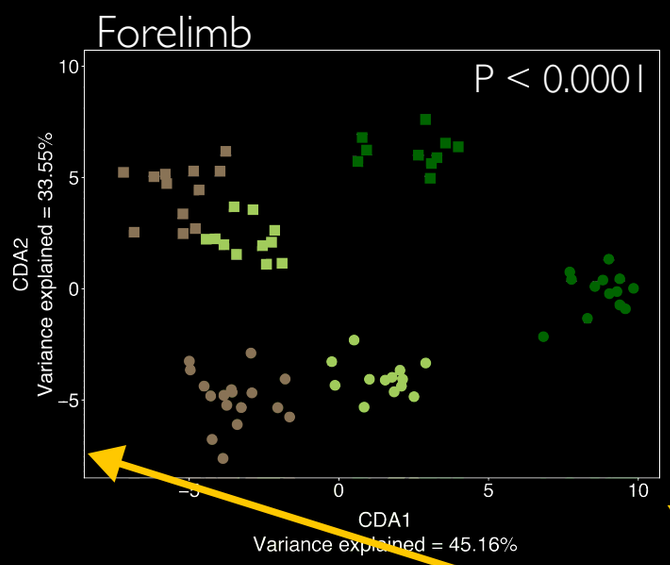


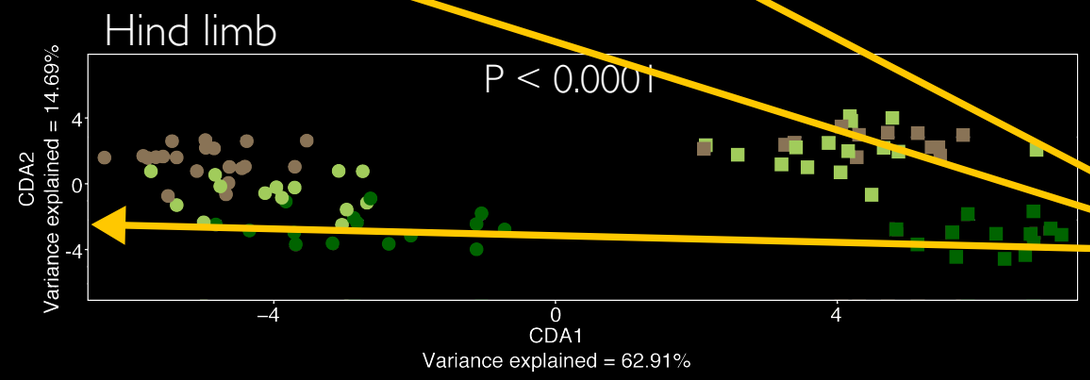
Canonical Discriminant Analysis

Important variables for group separation:

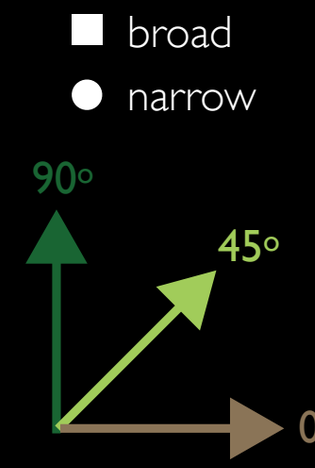
- Speed, duty factor, stride frequency
- Shoulder height, elbow angle/angular velocities, humerus depression/rotation
- Hip height, angles/angular velocities of femur depression/retraction & knee/ankle

Lizards run more slowly, are more crouched, and swing limbs faster to increase proportion of time feet are in contact with the ground





Linear combination of variables



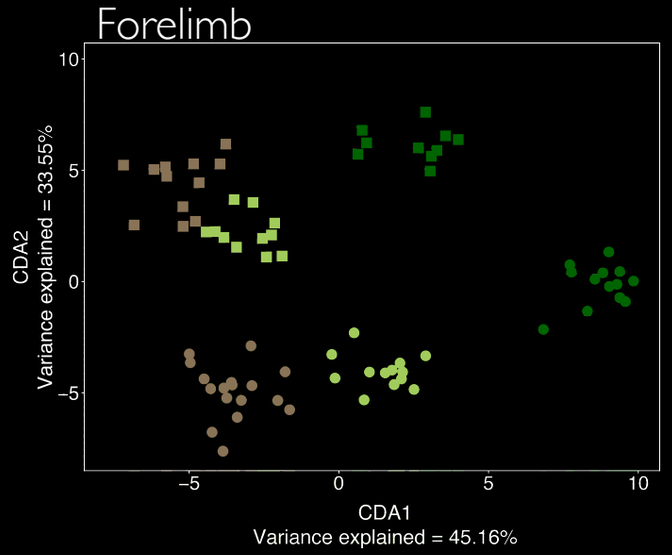
Canonical Discriminant Analysis
↓
Linear Discriminant Analysis

- CDA maximized the ratio of the between to within-group variation
- This gives a eigenvector corresponding to the direction of maximal separation between groups and the corresponding separation hyperplane
- This hyperplane can be used as a decision boundary for classification purposes
- Under assumptions of normality and equal variance between groups, this hyperplane corresponds to the classification boundary of LDA

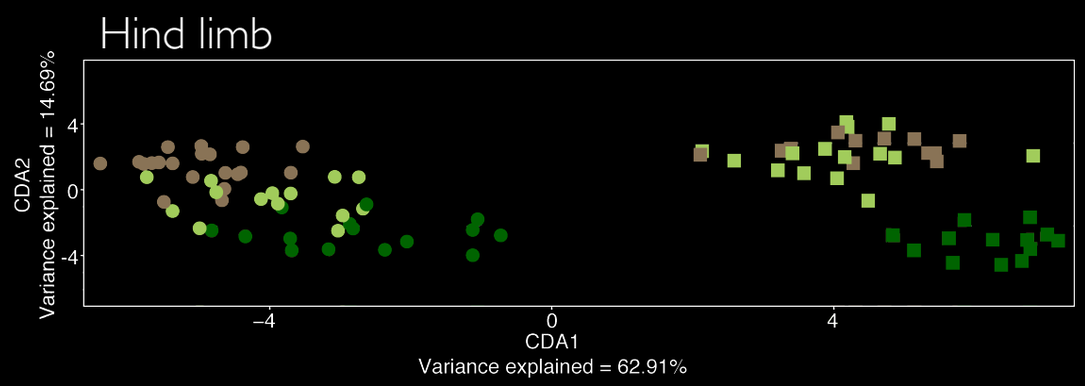
Linear Discriminant Analysis

- Goal: Classify the observations (strides) into the correct groups (environments)
- Classification model assessed via Leave-One-Out Cross-validation
- Classification accuracy + Confusion Matrix to assess the quality of the classification model both overall and by group

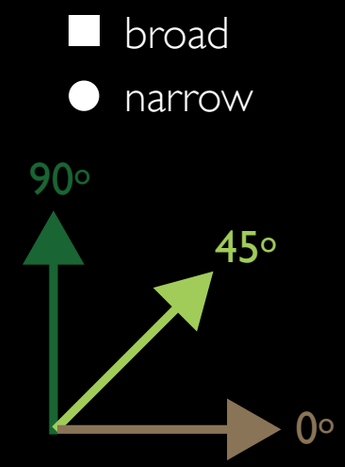
Linear Discriminant Analysis



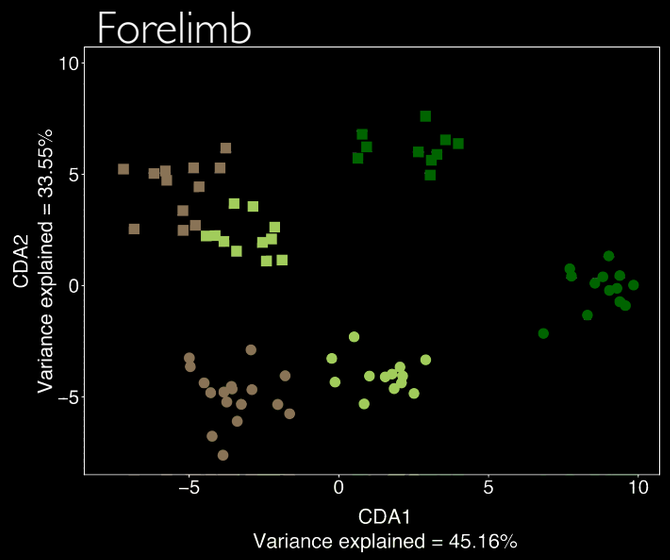
% observations correctly classified = 93.4%

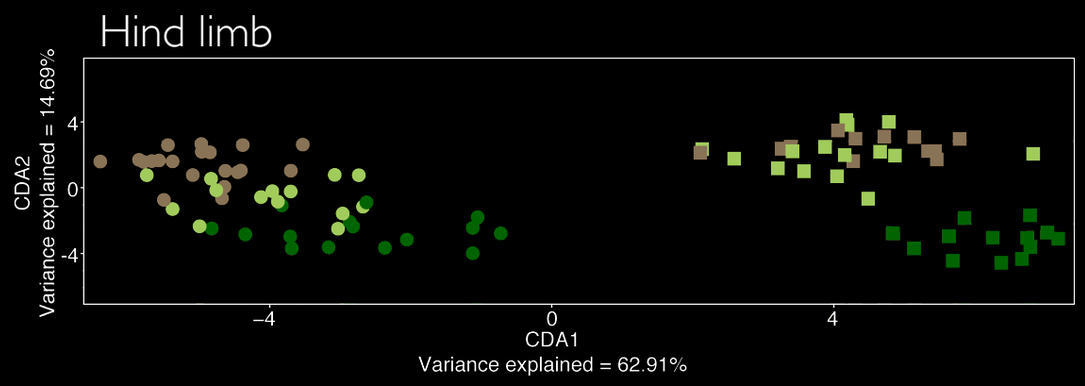


% observations correctly classified = 84.3%



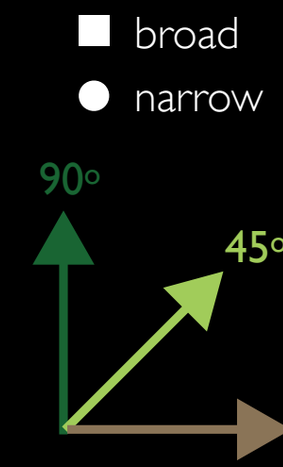
Linear Discriminant Analysis





% observations correctly classified

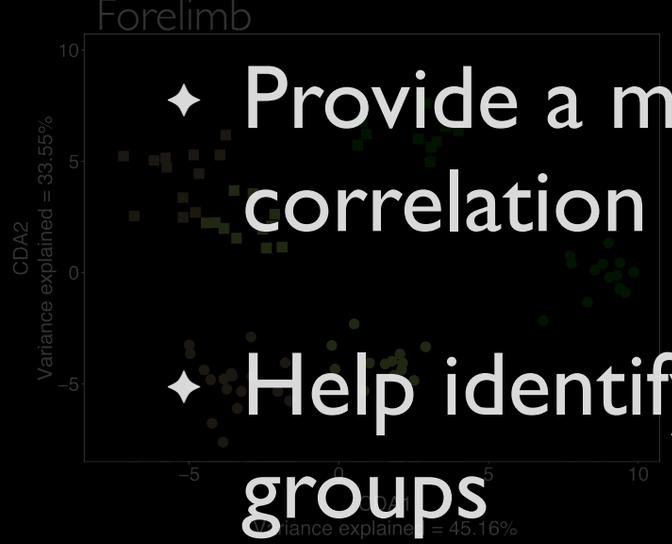
	Forelimb	Hind limb
Angles only	82.9%	84.3%
Angular velocities only	68.4%	58.4%
All variables	93.4%	84.3%

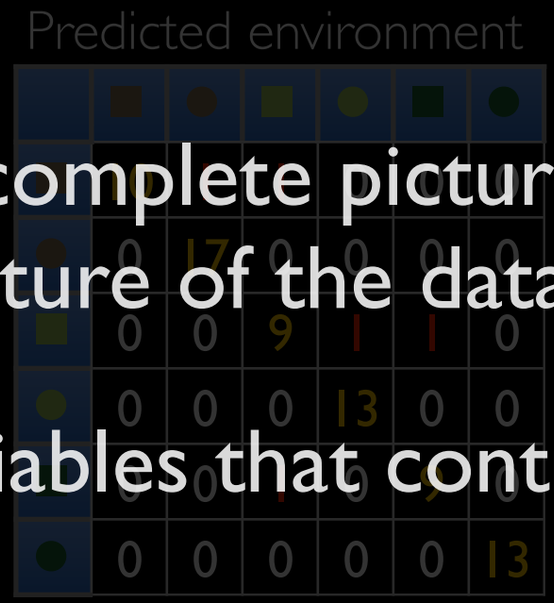


Linear Discriminant Analysis

Multivariate Analyses

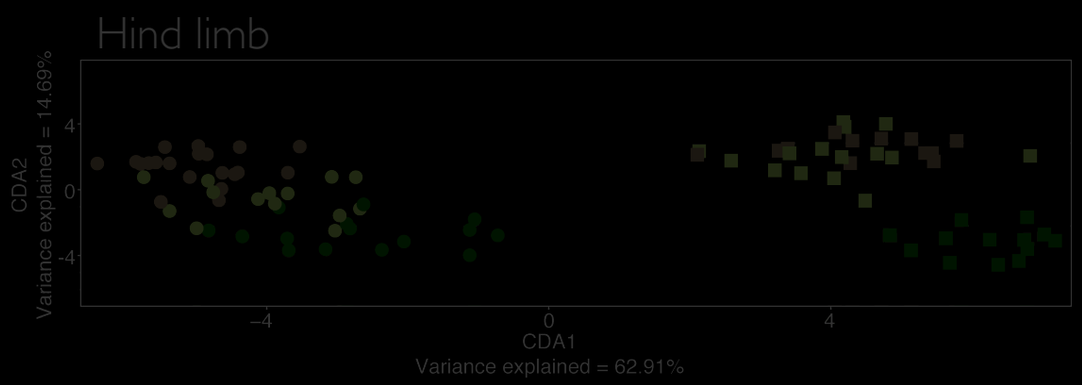
- Provide a more complete picture that incorporates correlation structure of the dataset
- Help identify variables that contribute to definitions of groups





classifies all observations well

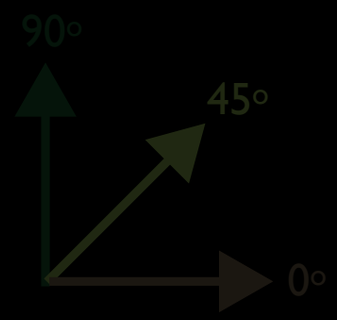
classifies diameter perfectly & only makes mistakes between inclines



Predicted environment

True environment						
	7	0	4	0	1	0
	0	18	0	2	0	0
	2	0	12	0	1	0
	0	1	0	12	0	1
	0	0	1	0	12	0
	0	0	0	1	0	14

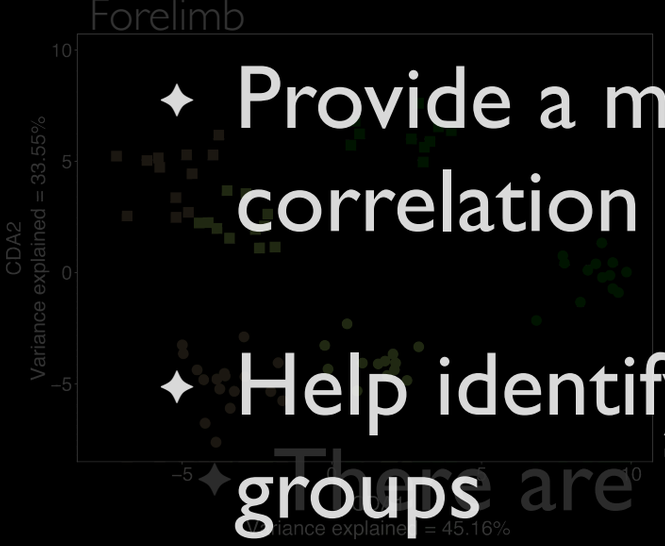
broad
narrow

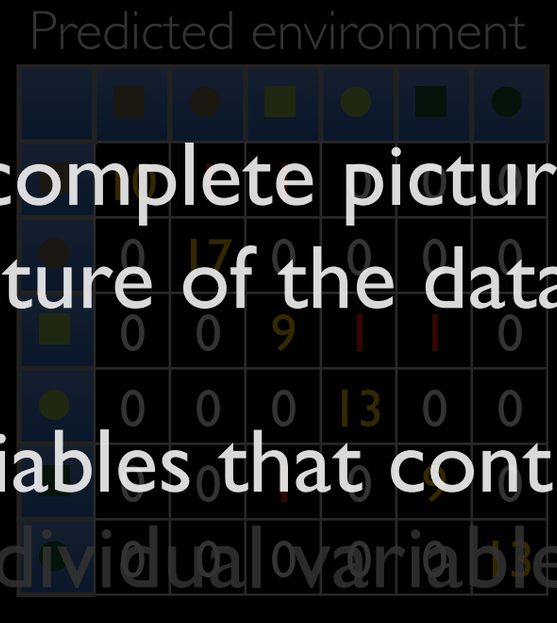


Linear Discriminant Analysis

Multivariate Analyses

- Provide a more complete picture that incorporates correlation structure of the dataset
- Help identify variables that contribute to definitions of groups

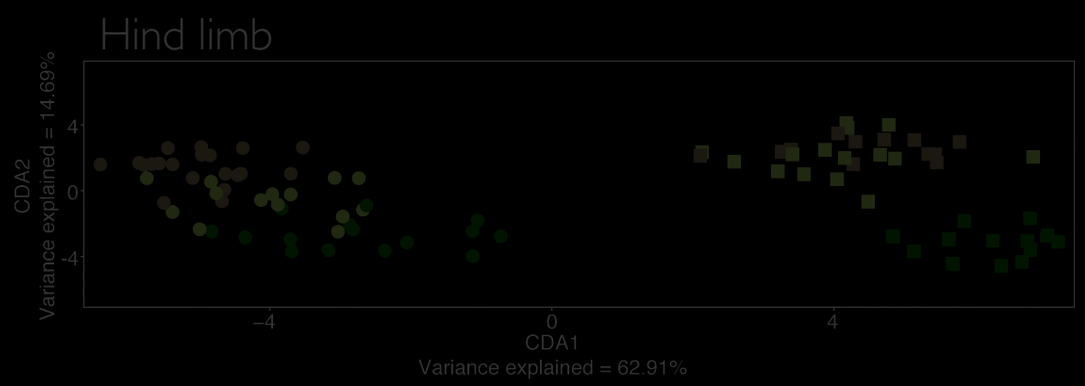


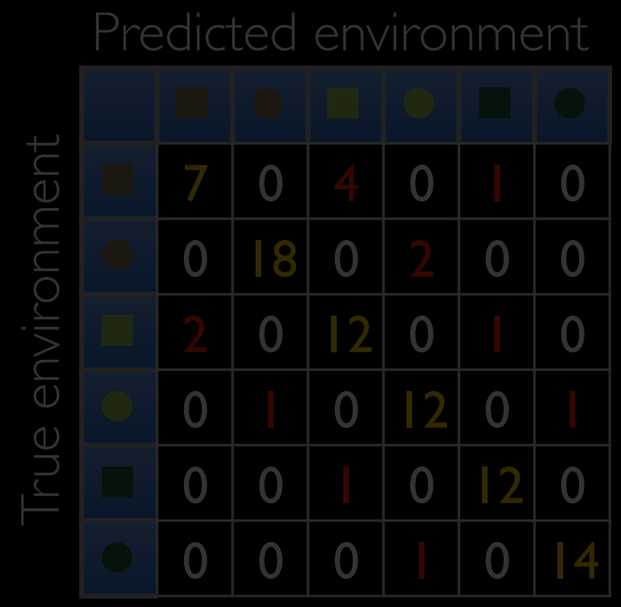


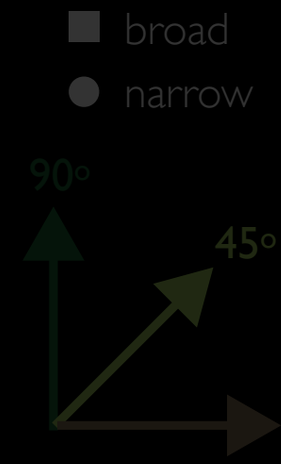
classifies all observations well

classifies diameter perfectly & only makes mistakes between inclines

BUT







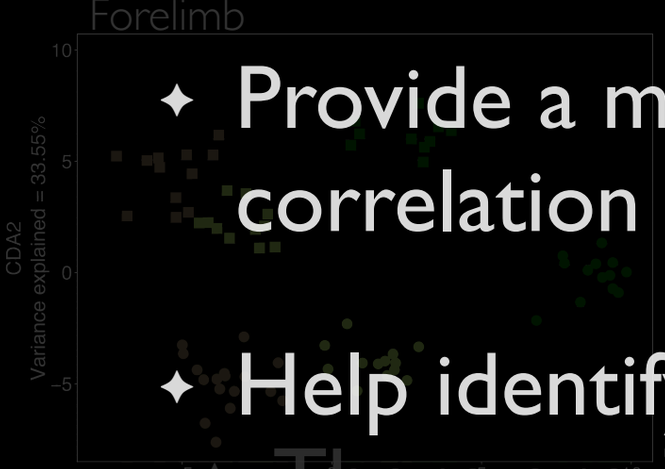
Linear Discriminant Analysis

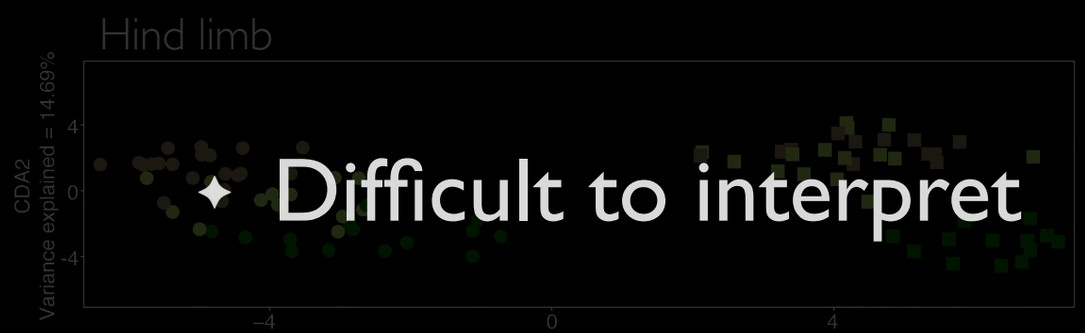
Multivariate Analyses

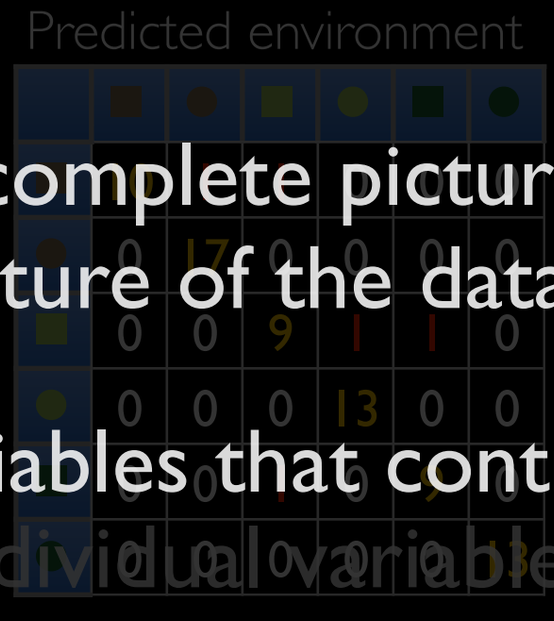
- Provide a more complete picture that incorporates correlation structure of the dataset
- Help identify variables that contribute to definitions of groups

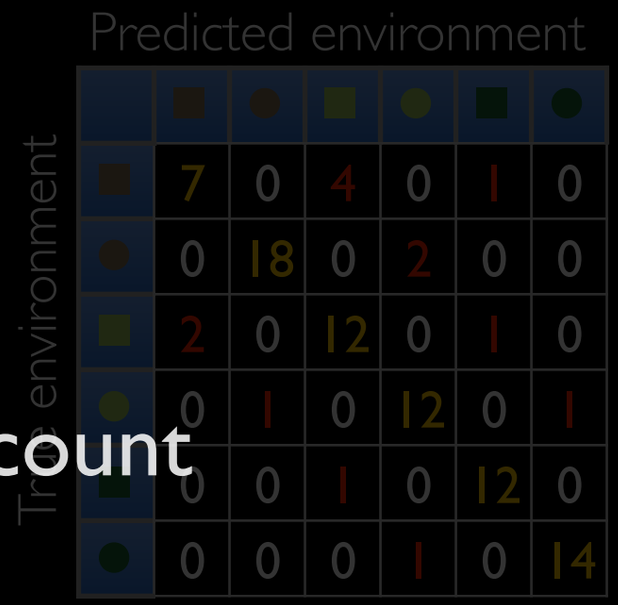
BUT

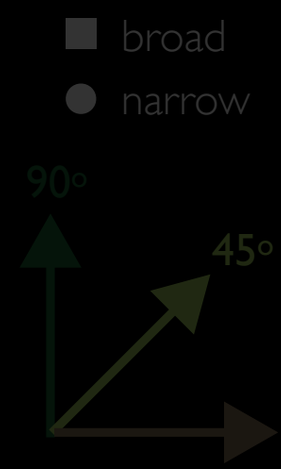
- Difficult to interpret
- Does not take Time into account





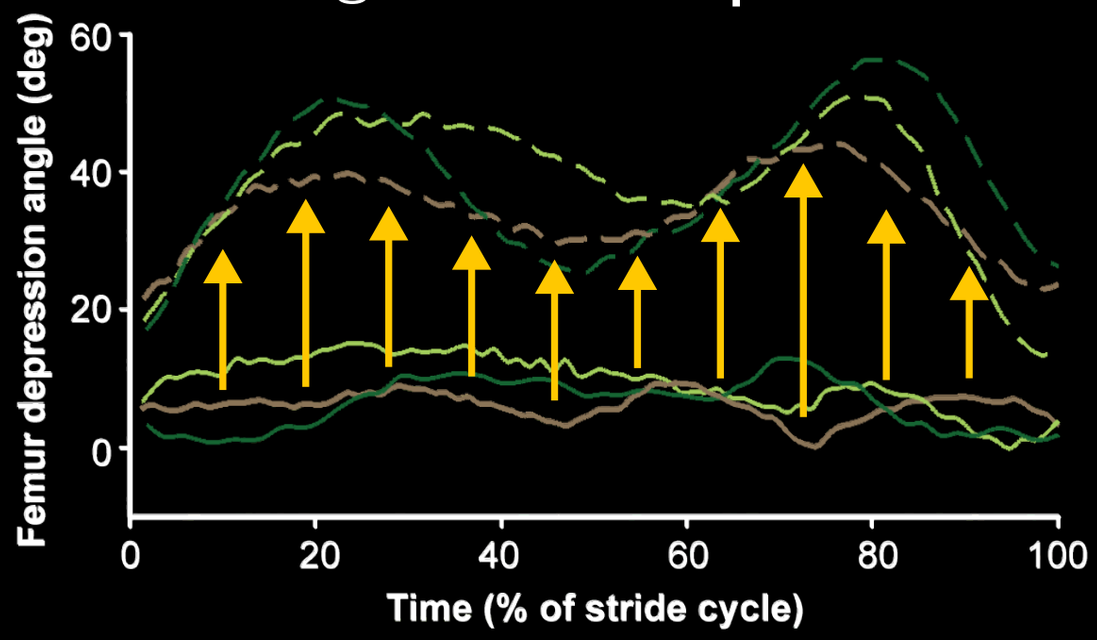




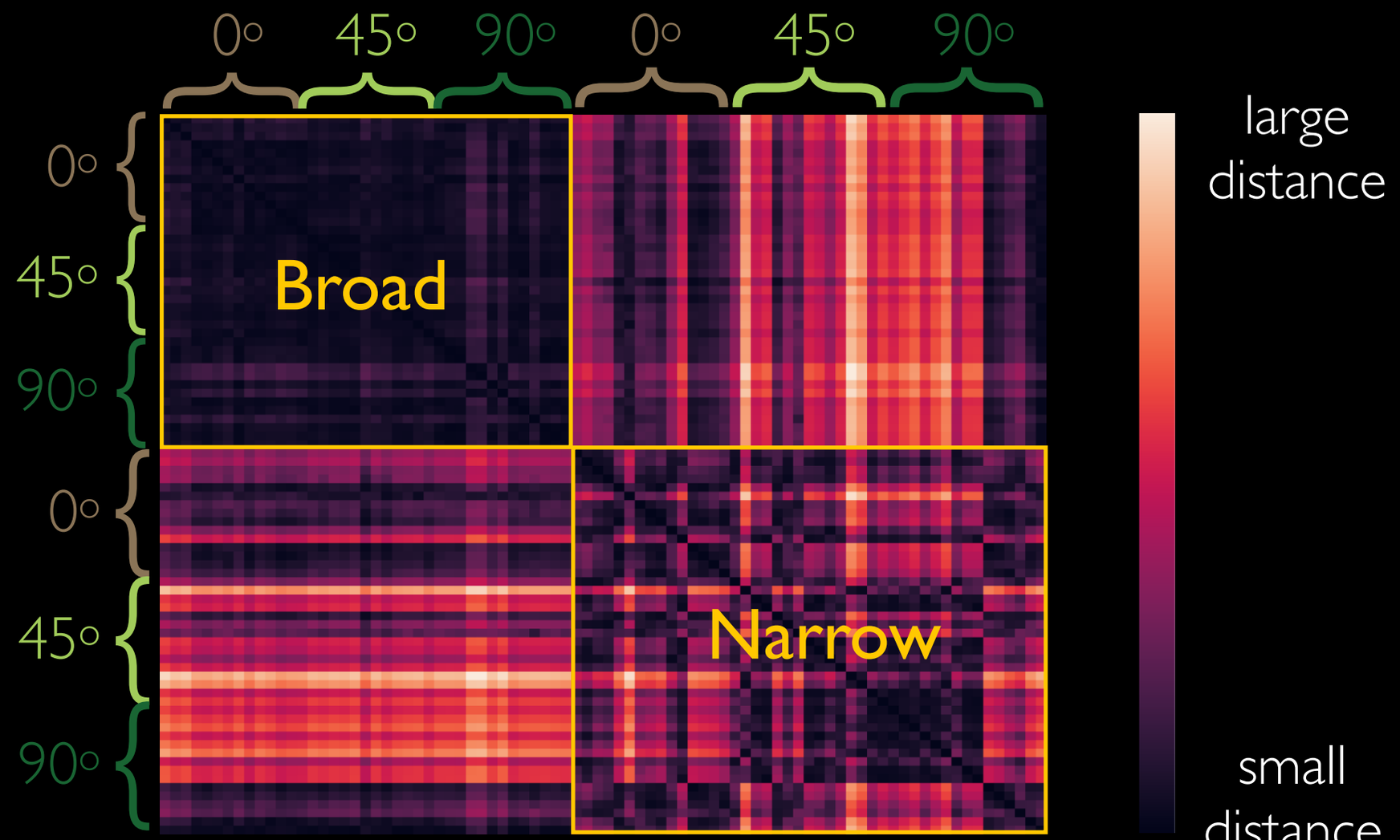


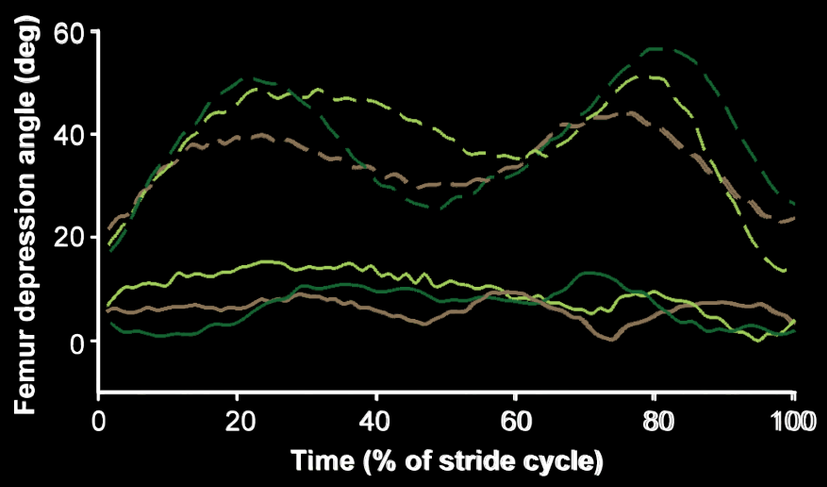
Dynamic Time Warping

- Goal: to measure & visualize the distance between time series of joint angles of each stride
- For each joint angle, calculate the distance between each instant in time during one stride and the corresponding time point in every other stride (allows non-linear alignment for time series of different length)
- The cumulative distance between time series of each stride can be visualized using a **Heat Map**



Dynamic Time Warping - Heat Map



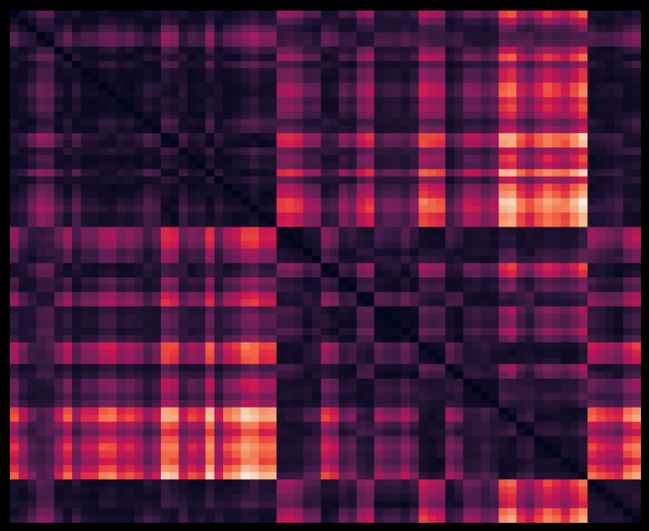


Dynamic Time Warping - Heat Map

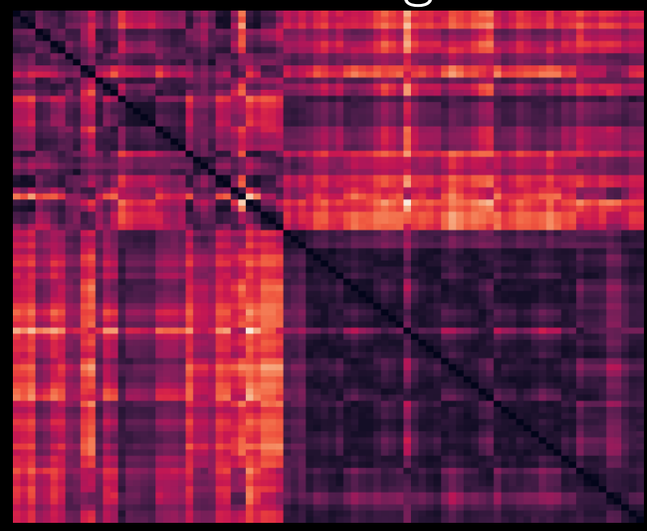
Femur depression

Femur rotation

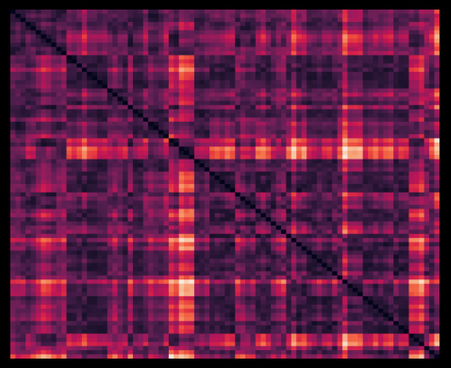
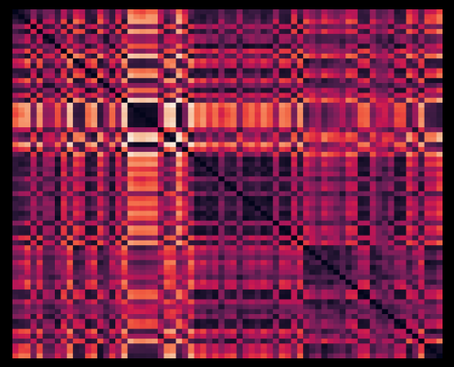
Humerus rotation

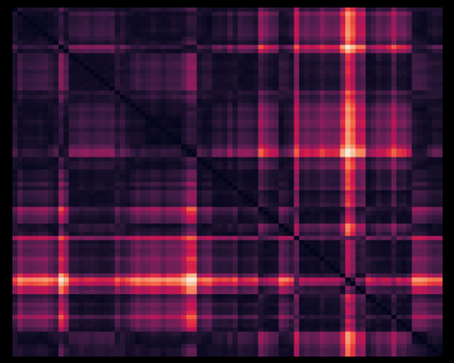


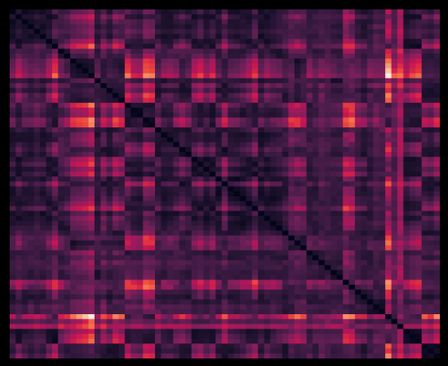
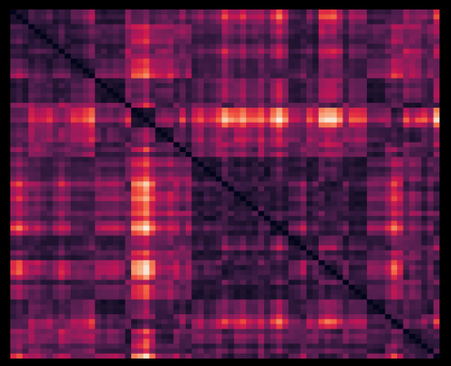
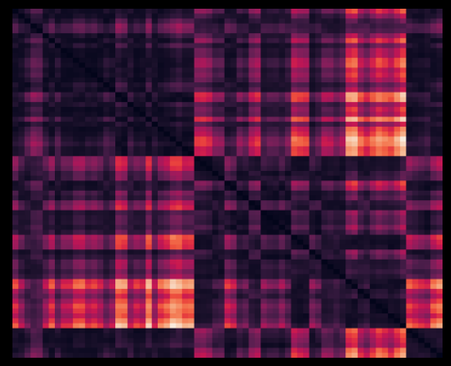
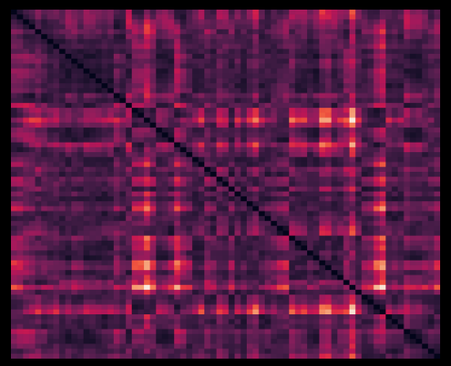
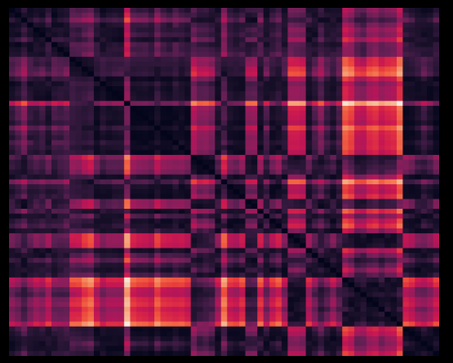
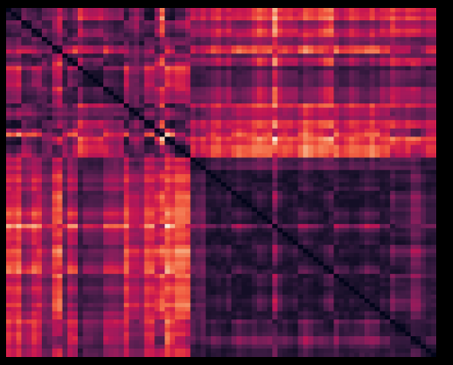
Knee angle



Dynamic Time Warping - Heat Map



Take home messages

- Environment clearly impacts the locomotor behavior of *Anolis* lizards
- The choice of statistical method has significant effect on the information we can extract from this dataset
- Although univariate methods can provide answers that are easily interpretable, more sophisticated and modern multivariate, classification, and time series methods can provide new, valuable insights that could change how biomechanists approach their data analysis

Questions?

website: www.comparativebiomechanics.com
email: klfoster@bsu.edu
twitter: @KathleenLFoster