

**TIRED OF ANALYSING
ALL THE VARIABLES
ONE BY ONE?**

MULTIVARIATE ANALYSIS!

- **Multivariate analysis is a good name**

Observation of variance and statistical analysis of **multiple** variables

MULTIVARIATE ANALYSIS

Canonical Correlation

- Multivariate Correlation: This is "just" a Correlation!
- Correlation of multiple Explanatory and multiple Response variables at once.

Discriminant Analysis

- Looks for the combination of variables that better characterise samples/observations according to different groups.
- Allows to make predictions and classify new samples in predefined groups

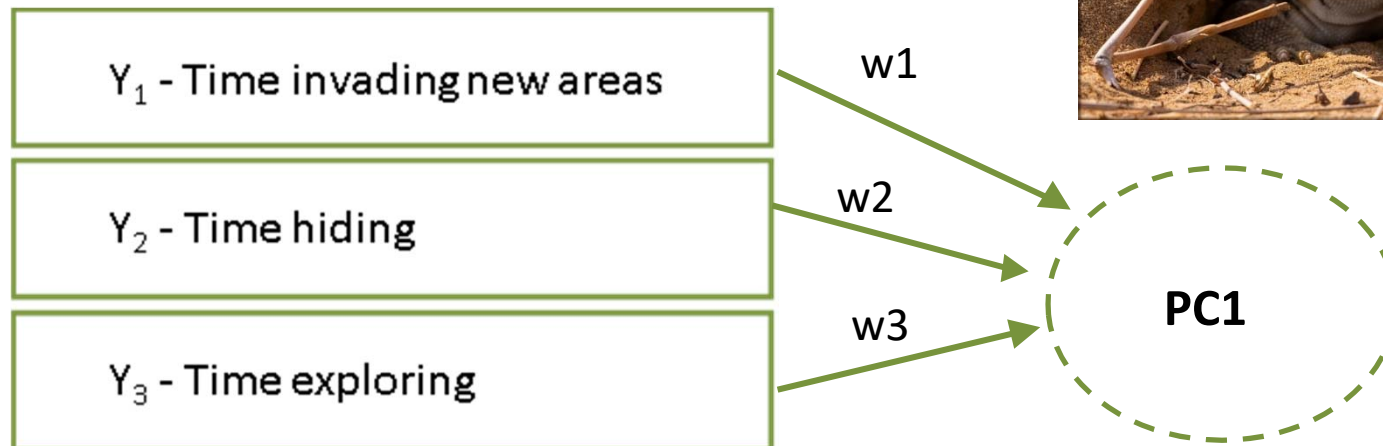
Multivariate ANOVA (MANOVA)

Allows to perform an ANOVA with multiple explanatory variables (groups/treatments) and multiple response variables

MULTIVARIATE ANALYSIS

Principal Component Analysis (PCA)

- Exploratory analysis tool
- Data reduction technique
- Reduces the dimensionality from many to few relevant independent variables
- The extracted independent variance (w) of each measure (Y) is added to the derived independent principal components (PC).



MULTIVARIATE ANALYSIS

Principal Component Analysis (PCA)

- Multivariate Variance-Covariance Analysis.
- Reduces dimensions by linear combinations of variables.
- Emphasises patterns by treating samples/observations as cases and then calculating the covariance matrices of all variables for each one

Correspondence Analysis (CA)

- Similar to PCA
- Treats samples and variables at the same level focusing on relative values.

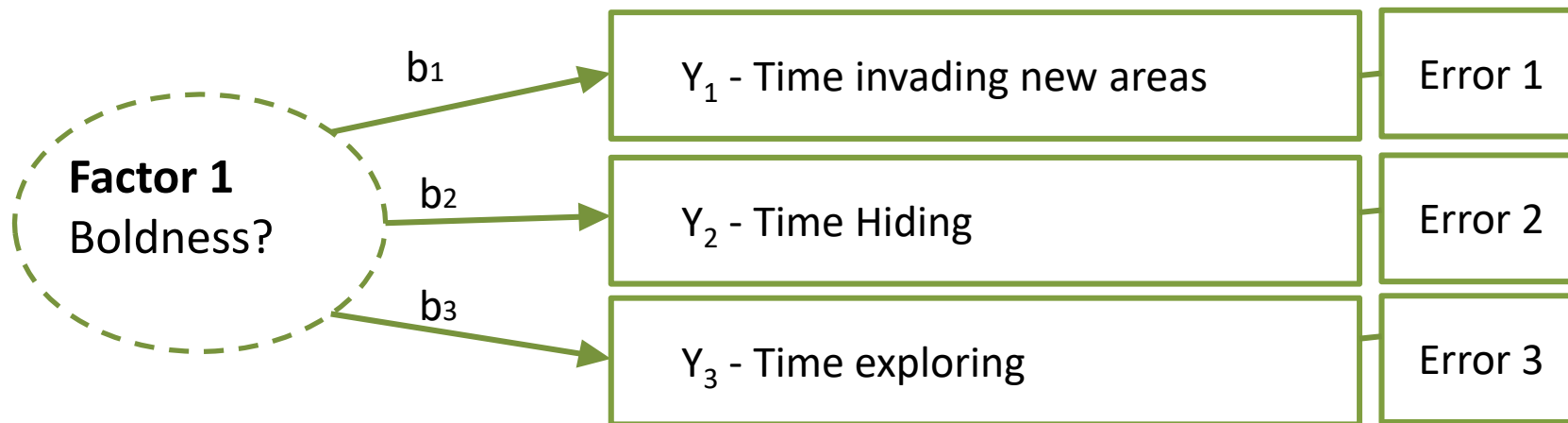
Principal Coordinate Analysis (PCoA)

- Similar to PCA but for non continuous variables (discrete or qualitative).

MULTIVARIATE ANALYSIS

Factor Analysis (EFA)

- Exploratory analysis tool
- Data reduction technique
- Model to find underlying/latent **factors** from which the observed variables were generated
- It analyses the relation (b) of each Factor with the variables (Y) and the error (variance) not explained by that underlying factor



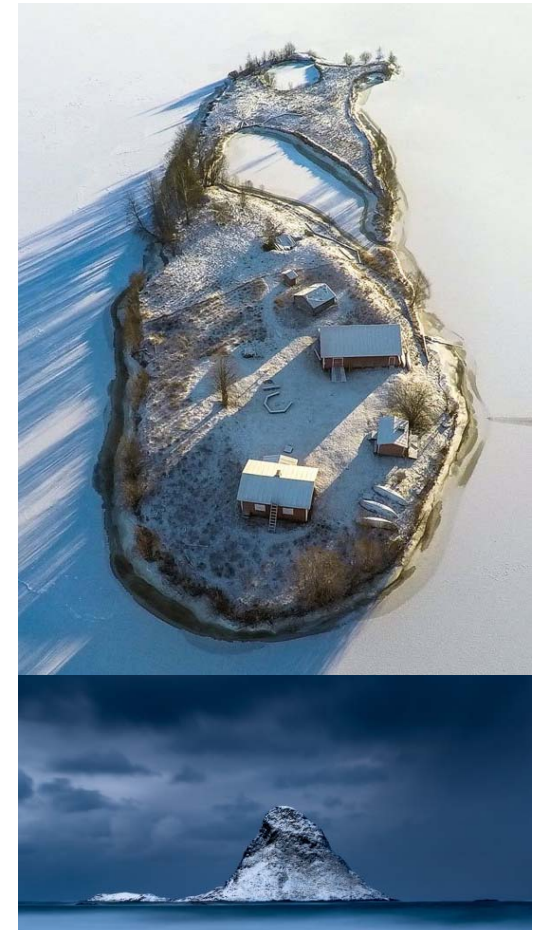
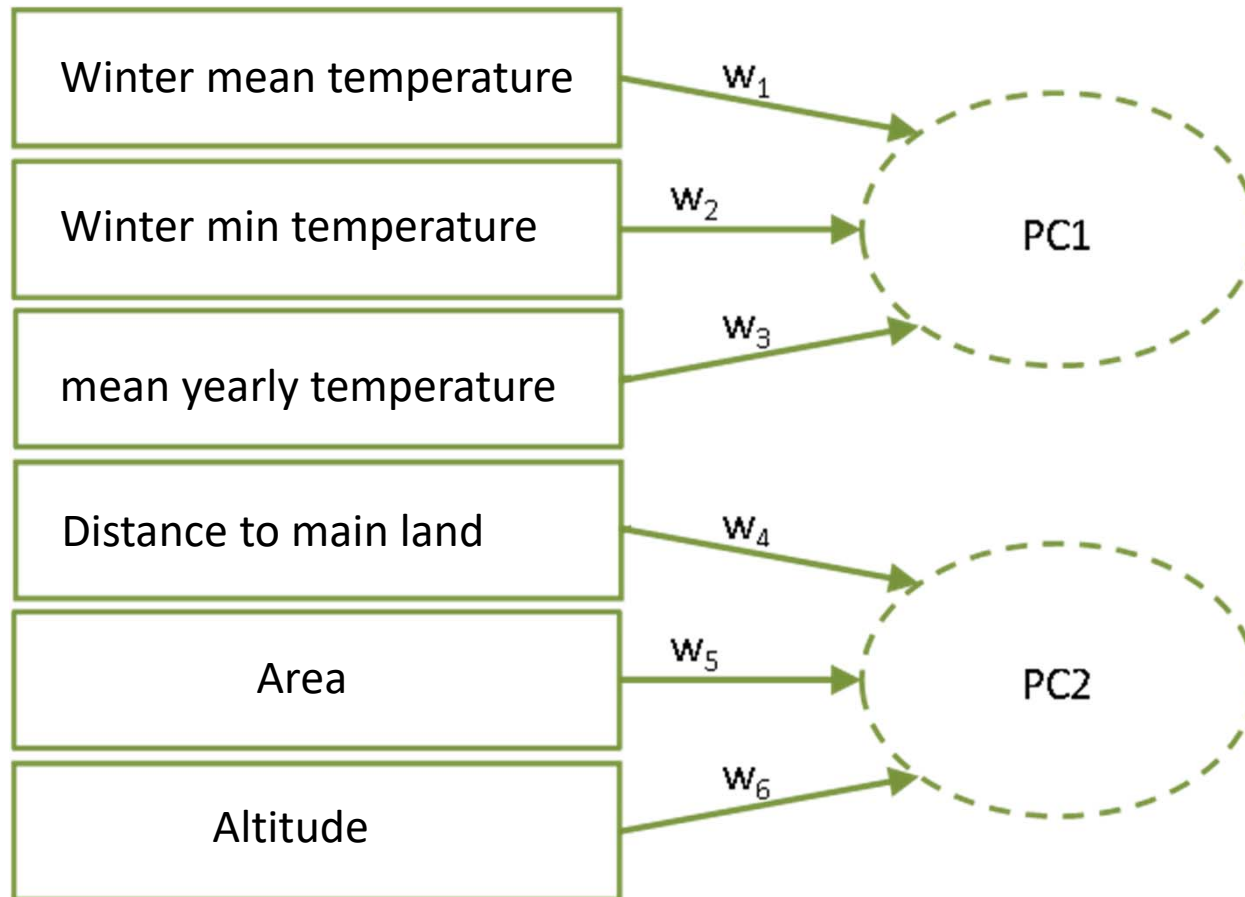
MULTIVARIATE ANALYSIS

Principal Component Analysis (PCA)

- Exploratory analysis tool
- Data reduction technique
- Reduces the dimensionality from many to few relevant independent variables
- The extracted independent variance (w) of each measure (Y) is added to the derived independent principal components (PC).

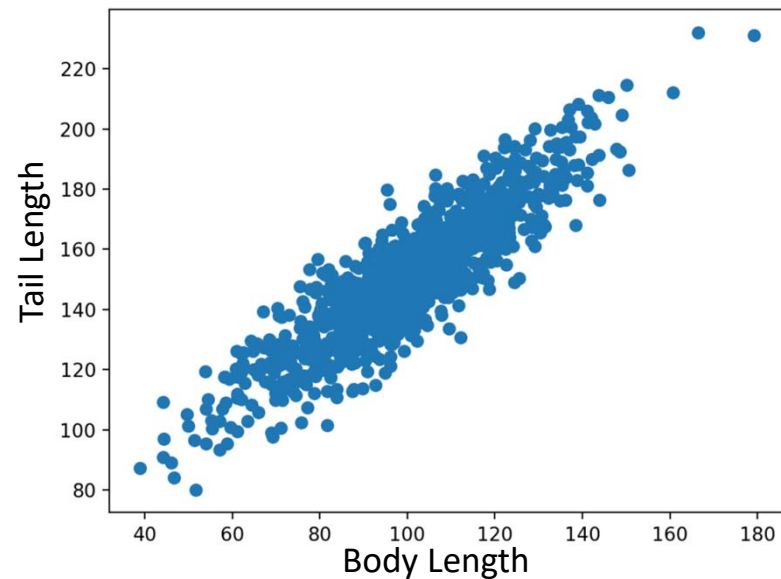


PRINCIPAL COMPONENT ANALYSIS

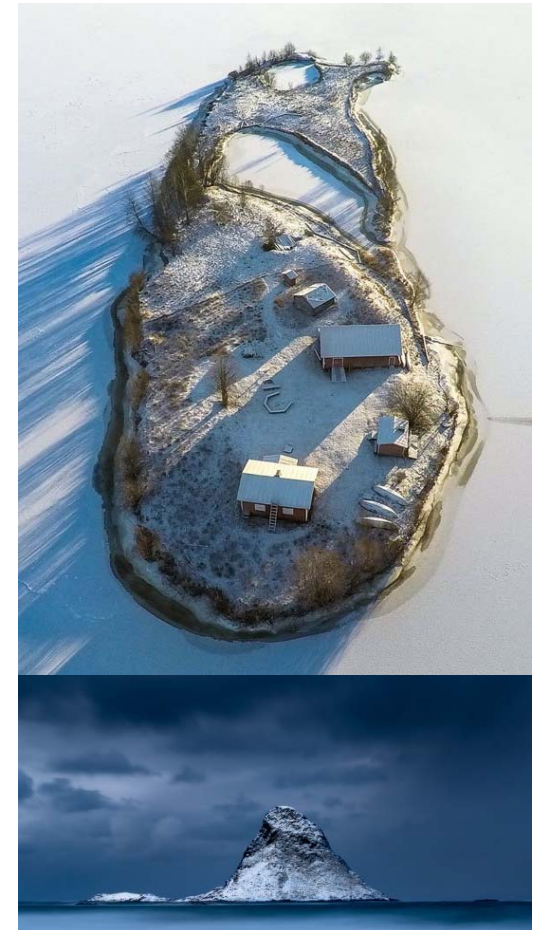
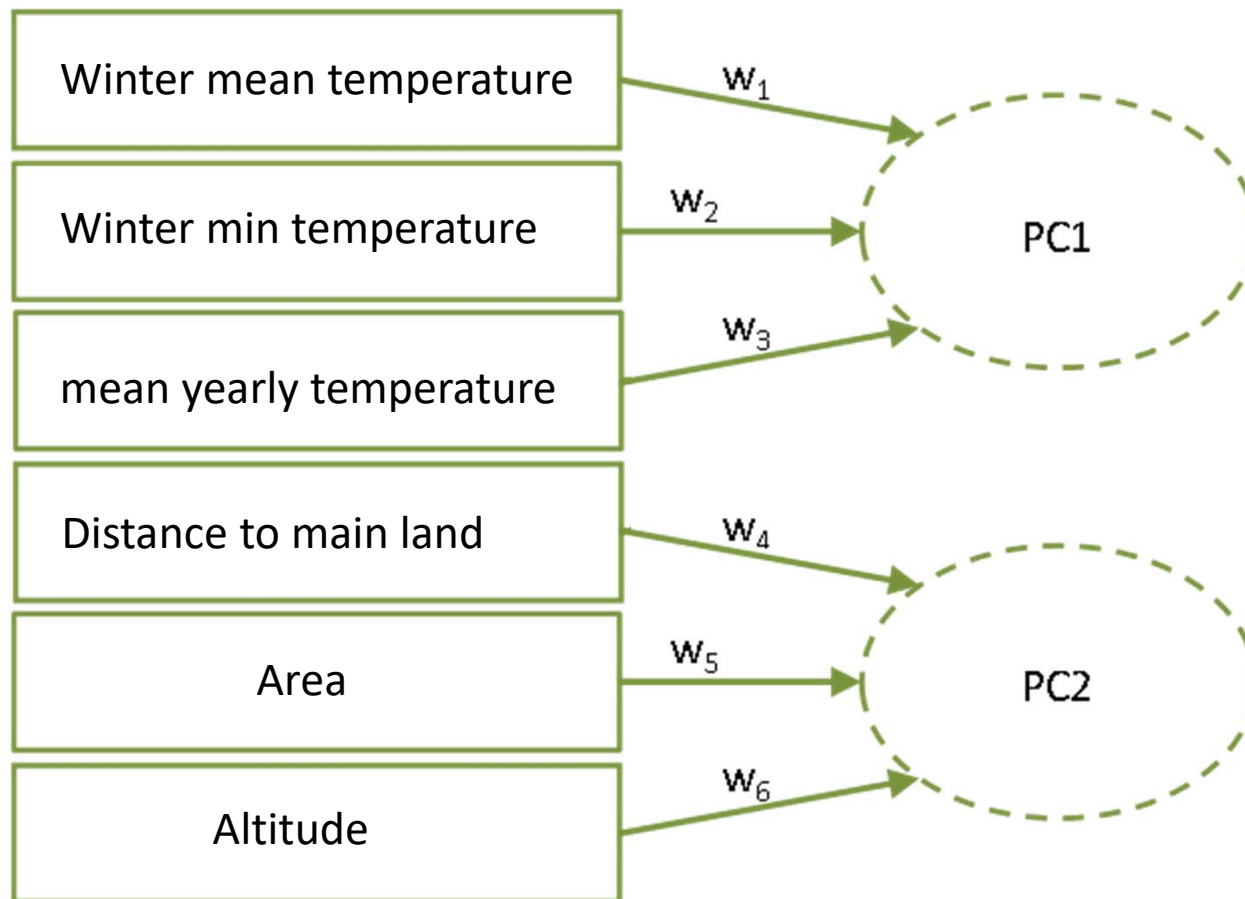


independent variables!

- choosing only independent variables we lose information
- Keeping all variables we overestimate their importance

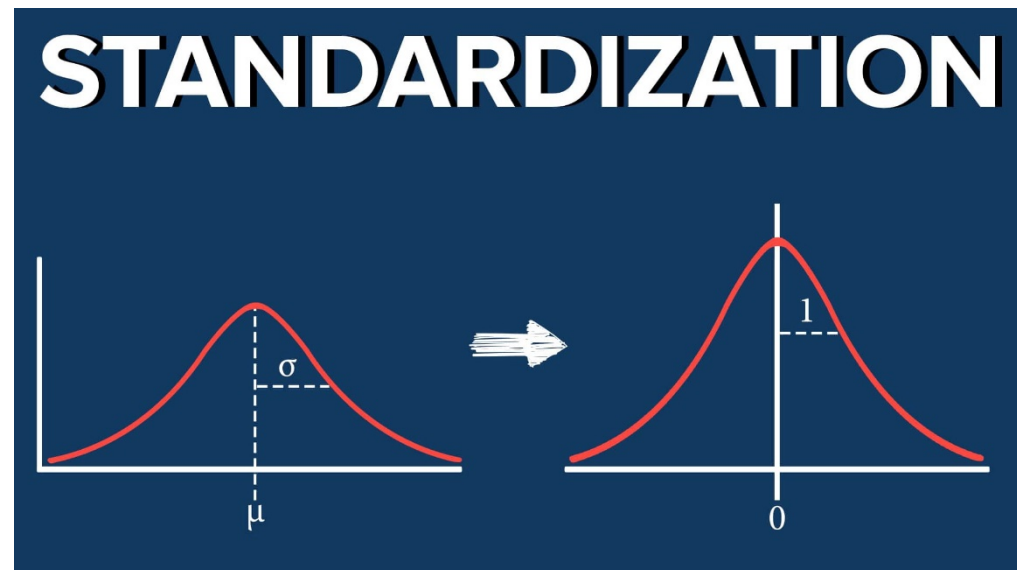


PRINCIPAL COMPONENT ANALYSIS



Magnitude

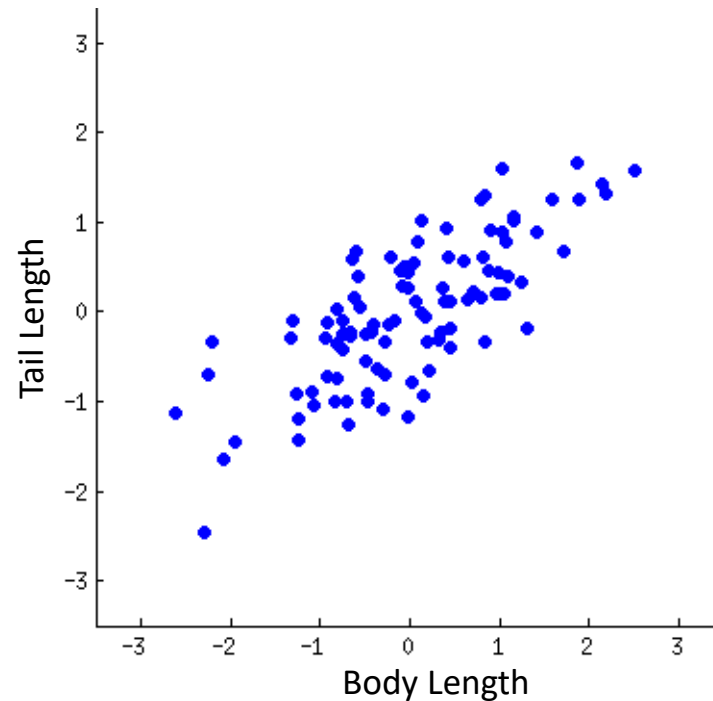
- We will analyse all variables together
- If magnitude of variables is different importance of bigger magnitudes will be overestimated
- We will need to standardize: *scale()*



PRINCIPAL COMPONENT ANALYSIS

Calculates a Covariance matrix (for each pair of variables)

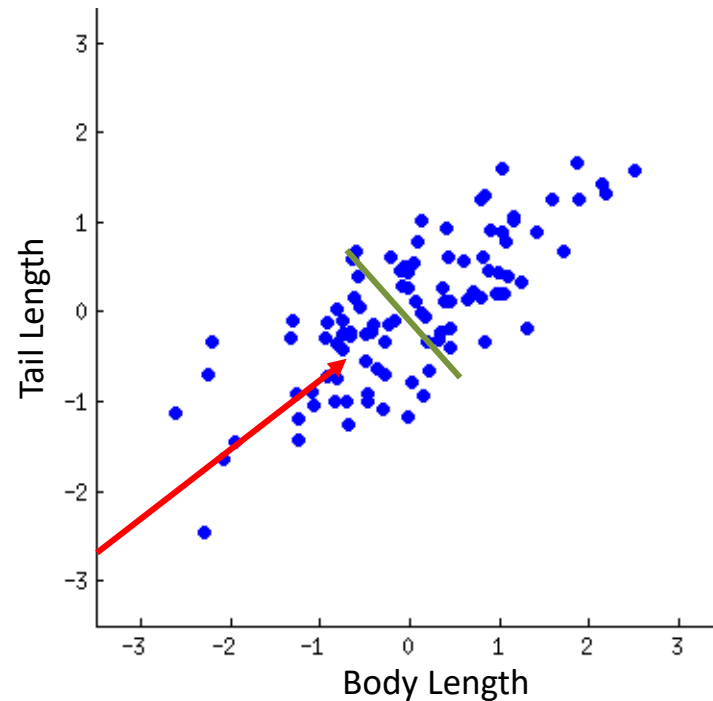
- Information about the variance: spread of the data
- Information about the covariance: in which direction



PRINCIPAL COMPONENT ANALYSIS

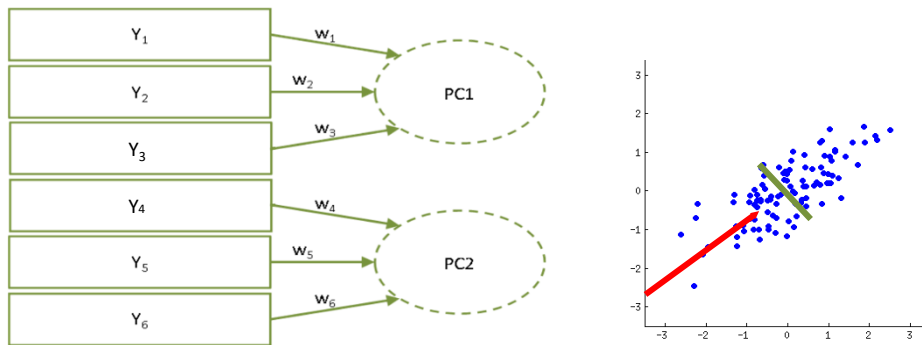
Covariance matrix summarized as:

- Eigenvalue: magnitude or scale of the **spread** of the data (variance)
- Eigenvector: in which **direction** is the data spread (covariance)

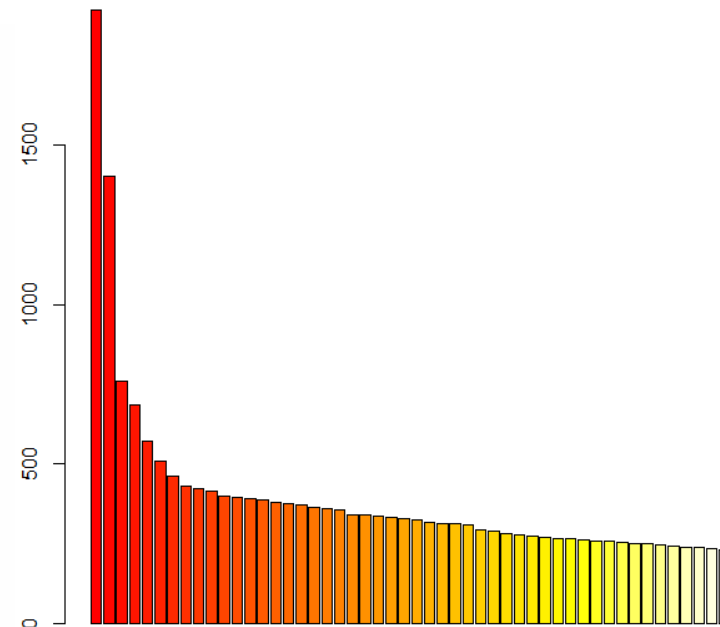
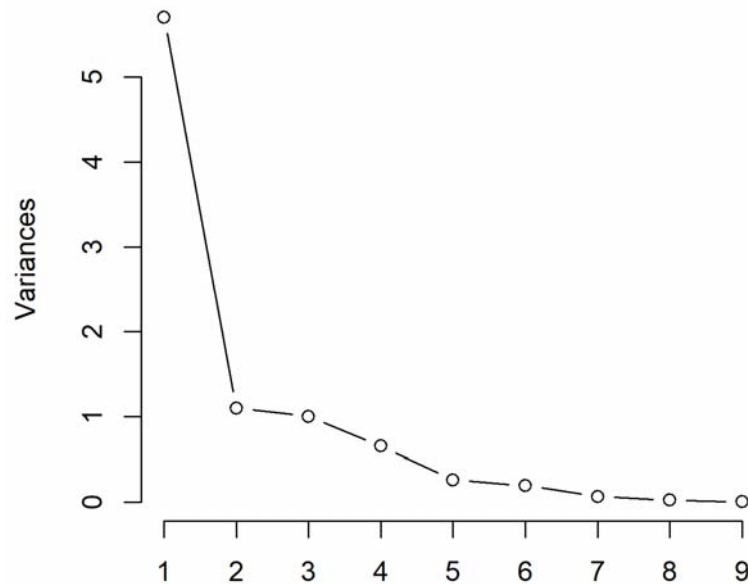


PCA: eigen values

Magnitude or scale of the **variance** of the data

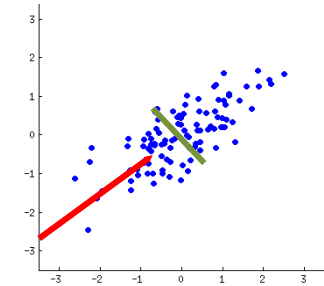
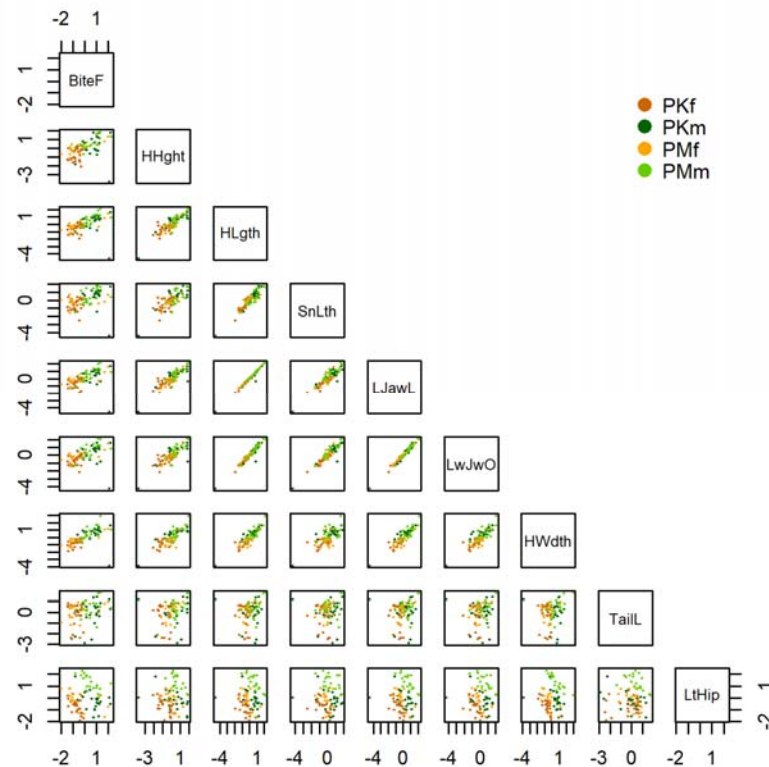
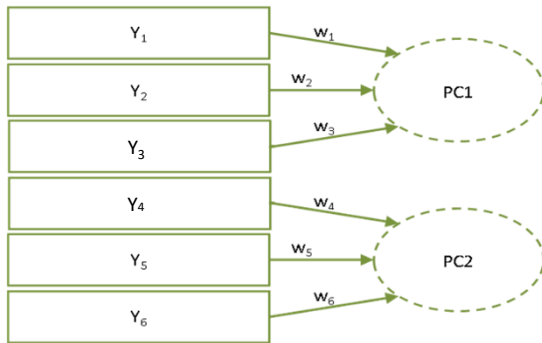


PCA eigenvalues



PCA: biplots

correlations are done for all combinations of variables

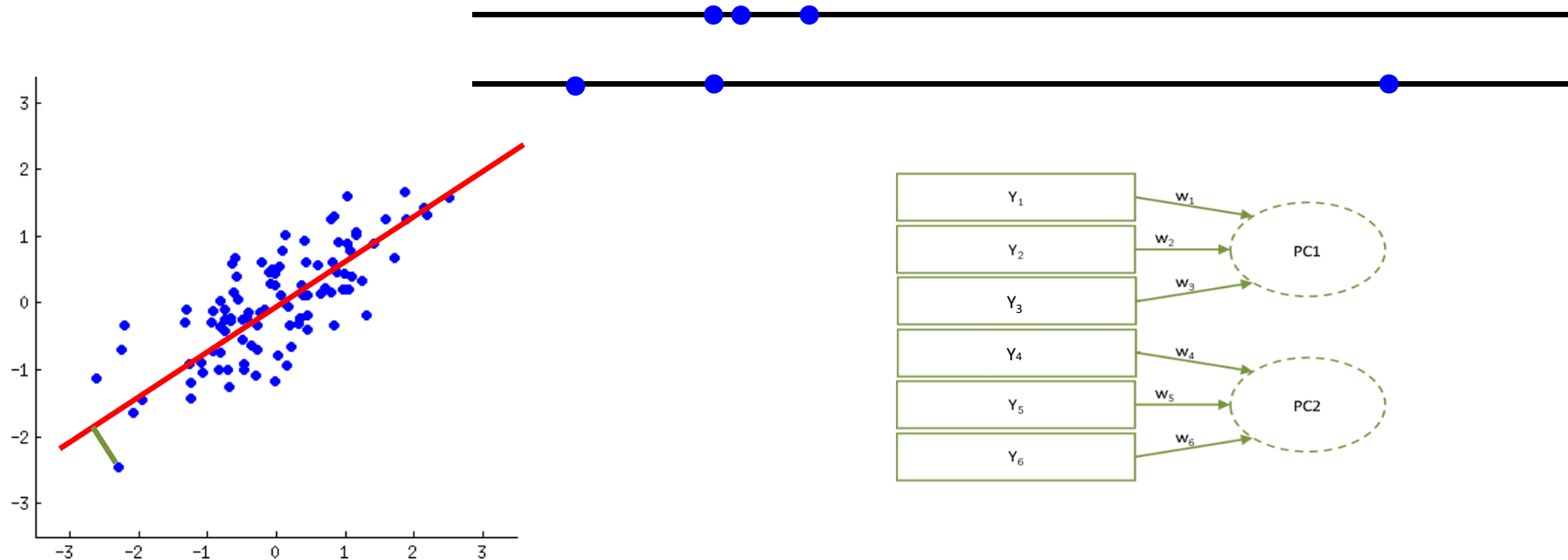


PRINCIPAL COMPONENT ANALYSIS

Principal Components

- Independent variables that summarize the covariance matrices of the data
- Maximise the **variance** and at the same time that minimize the **error**

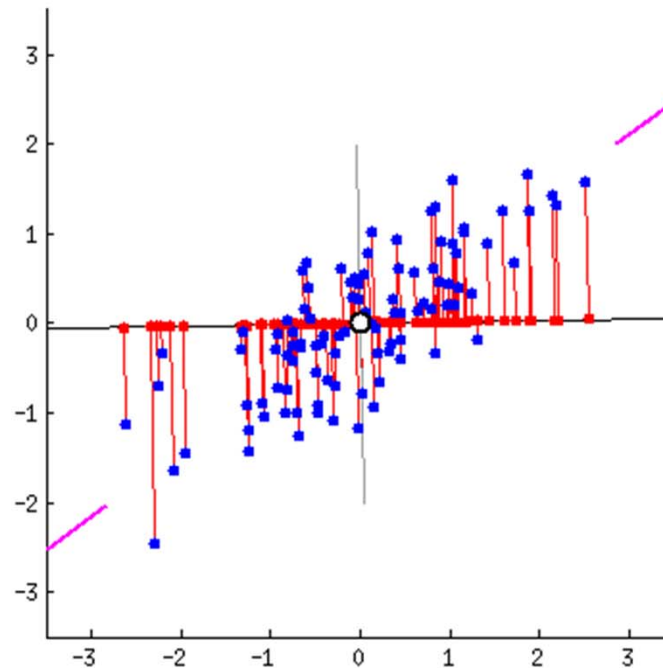
How does PCs at the same time maximise the variance and minimize the error?



PRINCIPAL COMPONENT ANALYSIS

Principal Components

How does PCs at the same time maximise the variance and minimize the error?



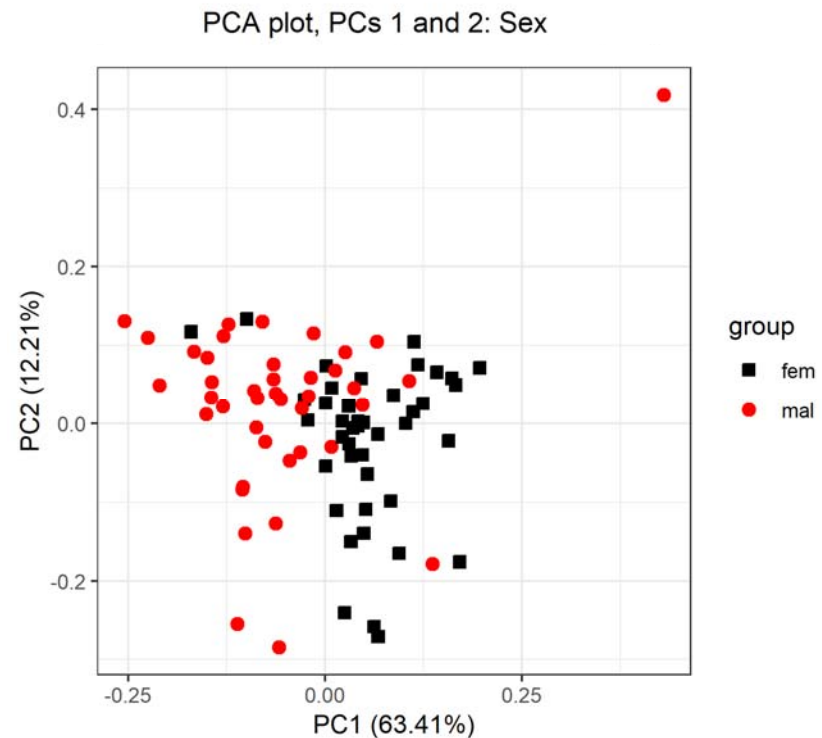
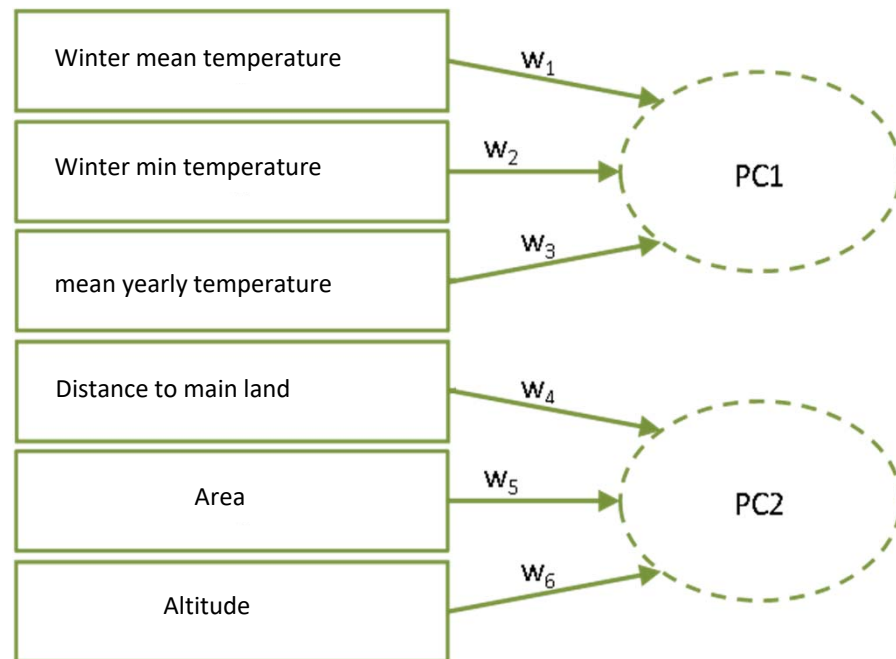
StackExchange
Cross Validated
Dennis

PRINCIPAL COMPONENT ANALYSIS

Scores

Transformed data values per observation according to the model

Coordinates of the samples in the new dimensional space (new variables).
Scores of each sample in each PC may allow us to sort/group them.

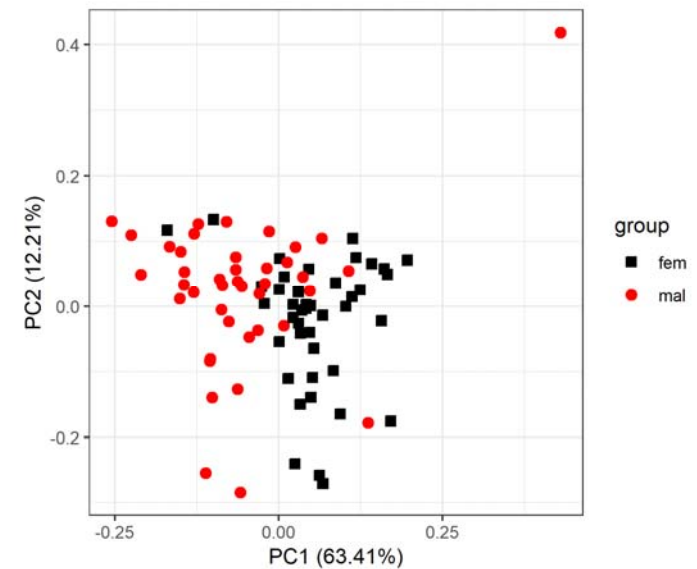
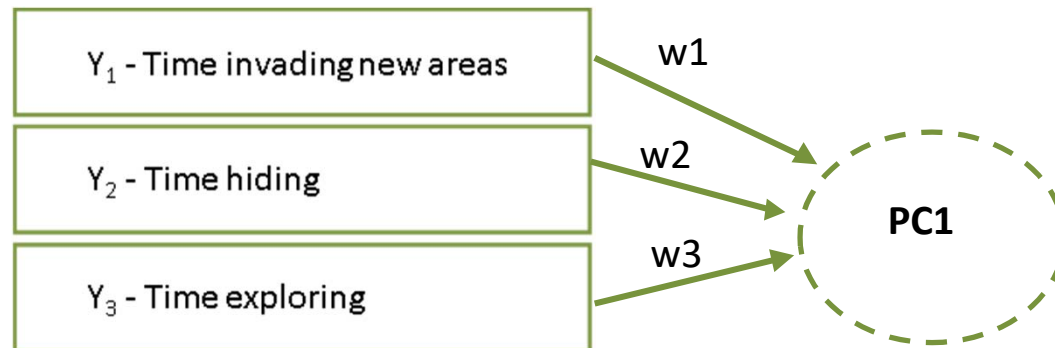


PRINCIPAL COMPONENT ANALYSIS

Loadings

covariances/correlations between the original variables and the PCs

How important is each variable? In which direction "pulls" the data?

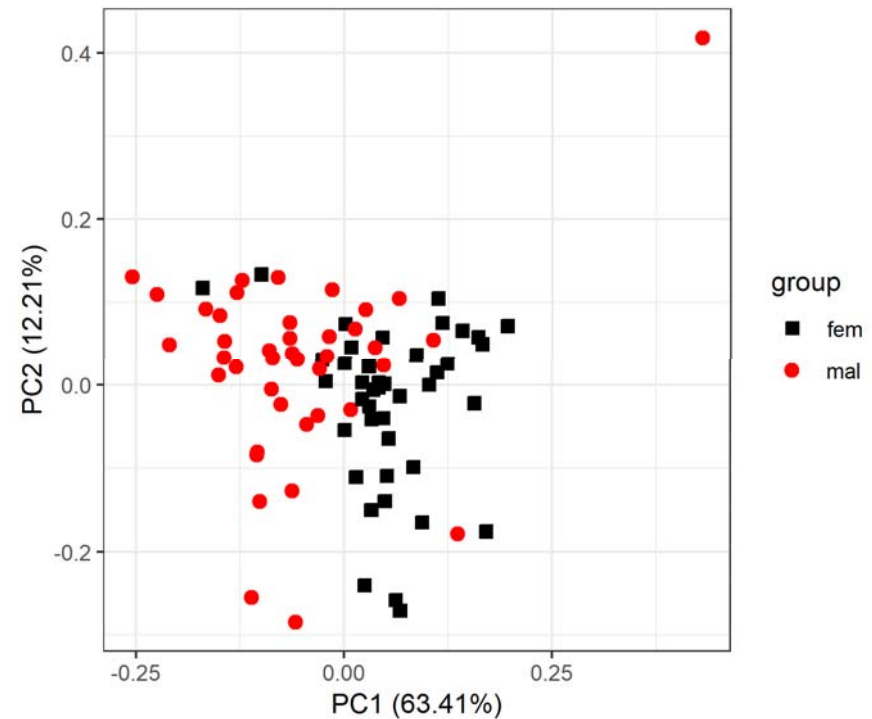
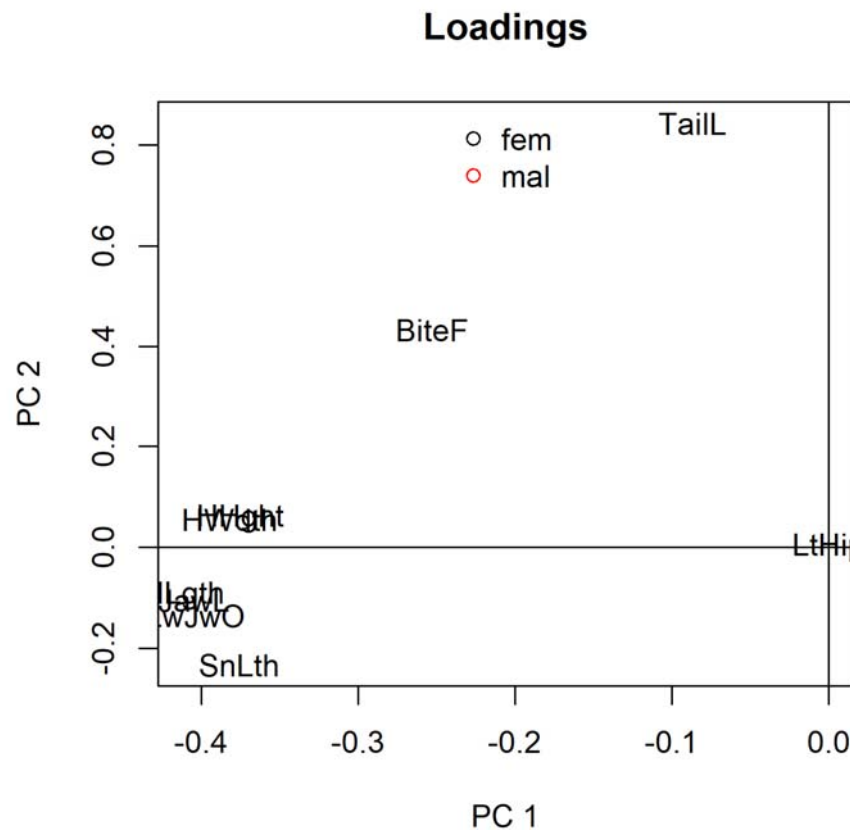


PRINCIPAL COMPONENT ANALYSIS

Loadings

covariances/correlations between the original variables and the PCs

How important is each variable? In which direction "pulls" the data?



PRINCIPAL COMPONENT ANALYSIS

Why?

See distribution of data according to the variance

Check if variables allow to group samples

Look for relevant independent variables

AGAIN:

- **PCA is a blind exploratory test.**
- **No assumptions about our groups.**
- **Summarizes all the variable's variance**

Requirements: Continuous data, no missing values, similar scale

Very robust to: Over-fitting, over-representation

Robust to: No normality of the data (but it is recommended)

PRINCIPAL COMPONENT ANALYSIS

Procedure

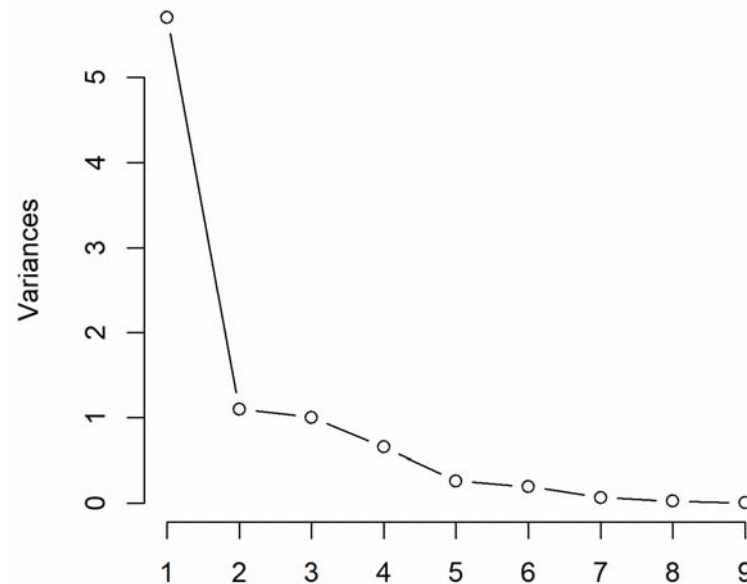


1. Reduce the number of dimensions (relevant PCs)
2. See if there are groups
3. Account the influence of each variable in the PCs (loadings)
- (4. Check significance of the groups)

PRINCIPAL COMPONENT ANALYSIS

Procedure

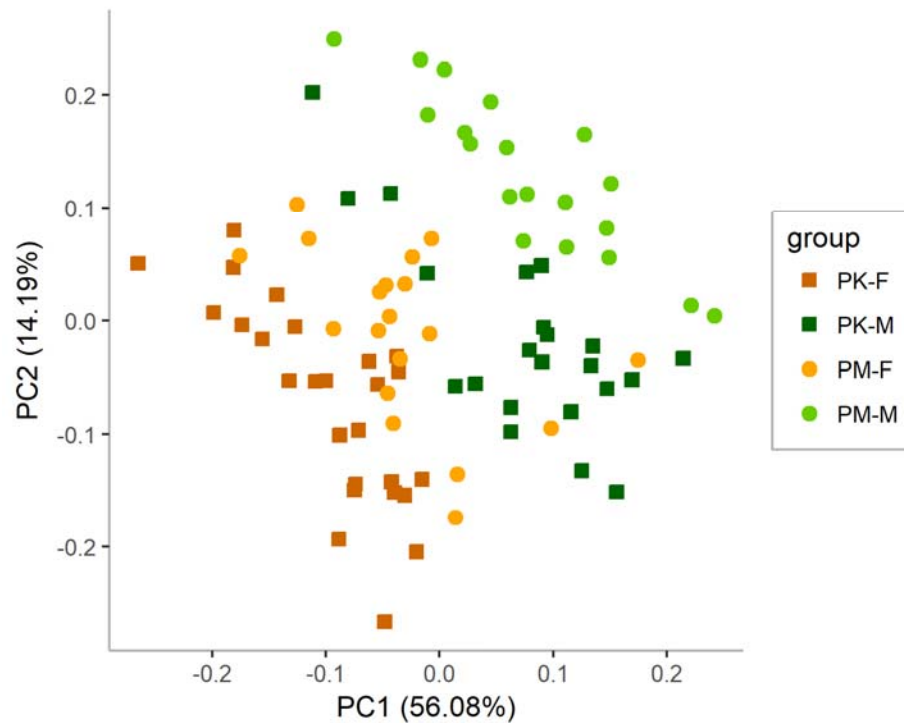
- 1. Reduce the number of dimensions (relevant PCs)**
2. See if there are groups
3. Account the influence of each variable in the PCs (loadings)
- (4. Check significance of the groups)



PRINCIPAL COMPONENT ANALYSIS

Procedure

1. Reduce the number of dimensions (relevant PCs)



Variance explained by each PC?

Choose PCs to plot

Usually PC1 and 2 is enough

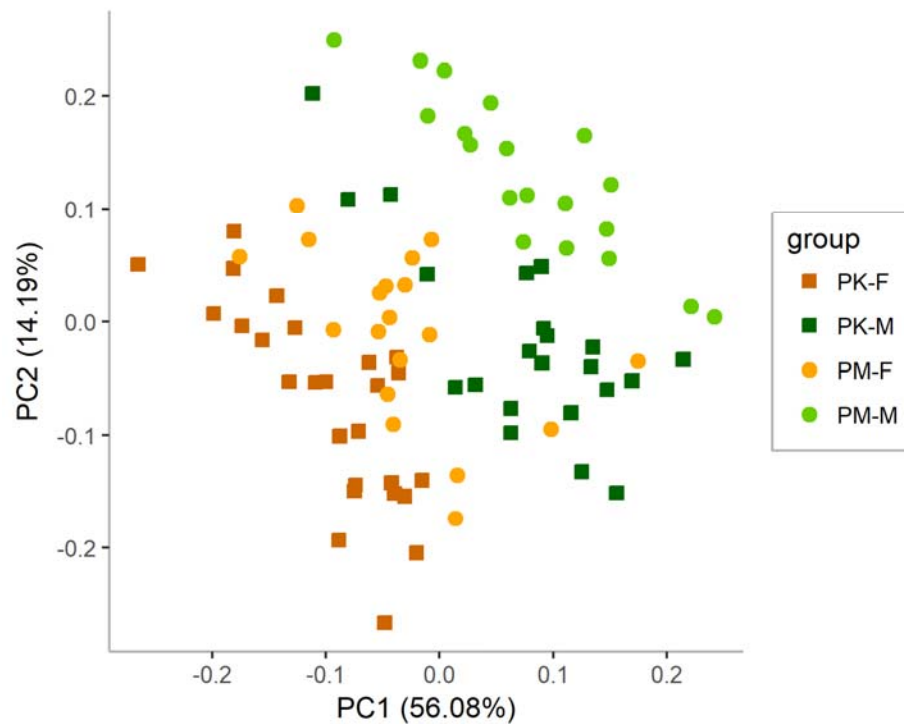
May be interesting to plot more

We could do a 3D plots with 3 PCs

PRINCIPAL COMPONENT ANALYSIS

Procedure

2. See if there are groups



Do the PCs that explain most of the variance define any biologically meaningful group?

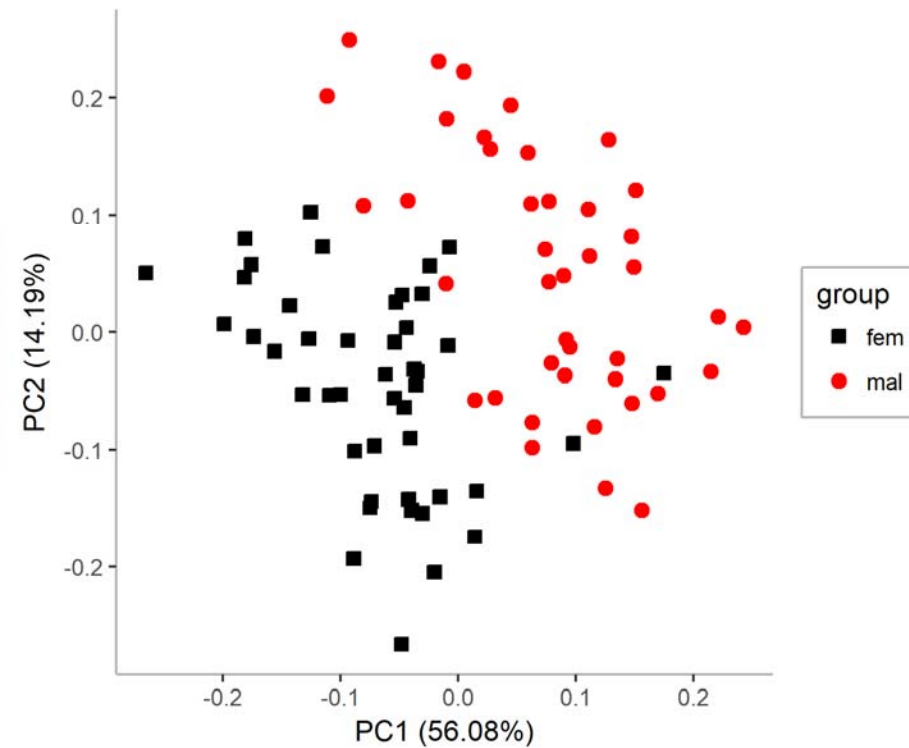
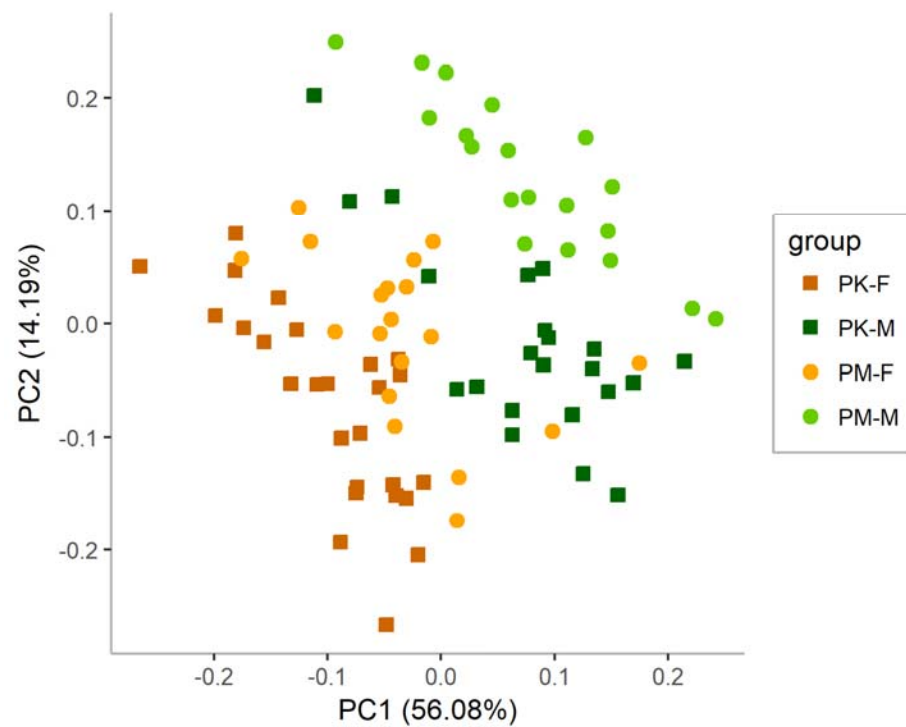
Colour the plot according to different grouping variables

Check if new patterns / groups emerge

PRINCIPAL COMPONENT ANALYSIS

Procedure

2. See if there are groups



PRINCIPAL COMPONENT ANALYSIS

Procedure

3. Account the influence of each variable in the PCs (loadings)

Check loadings of each variable in PCs that show a pattern / group your samples

Variables that contribute the most to relevant PCs are worth to be analysed further

Variables that contribute to the same PC may be related

