A tutorial on principal component analysis

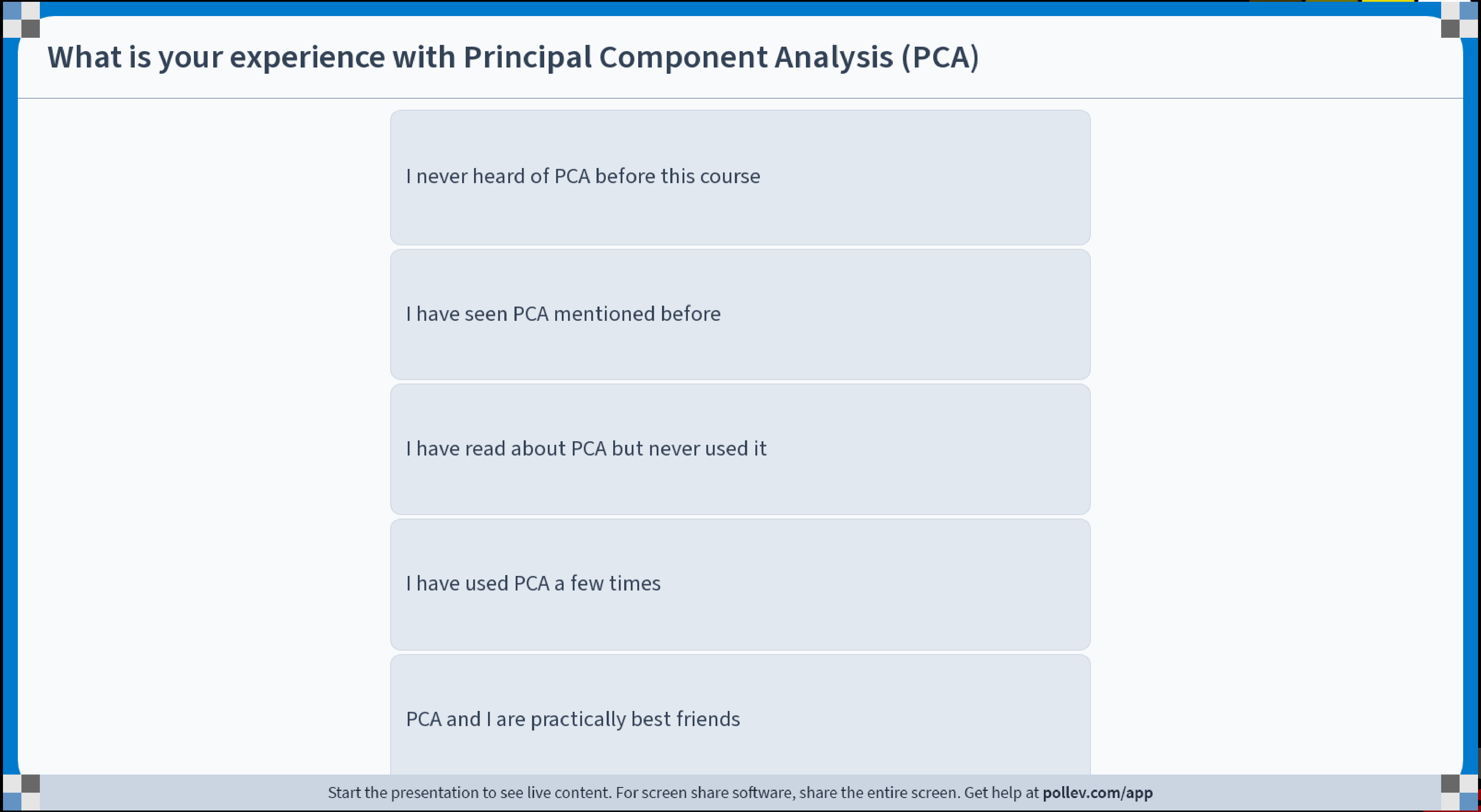
Rasmus R. Paulsen

DTU Compute

Based on

Jonathan Shlens: A tutorial on Principal Component Analysis (version 3.02 – April 7, 2014)

http://compute.dtu.dk/courses/02515



What is your experience with Principal Component Analysis (PCA)

I never heard of PCA before this course

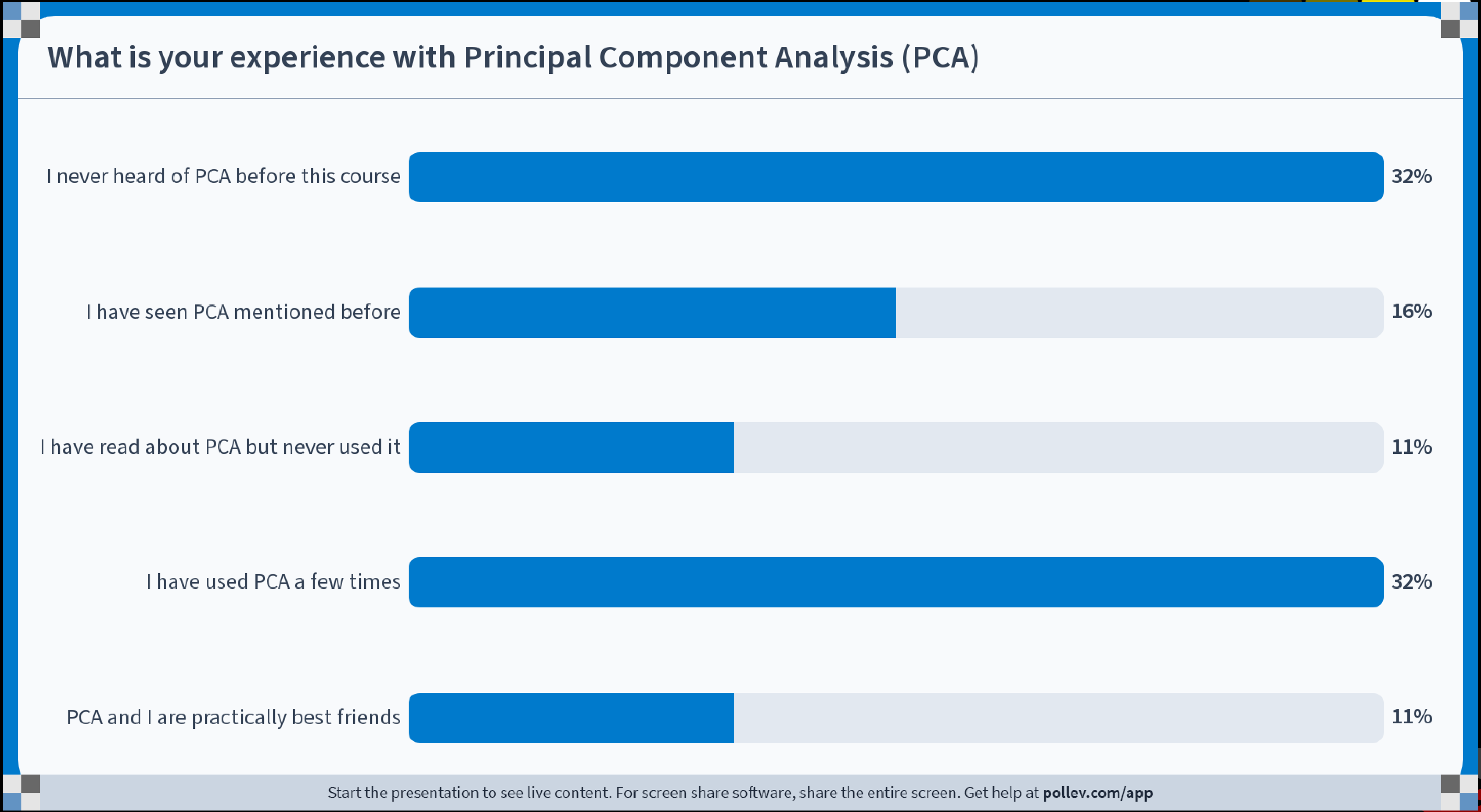
I have seen PCA mentioned before

I have read about PCA but never used it

I have used PCA a few times

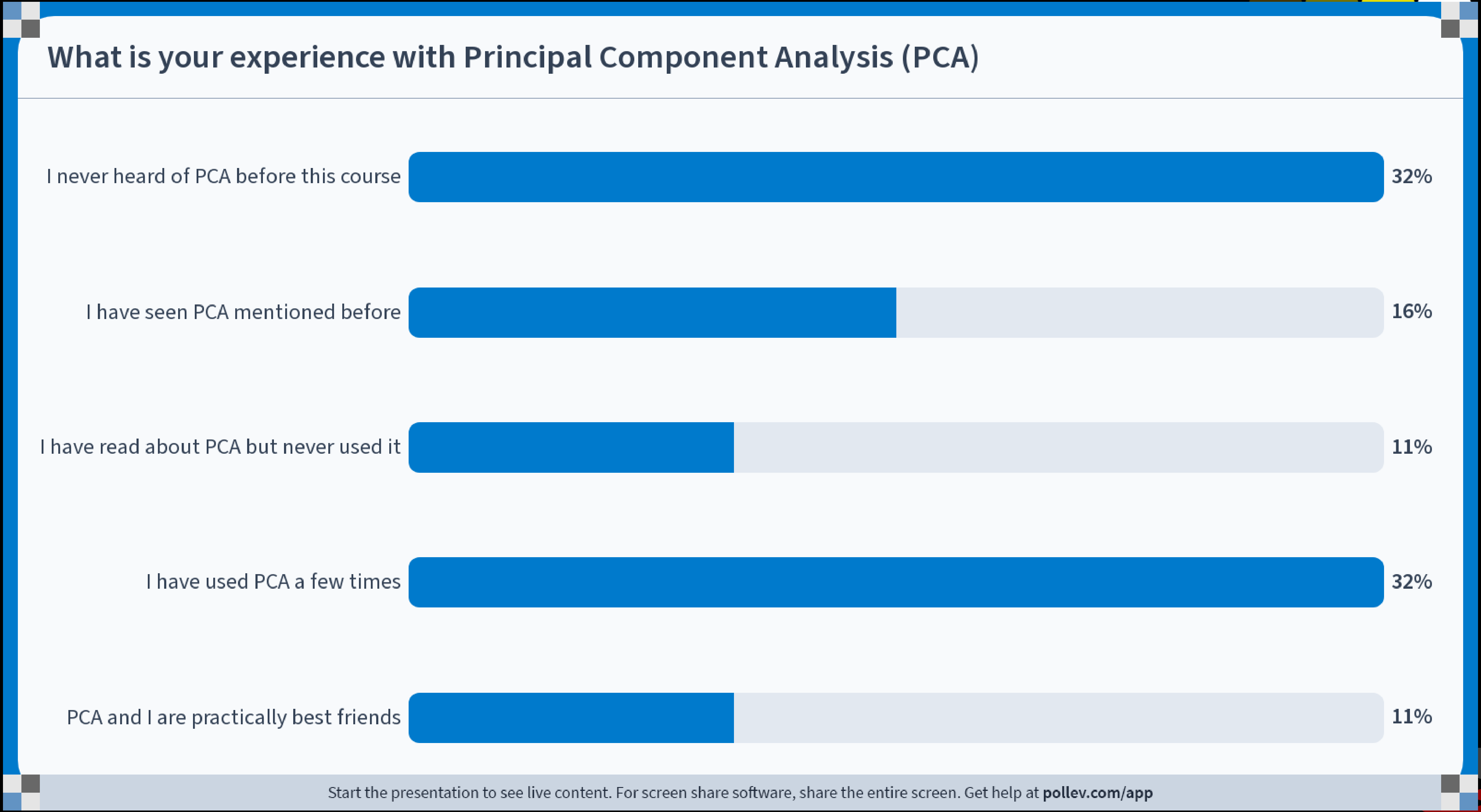
PCA and I are practically best friends

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**



What is your experience with Principal Component Analysis (PCA)

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**



What is your experience with Principal Component Analysis (PCA)

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

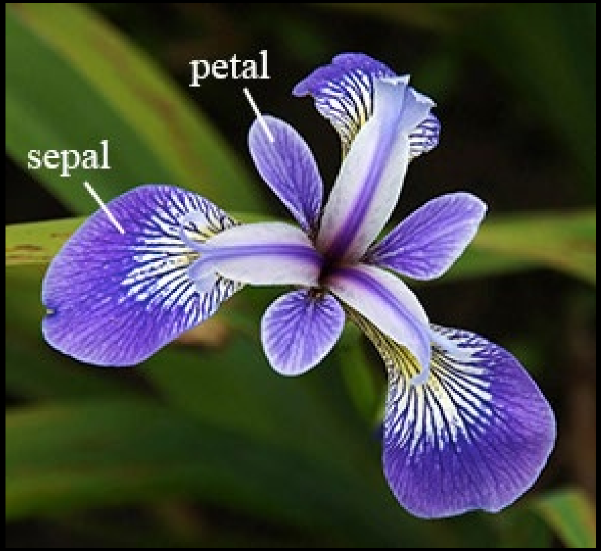
Principal Component Analysis (PCA) learning objectives

- Describe the concept of principal component analysis
- Explain why principal component analysis can be beneficial when there is high data redundancy
- Arrange a set of multivariate measurements into a matrix that is suitable for PCA analysis
- Compute the covariance of two sets of measurements
- Compute the covariance matrix from a set of multivariate measurements
- Compute the principal components of a data set using Eigenvector decomposition
- Describe how much of the total variation in the data set that is explained by each principal component

Iris data

The **Iris flower data** set or Fisher's Iris data set is a data set introduced by Ronald Fisher in his 1936 paper *The use of multiple measurements in taxonomic problems*

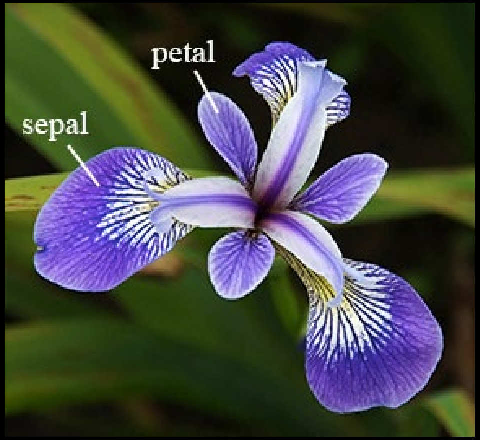
Iris data



- 3 Iris types
  - 50 flowers of each type
- For each flower
  - Sepal length
  - Sepal width
  - Petal length
  - Petal width
- We use one type as example
  - 50 measured flowers

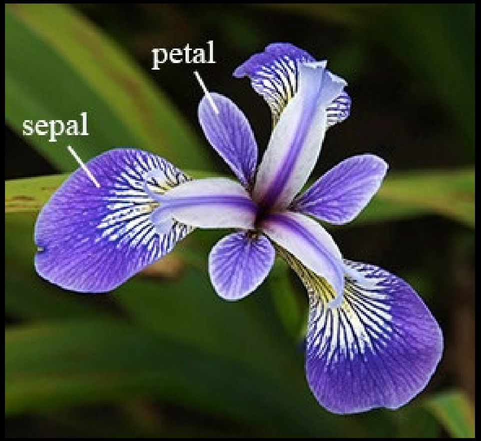
Iris Data Matrix

- One column is one flower
- One row is all measurements of one type



$$\mathbf{X} = \begin{array}{c} \\ \\ \\ \\ \\ \end{array}\begin{array}{ccc} 1 & & 50 \\ \left[\begin{array}{c} \text{Sepal length}_1 \\ \text{Sepal width}_1 \\ \text{Petal length}_1 \\ \text{Petal width}_1 \end{array}\right. & \begin{array}{c} \cdots \\ \cdots \\ \cdots \\ \ldots \end{array} & \left.\begin{array}{c} \text{Sepal length}_{50} \\ \text{Sepal width}_{50} \\ \text{Petal length}_{50} \\ \text{Petal width}_{50} \end{array}\right] \end{array}$$

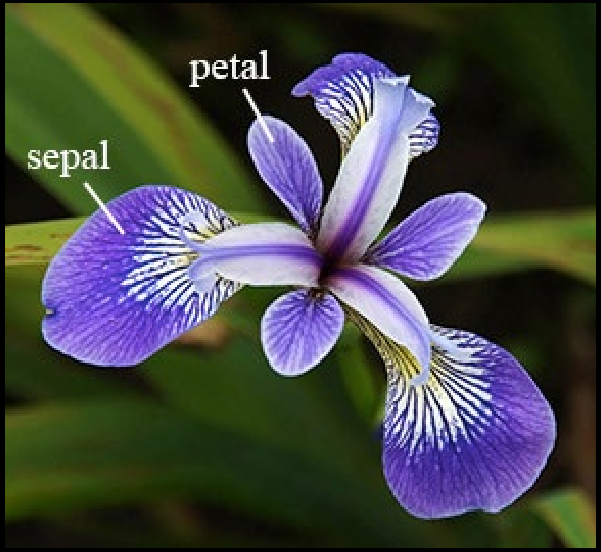
What can we use these data for?

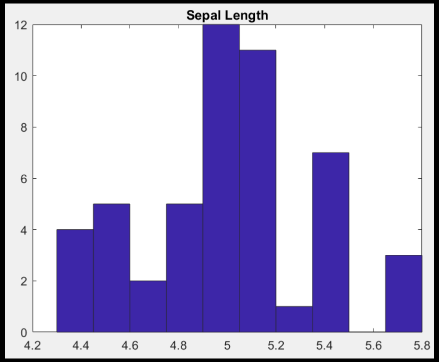


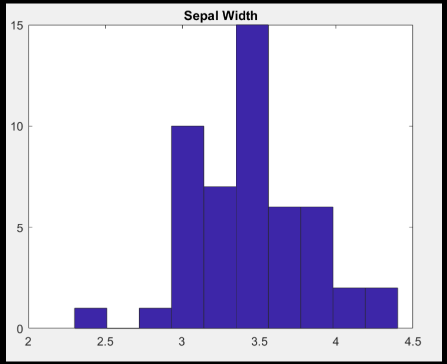
- The measurements can be used to:
  - Recognize a species of flowers
  - Classify flowers into groups
  - Describe the characteristics of the flower
  - Quantify growth rates
  - …
- Do we need all the measurements?
  - Can we *boil down* or *combine* some measurements?
- Are some measurements *redundant?*

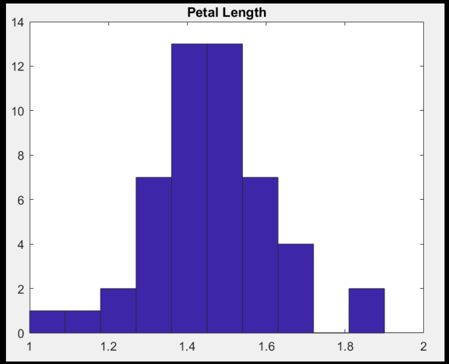
Variance

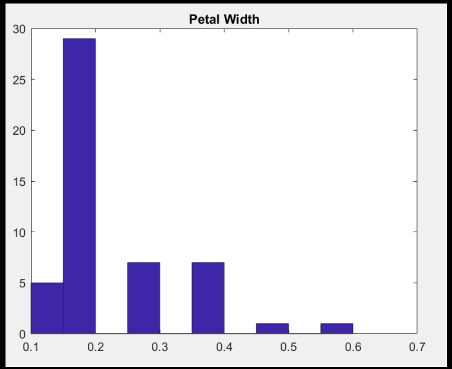
$$\begin{aligned}\sigma^2_{SL} &= 0.1242\\ \sigma^2_{SW} &= 0.1437\\ \sigma^2_{PL} &= 0.0302\\ \sigma^2_{PW} &= 0.0111\end{aligned}$$



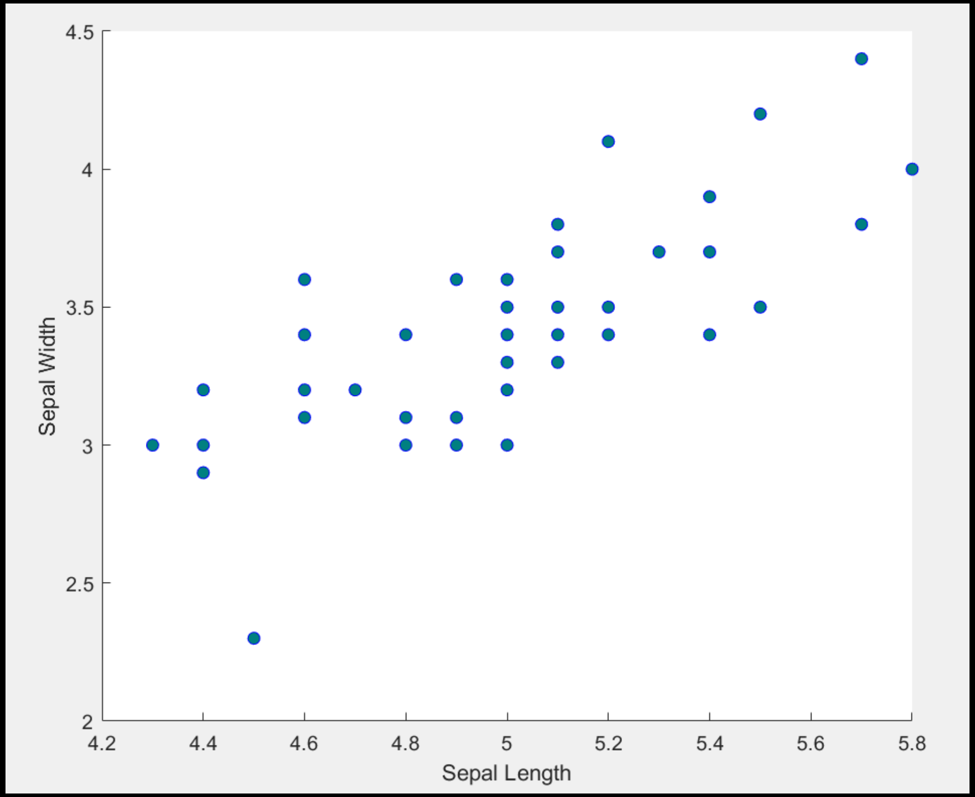




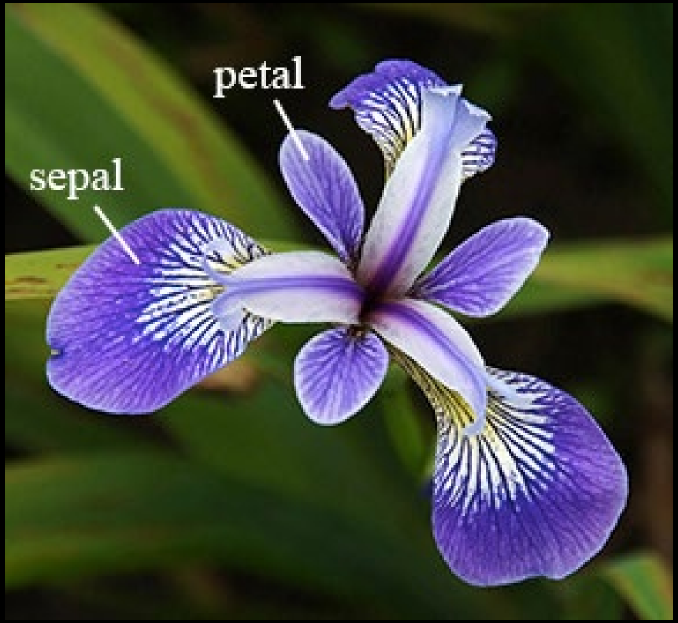




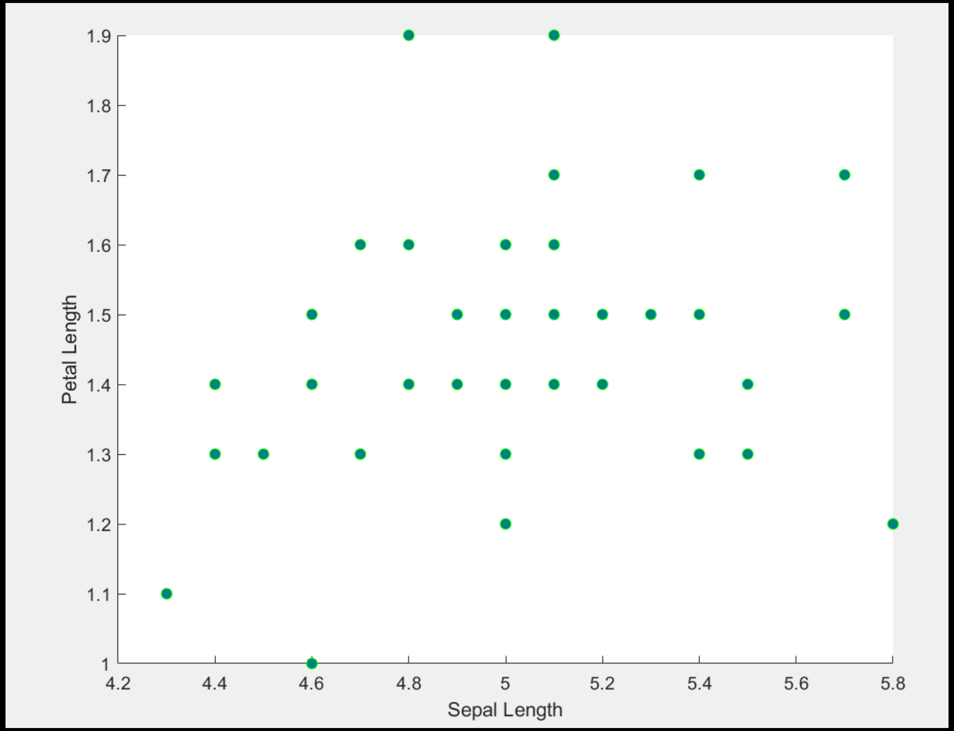
High Redundancy



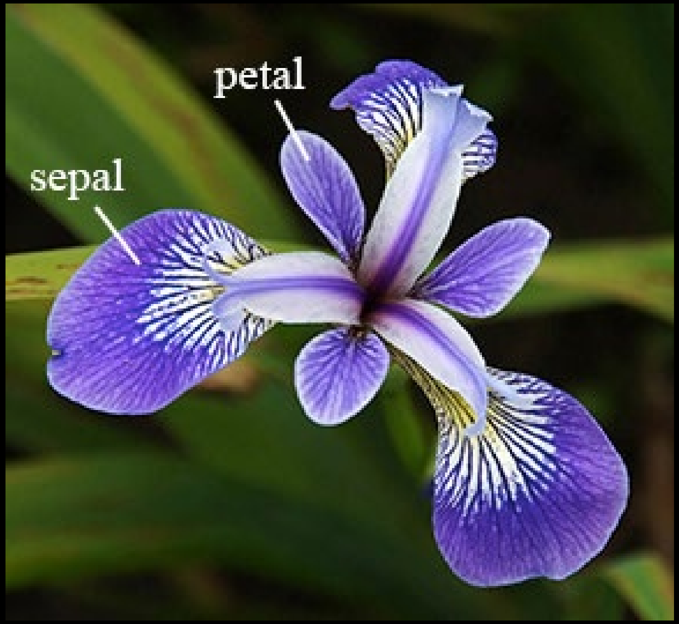
Observation: We can explain quite a lot of the sepal width if we know the sepal lengths



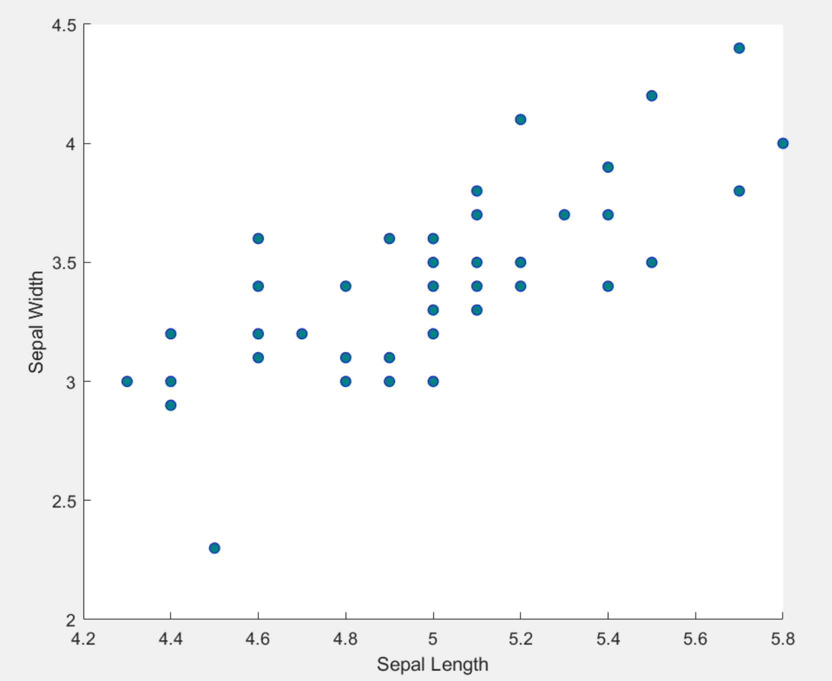
Low Redundancy

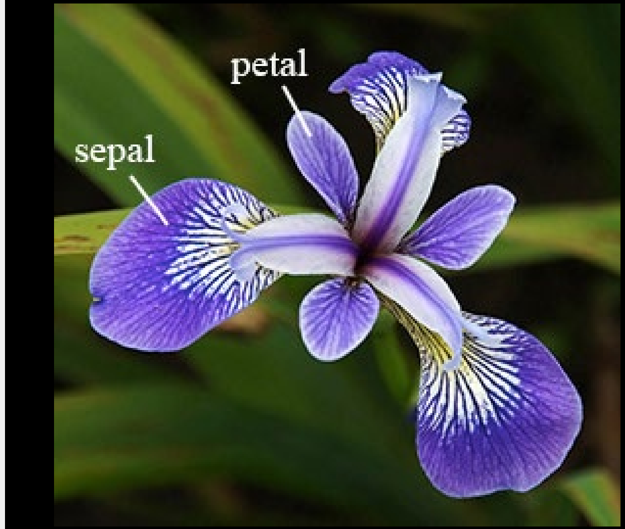


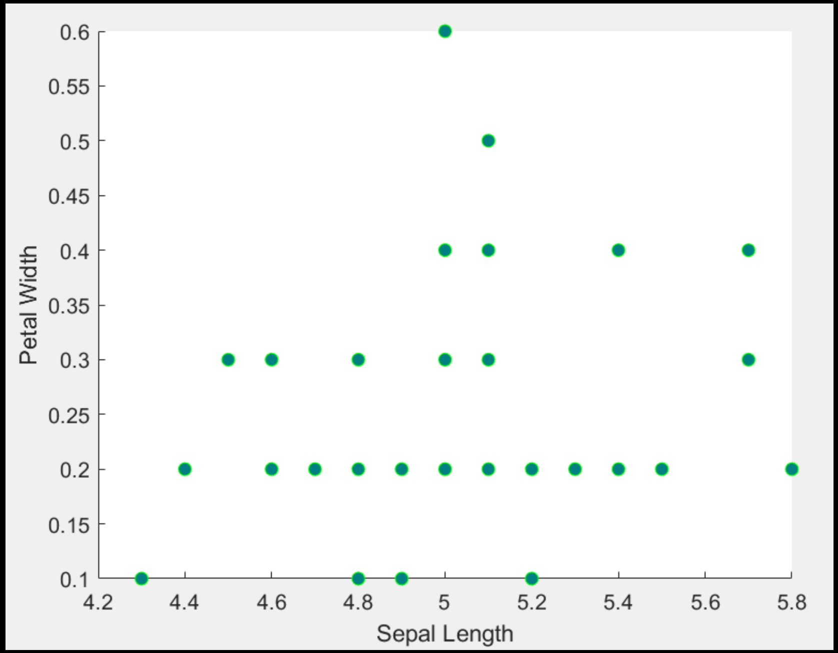
Observation: We can NOT explain the petal length if we know the sepal lengths



Covariance

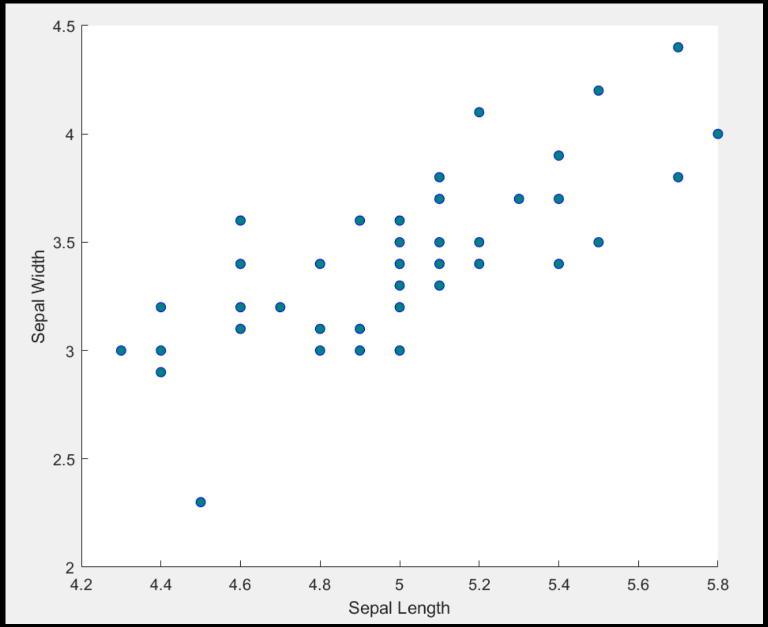


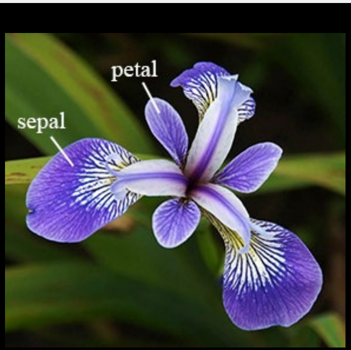




Covariance measures the *relationship* between measurements

High Covariance





Sepal length and sepal width

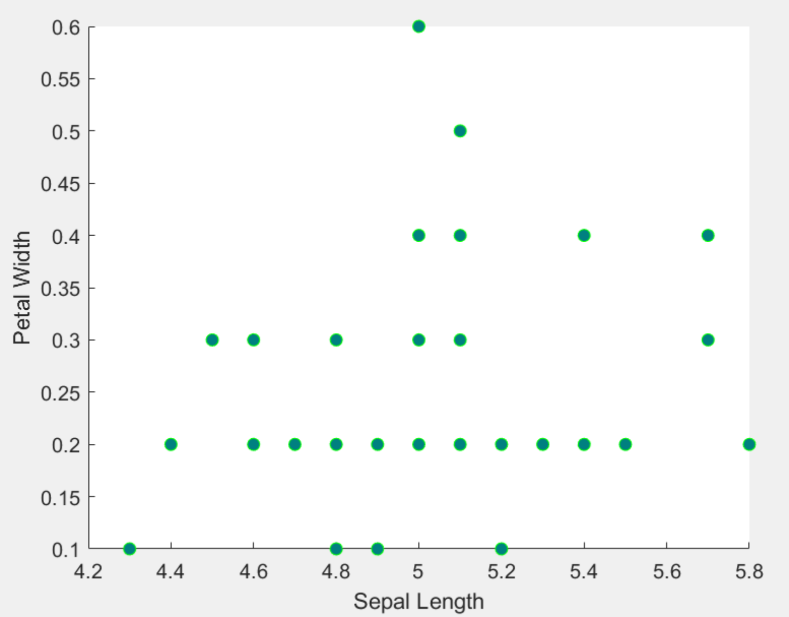
$$a_i = \mathrm{SL} = \{5.1, 4.9 \ldots, 5\}$$

$$\mathrm{b_i} = \mathrm{SW} = \{3.5, 3, \ldots, 3.3\}$$

$$\sigma^2_{\mathrm{SL,SW}} = \frac{1}{n} \sum_i a_i b_i = 17.2578$$

Note that in practice n-1 is used instead of n

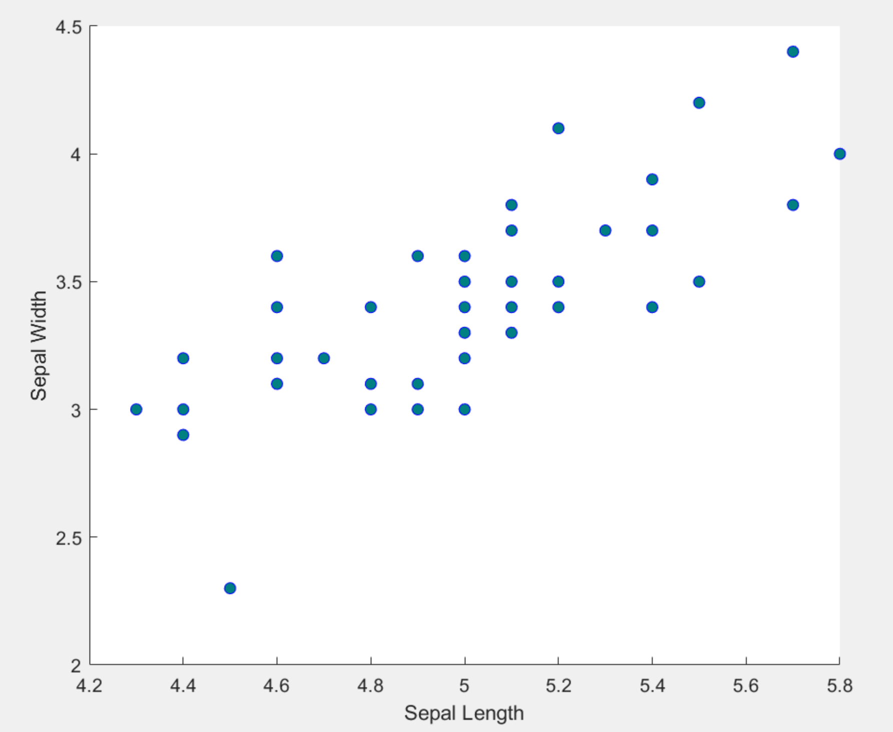
Low covariance



Sepal length and petal width

$$\sigma^2_{\mathrm{SL,PW}} \quad = \quad \frac{1}{n}\sum_i a_i b_i = 1.2416$$

Vector notation for covariance



Sepal length and sepal width

$$\mathbf{a} = \mathrm{SL} = [5.1, 4.9 \ldots, 5]$$

$$\mathbf{b} = \mathrm{SW} = [3.5, 3, \ldots, 3.3]$$

$$\sigma^2_{\mathrm{SL,SW}} \quad = \quad \frac{1}{n}\mathbf{ab}^T$$

Matrix notation for covariance

m x n matrix (m=4 and n=50)

$$\mathbf{X} = \begin{bmatrix} \text{Sepal length}_1 & \cdots & \text{Sepal length}_{50} \\ \text{Sepal width}_1 & \cdots & \text{Sepal width}_{50} \\ \text{Petal length}_1 & \cdots & \text{Petal length}_{50} \\ \text{Petal width}_1 & \ldots & \text{Petal width}_{50} \end{bmatrix}$$

$$\mathbf{C_X} \equiv \frac{1}{n}\mathbf{X}\mathbf{X}^T$$

m x m square matrix (m=4)

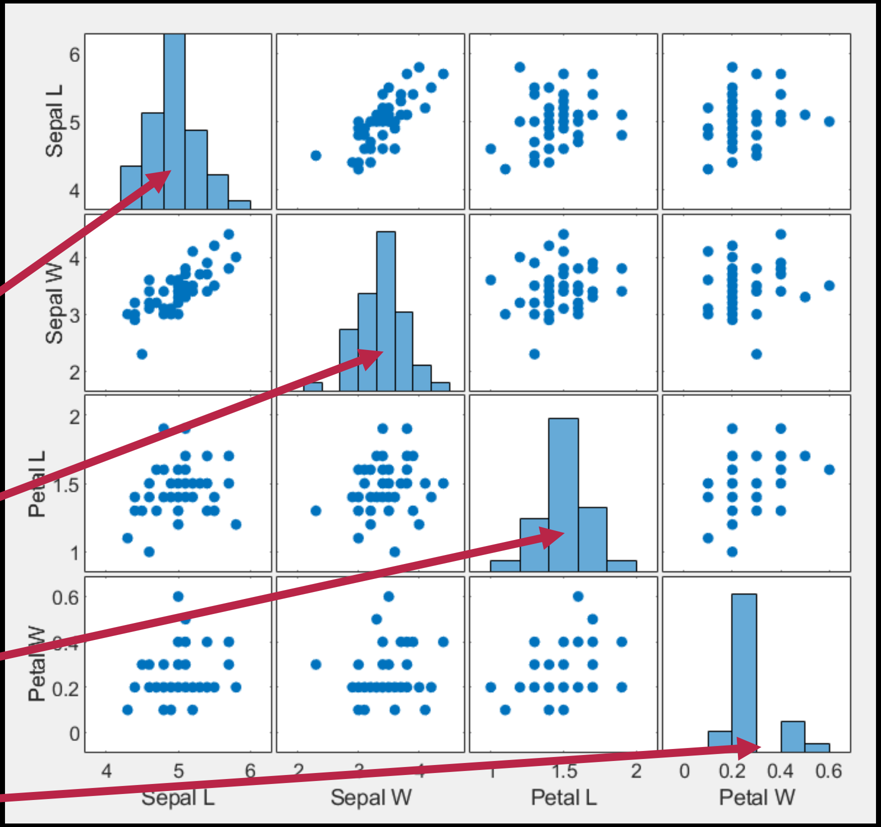
Note that in practice n-1 is used instead of n

Covariance matrix autopsy

$$\mathbf{C_X} = \frac{1}{n}\mathbf{X}\mathbf{X}^T$$

The diagonal elements are the variances

$$\begin{aligned}\sigma_{SL}^2 &= 0.1242\\ \sigma_{SW}^2 &= 0.1437\\ \sigma_{PL}^2 &= 0.0302\\ \sigma_{PW}^2 &= 0.0111\end{aligned}$$



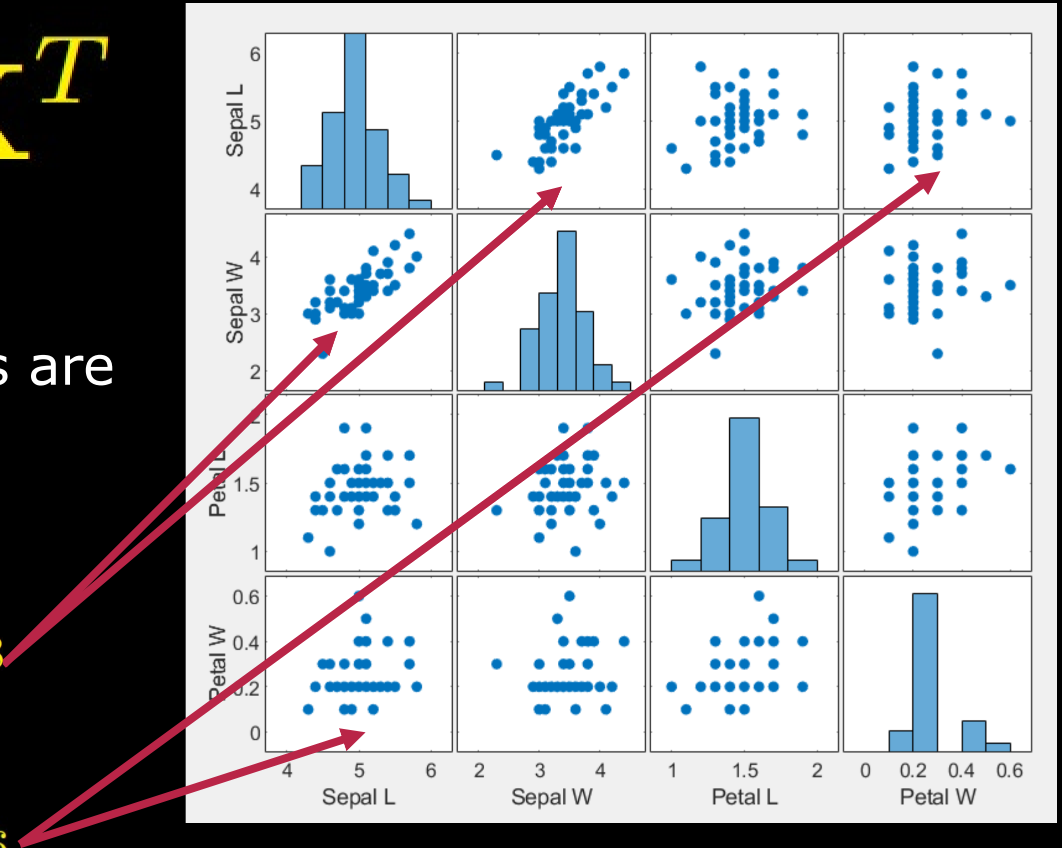
Covariance matrix autopsy II

$$\mathbf{C_X} \equiv \frac{1}{n}\mathbf{X}\mathbf{X}^T$$

The off-diagonal elements are the covariance

$$\sigma^2_{\mathrm{SL,SW}} = \frac{1}{n}\sum_i a_i b_i = 17.2578$$

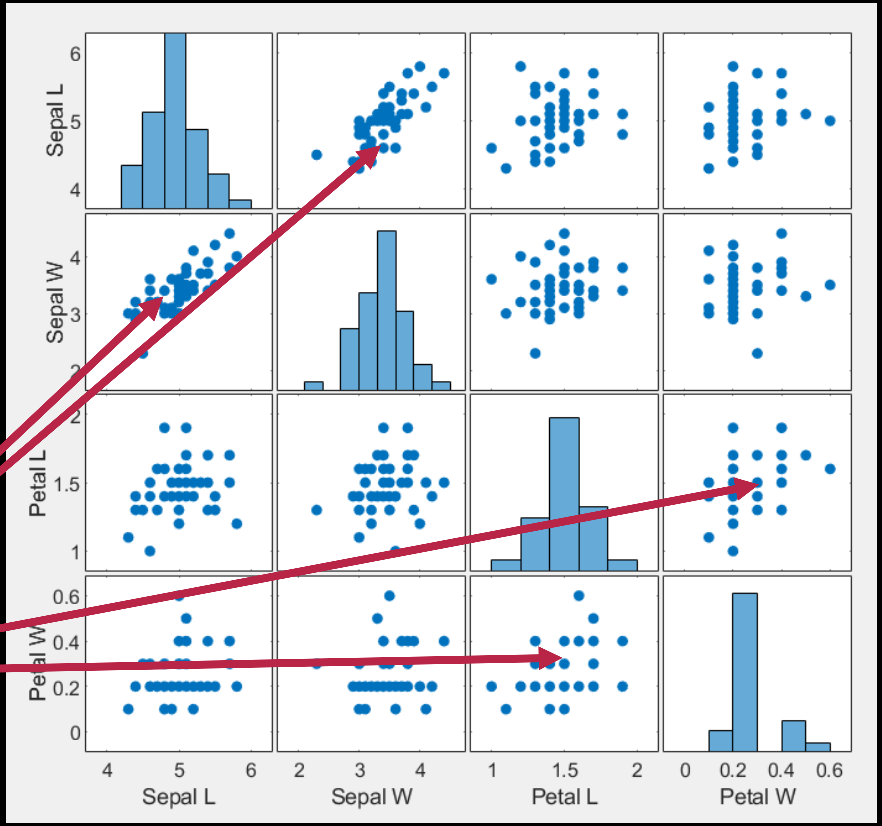
$$\sigma^2_{\mathrm{SL,PW}} = \frac{1}{n}\sum_i a_i b_i = 1.2416$$



Symmetric!

Covariance matrix autopsy III

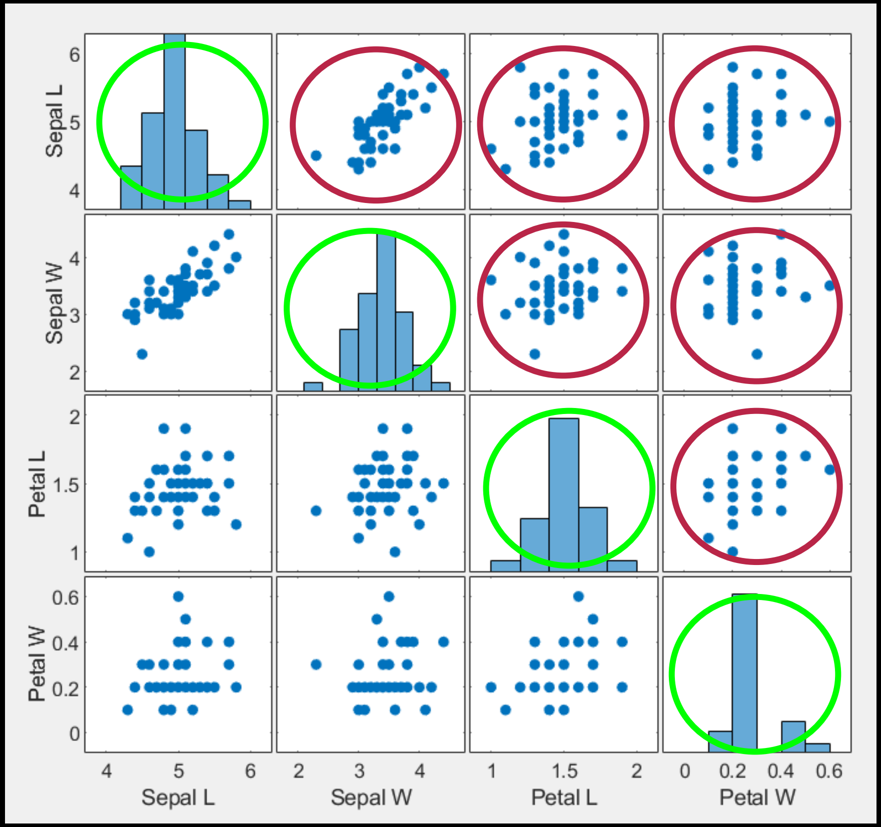
$$\mathbf{C_X} \equiv \frac{1}{n}\mathbf{X}\mathbf{X}^T$$



High redundancy

Symmetric!

Goals

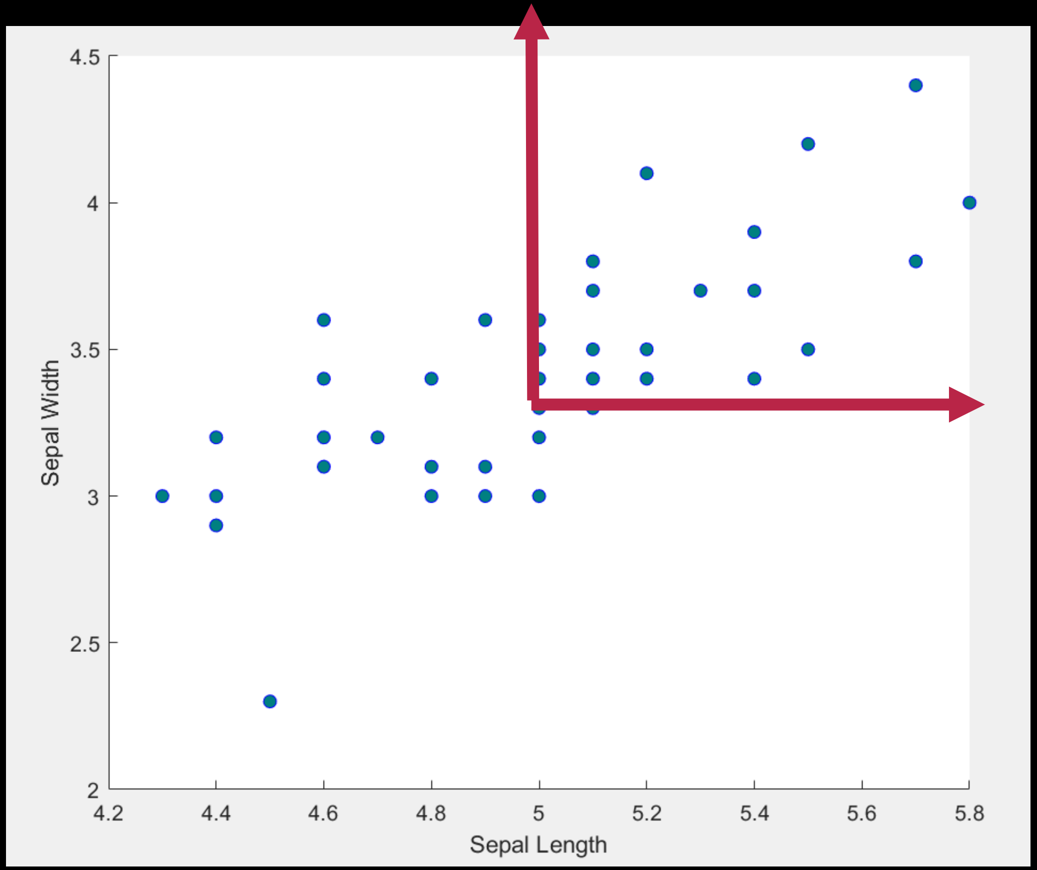


- Minimize redundancy
  - Covariance should be small
- Maximize signal
  - Variance should be large

Signal to noise ratio:

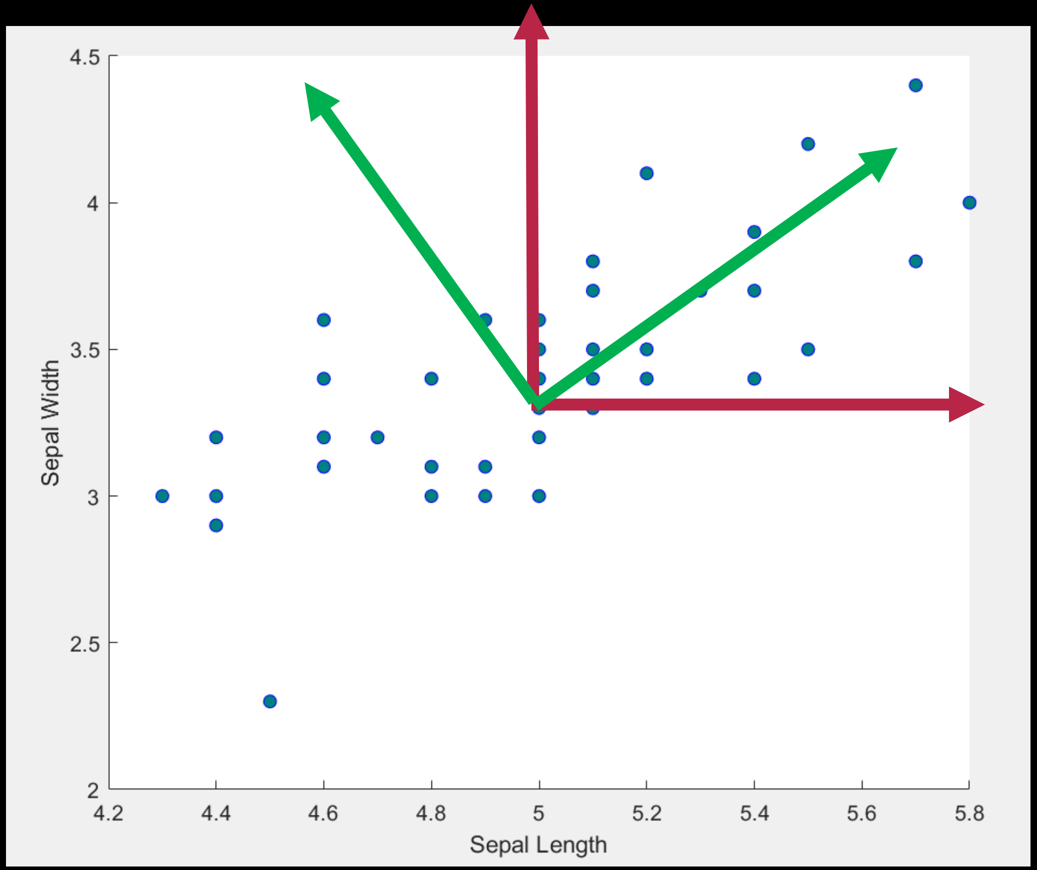
$$\mathrm{SNR} = \frac{\sigma^2_{\mathrm{signal}}}{\sigma^2_{\mathrm{noise}}}$$

Changing basis

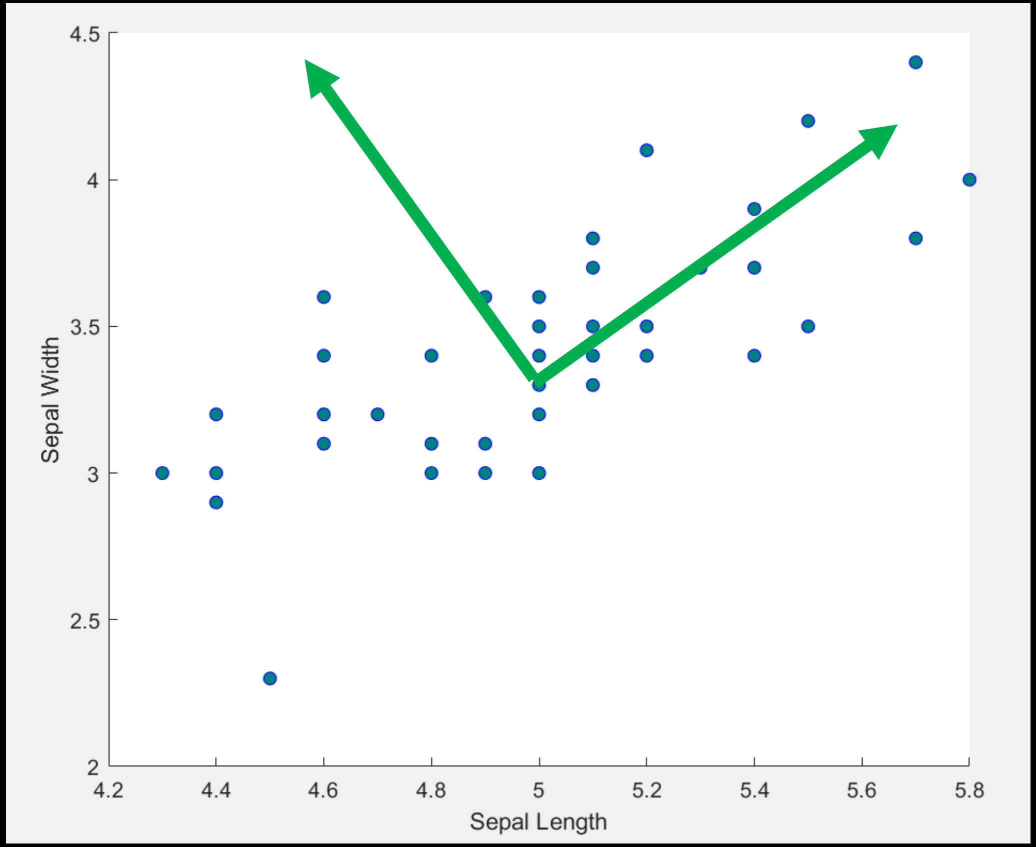


- We start by subtracting the mean
  - Centering data
- Red lines are the default basis

Changing basis

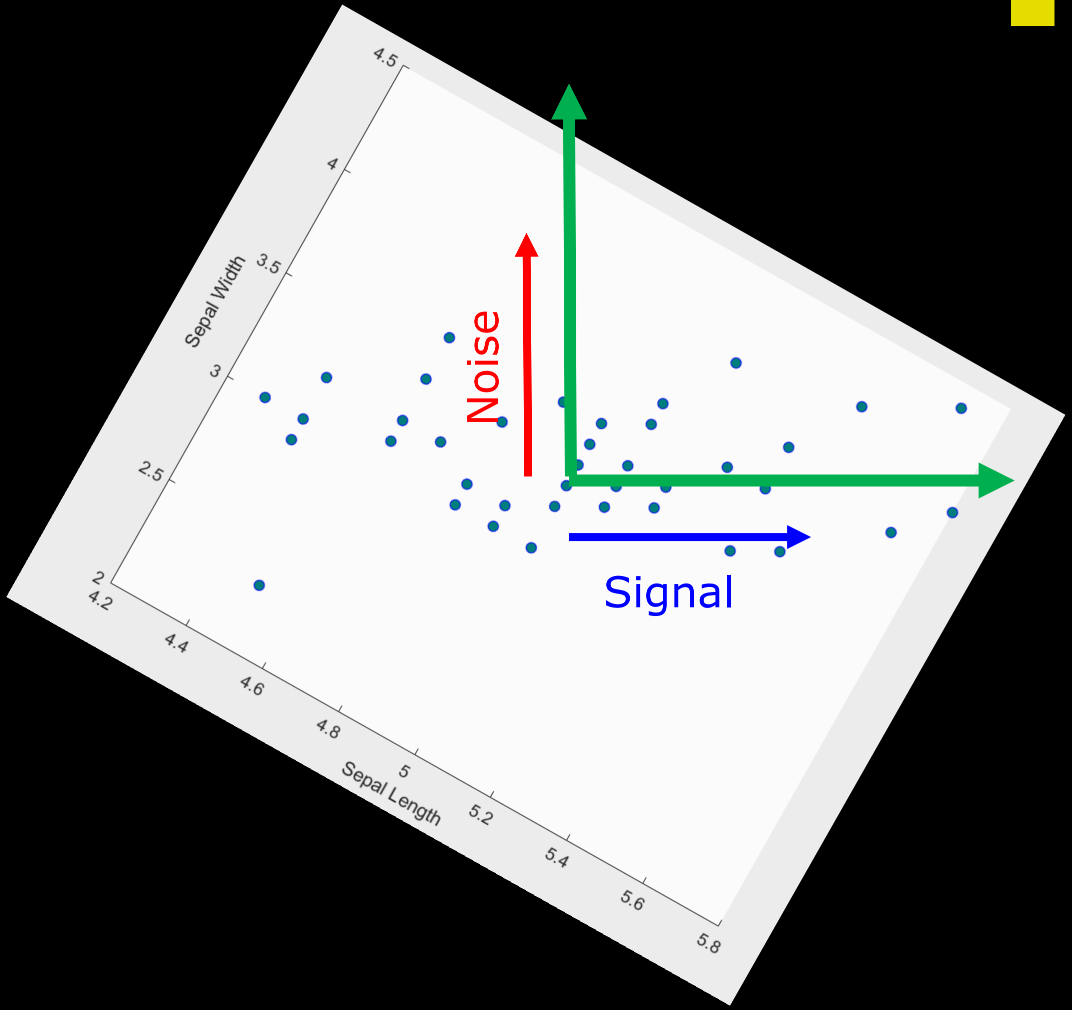


Changing basis



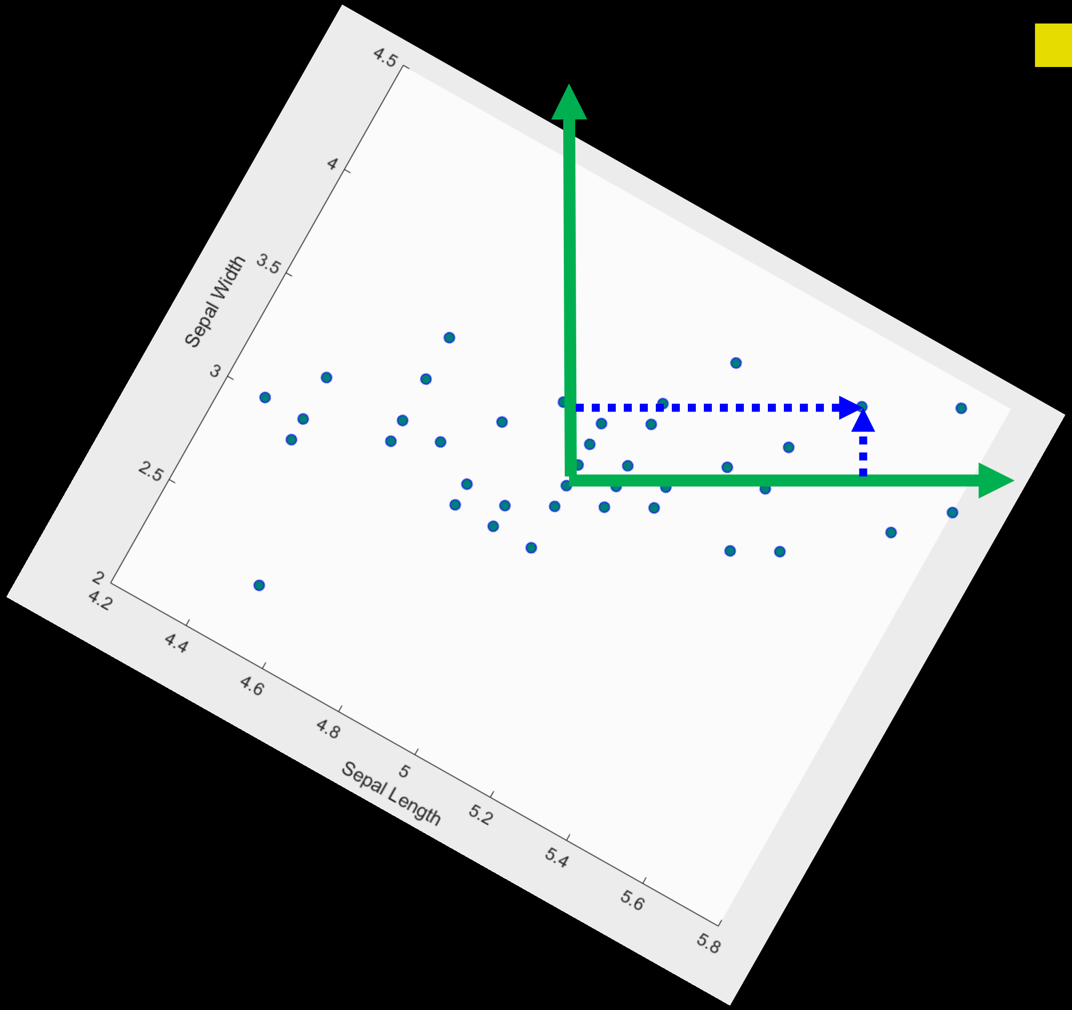
- A new basis that follows the *covariance* in the data

Changing basis



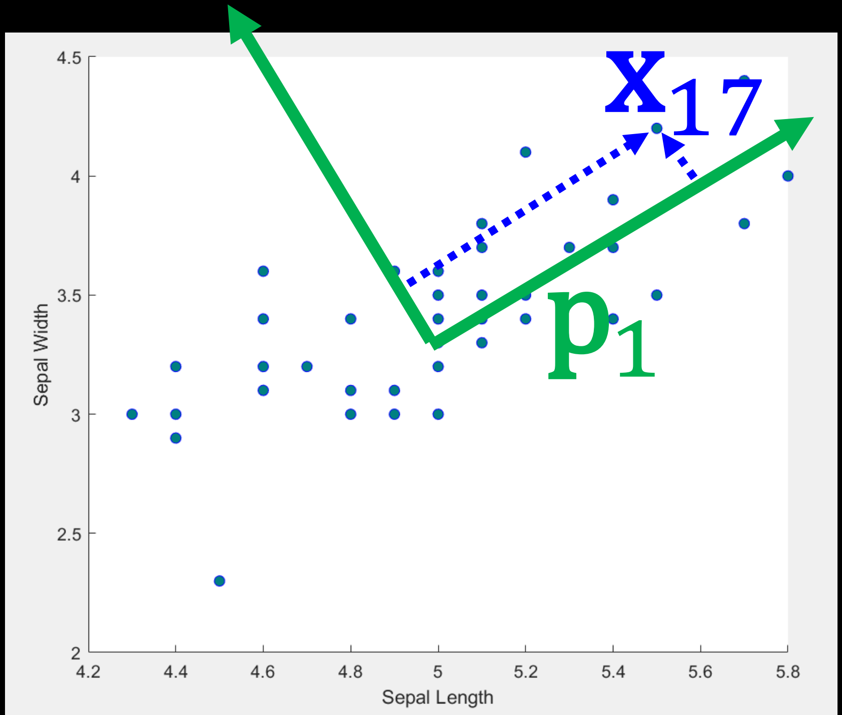
- Lets try to rotate the data – for visualisation

Changing basis



- Finding the measurement values in the new basis

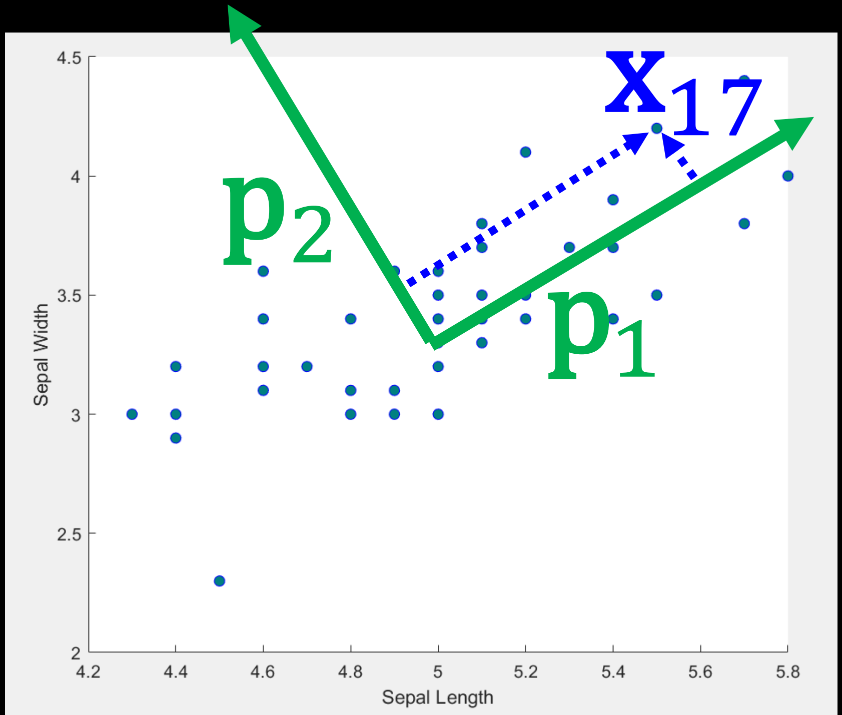
Changing basis



- The dot product projects a point down to a new axis

$$\mathbf{x}_{17,\mathrm{new}} = x_{17} \cdot p_1$$

Changing basis



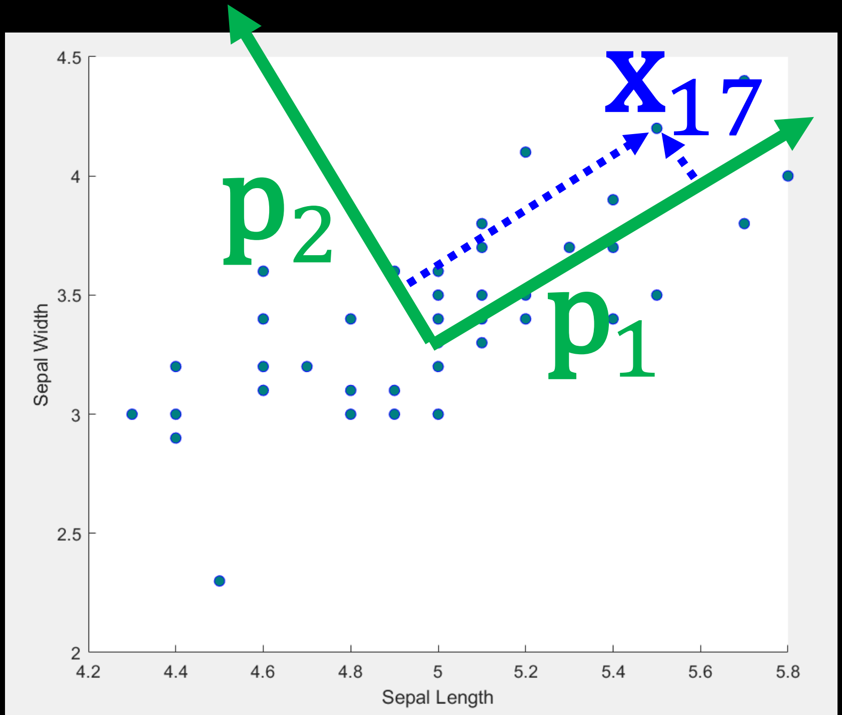
- The dot product projects a point down to a new axis

$$\mathbf{PX} = \mathbf{Y}$$

- $\mathbf{p}_1$ and $\mathbf{p}_2$ are the rows of $\mathbf{P}$

$$\mathbf{X} = \begin{bmatrix} \text{Sepal length}_1 & \cdots & \text{Sepal length}_{50} \\ \text{Sepal width}_1 & \cdots & \text{Sepal width}_{50} \\ \text{Petal length}_1 & \cdots & \text{Petal length}_{50} \\ \text{Petal width}_1 & \ldots & \text{Petal width}_{50} \end{bmatrix}$$

Changing basis

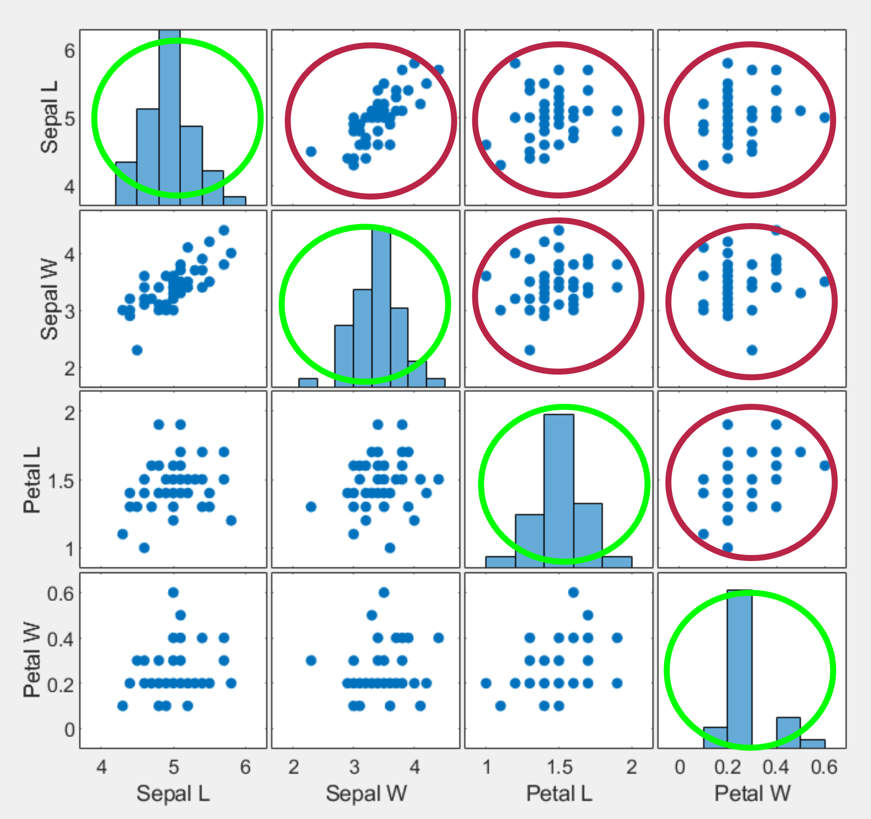


- The dot product projects a point down to a new axis

$$\mathbf{PX} = \mathbf{Y}$$

- Here **Y** contains the new coordinates/measurements per sample

Goals



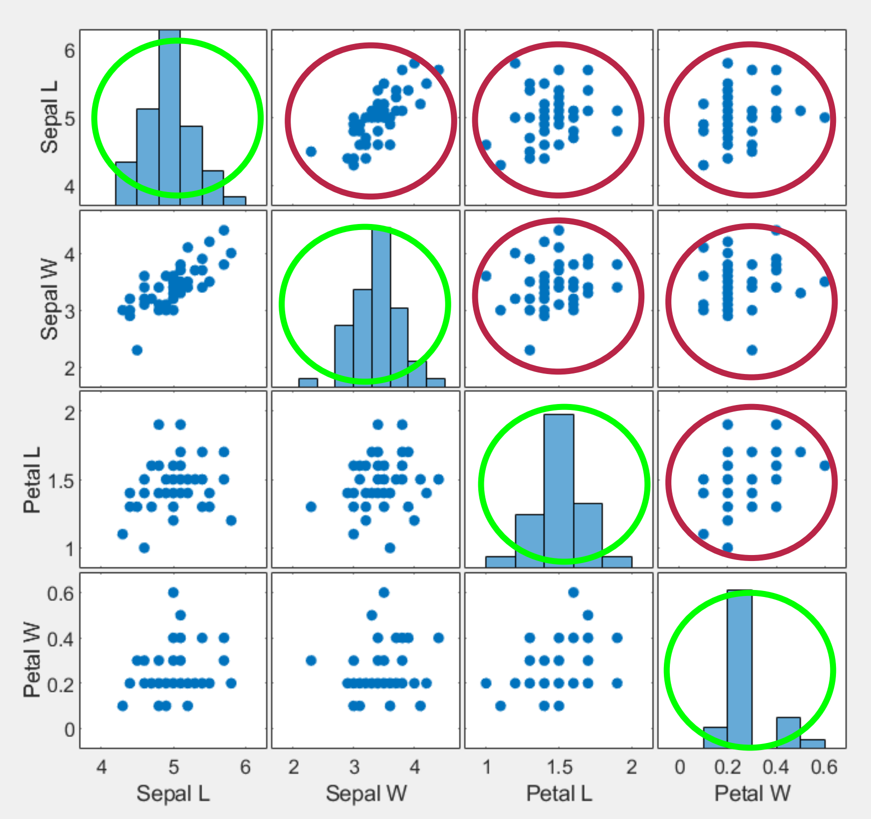
- Minimize redundancy
  - Covariance should be small
- Maximize signal
  - Variance should be large
- Transform our data
  - Rotating and scaling the basis

$$\mathbf{Y} = \mathbf{P}\mathbf{X}$$

- So it will have

$$\mathbf{C_Y} \equiv \frac{1}{n}\mathbf{Y}\mathbf{Y}^T$$

Goals



- The covariance matrix

$$\mathbf{C_Y} \equiv \frac{1}{n}\mathbf{YY}^T$$

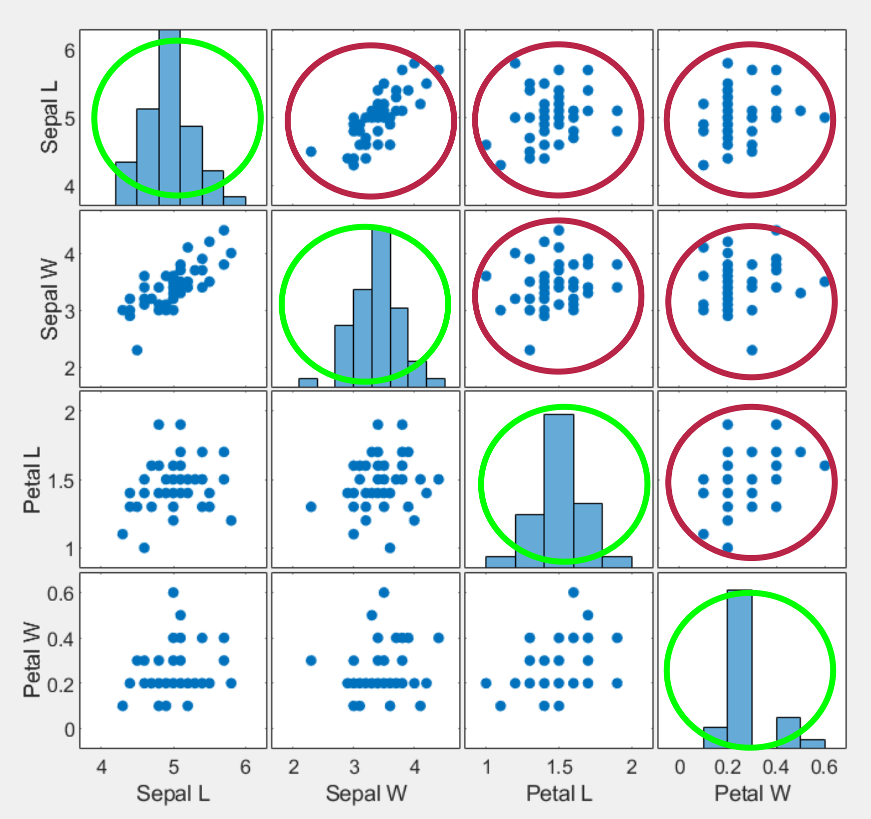
- Should be *as diagonal as possible*

- We do this by

$$\mathbf{Y} = \mathbf{PX}$$

  - Where **P** are the principal components

Computing the principal components

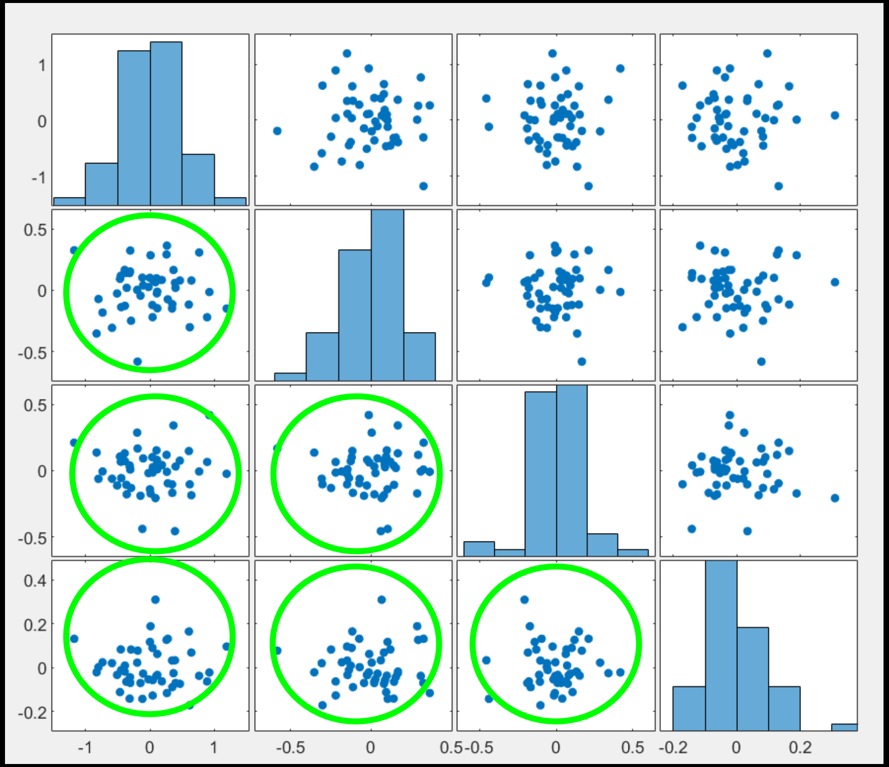


- The Principal Components of $\mathbf{X}$ are the eigenvectors of

$$\mathbf{C_X} = \frac{1}{n}\mathbf{X}\mathbf{X}^T$$

- The i'th diagonal value of $\mathbf{C_Y}$ is the variance along principal component number i

New covariance matrix for Iris data



Covariance: 0

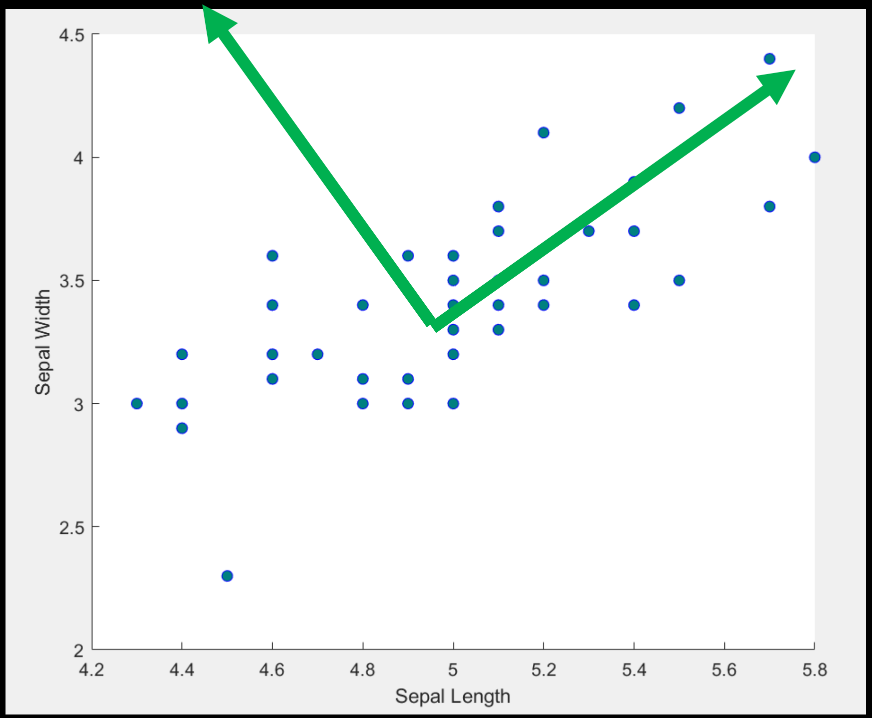
- The principal component are found and

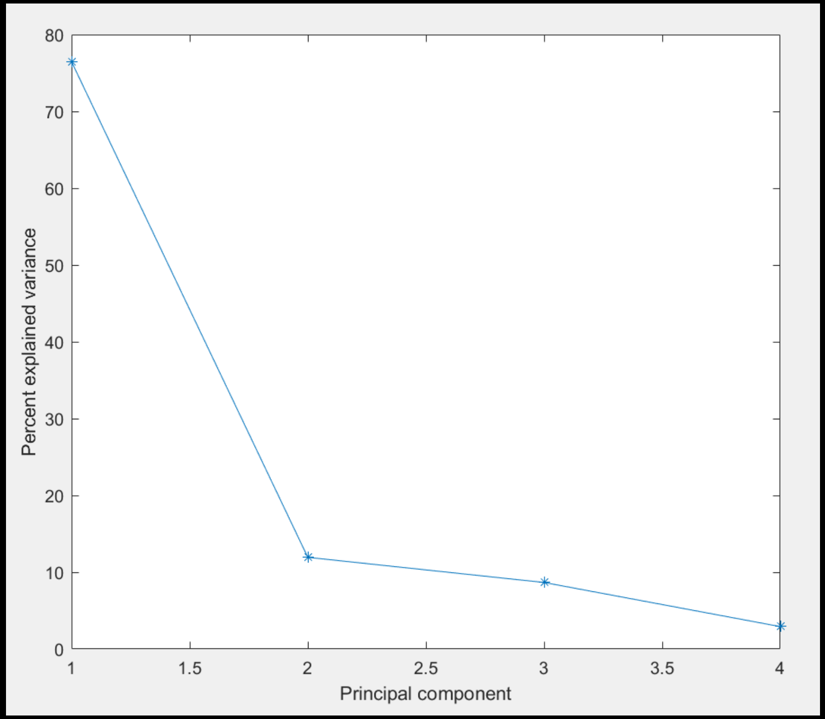
$$\mathbf{Y} = \mathbf{P}\mathbf{X}$$

- With the covariance matrix

$$\mathbf{C_Y} \equiv \frac{1}{n}\mathbf{Y}\mathbf{Y}^T$$

Explained variance

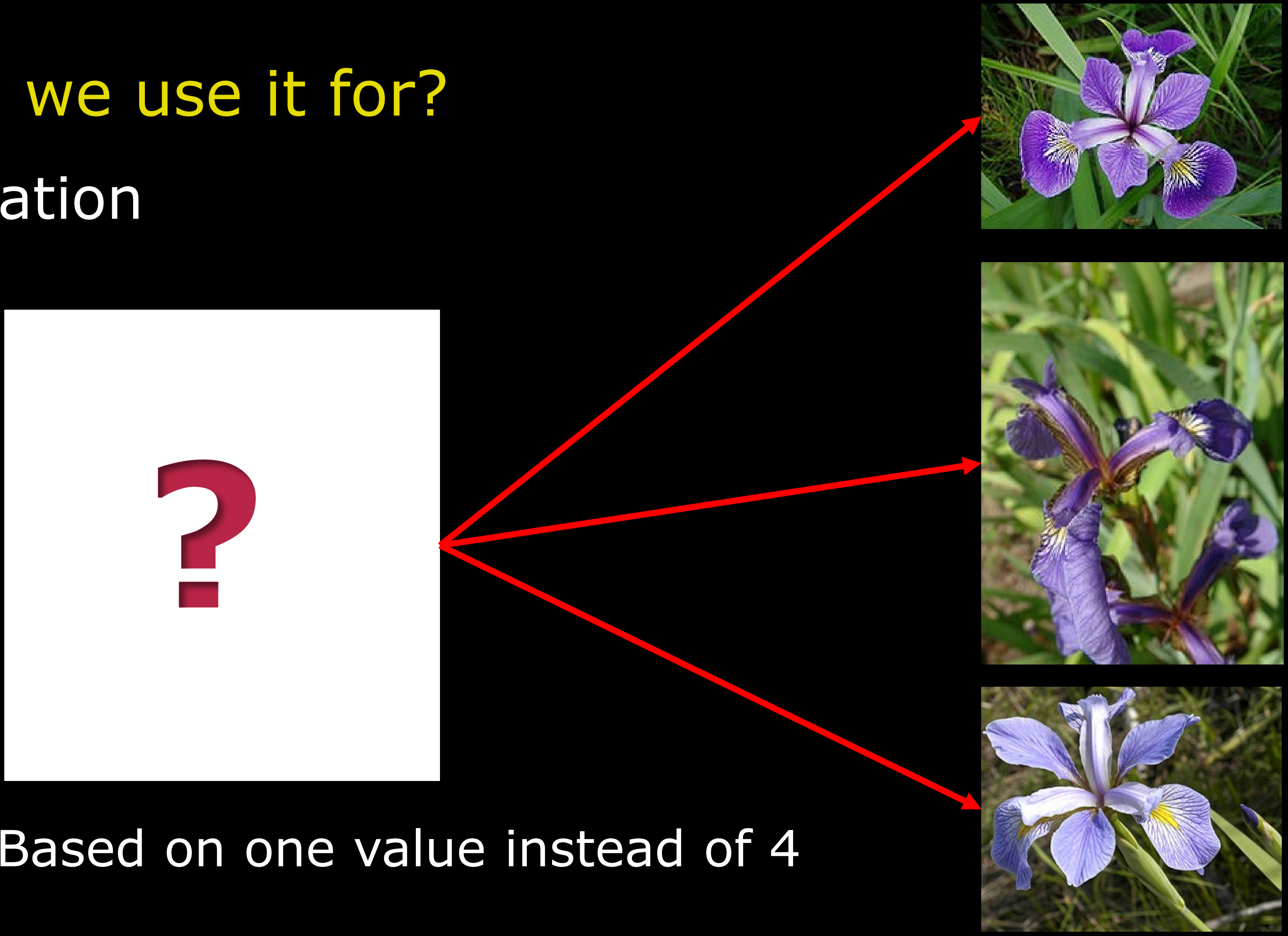




One component explains 75% of the total variation – so for each flower we can have one number that explains 75% percent of the 4 measurements!

What can we use it for?

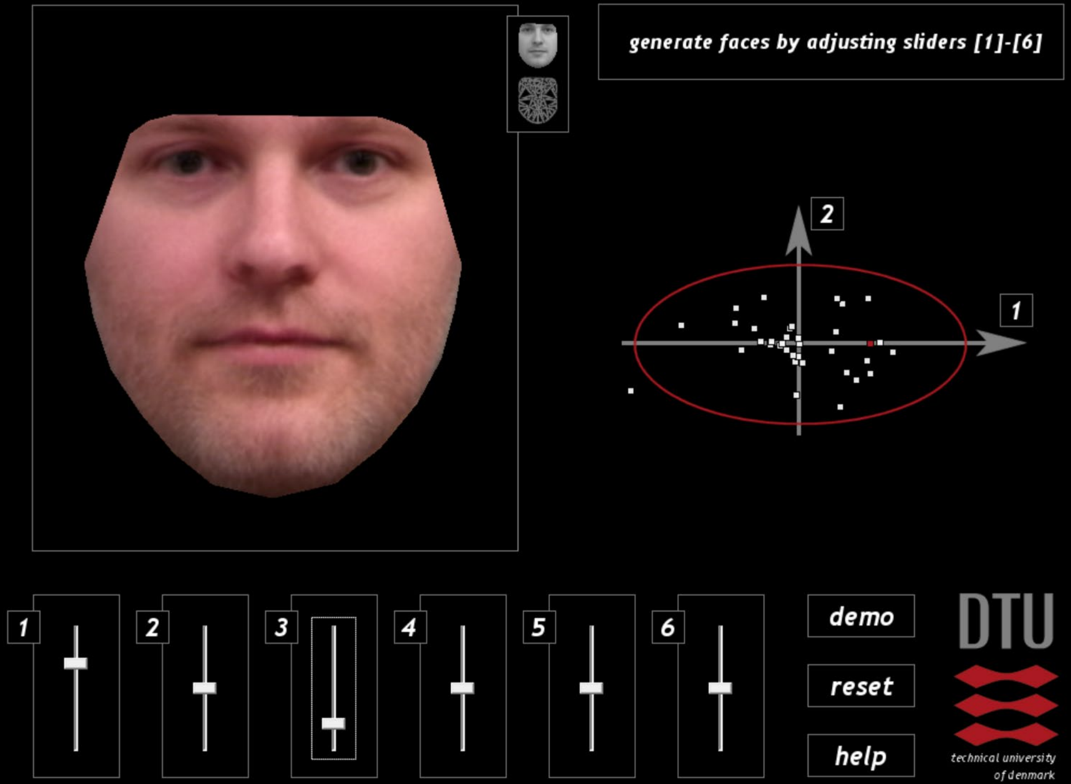
- Classification

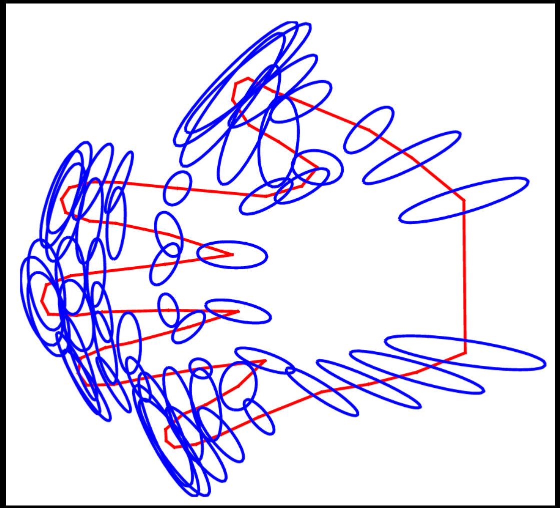


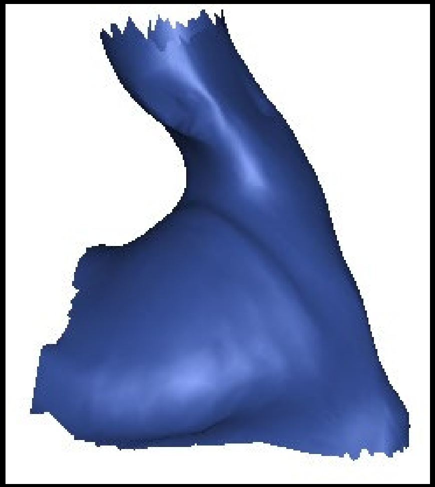
Based on one value instead of 4

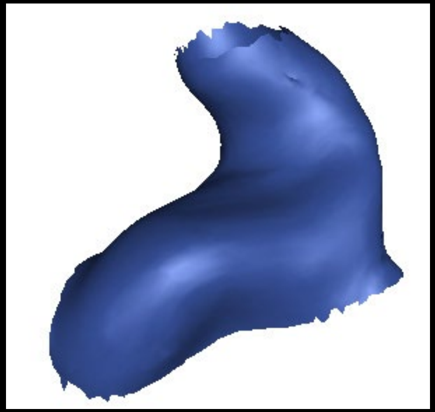
What can we use it for?

- Many more examples in the course


generate faces by adjusting sliders [1]-[6]
2
1
1
2
3
4
5
6
demo
reset
help
DTU
technical university of denmark







Final note – practical estimation of covariance matrix

$$\mathbf{C_x} \equiv \frac{1}{n}\mathbf{X}\mathbf{X}^T$$

In practice n-1 is used instead of n for exercises and in the exam.

$$\mathbf{C_x} \equiv \frac{1}{n-1}\mathbf{X}\mathbf{X}^T$$