# Dimensionality reduction

BIPN 162

By the end of today’s lecture you will be able to:

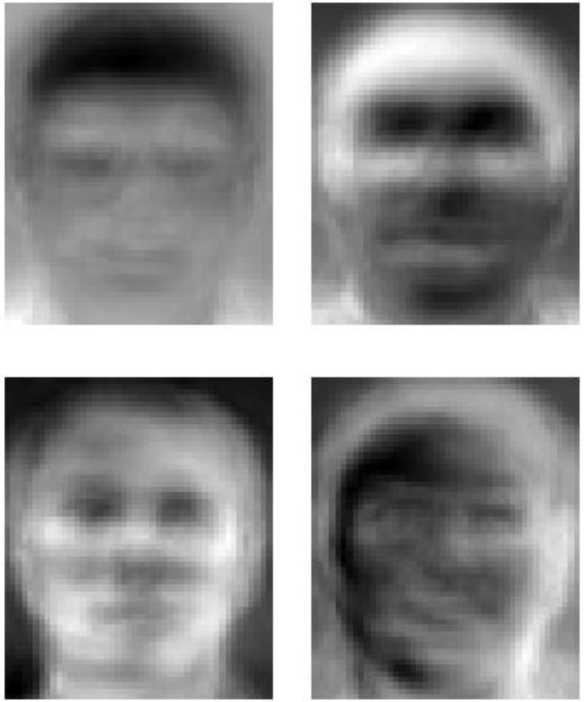
- Motivate the use of dimensionality reduction in neuroscience
- Identify several popular methods and their limitations
- Define what eigenvalues and eigenvectors are and determine them using Python
- Describe the steps of PCA and implement it in Python
    - Define and generate a **covariance matrix**

**Dimensionality reduction** assumes that the messy & complex world comes from the interplay of a few “latent” factors

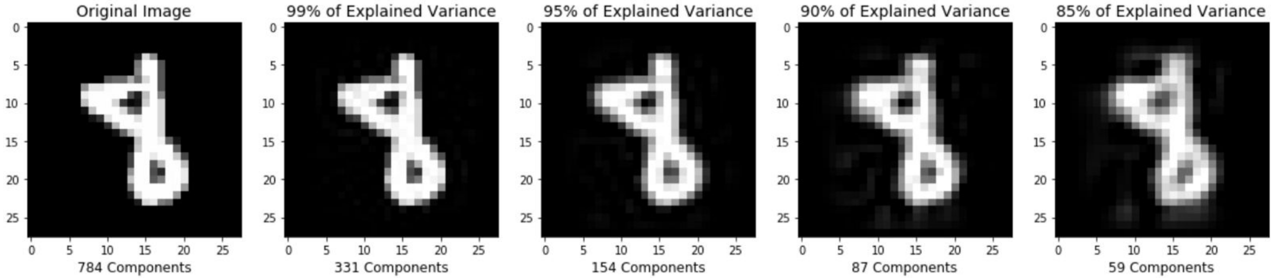
> “Describing a complex signal, such as a visual scene or a pattern of neural activity in terms of just a few summarizing features”
> - Pang et al. 2016

# Implementations of dimensionality reduction *beyond* neuroscience

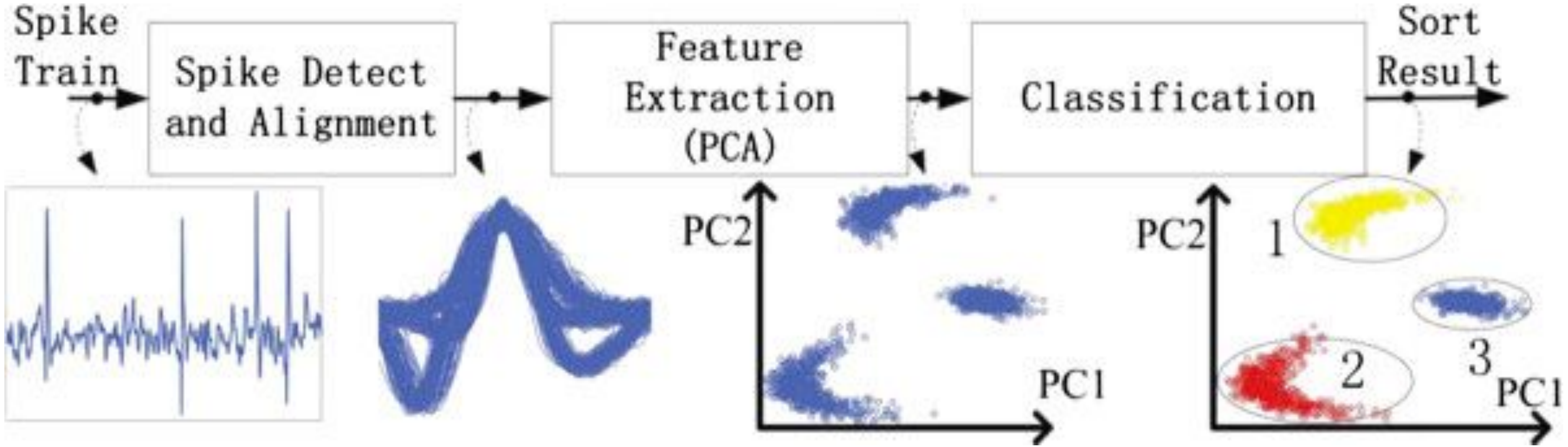
- Personality (five-factor OCEAN model; Costa & McCrae, 1992)
- Facial recognition
- Handwriting recognition



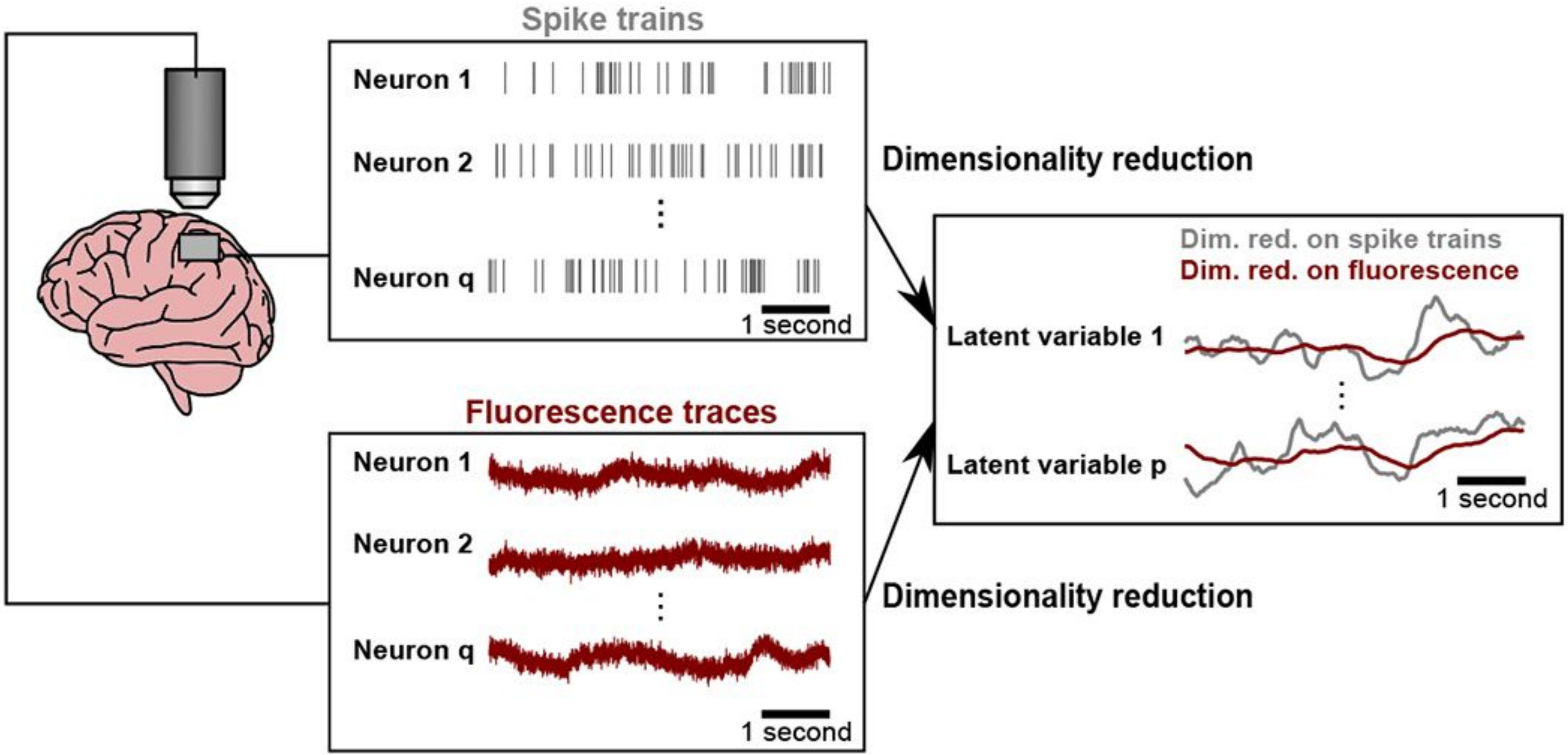
[Eigenfaces](); [try it yourself]()



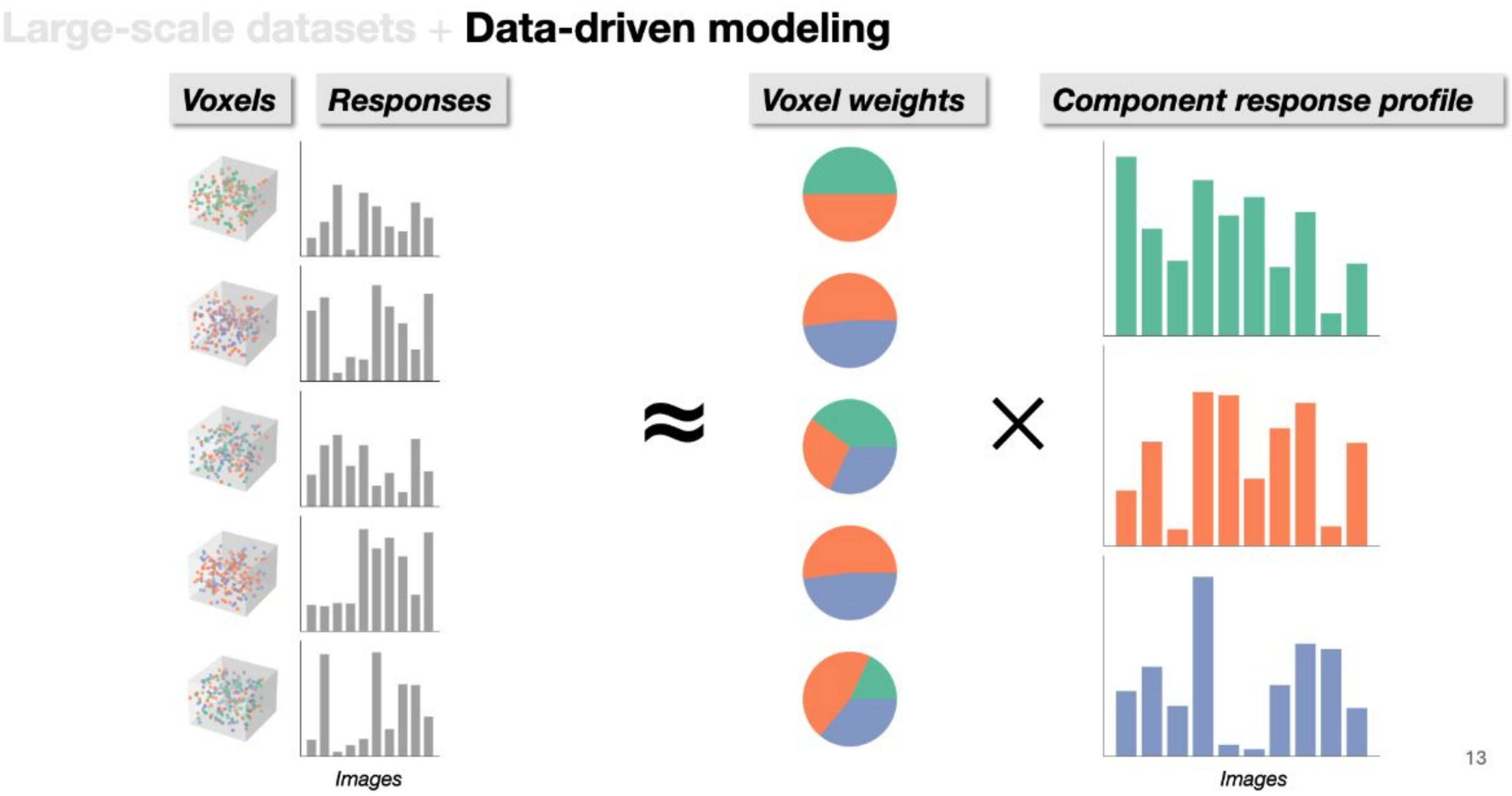

[Image source]()




We've seen dimensionality reduction before...




The use of dimensionality reduction on spike trains or fluorescence
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10358781/



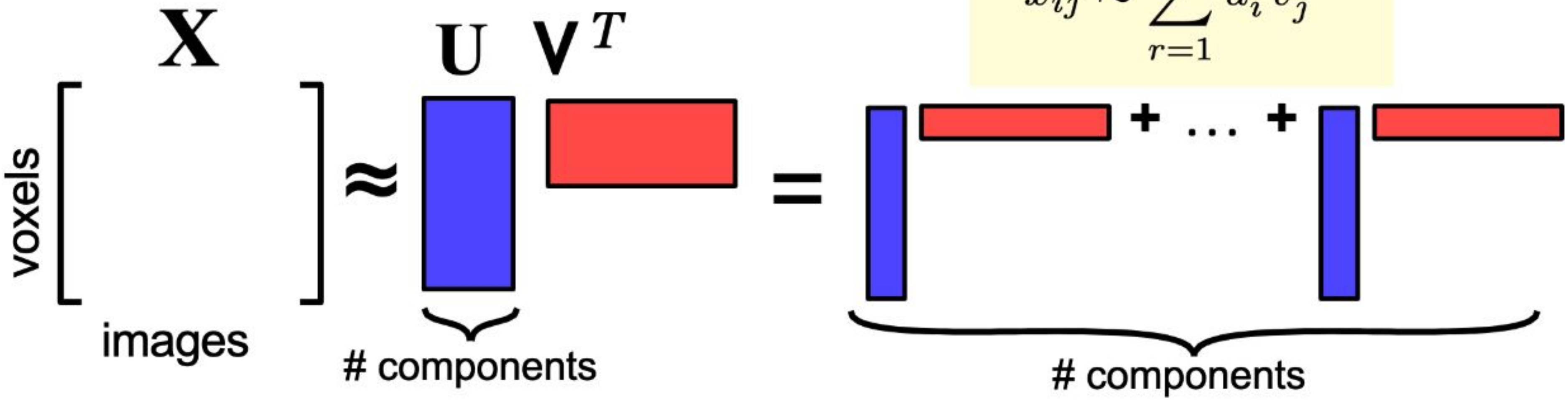

We've seen dimensionality reduction before... (slide: Dr. Khosla)

# Matrix factorization

Often **non-negative matrix factorization (NMF)**, which enforces positive values.

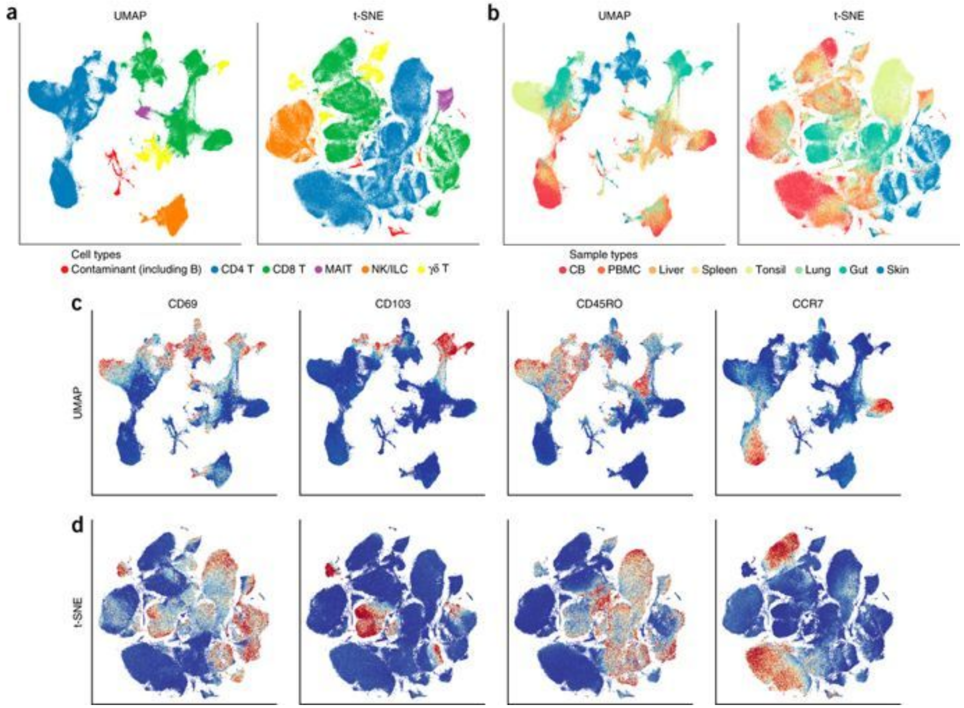
A simple and general framework for extracting correlations and low-dimensional structure from matrix-coded datasets

$$x_{ij} \approx \sum_{r=1}^{R} u_i^r v_j^r$$




Slide: Dr. Meekanshki Khosla

# Implementations of dimensionality reduction




Single-cell techniques (paper)

t-SNE: https://lvdmaaten.github.io/tsne/

# Why do we need dimensionality reduction in neuroscience?

- Neural systems and behaviors are complex
- We need many numbers to characterize systems like networks of neurons or even a single neuron — in other words, our data is **multivariate**

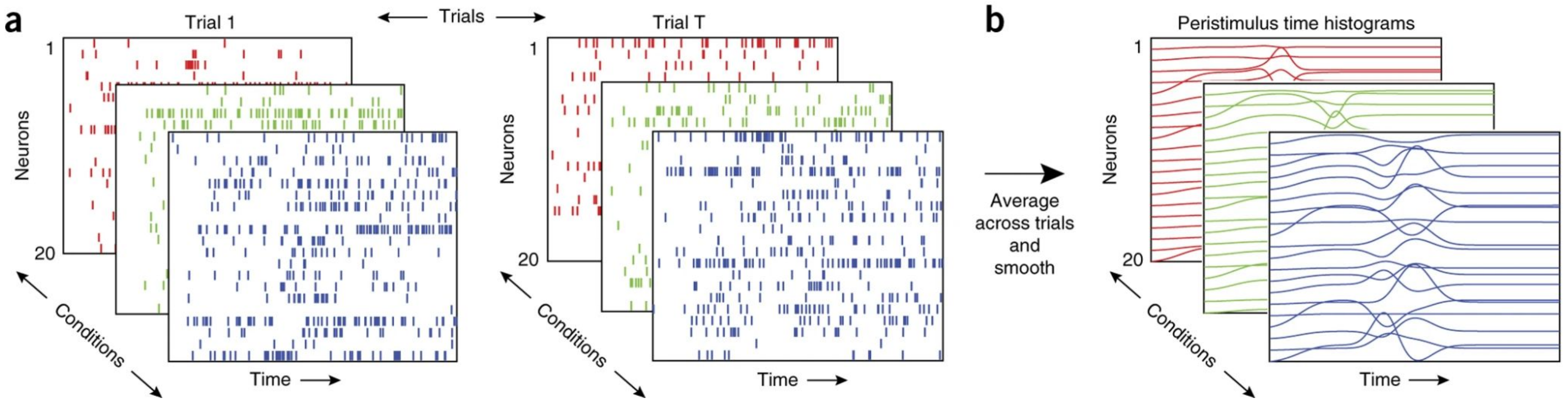
➔ Dimensionality reduction can help reduce the complexity of systems/data

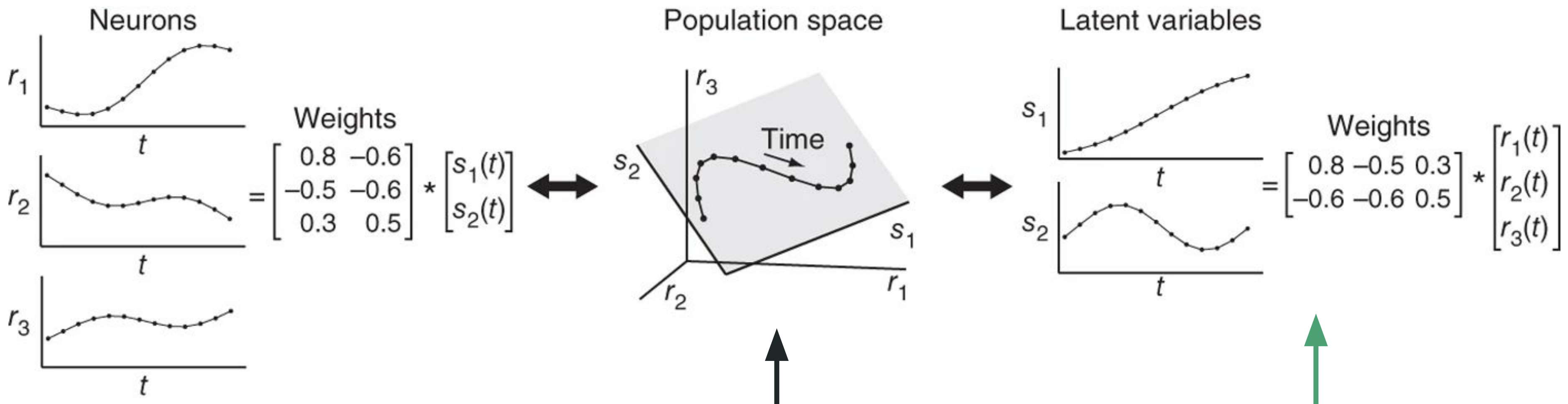
“[W]e ought to seek representations, both of the stimulus and of brain activity, that are **concise**, **complete**, and **informative** about the workings of the nervous system, and yet which are not biased by an experimenter’s arbitrary choice. **Considering this task from the perspective of dimensionality reduction provides an entry point into principled mathematical techniques that let us discover these representations directly from experimental data**, a key step to developing rich yet comprehensible models for brain function.”

Pang et al. (2016)

# Three primary uses of dimensionality reduction for large-scale recordings in neuroscience:

1. Retaining population structure for single-trial analyses
2. Extracting hidden “latent” features from data to test hypotheses about neural population structure
3. Exploratory data analysis



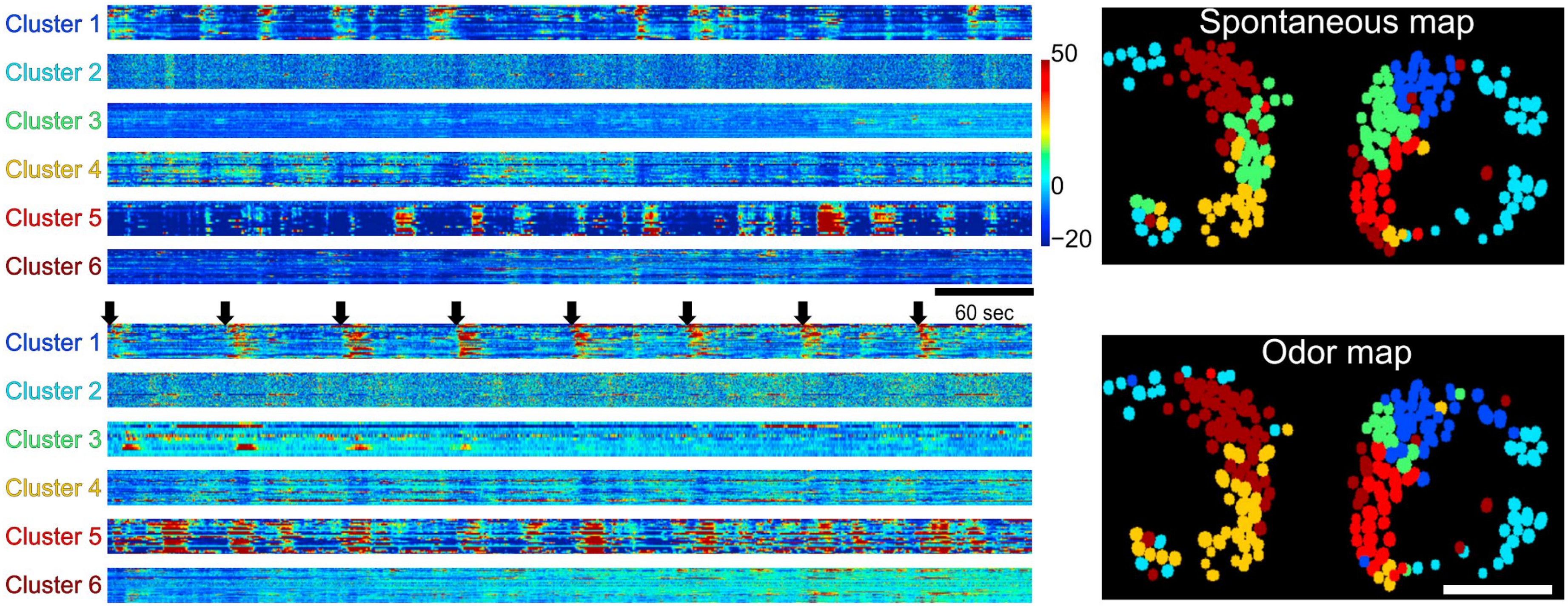



Population activity r1, r2 and r3 can be reconstructed by taking a weighted combination of the latent variables, where the weights are specified by the matrix shown.

Population activity (black points) lies in a plane (shaded gray). Each point represents the population activity at a particular time and can be equivalently referred to using its high-dimensional coordinates [r1, r2, r3] or low-dimensional coordinates [s1, s2]. The points trace out a trajectory over time (black curve).

The latent variables s1 and s2 can be obtained by taking a weighted combination of the population activity, where the weights are specified by the matrix shown.

[Dimensionality reduction for large-scale neural recordings | Nature Neuroscience]




PCA and K-Means identified six different clusters of neurons, which are also spatially-organized (right). These clusters are similar during both spontaneous and odor-evoked activity.

Current Biology 24, 434–439, February 17, 2014 ©2014 Elsevier Ltd All rights reserved http://dx.doi.org/10.1016/j.cub.2014.01.015

Report

# Spontaneous Activity Governs Olfactory Representations in Spatially Organized Habenular Microcircuits

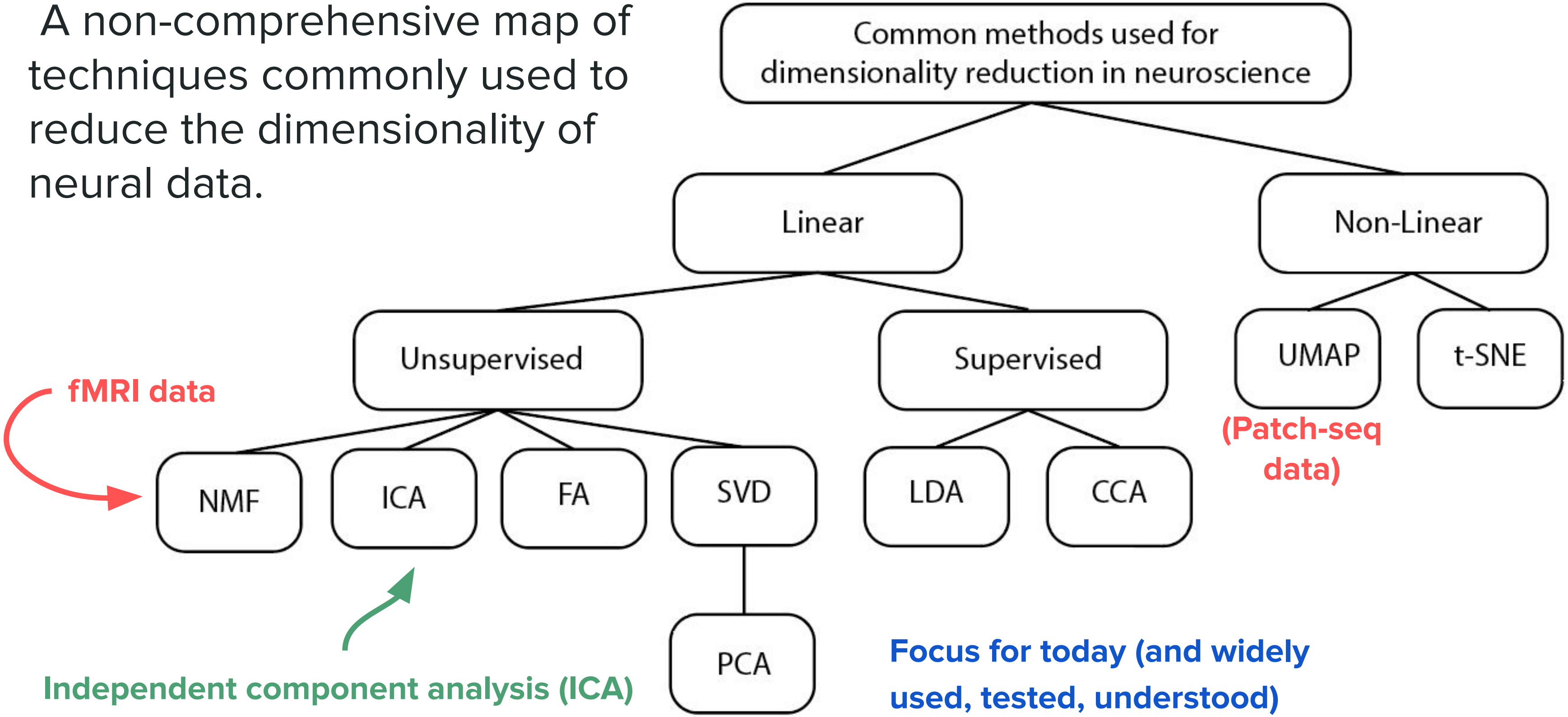
Suresh Kumar Jetti,[1,2,4] Nuria Vendre
and Emre Yaksi[1,2,3,*]
[1]NERF, Kapeldreef 75, 3001 Leuven, B
[2]KU Leuven, Kapeldreef 75, 3001 Leuv
[3]VIB, Kapeldreef 75, 3001 Leuven, Bel

Moreover, we show that the spontaneous dHb activity is not random but structured into functionally and spatially organized clusters of neurons, which reflects the favored states of the dHb network. These dHb clusters are also preserved during odor stimulation and govern olfactory responses. Finally, we show that functional dHb clusters overlap with genetically defined dHb neurons [4], which regulate experience-dependent fear. Thus, we propose that the dHb is composed of functionally, spatially, and genetically distinct microcircuits that regulate different behavioral programs.

# By the end of today's lecture you will be able to:

- Motivate the use of dimensionality reduction in neuroscience
- Identify several popular methods and their limitations
- Define what eigenvalues and eigenvectors are and determine them using Python
- Describe the steps of PCA and implement it in Python
    - Define and generate a **covariance matrix**

A non-comprehensive map of techniques commonly used to reduce the dimensionality of neural data.




Independent component analysis (ICA) doesn’t enforce an orthogonal principle component, and is therefore useful for sources that may not be orthogonal — for example, voices at a cocktail party

Focus for today (and widely used, tested, understood)

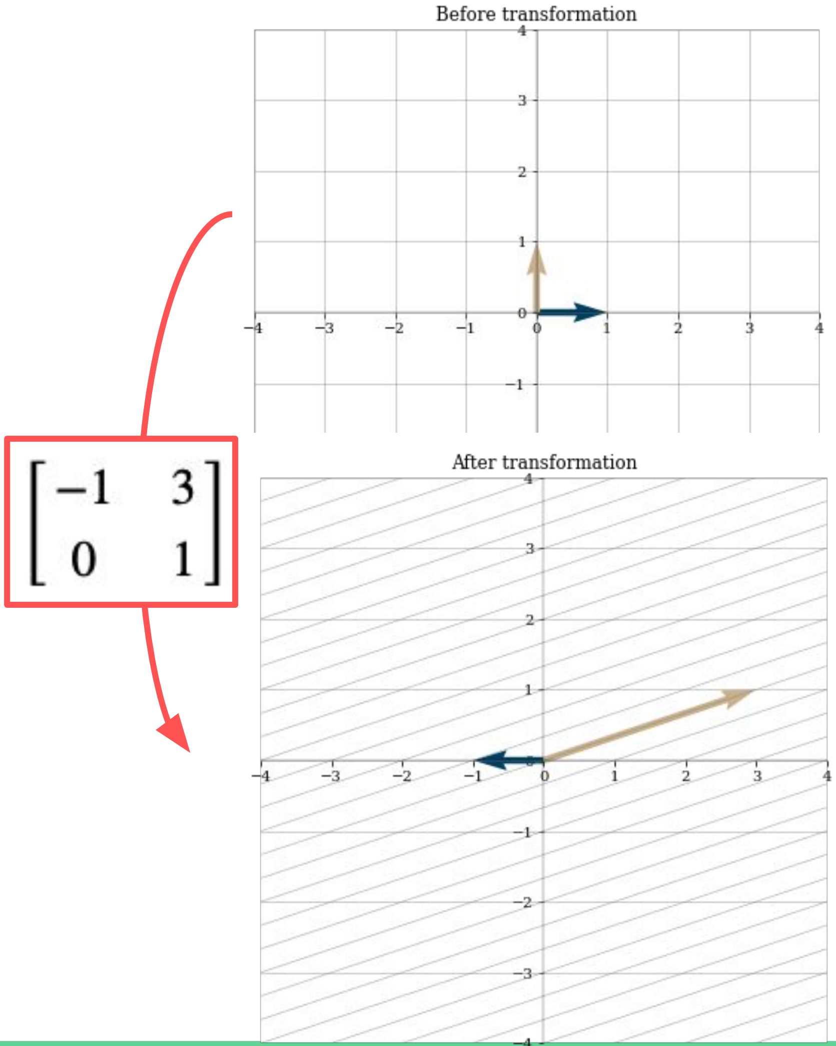
**Principal components analysis (PCA)** allows us to find axes in high dimensional space that express the information contained in the data with less variables than the original dataset.

# By the end of today's lecture you will be able to:

- Motivate the use of dimensionality reduction in neuroscience
- Identify several popular methods and their limitations
- **Define what eigenvalues and eigenvectors are and determine them using Python**
- Describe the steps of PCA and implement it in Python
  - Define and generate a **covariance matrix**

# Reminders

- Matrices apply linear transformations and can be used to change basis vectors
- Every vector has a span, a line that passes through the origin and its tip



# Eigenvectors

Most vectors will get knocked off their span in a linear transformation — those that don’t are **eigenvectors**

*(And so, we can also think about linear transformations in terms of eigenvectors).*

Eigenvectors don’t move off their span, but they can change length.

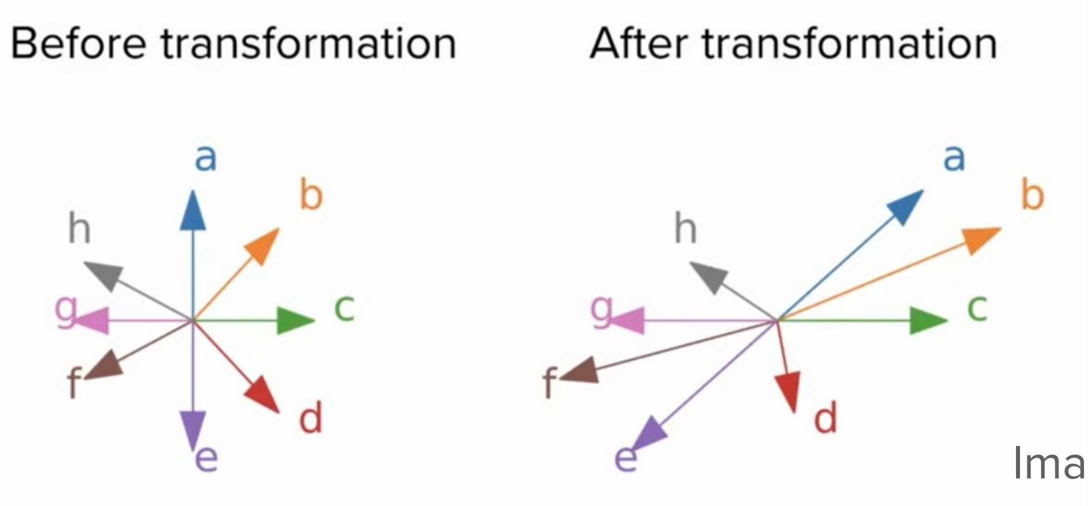



Image: NMA

# Eigenvectors

Most vectors will get knocked off their span in a linear transformation
— those that don't are **eigenvectors**

*(And so, we can also think about linear transformations in terms of eigenvectors).*

Eigenvectors don't move off their span, but they can change length.

$$W\mathbf{v} = \lambda \mathbf{v}$$

Matrix x Eigenvector = Eigenvalue x Eigenvector

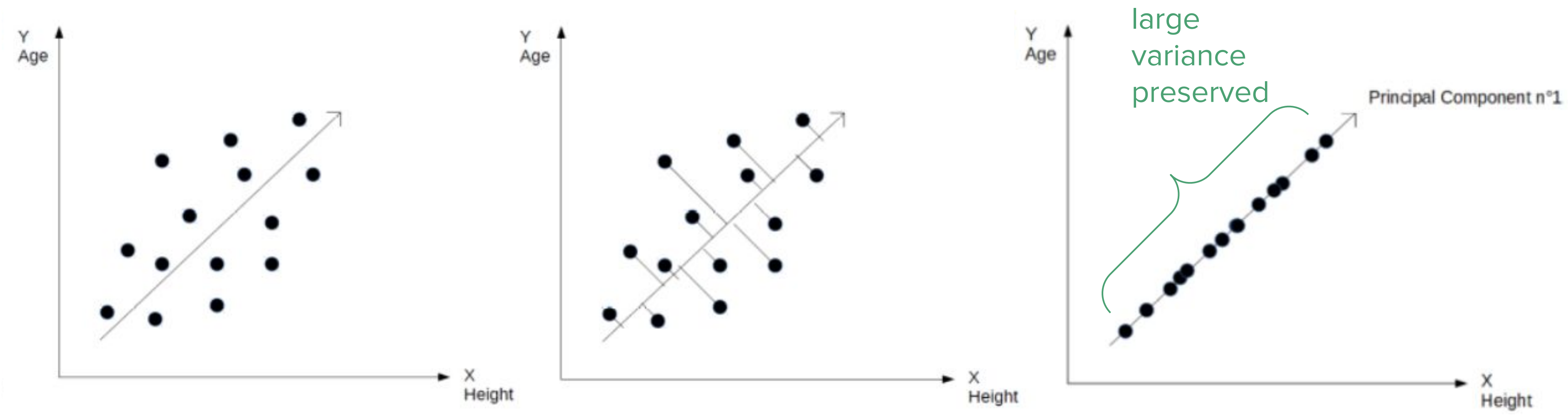
Solve for lambda (λ) and v

The **eigenvalue** is a **scalar** that tells you how much!

Image: NMA; another explanation: https://www.youtube.com/watch?v=PFDu9oVAE-g

# By the end of today's lecture you will be able to:

- Motivate the use of dimensionality reduction in neuroscience
- Identify several popular methods and their limitations
- Define what eigenvalues and eigenvectors are and determine them using Python
- **Describe the steps of PCA and implement it in Python**
    - Define and generate a **covariance matrix**



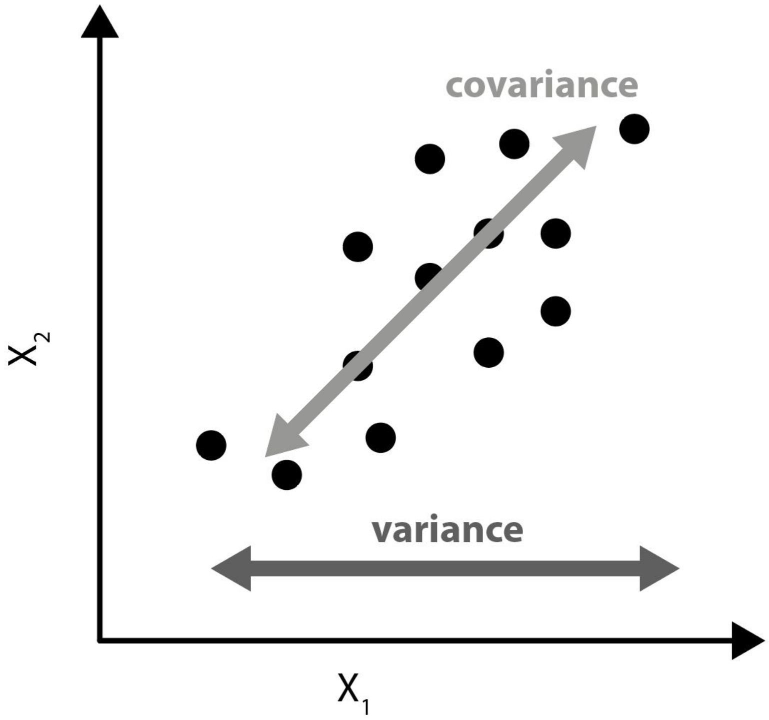

The arrow captures the direction where the **variance is maximized**.
Points that are furthest away in the first panel are *still* the furthest away.

We can take two variables (age & height),
and describe them by one vector.

# Performing PCA involves five steps

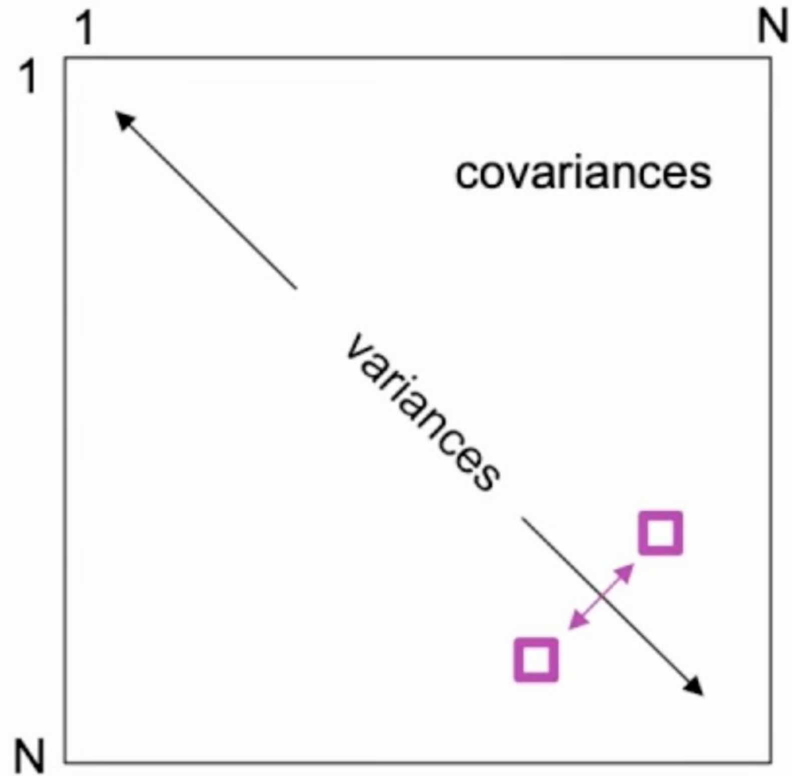
(0) Subtract the mean (and optionally, normalize)

1. Calculate the **covariance matrix** between the variables.



# Covariance matrix ($\boldsymbol{\Sigma}$)

$\boldsymbol{\Sigma} =$



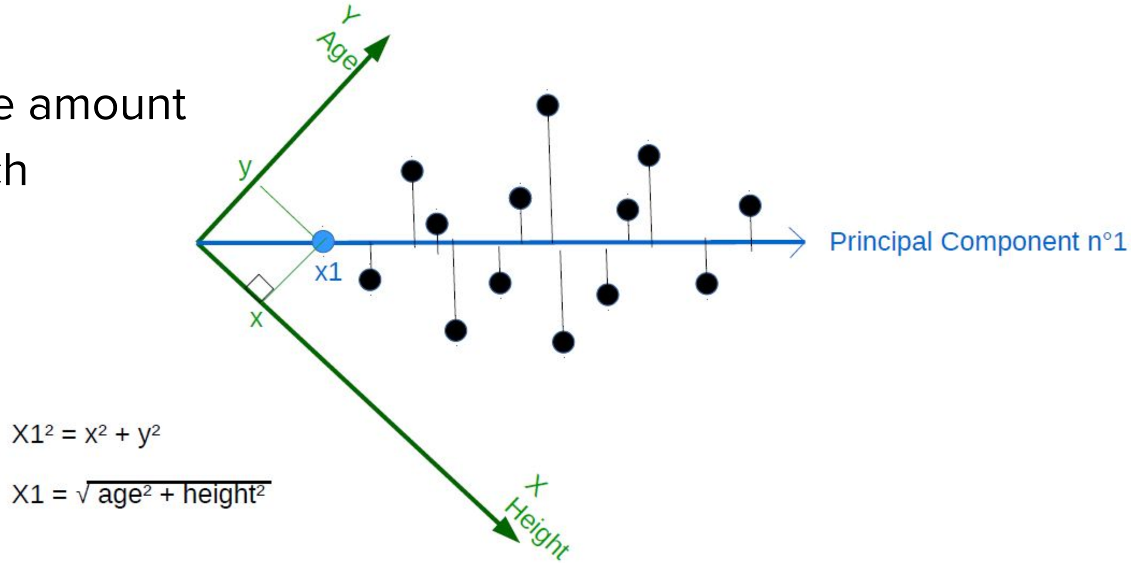

**Variance** on diagonal, **covariance** on off-diagonal

Symmetric!

$$\boldsymbol{\Sigma}_{i,j} = \boldsymbol{\Sigma}_{j,i}$$

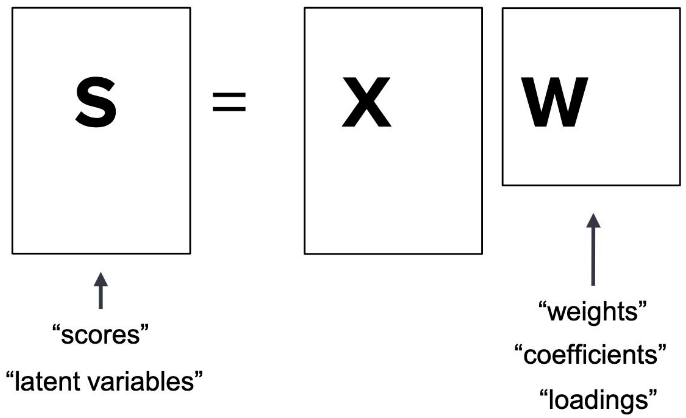
# The **covariance matrix** helps us find the dimension where the variance is maximal

- The **eigenvectors** of the covariance matrix are the Principal Components — the direction which the data is stretched
- The **eigenvalues** quantify the amount of variance explained by each Principal Component

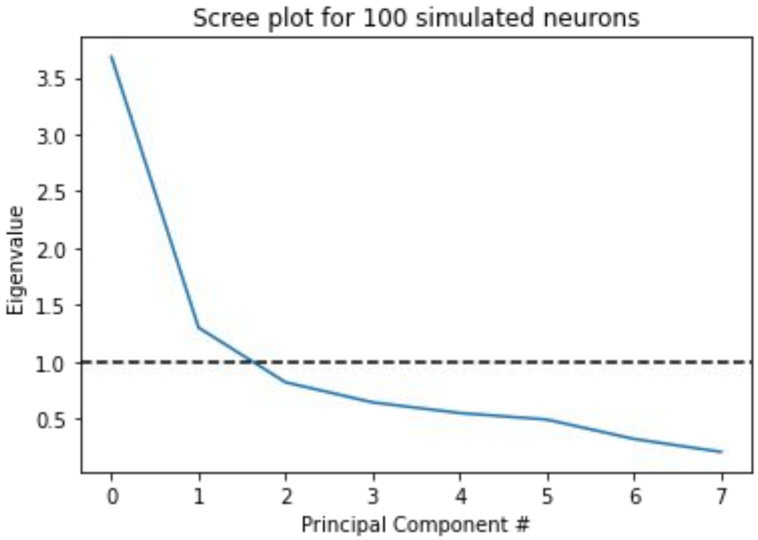


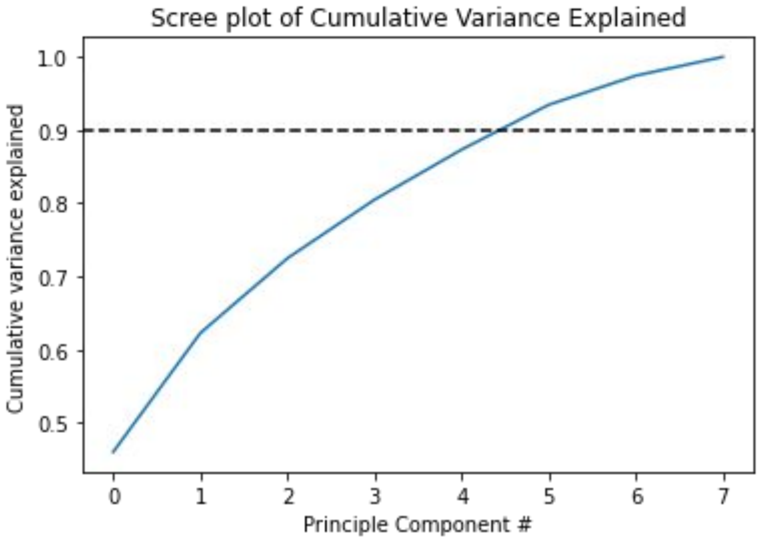
# Performing PCA involves five steps

1. Subtract the mean (and optionally, normalize)
2. Calculate the **covariance matrix** between the variables.
3. Extracting the factors by **rotation** (eigenvector decomposition).
4. Projecting mean-centered data onto the eigenvectors, using matrix multiplication



# How do we decide how many components are enough?



# Performing PCA involves five steps

1. Subtract the mean (and optionally, normalize)
2. Calculate the **covariance matrix** between the variables.
3. Extracting the factors by **rotation** (eigenvector decomposition).
4. Projecting mean-centered data onto the eigenvectors
5. Interpreting:
    a. Determining the **number** of factors needed
    b. Interpreting the **meaning** of factors.

*(we'll work through these steps in the notebook)*

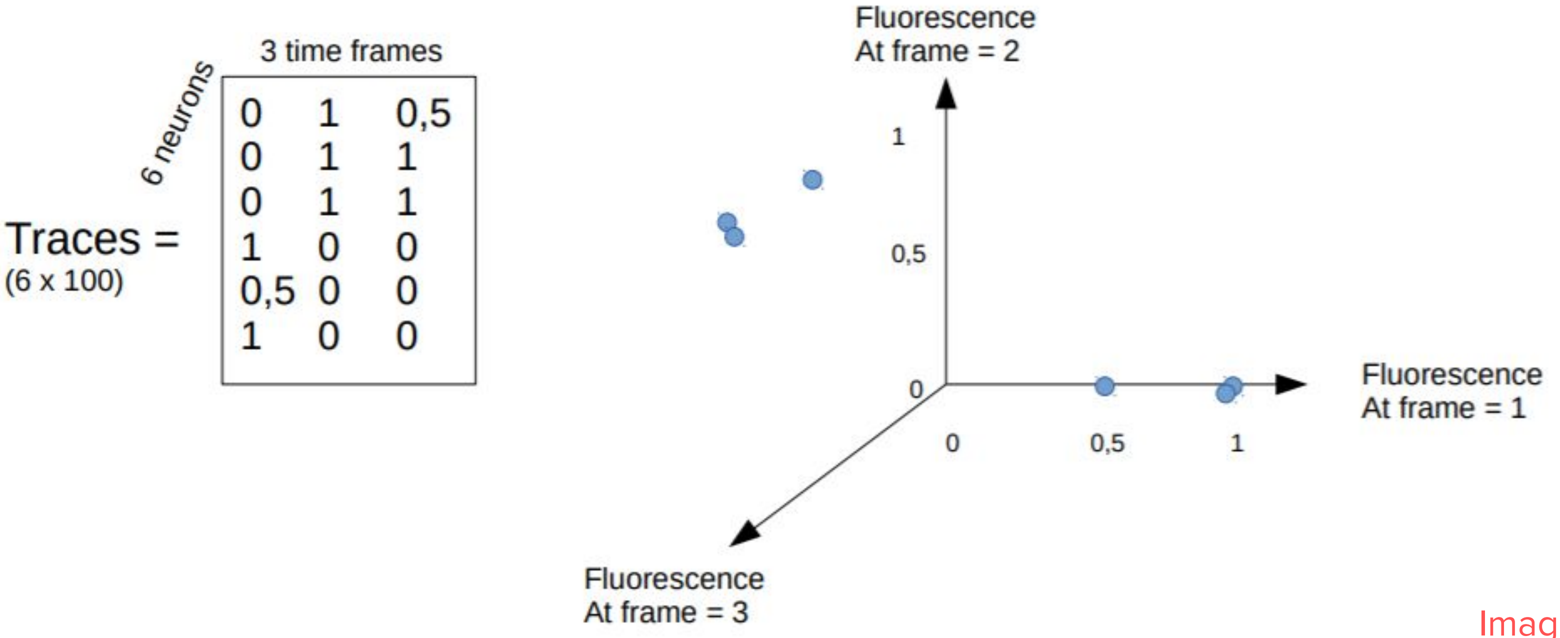



Image source

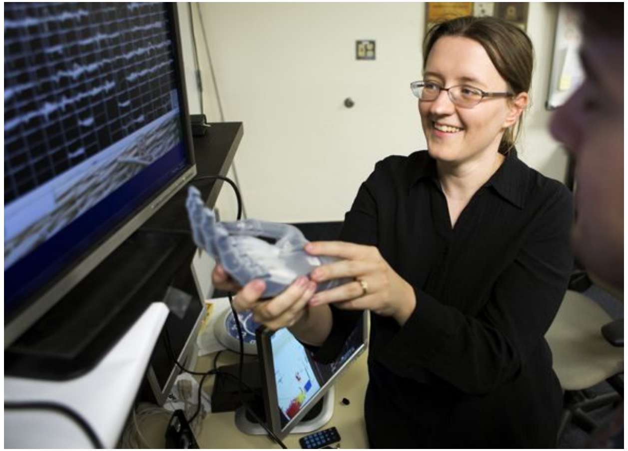
**We can also perform PCA on *time series***

Here, there are 6 neurons recorded over 3 time frames, creating 6 data points in a 3D space. (In reality, there are more like 54000 recorded frames, so this plot should have *50,000* dimensions)

# Dataset for today

First author:
**Cynthia Chestek**

(Read about her work!)



10742 • The Journal of Neuroscience, October 3, 2007 • 27(40):10742–10750

Behavioral/Systems/Cognitive

## Single-Neuron Stability during Repeated Reaching in Macaque Premotor Cortex

**Cynthia A. Chestek,[1*] Aaron P. Batista,[1,2*] Gopal Santhanam,[1] Byron M. Yu,[1] Afsheen Afshar,[1,3] John P. Cunningham,[1] Vikash Gilja,[4] Stephen I. Ryu,[1,5] Mark M. Churchland,[1,2] and Krishna V. Shenoy[1,2]**
[1]Department of Electrical Engineering, [2]Neurosciences Program, [3]Medical Scientists Training Program, [4]Department of Computer Science, and [5]Department of Neurosurgery, Stanford University, Stanford, California 94305

https://www.jneurosci.org/content/27/40/10742

PI: **Krisha Shenoy**

(Read about his life & work)

# Resources

Python Data Science Handbook, In Depth: PCA & In-Depth K-Means

StatQuest PCA overview and in more detail.

Dimensionality Reduction, *Neural Data Science* Chapter 8

Dimensionality reduction for large-scale neural recordings | Nature Neuroscience

An introduction to machine learning with scikit-learn

Steve Brunton's SVD Lecture Series