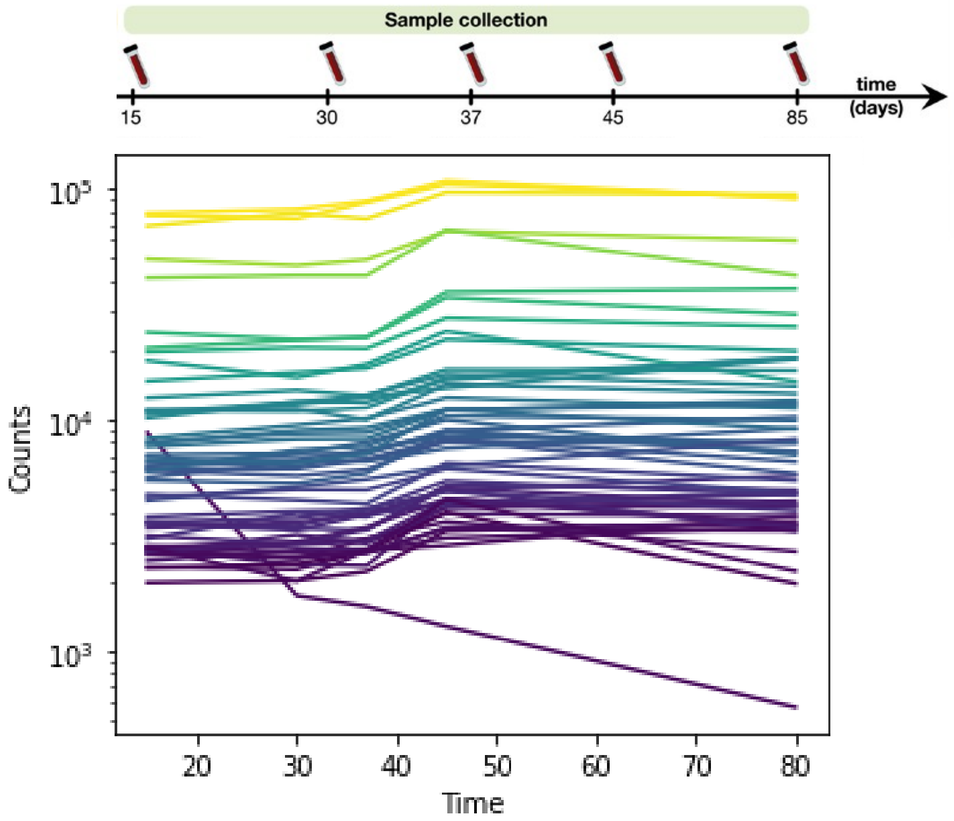
Quantitative Viral Dynamics Across Scales, 2022

Principal Component Analysis for trajectories

M. Bensouda Koraichi, A. Mazzolini
T. Mora, A. Walczak,
Q. M. Nguyen

Why PCA for trajectories



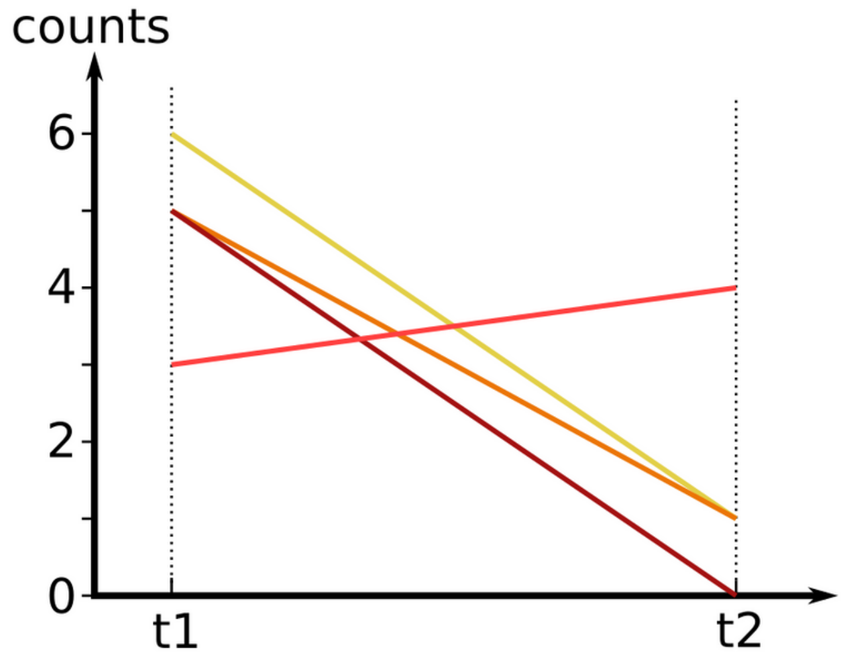
Our trajectory is a **5d** object (we have 5 time points)

We want to use PCA to do a **dimensional reduction** and represent the trajectory in a **2d** space

In this space we can more easily understand **global trends**

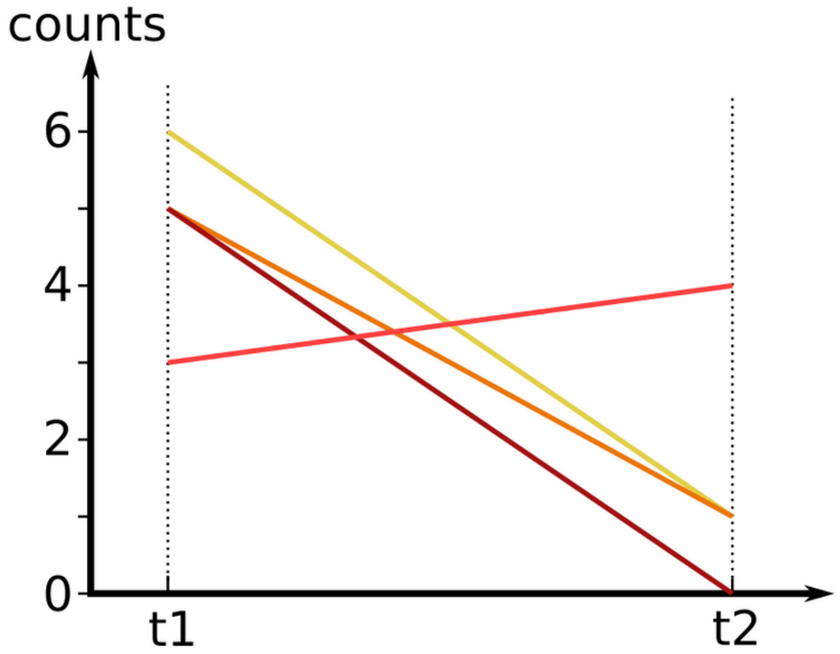
A toy example to understand PCA

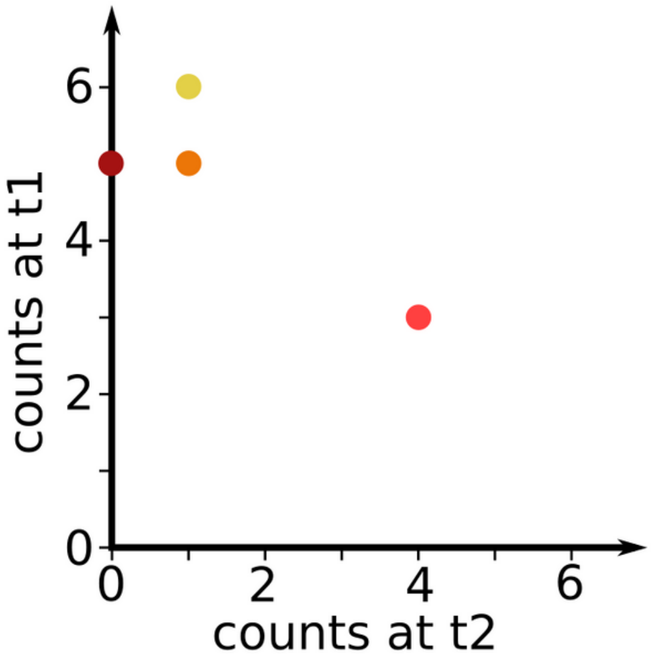
	Counts	
	t1	t2
CDR3 1	3	4
CDR3 2	5	0
CDR3 3	5	1
CDR3 4	6	1



A toy example to understand PCA

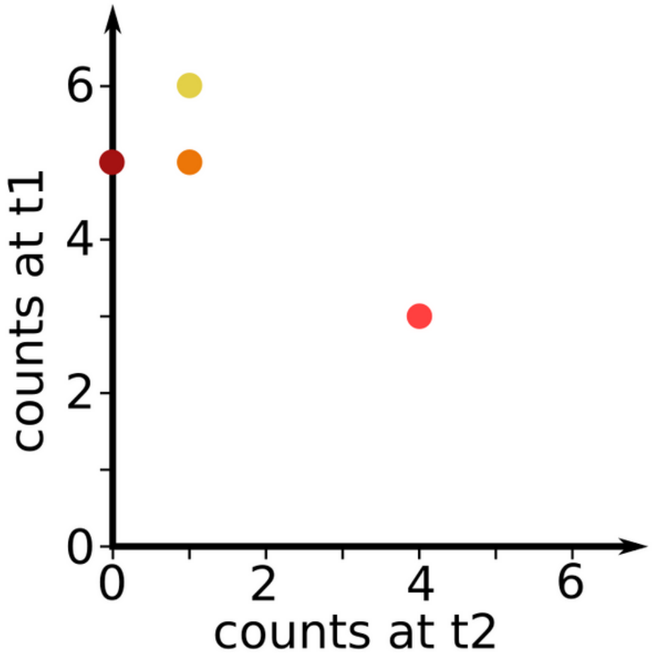
1 - Representation on the Cartesian plane

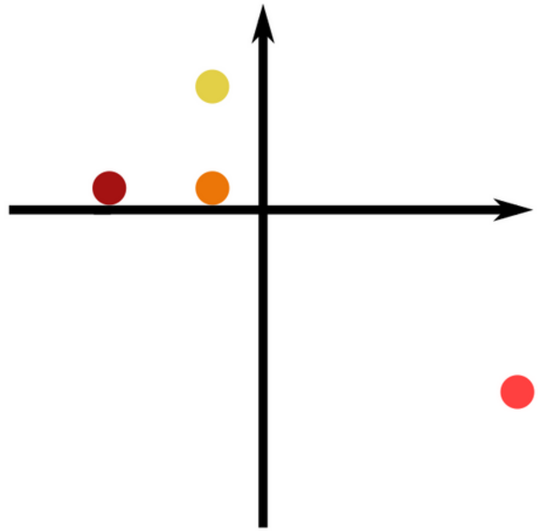




A toy example to understand PCA

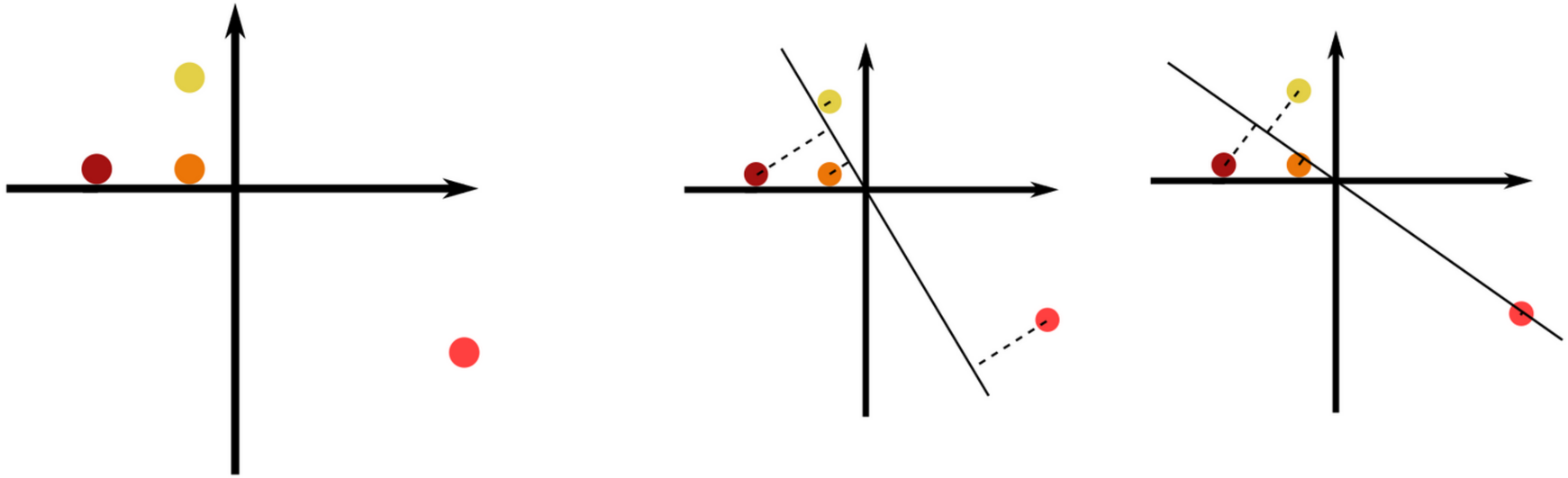
2 - Shift the origin to the center of mass





A toy example to understand PCA

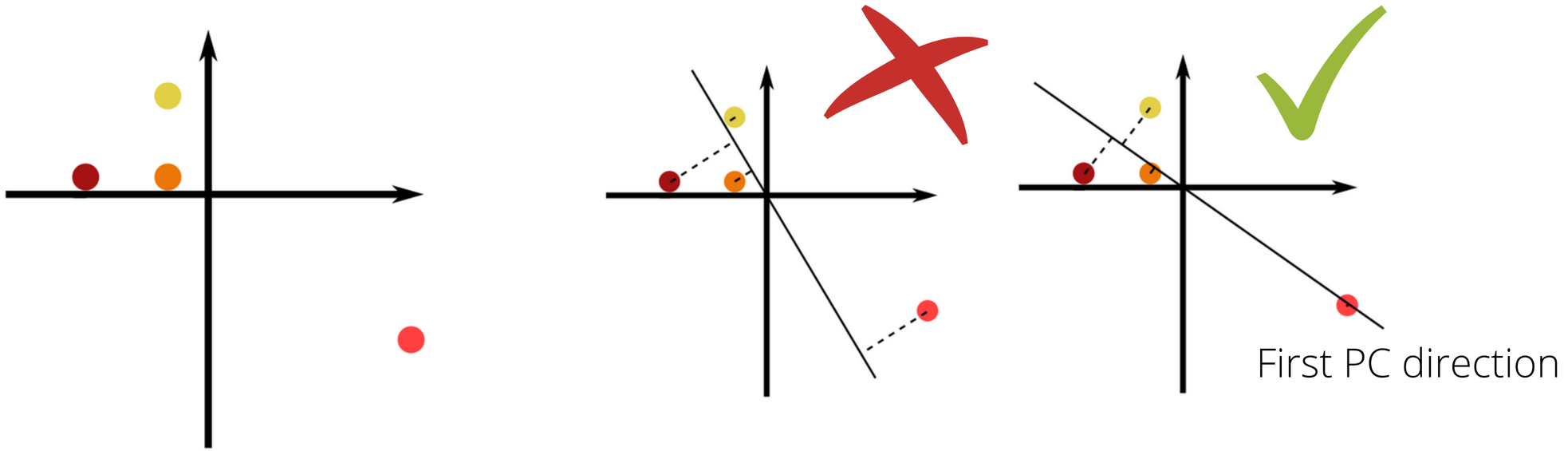
3 - Find the best linear fit passing from the origin



i.e. minimization of the summation of the squared distances of the points from the line

A toy example to understand PCA

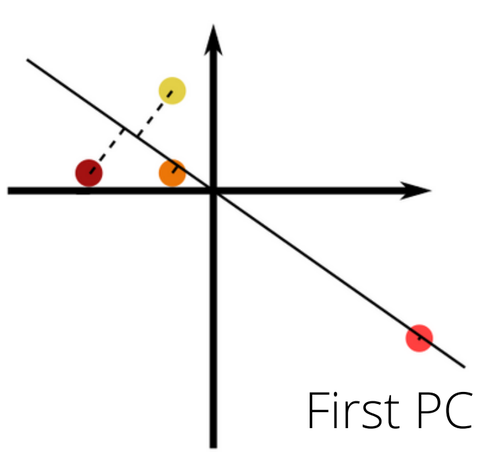
3 - Find the best linear fit passing from the origin



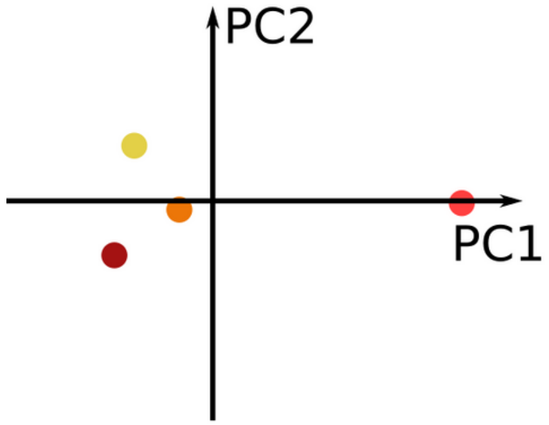
i.e. minimization of the summation of the squared distances of the points from the line

A toy example to understand PCA

4 - Rotation towards the first Principal Component

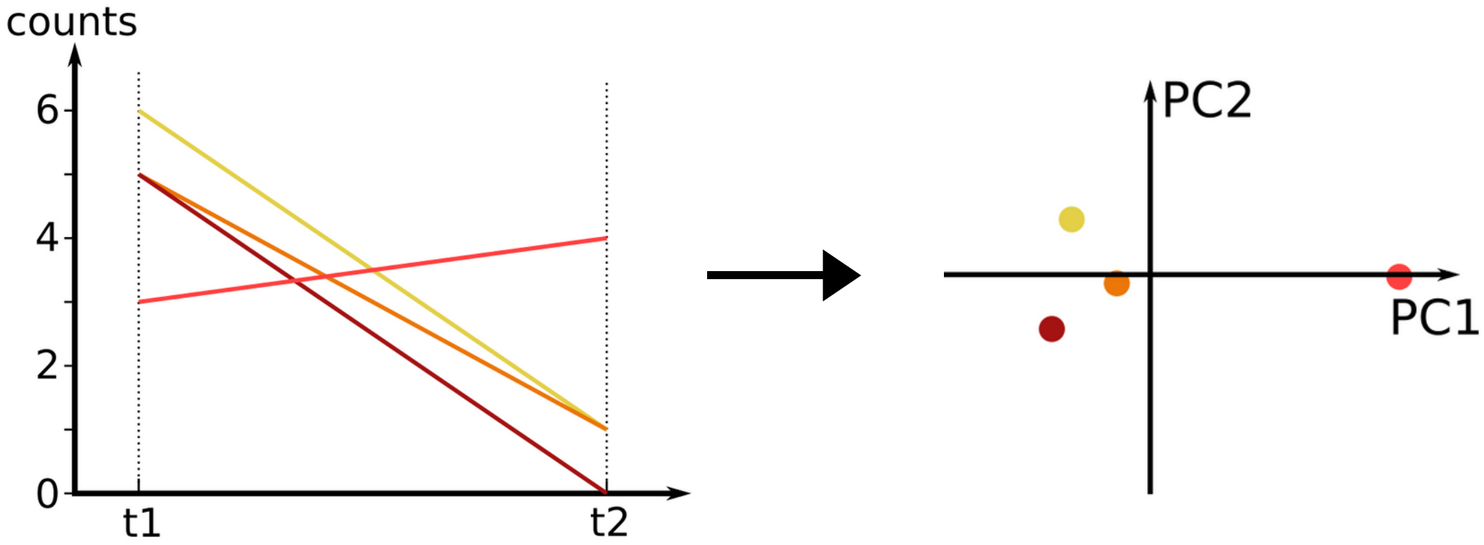


First PC direction



In 2d, the PC2 is the perpendicular line to PC1

A toy example to understand PCA

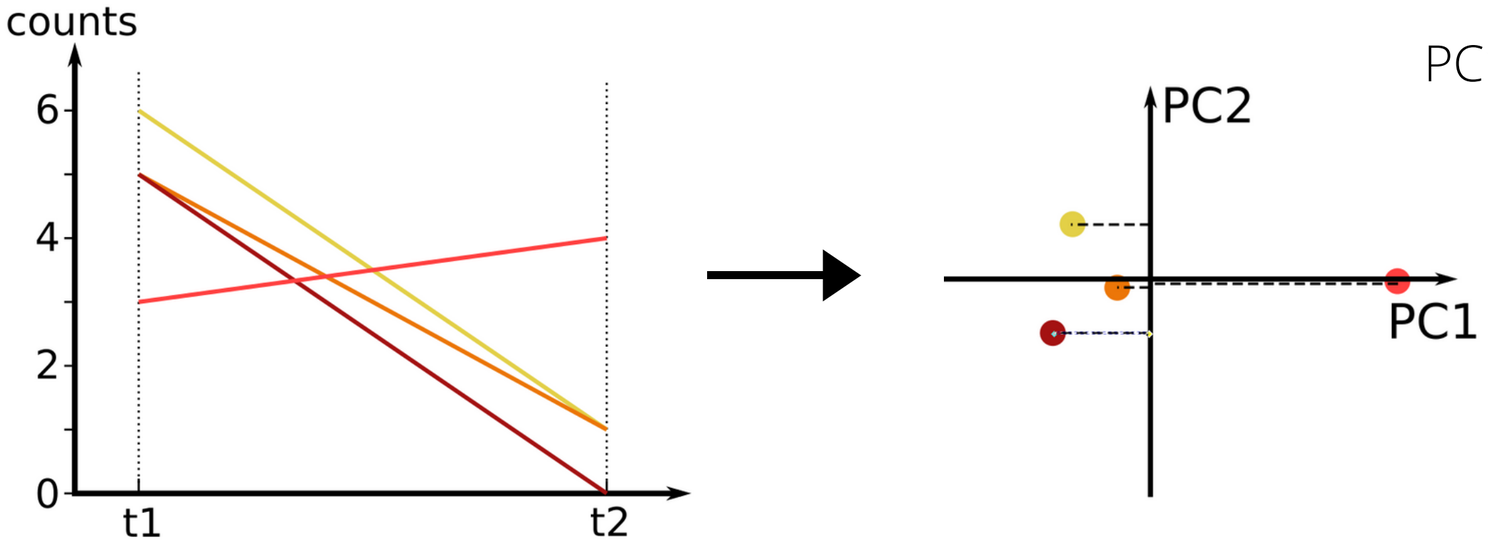


What we gain from this?

I'm ordering the axis from the **most informative** to the less

(true also for higher dimension)

A toy example to understand PCA



PC1 variation: 4.25

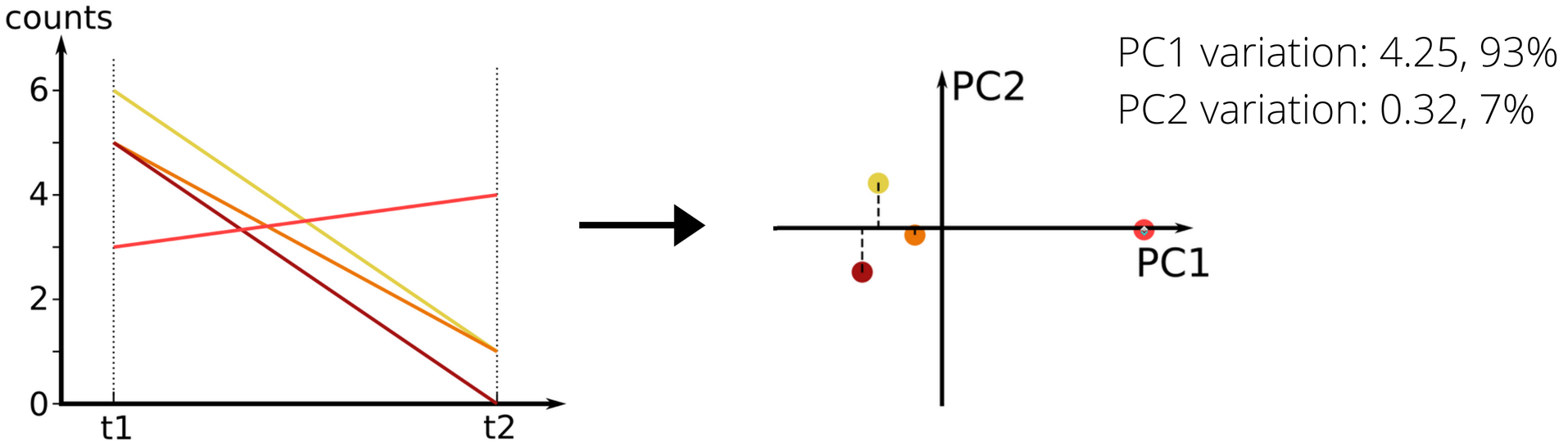
What we gain from this?

I'm ordering the axis from the **most informative** to the less

(true also for higher dimension)

Most informative: maximize data variation

A toy example to understand PCA



What we gain from this?

I'm ordering the axis from the **most informative** to the less

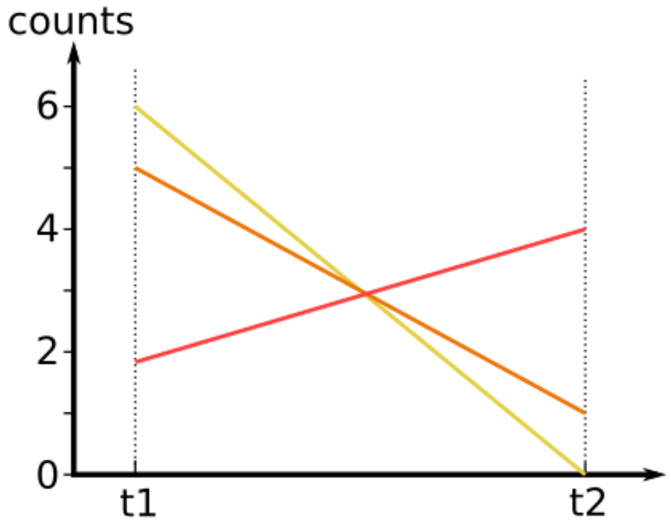
(true also for higher dimension)

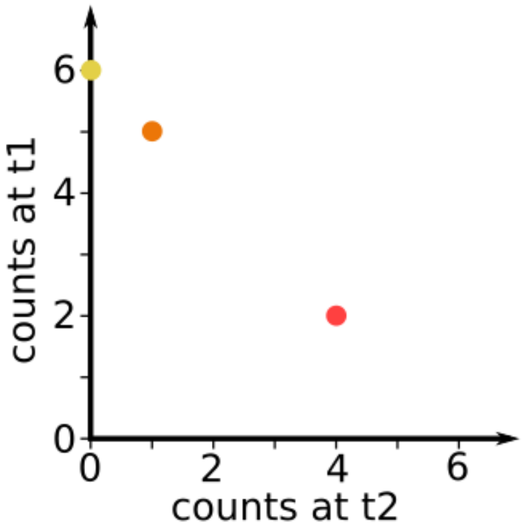
Most informative: maximize data variation

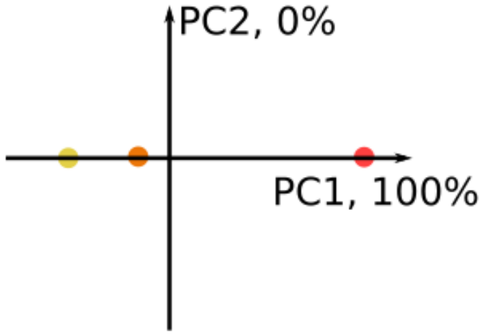
A toy example to understand PCA

Extreme case: data are effectively 1d

	Counts	
	t1	**t2**
CDR3 1	2	4
CDR3 2	6	0
CDR3 3	5	1

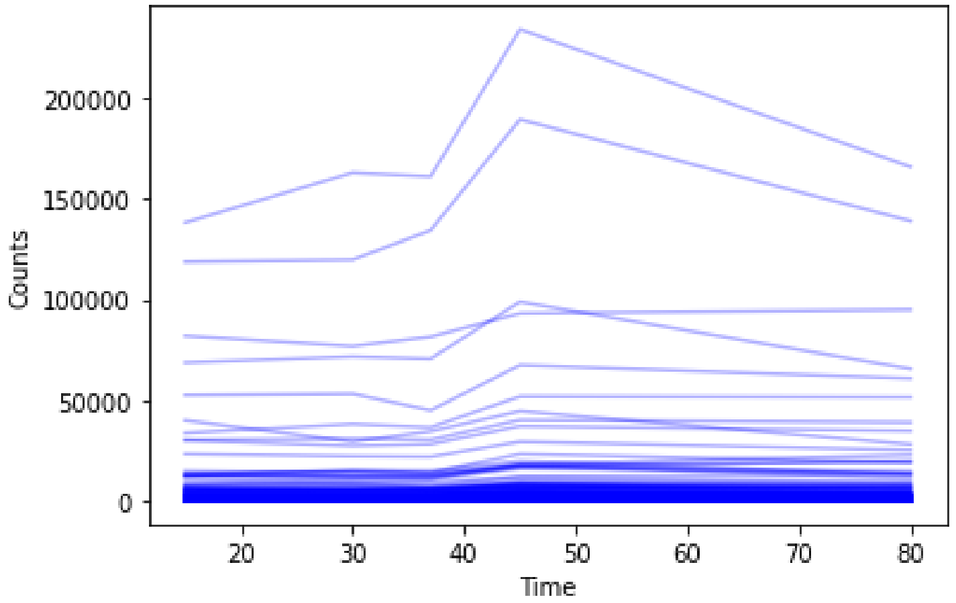






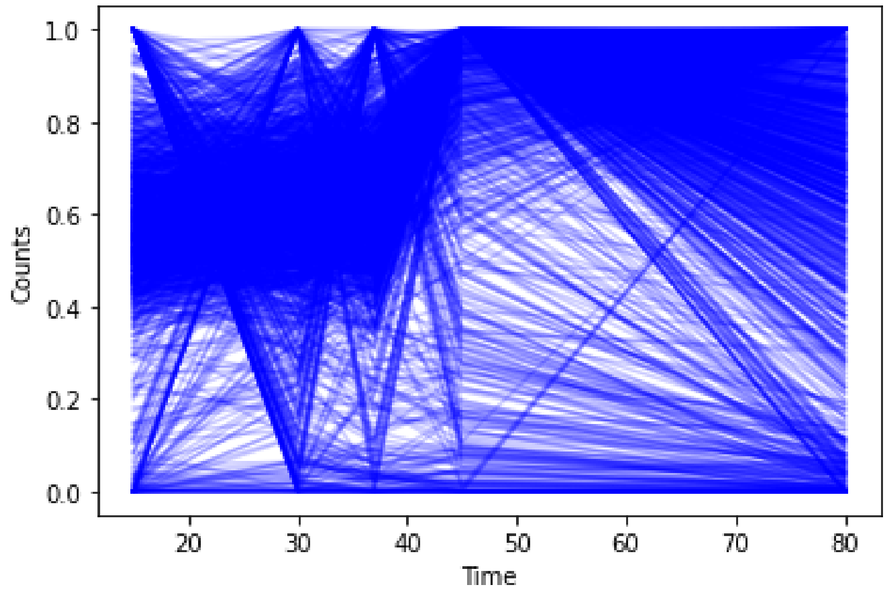
PCA for our 5d trajectories

We consider the CDR3 that have at least 1000 counts in at least 1 time point



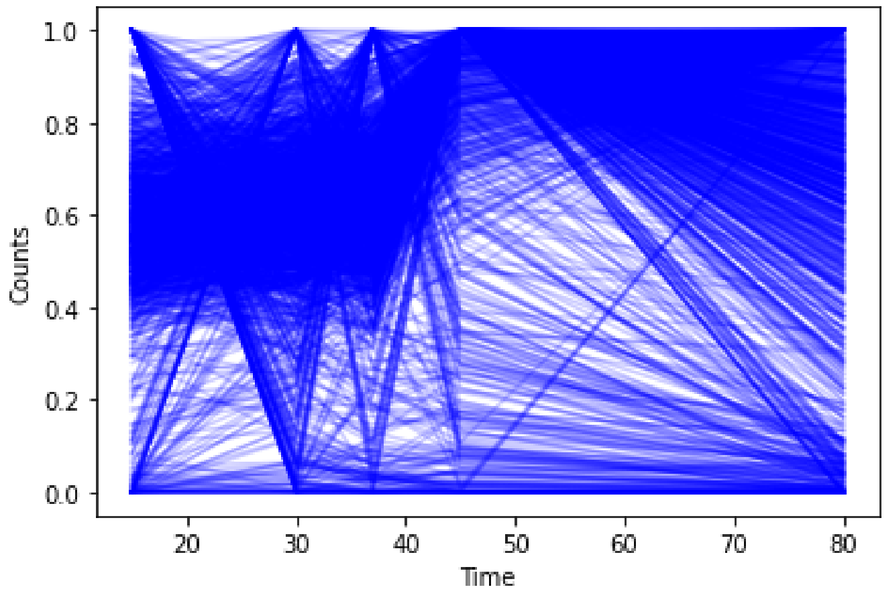
PCA for our 5d trajectories

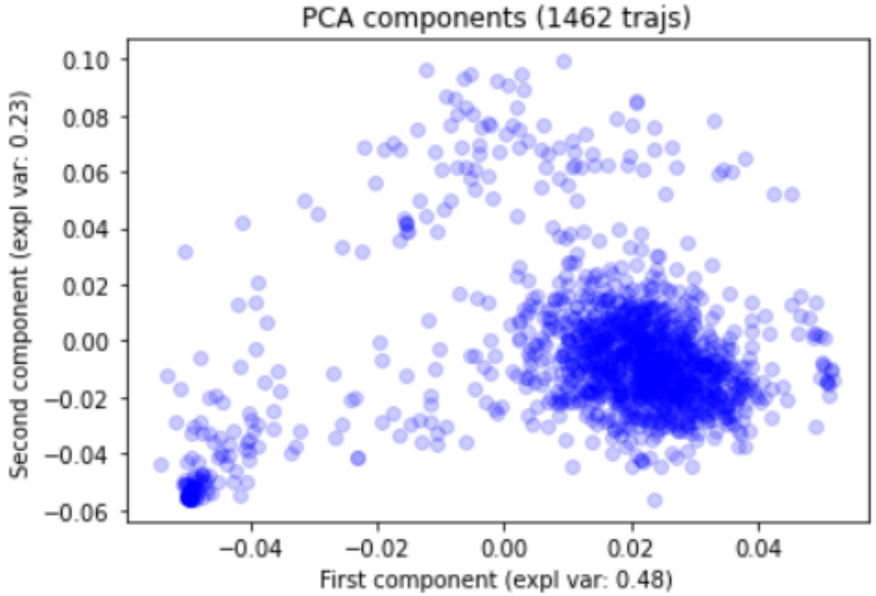
We consider the CDR3 that have at least 1000 counts in at least 1 time point and we normalize by the max count



PCA for our 5d trajectories

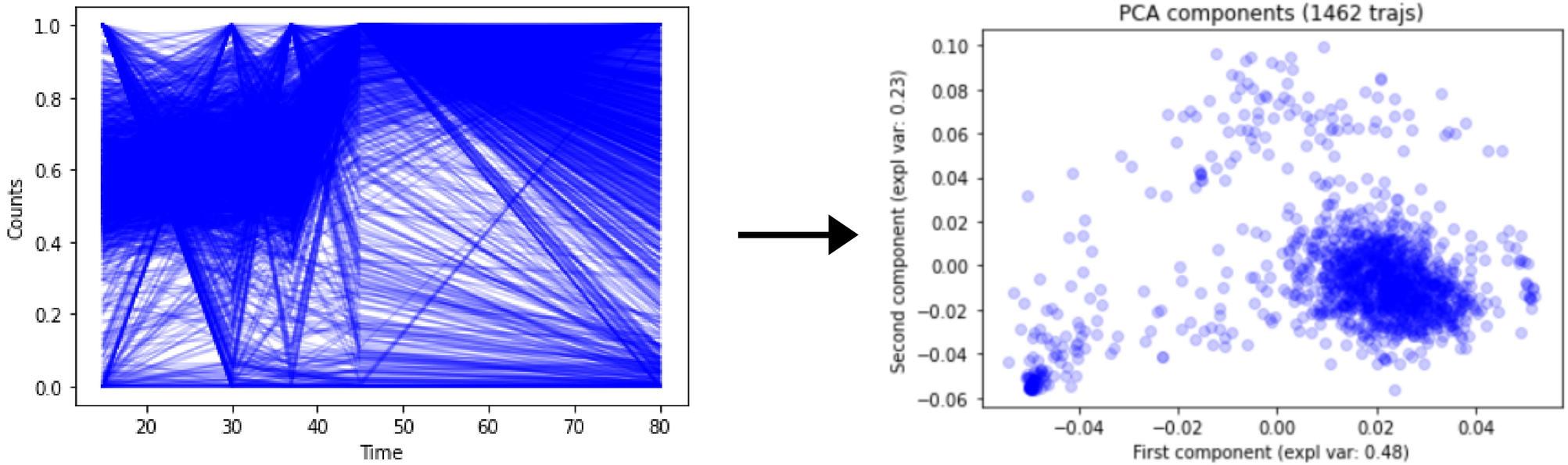
We consider the CDR3 that have at least 1000 counts in at least 1 time point and we normalize by the max count





PCA for our 5d trajectories

We consider the CDR3 that have at least 1000 counts in at least 1 time point and we normalize by the max count



The best 2d representation of my 5d data. There is interesting structure!