UNIVERSITY OF WARWICK

Lecture 10

Principal Component Analysis (PCA)

Gabriele Pergola and Amina Asif

Machine Learning Algorithms and Practice
CS429

October 2026

Credits: Some slides adapted from Fayyaz M

Part 1: the main idea

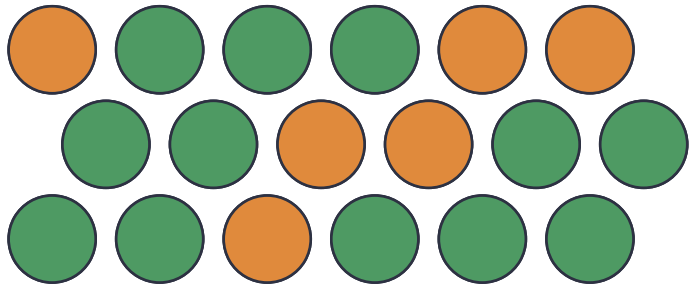
What a PCA plot shows us

Seeing the structure in many measurements at once

Patients who look alike

Imagine we have a group of patients who all came to the clinic with a respiratory infection…

Some of them might have **flu**, some a **common cold**, and some a **chest infection**…



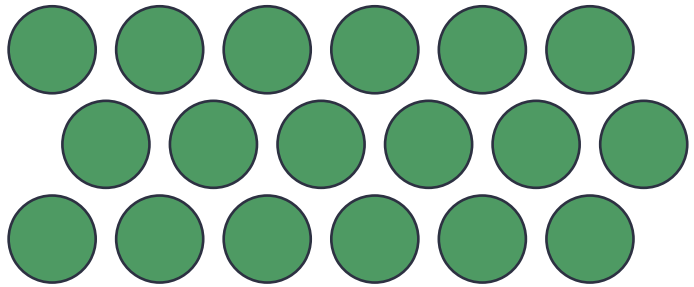
Patients who look alike

Imagine we have a group of patients who all came to the clinic with a respiratory infection…

Some of them might have **flu**, some a **common cold**, and some a **chest infection**…

…but, unfortunately, we cannot see the differences from the outside.

So we take some measurements from each patient: coughs per hour, body temperature, hours of sleep, heart rate, sneezes per hour, and the number of days they have been unwell.



Here are the data

Here are the data, for the first 10 of our **18** patients.

Each **row** is a patient…

	Coughs /hour	Temp. (°C)	Sleep (h)	Heart rate	Sneezes /hour	Days unwell
Patient 1	15	39.4	4.6	97	3	5
Patient 2	5	37.4	6.9	79	12	2
Patient 3	20	38.3	2.6	82	1	9
Patient 4	23	38.2	5.0	95	1	10
Patient 5	10	38.9	4.2	96	4	4
Patient 6	15	39.1	4.8	96	1	4
Patient 7	23	38.7	3.7	90	2	10
Patient 8	24	38.9	3.3	89	1	11
Patient 9	9	38.6	3.5	99	0	5
Patient 10	15	38.7	5.0	103	0	4
…	…	…	…	…	…	…

Here are the data

Here are the data, for the first 10 of our **18** patients.

Each **row** is a patient…

…and each **column** is one measurement, taken from every patient.

	Coughs /hour	Temp. (°C)	Sleep (h)	Heart rate	Sneezes /hour	Days unwell
Patient 1	15	39.4	4.6	97	3	5
Patient 2	5	37.4	6.9	79	12	2
Patient 3	20	38.3	2.6	82	1	9
Patient 4	23	38.2	5.0	95	1	10
Patient 5	10	38.9	4.2	96	4	4
Patient 6	15	39.1	4.8	96	1	4
Patient 7	23	38.7	3.7	90	2	10
Patient 8	24	38.9	3.3	89	1	11
Patient 9	9	38.6	3.5	99	0	5
Patient 10	15	38.7	5.0	103	0	4
…	…	…	…	…	…	…

Here are the data

Here are the data, for the first 10 of our **18** patients.

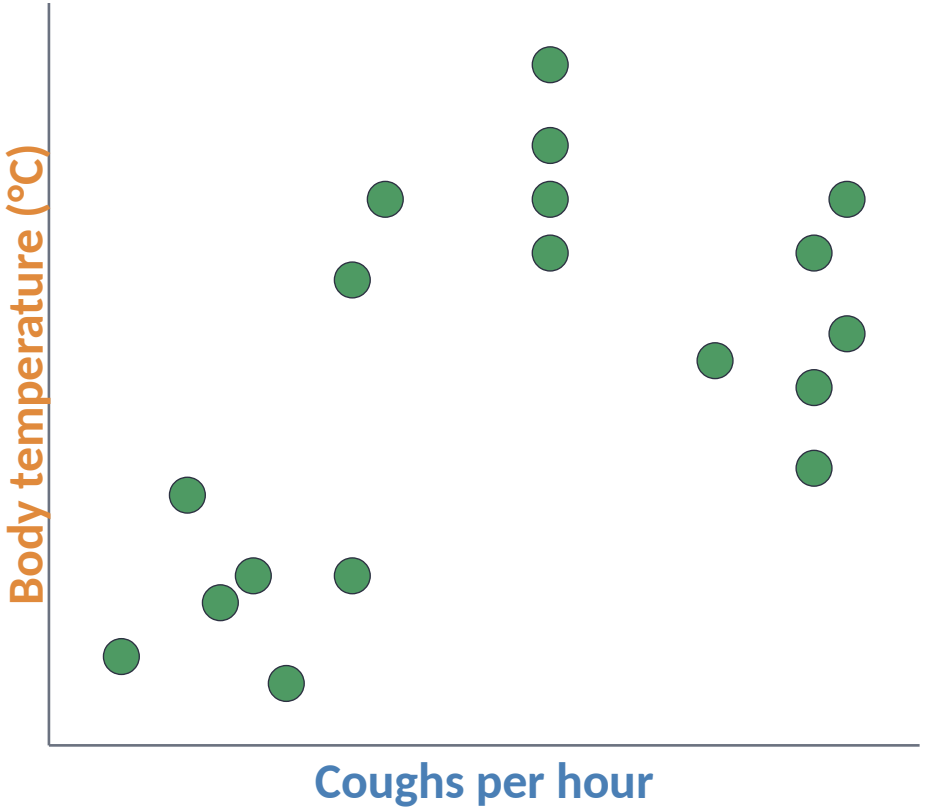
Each **row** is a patient…

…and each **column** is one measurement, taken from every patient.

	Coughs /hour	**Temp.** (°C)	**Sleep** (h)	**Heart** rate	**Sneezes** /hour	**Days** unwell
Patient 1	15	39.4	4.6	97	3	5
Patient 2	5	37.4	6.9	79	12	2
Patient 3	20	38.3	2.6	82	1	9
Patient 4	23	38.2	5.0	95	1	10
Patient 5	10	38.9	4.2	96	4	4
Patient 6	15	39.1	4.8	96	1	4
Patient 7	23	38.7	3.7	90	2	10
Patient 8	24	38.9	3.3	89	1	11
Patient 9	9	38.6	3.5	99	0	5
Patient 10	15	38.7	5.0	103	0	4
…	…	…	…	…	…	…

Two measurements at a time

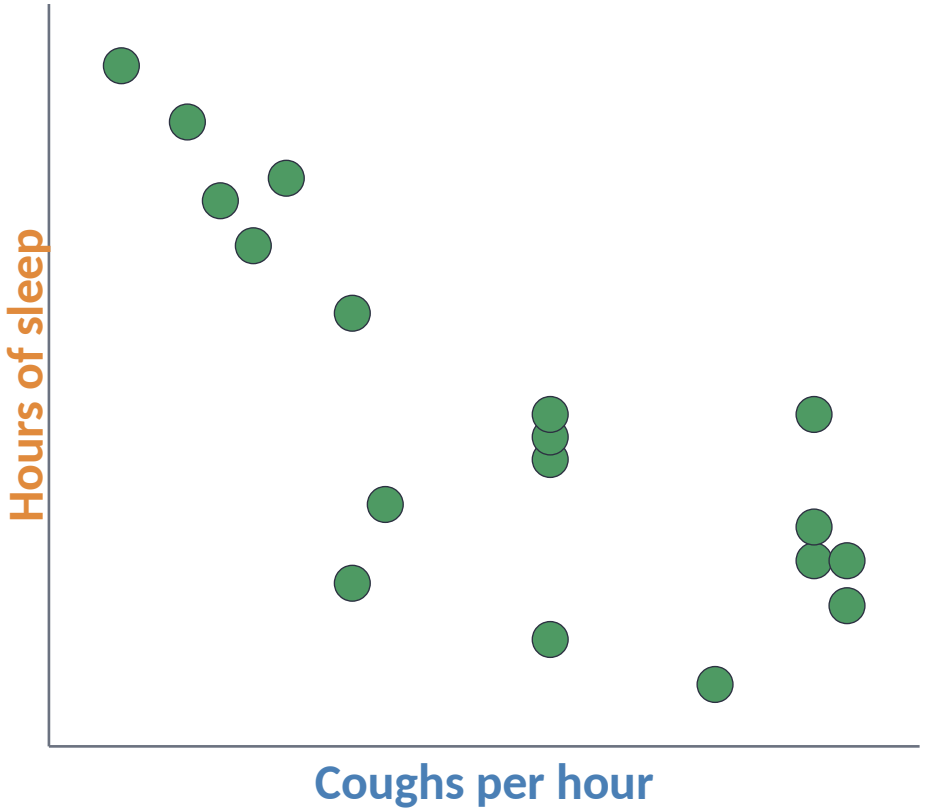
If we only had two measurements, we already know what to do: plot one against the other, and look at how closely they are related, using the **correlation**.



Two measurements at a time

If we only had two measurements, we already know what to do: plot one against the other, and look at how closely they are related, using the **correlation**.

We could also compare coughs with hours of sleep: here, the relationship is negative…

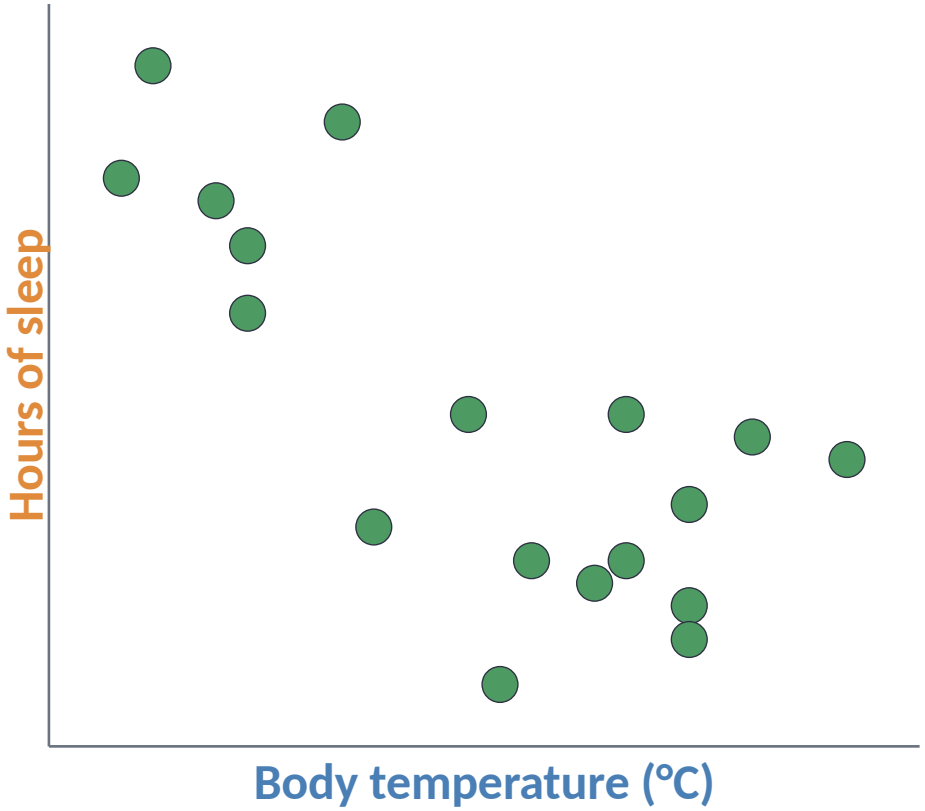


Two measurements at a time

If we only had two measurements, we already know what to do: plot one against the other, and look at how closely they are related, using the **correlation**.

We could also compare coughs with hours of sleep: here, the relationship is negative…

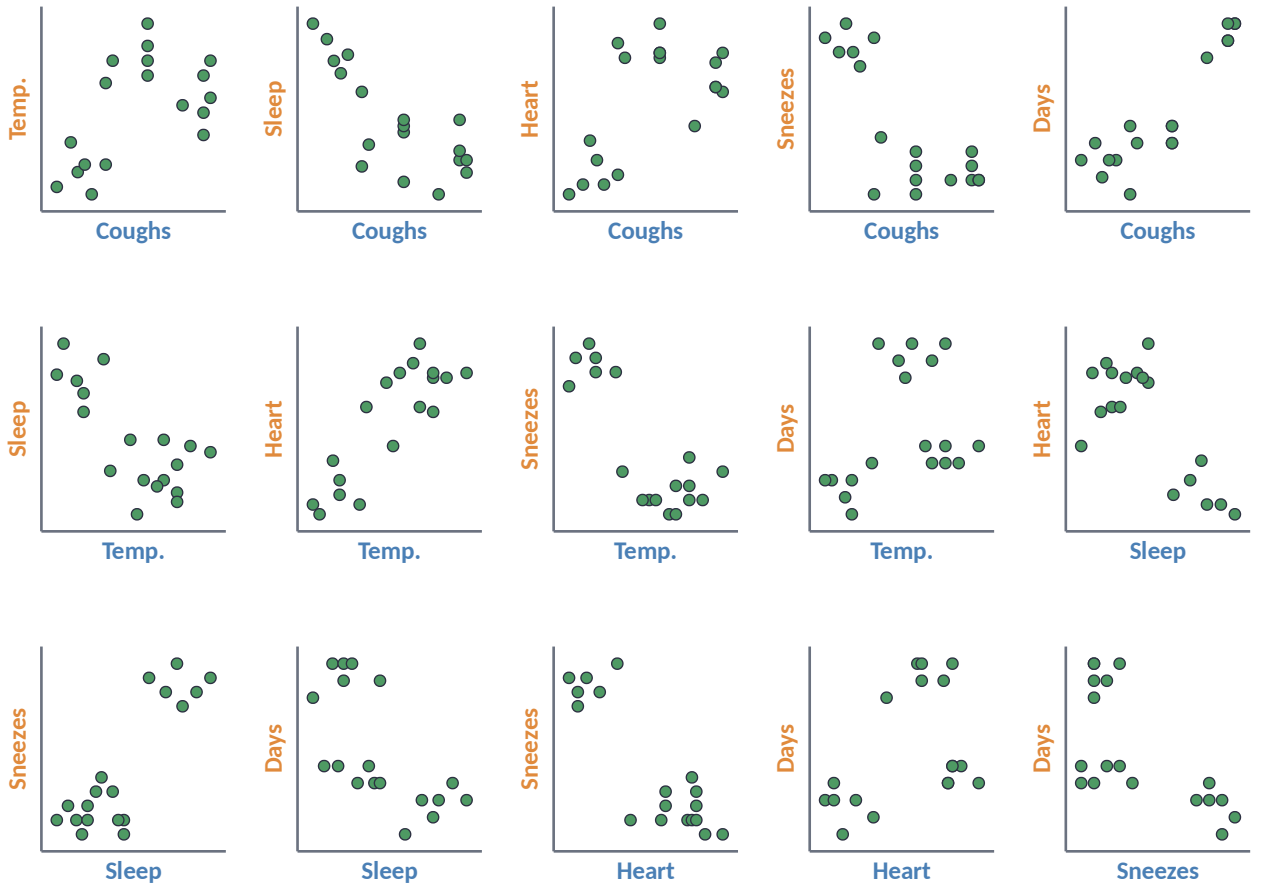
…and, lastly, temperature with sleep. Each plot tells us something, but only about two measurements at a time.



Three measurements, or more?

With 3 measurements, we could draw a 3D graph, but it would already be hard to read.

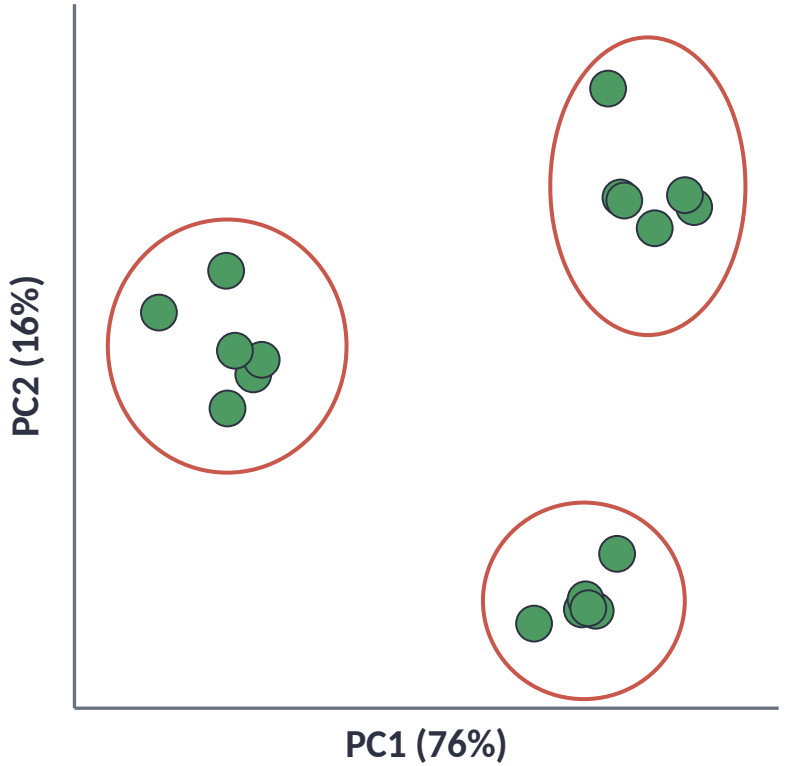
With all **6** measurements, should we draw lots and lots of 2D plots (15 of them!) and try to make sense of them all?



A PCA plot

A **PCA plot** converts the correlations (or lack of them) among *all* of the measurements into a single 2-D graph.

Patients whose measurements are highly correlated cluster together. Here, we can see three clusters.



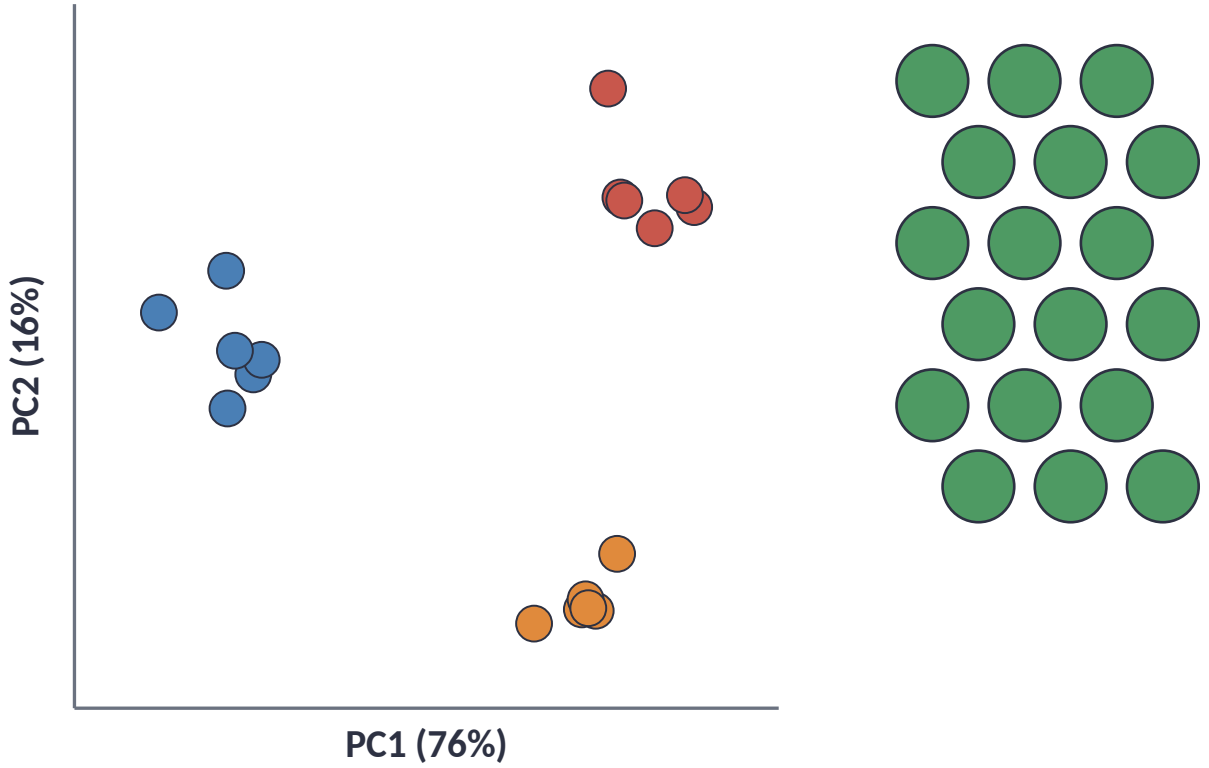
A PCA plot

A **PCA plot** converts the correlations (or lack of them) among *all* of the measurements into a single 2-D graph.

Patients whose measurements are highly correlated cluster together. Here, we can see three clusters.

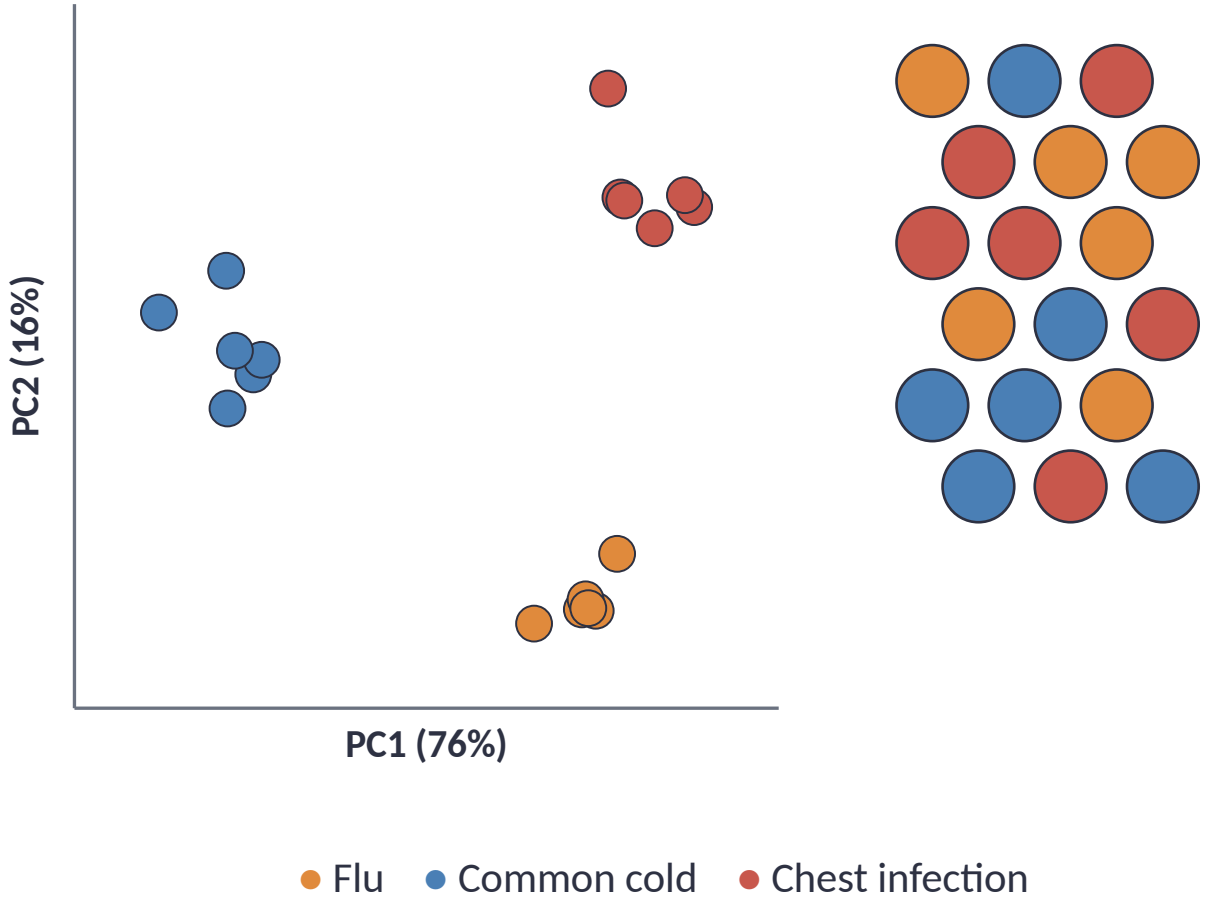
To make the clusters easier to see, we can colour-code them.

Once we have identified the clusters, we can go back to the original patients…



A PCA plot

…and see that they correspond to three different illnesses, each affecting the measurements in its own way. PCA never saw the diagnoses: it found the groups from the measurements alone.

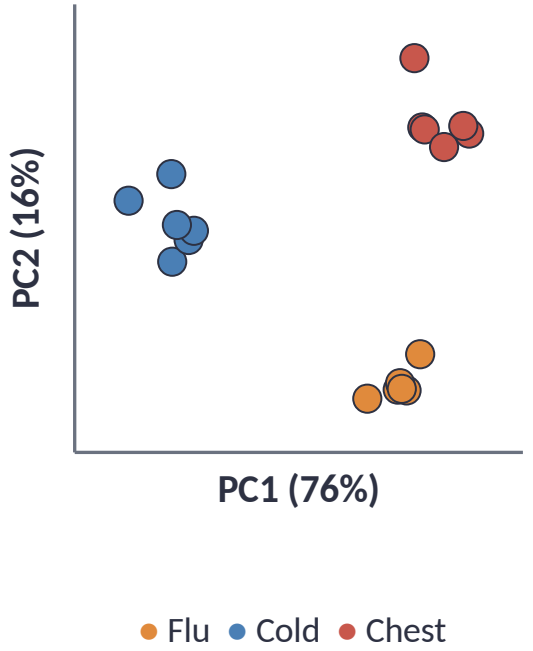


What do the axes mean?

PCA also tells us what each axis means, through the **loading scores**: how much each measurement contributes to each principal component. Large values, positive or negative, matter most.

PC1 combines more coughs, a higher temperature, a faster heart rate and more days unwell, with less sleep and fewer sneezes. Patients on the right are acutely unwell; those on the left look more like a common cold.

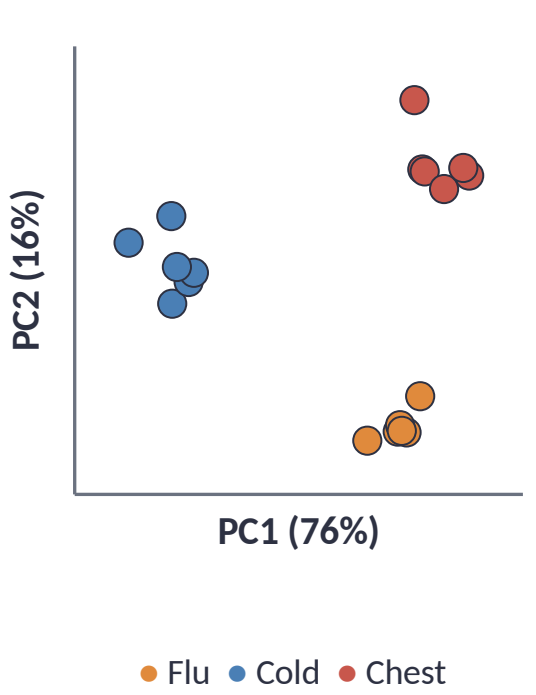
	PC1	PC2
Coughs per hour	0.41	0.43
Temperature	0.39	−0.48
Hours of sleep	−0.43	0.01
Heart rate	0.41	−0.41
Sneezes per hour	−0.45	0.11
Days unwell	0.35	0.64



What do the axes mean?

PC2 sets coughs and days unwell against temperature and heart rate. It separates a **long, chesty illness** with a moderate fever (chest infection, at the top) from a short, feverish one (flu, at the bottom).

	PC1	PC2
Coughs per hour	**0.41**	**0.43**
Temperature	**0.39**	−0.48
Hours of sleep	−0.43	0.01
Heart rate	**0.41**	−0.41
Sneezes per hour	−0.45	0.11
Days unwell	**0.35**	**0.64**

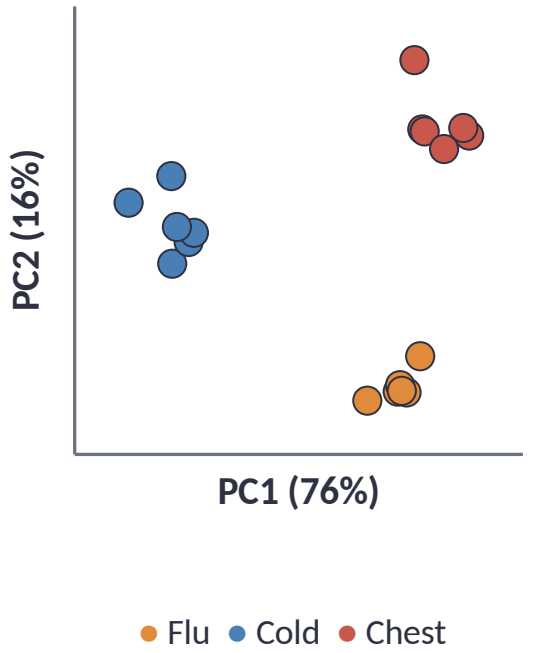


What do the axes mean?

PC2 sets coughs and days unwell against temperature and heart rate. It separates a long, chesty illness with a moderate fever (chest infection, at the top) from a short, feverish one (flu, at the bottom).

This is how PCA tells us which measurements are the most valuable for separating the patients.

	PC1	PC2
Coughs per hour	**0.41**	**0.43**
Temperature	**0.39**	−0.48
Hours of sleep	−0.43	0.01
Heart rate	**0.41**	−0.41
Sneezes per hour	−0.45	0.11
Days unwell	**0.35**	**0.64**

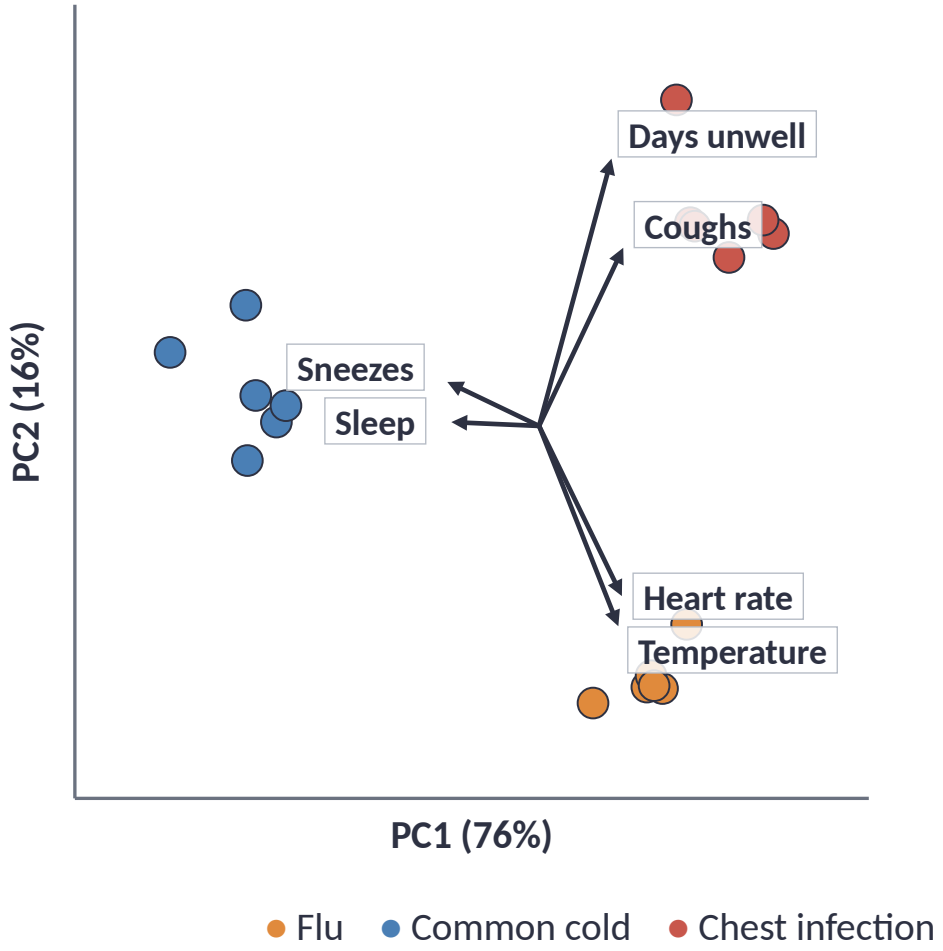


A biplot: patients and measurements together

A **biplot** draws the patients (their scores) and the measurements (their loadings) on the same PCA plot.

Each arrow is one measurement: its coordinates are the measurement's loadings on PC1 and PC2 (scaled up, so that the arrows are visible).

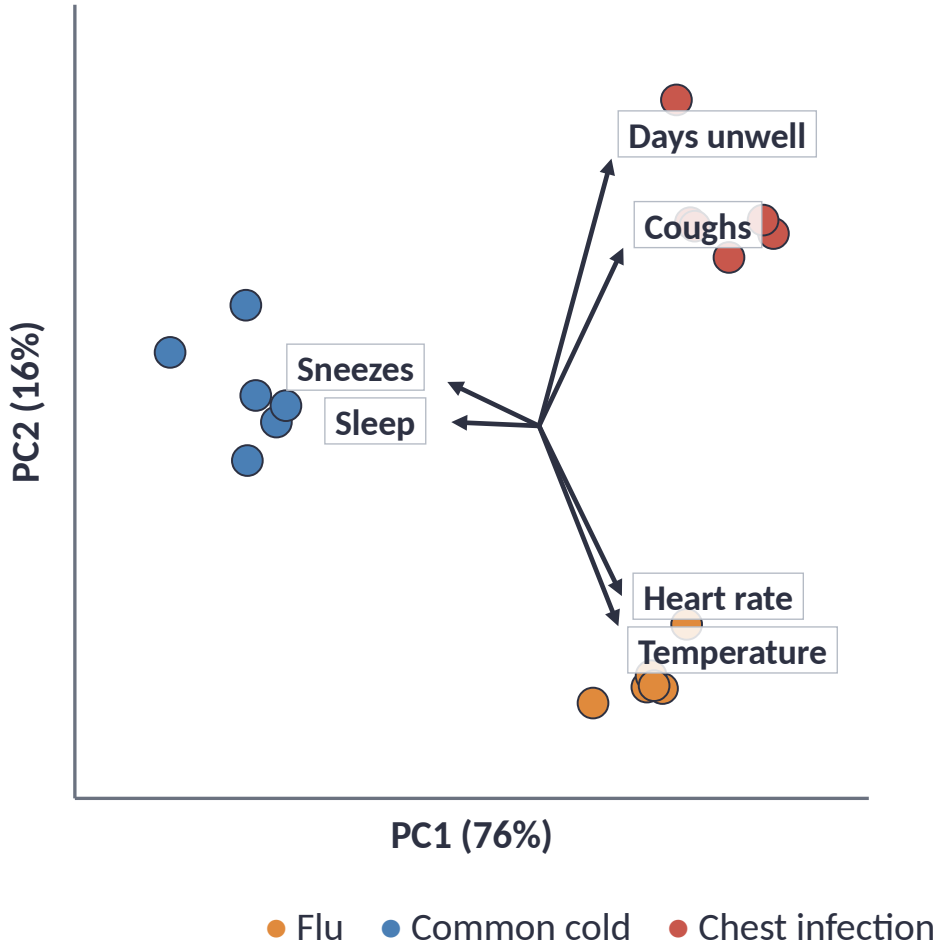
An arrow points to where that measurement is high. The flu patients lie towards **temperature** and **heart rate**; the chest infections towards **days unwell** and **coughs**; the colds towards **sneezes** and **sleep**.



A biplot: patients and measurements together

The angles between arrows also tell a story: arrows pointing the same way are positively correlated (temperature and heart rate, r = 0.85), opposite arrows are negatively correlated (coughs and sleep, r = −0.78), and arrows at a wide angle are only weakly correlated (temperature and days unwell, r = 0.36).

Note: these readings are approximate, because the biplot shows only the first 2 PCs; here they hold 92% of the variation, so the picture is a faithful one.



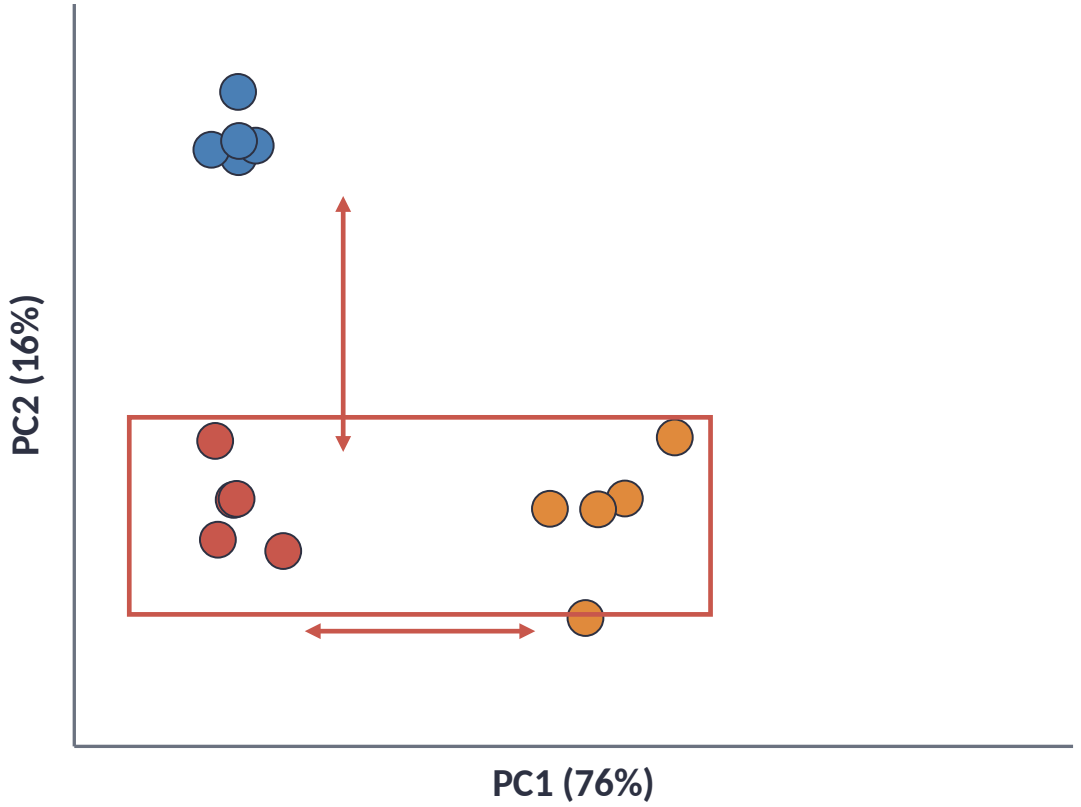
The axes are ranked by importance

Here is one last main idea about how to interpret PCA plots: **the axes are ranked in order of importance**.

Differences along the first principal component axis (**PC1**) are more important than differences along the second (**PC2**).

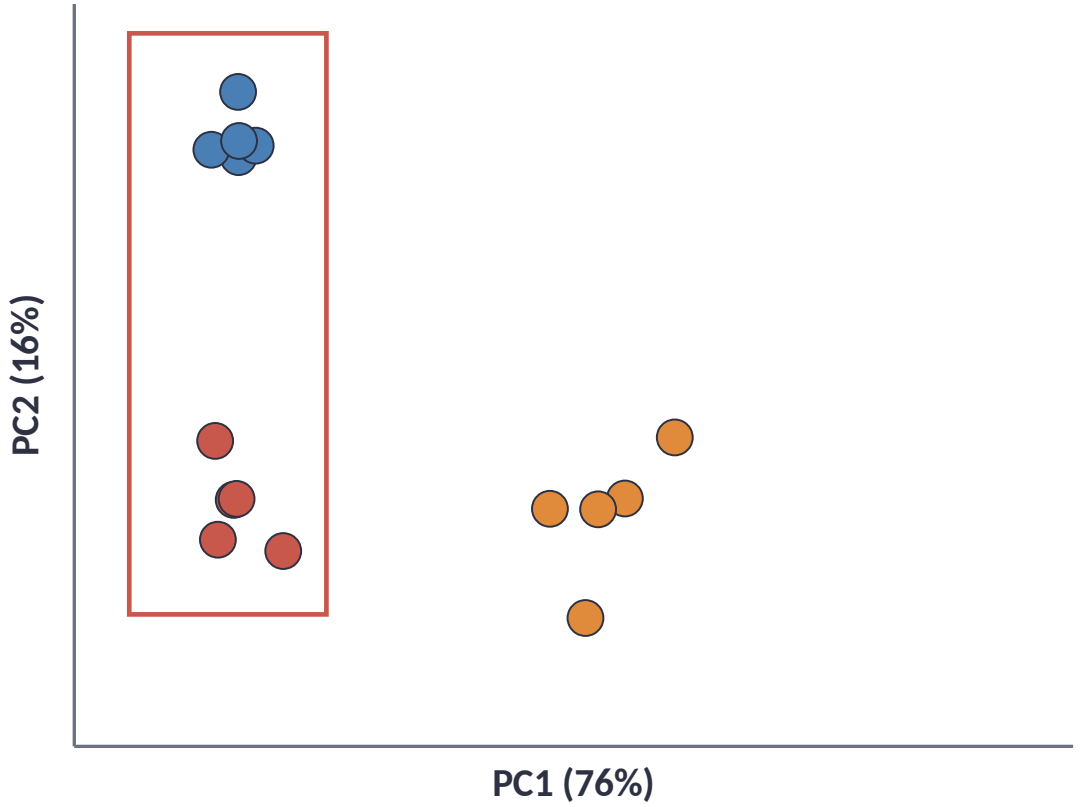
So, even if the distance between these two clusters (along PC2) looks the same as the distance between these two (along PC1)…

…the two clusters separated along PC1 are more different from each other…



The axes are ranked by importance

…than the two clusters separated along PC2. In our plot, PC1 accounts for 76% of the variation and PC2 for only 16%.



Why PCA matters in machine learning

So far, we have used PCA to *look* at data.

In machine learning, it has several further uses.

Visualisation: seeing high-dimensional data in 2-D, as we have just done.

Dimensionality reduction: describing each patient with only the first k PCs instead of all the measurements, as our PCA plot does: here, 2 PCs keep 92% of the variation in 6 measurements.

Noise reduction: the last PCs often capture little more than measurement noise, so dropping them can clean the data.

Feature preprocessing: a classifier, such as the flu classifier from earlier in the module, can use a few uncorrelated PC scores instead of many correlated measurements, which can speed up training and reduce overfitting.