PRINCIPAL COMPONENT ANALYSIS

One MDS method
[Multidimensional scaling]

21 · there are 21 of us · we have different preferences

12:00 am

Guys, I want to find someone to share the weekend, but it's hard to decide whom I may invite to share with.

Yea, everyone has different interests. It is hard to find someone more like-minded, and we can't make friends through appearances.

Well, I happen to know there's an analysis technique that can satisfy this problem.
Have you ever heard about PCA ?

It's a statistical technique for analysing large datasets that may contain many dimensions. Mainly used to define the relationship of unrelated variables.

I see, so we can use it to calculate the everyone's interests and to find the most fitted partners in our group.

Wow, let's see how it performs!

## which of us are better to spend a lovely weekend together ?

Testing time!

Analysis Process

Before the final application of data, we want to make a small internal experiment to help you comprehend the procedure of PCA and the feasibility of the analysis.

we made a small questionnaire of five questions identifying their wants independently. In this part, we will pick up two questions in these five to make the procedure clearly for you.

Q1 Do you like having exhibitions?
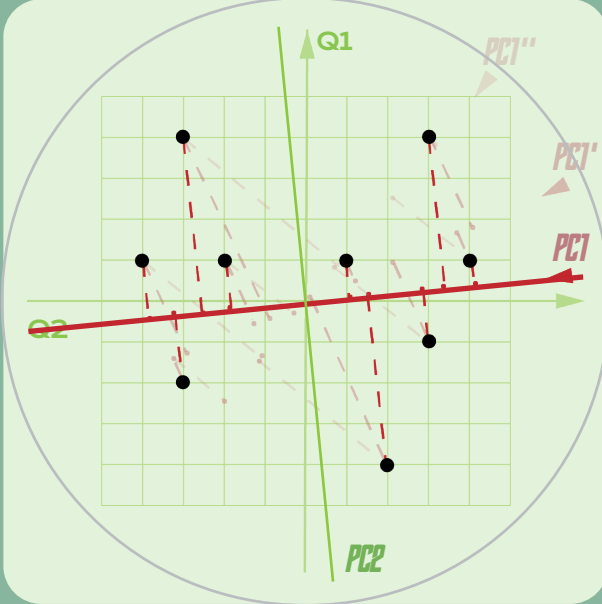**Q2** Do you like going cycling?

*Some data are overlapping
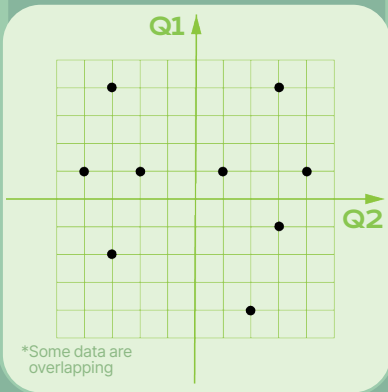
STEP 1

1. Calculating the covariance matrix

2. Calculate the eigenvalues and eigenvectors of the covariance matrix

This sounds complicated, but we can understand the process as:

1 Rotate a line through the origin
2 Find the shortest distance

STEP 2

Sort the eigenvalues from largest to smallest

PC1 96.5% eigenvalues
PC2 3.43% eigenvalues

k-PC1 ✓ > k-PC2

Keep the top k eigenvectors as new space

However, it is important to note that when discarding eigenvectors with smaller eigenvalues, part of the information is lost along with them :

Eigenvalues = Information

In this example, we dropped PC2 and also the information carried on PC2, now the data after dimensionality reduction has only 96.5% of the information of the original data.

STEP 3

Transform the data into new space constructed by the above k eigenvectors.

*Some data are overlapping
PC1

NOW! It's time to address the real issues

Start from here

Take me as an example

Q1 Do you like having exhibitions?	Q2 Do you like going cycling?	Q3 Do you like a forest walk?	Q4 Do you like spending an afternoon sipping a latte at a beautiful café?	Q1 Do you like having exhibitions?
♥♥♥♡♡	♥♥♡♡♡	♥♥♥♥♡	♥♥♥♡♡	♥♥♡♡♡

Questionnaire with 5 questions

5 Questions

5 Dimensions
100% INFOMATION

In the process of dimensionality reduction, part of the information will be lost with the discarded PC

PC1 79.9% ✓
PC2 10.8% ✓
PC3 5.07%
PC4 2.66% ✓
PC5 1.58%

⚠ The new PCs have different meanings from the original five questions

After selecting four types of PCs, the data came to 4 dimensions, but we still can't see what the data looks like right now.
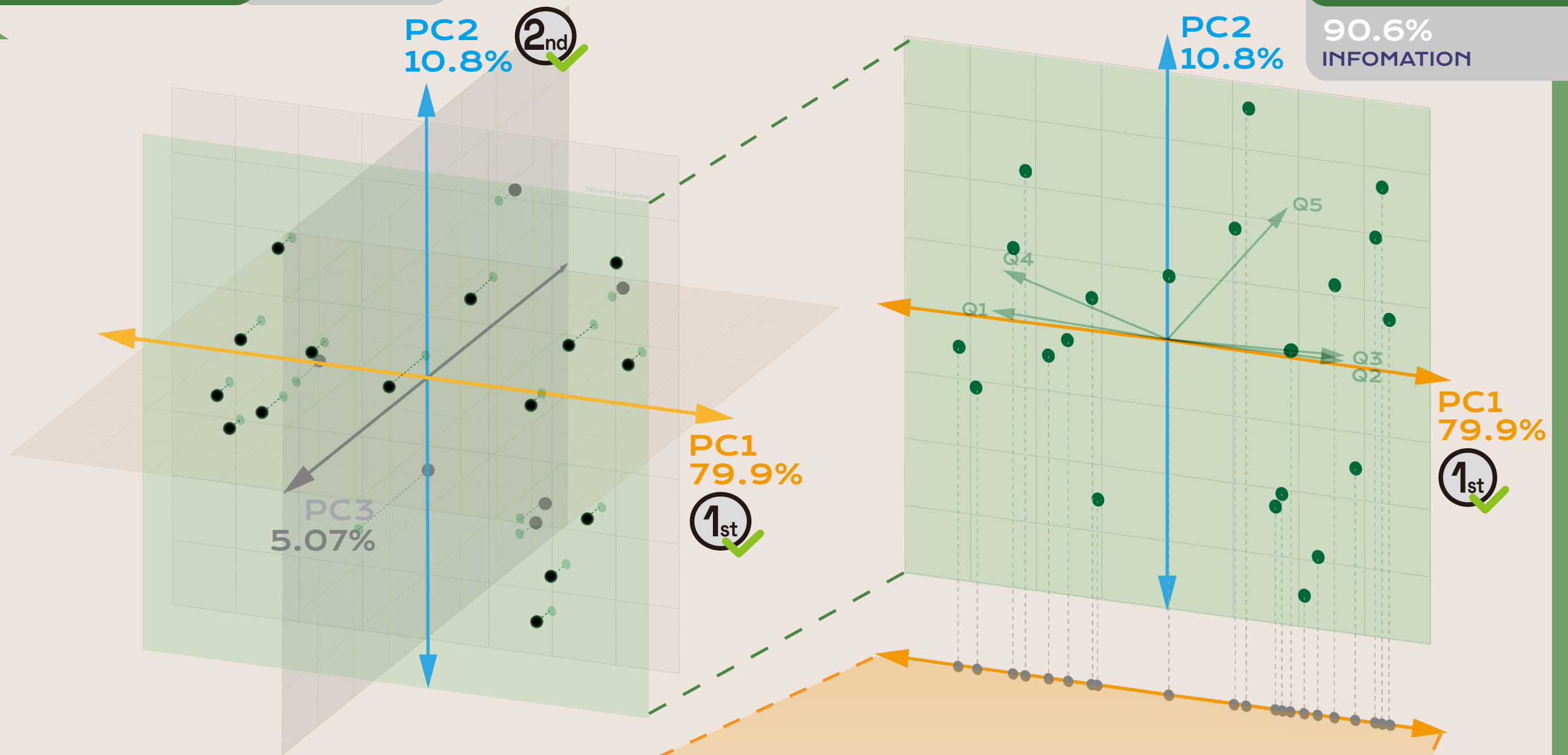
4 Dimensions
98.4% INFOMATION

PC1 79.9% ✓
PC2 10.8% ✓
PC3 5.07% ✓
PC4 2.66%

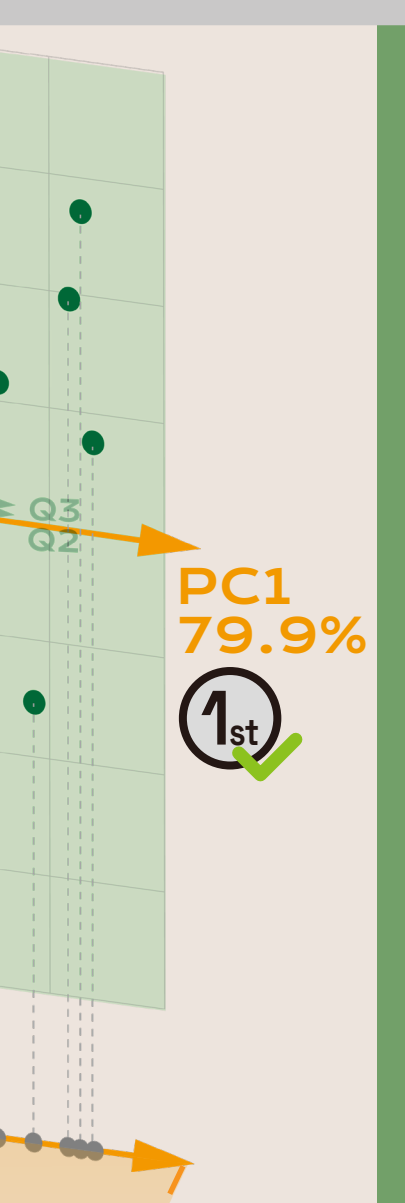After selecting the three types of PCs, the data came to 3 dimensions, now we can start to see the data

3 Dimensions
95.7% INFOMATION

PC2 10.8% 2nd ✓
PC1 79.9% 1st ✓
PC3 5.07%

2 Dimensions
90.6% INFOMATION

PC2 10.8%
PC1 79.9% 1st ✓

Q5 Q4 Q1 Q3 Q2

Ow, we can be familiar with using PCA as a statistical technique to define the perfect partners for sharing the weekends!

Also, if you want to portray the characteristics of a large number of people through data reduction, the PCA algorithm is essential to this purpose.

Wish you have a lovely weekend~
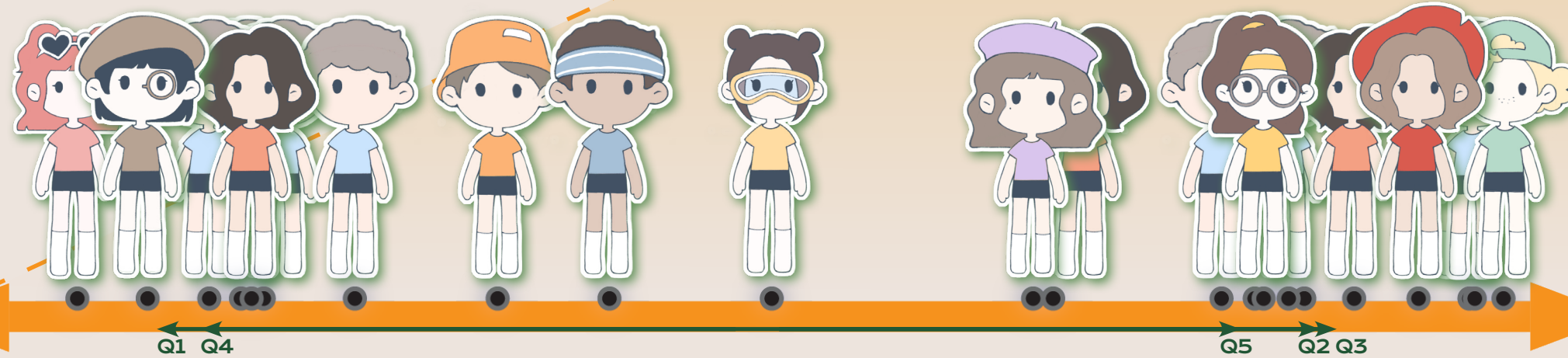
1 Dimension
79.9% INFORMATION

In the end only the information contained in PC1 was retained.

CITY Quiet Culture
Q1 Q4 ... Q5 Q2 Q3
NATURE ENERGETIC SPORTS

VISUAL EXPLANATIONS OF STATISTICAL METHODS

Principal Component Analysis

AUTHORS
Sun Litong
Lin Ran
Li Manman
Li Ye
Zhong Yiheng
Gao Yijia
Spapperi Gestri Lucrezia