

# Practice Example: Grouping Students by Study Habits

A Simple Clustering Exercise

## The Problem

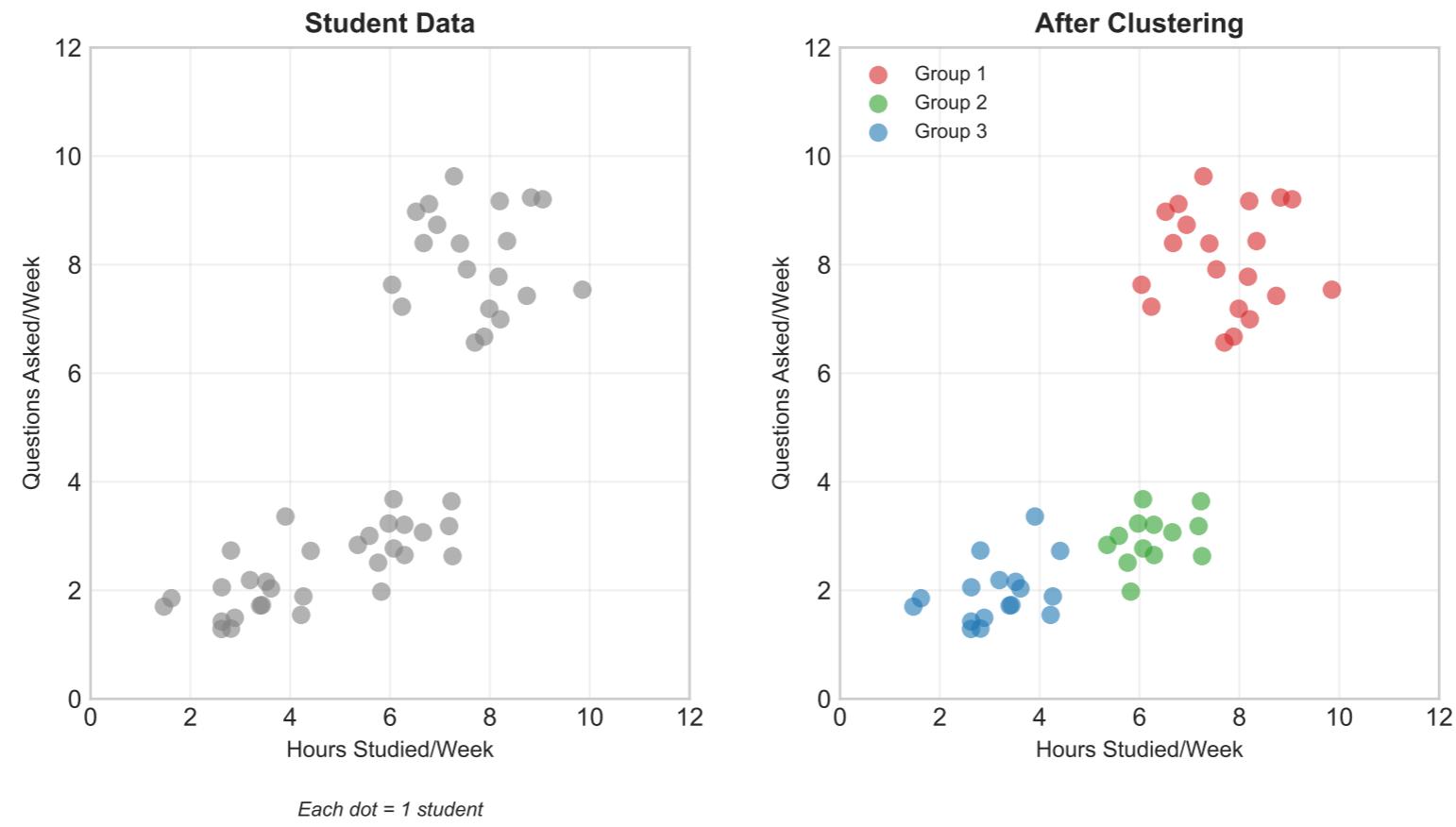
A teacher wants to understand different study patterns in class.

### Data collected:

- Hours studied per week
- Number of questions asked

50 students total

### Goal: Find study groups



## Your Turn: Step 1

Load the data:

```
import pandas as pd
data = pd.read_csv("students.csv")
```

Look at it:

```
print(data.head())
```

You should see:

	hours	questions
0	3.2	2.1
1	8.5	7.9
...		

Prepare the data:

```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
X_scaled = scaler.fit_transform(data)
```

Why scale?

Makes features comparable

Apply clustering:

```
from sklearn.cluster import KMeans
kmeans = KMeans(n_clusters=3)
labels = kmeans.fit_predict(X_scaled)
```

## Your Turn: Step 2

See the results:

```
import matplotlib.pyplot as plt
plt.scatter(data["hours"],
            data["questions"],
            c=labels)
plt.xlabel("Hours")
plt.ylabel("Questions")
plt.show()
```

Count groups:

```
for i in range(3):
    count = (labels == i).sum()
    print(f"Group {i}: {count} students")
```

## Your Turn: Step 3

## What We Found

### Group 1:

- Need extra help
- Low hours, few questions

### Group 2:

- Independent learners
- Good hours, few questions

### Group 3:

- Highly engaged
- Many hours, many questions

## Check Your Work

- Did you find 3 groups?
- Are groups visually separated?
- Do groups make sense?

### What to look for:

- Clear differences between groups
- Similar students in same group
- Groups tell a story

### Next: Try with K=2 or K=4

Which is better? Why?