# Discovering Patterns
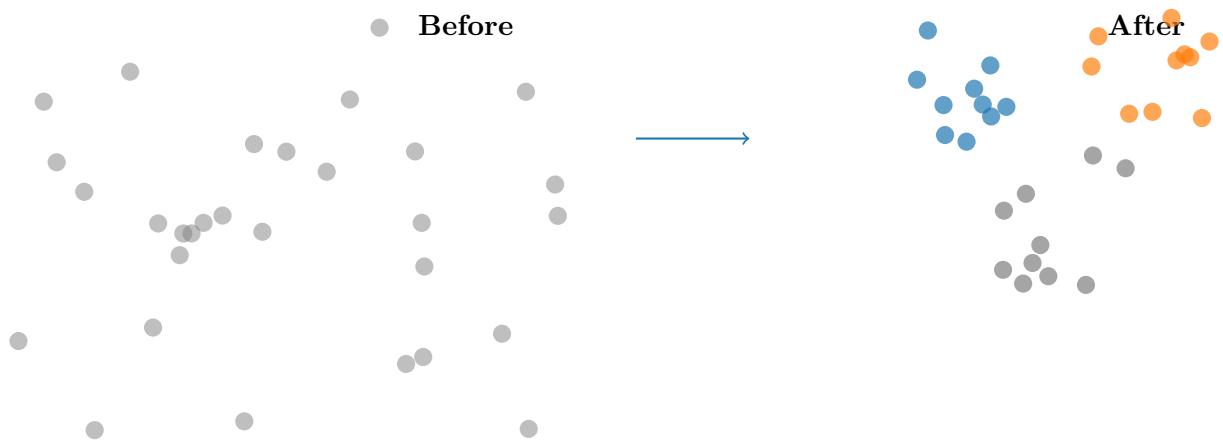
What you'll discover before we begin
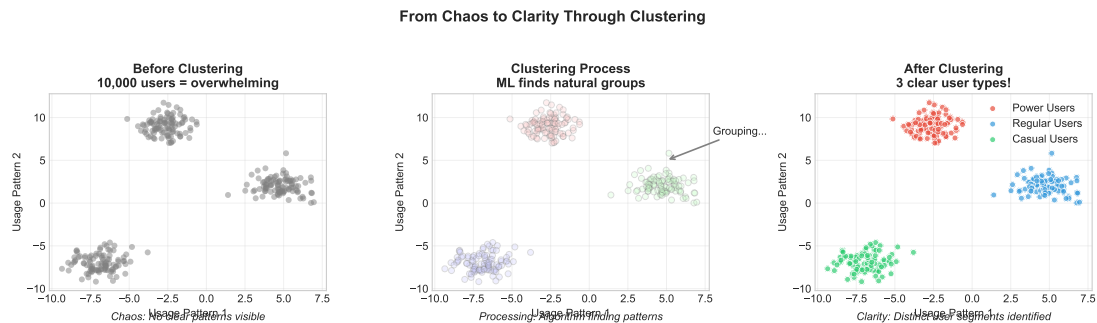


**Question:** How do computers find groups in data?

Let's discover the answer together...

## Discovery 1: Do You See the Groups?

### From messy to organized

**From Chaos to Clarity Through Clustering**



**Before Clustering**
**10,000 users = overwhelming**
*Chaos: No clear patterns visible*

**Clustering Process**
**ML finds natural groups**
*Processing: Algorithm finding patterns*

**After Clustering**
**3 clear user types!**
*Clarity: Distinct user segments identified*

- Power Users
- Regular Users
- Casual Users

**Look carefully and answer:**

1. Look at the LEFT side. Do the dots seem random or organized?

   _____

2. Look at the RIGHT side. What changed?

   _____

3. Count how many separate groups you can see on the right:
   Circle one:    2    3    4    5    More than 5

4. If you had to sort 1000 dots like this by hand, how long would it take you?

   _____

---

**Think about this:**
The computer found these groups in less than 1 second.
How did it know which dots belong together?

## Discovery 2: Things That Belong Together

### Look at these dots

**Clustering: Finding Groups in Data (No Math Required!)**
*A Visual Introduction for Beginners*

| | | |
|---|---|---|
| **What is Clustering?** | **After Clustering** | **Example: Student Projects** |

Before: Mixed Data — Random points - no groups

After: Organized Groups — Points grouped by similarity

Tech Projects *(Apps, Websites)*

Art Projects *(Paintings, Music)*

Science Projects *(Lab work, Research)*

**How Computers See Data**

| Item | Feature 1 | Feature 2 |
|------|-----------|-----------|
| A | 3.2 | 1.5 |
| B | 3.1 | 1.6 |
| C | 8.5 | 7.2 |
| D | 8.7 | 7.1 |

Similar numbers = Same group

**Distance = Similarity**

Point B

Distance

Point A — Close = Similar

**Clustering Process**

1. Start with data points
2. Measure distances
3. Find close neighbors
4. Form groups
5. Check if good groups
6. Done!

Key Idea: Clustering finds natural groups by looking at which items are similar (close together)
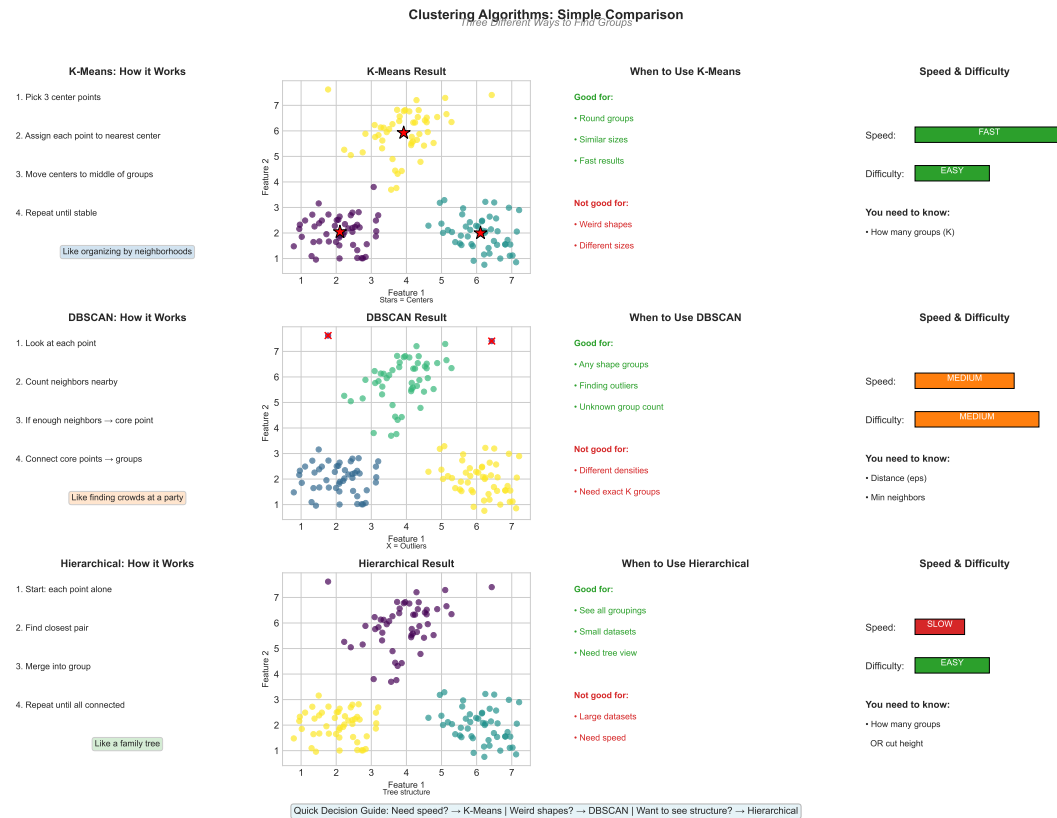
**Your turn to group them:**

1. Draw circles around dots that seem to belong together (use a pencil)

2. What rule did you use to decide? (Check all that apply)

   ☐ Dots that are close together

   ☐ Dots in the same area

   ☐ Dots that form a shape

   ☐ Other: _____

3. Would everyone group them the same way you did?  Yes ☐  No ☐

4. What if you could only see horizontal position (left-right)? Would groups change?

   _____

**Discovery moment:**
There's no single "correct" way to group things. It depends on what you're looking for!

## Discovery 3: Same Dots, Different Stories

### Three different ways to group the same dots

**Clustering Algorithms: Simple Comparison**
*Three Different Ways to Find Groups*

**K-Means: How it Works**

1. Pick 3 center points

2. Assign each point to nearest center

3. Move centers to middle of groups

4. Repeat until stable

Like organizing by neighborhoods

**K-Means Result**

Feature 2 / Feature 1
Stars = Centers

**When to Use K-Means**

Good for:
• Round groups
• Similar sizes
• Fast results

Not good for:
• Weird shapes
• Different sizes

**Speed & Difficulty**

Speed: FAST

Difficulty: EASY

You need to know:
• How many groups (K)

**DBSCAN: How it Works**

1. Look at each point

2. Count neighbors nearby

3. If enough neighbors → core point

4. Connect core points → groups

Like finding crowds at a party

**DBSCAN Result**

Feature 2 / Feature 1
X = Outliers

**When to Use DBSCAN**

Good for:
• Any shape groups
• Finding outliers
• Unknown group count

Not good for:
• Different densities
• Need exact K groups

**Speed & Difficulty**

Speed: MEDIUM

Difficulty: MEDIUM

You need to know:
• Distance (eps)
• Min neighbors

**Hierarchical: How it Works**

1. Start: each point alone

2. Find closest pair

3. Merge into group

4. Repeat until all connected

Like a family tree

**Hierarchical Result**

Feature 2 / Feature 1
Tree structure

**When to Use Hierarchical**

Good for:
• See all groupings
• Small datasets
• Need tree view

Not good for:
• Large datasets
• Need speed

**Speed & Difficulty**

Speed: SLOW

Difficulty: EASY

You need to know:
• How many groups
OR cut height

Quick Decision Guide: Need speed? → K-Means | Weird shapes? → DBSCAN | Want to see structure? → Hierarchical

**Compare the three approaches:**

1. Which grouping makes most sense to you? Circle one:    A    B    C
Why? _____

2. Count the groups in each:

Approach A has ___ groups

Approach B has ___ groups

Approach C has ___ groups

3. Think of a real situation where each grouping would be useful:
**A would be good for:** _____
**B would be good for:** _____
**C would be good for:** _____

## What You Noticed

### Your discoveries

**1. The Challenge**
Why is it hard to group things when there are many of them?
_____
_____

**2. The Pattern**
What happened when dots were organized into groups? How did it help you understand the data?
_____
_____

**3. The Surprise**
What surprised you about the different grouping methods?
_____
_____

**4. The Question**
If a computer could instantly group things for you, what would you want it to organize?
_____
_____

**Ready to learn more?**
In our first class, you'll discover how these simple ideas help companies understand their customers, products, and innovations.