

Explanatory Notes for 6.390

Shauntclair Ruiz (Current TA)

Fall 2022

Hierarchical Clustering

That last example reveals something: not all types of groups are the same! Some are much broader than others, for example.

If two types of groups are different, then why do we only have to have one type in our clustering? We don't have to restrict ourselves to a single k .

Instead, we could treat some groups as inside of other groups: we call this a *hierarchy*, because some groups are "higher" on the scale.

Definition 1

Hierarchical Clustering is when we cluster at multiple different **levels**.

Some groups are **high-level**, or **coarse**: they are groupings that contains **more elements**: items in the same group can be more **different**.

Some groups are **low-level**, or **fine**: they are groupings that contain **fewer elements**: items in the same group have to be very **similar**.

Example: Categorizing living things is done using hierarchical clustering: some groupings **contain** other groupings.



The top row of clusters are more **coarse**, while the bottom row is more **fine**.

We can **split** our groupings into **smaller** and smaller ones, to be as **fine-grained** as we need. Useful!