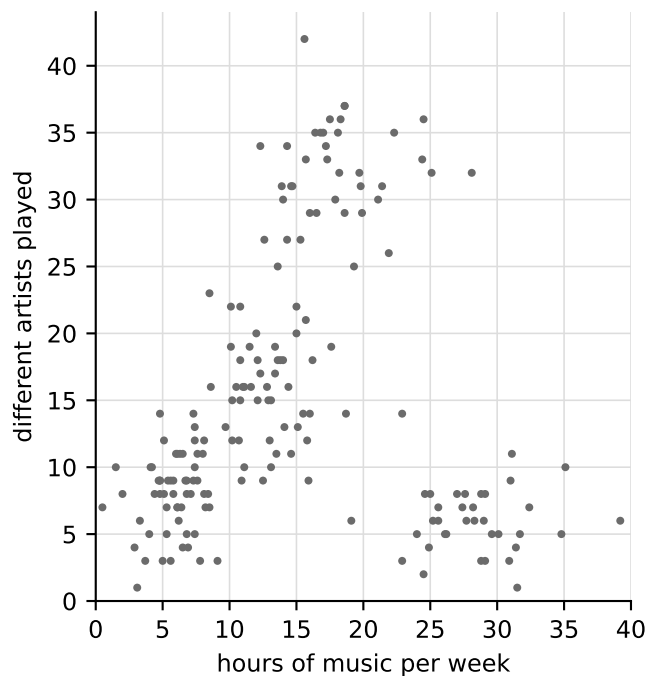


DATA 202 — Clustering · Week 5 Monday

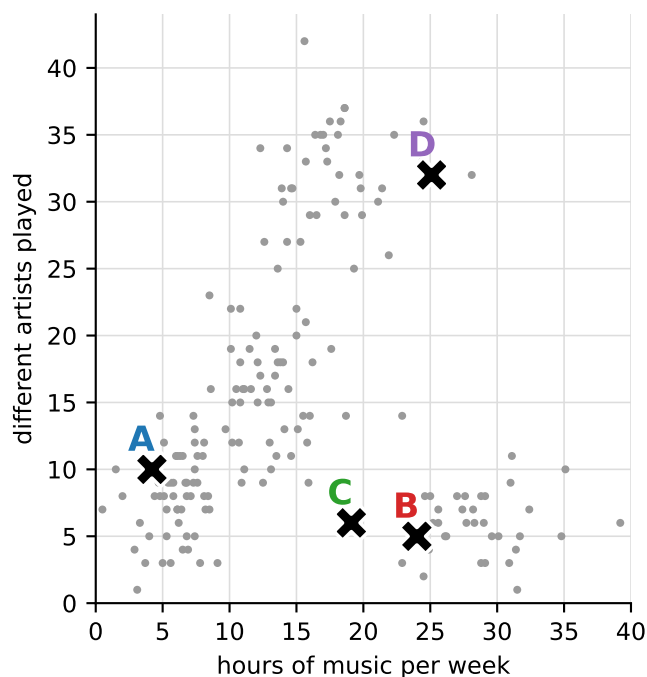
Draw right on the plots. Every plot shows the **same 180 listeners** as today's slides (one dot per listener). Your professor will tell you when to do each part.

1 · How many groups?



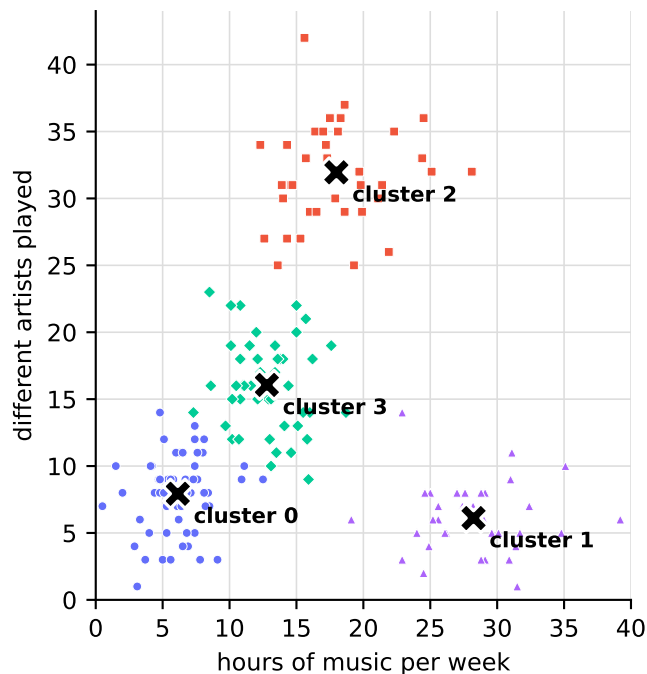
1. **Circle** every group of listeners you see.
2. How many groups did you circle? _____
3. Compare with a neighbor: where do your circles disagree?

2 · k-means by hand: one round



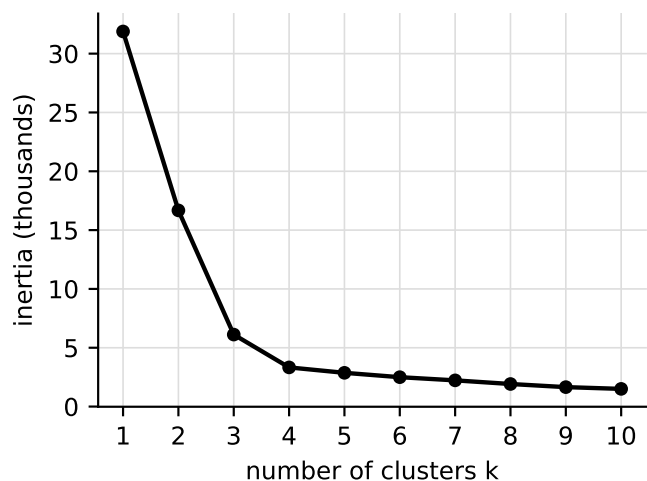
1. The four X's (**A–D**) are the **starting centroids**: listeners 38, 71, 129 and 158.
2. **Assign.** Draw lines that split the plot so every listener goes with its **nearest X**. *Tip: a boundary runs halfway between two X's.*
3. **Update.** In each of your groups, draw a **new X** at the middle of its points.
4. Compare with the *Round 1* panels on screen. Which X moved the most? Which listeners changed group?

3 · Name the groups



1. This is where k-means ($k = 4$) ended up. The cluster numbers mean nothing.
2. Next to each group, write a **short name** for the kind of listener it holds.
3. Put a ★ next to one listener who could belong to **two** groups.
4. Compare names with a neighbor. Same names? Why did you choose yours?

4 · How many groups?



1. This is the inertia for $k = 1$ to 10. **Circle the elbow.**
2. From the screen: elbow says $k = \underline{\hspace{2cm}}$ and silhouette says $k = \underline{\hspace{2cm}}$.
3. I'd choose $k = \underline{\hspace{2cm}}$ because
