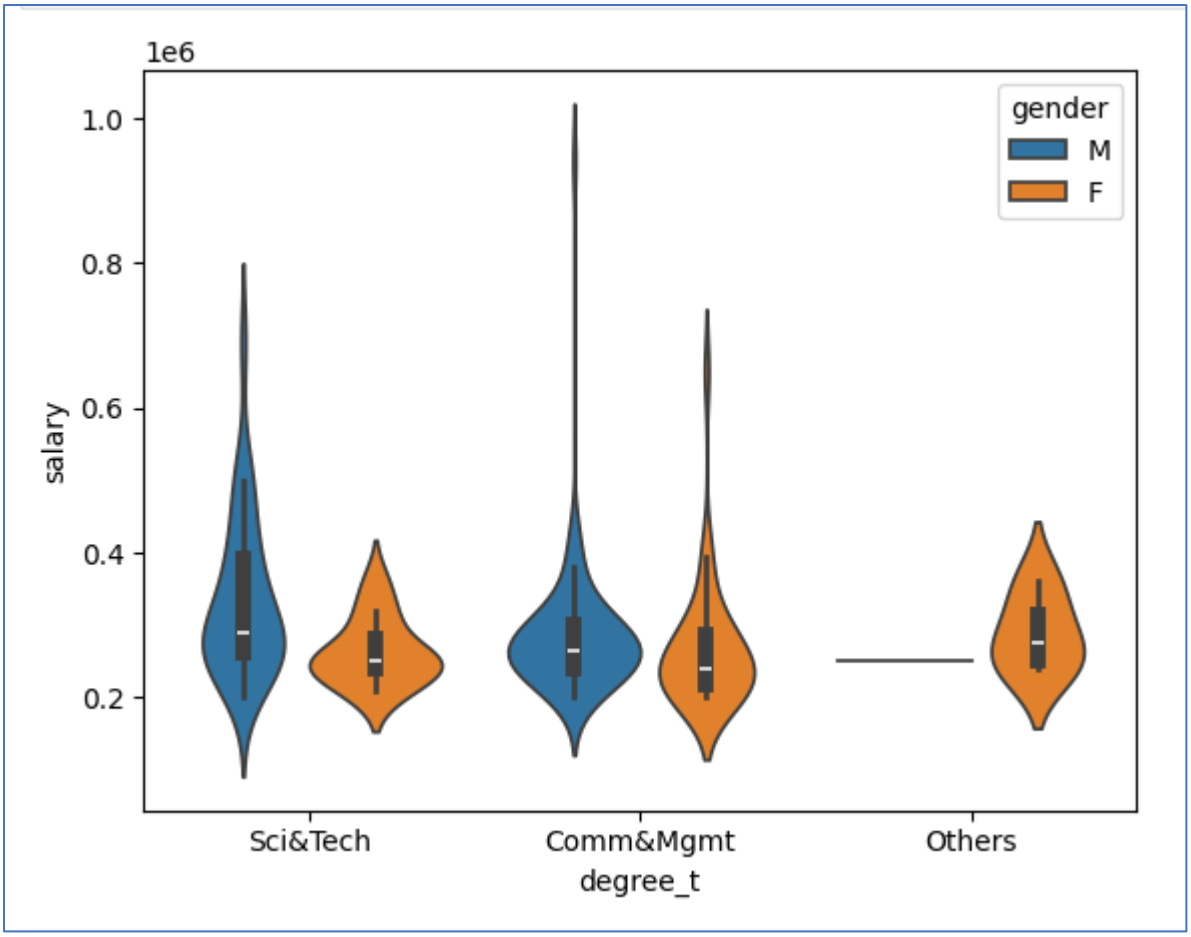
Seaborn-ASG

What is swarmplot?

- swarmplot is used to show the **distribution of data points** along a categorical axis.
- It is similar to stripplot (which just plots points), **but smarter**:
  - It **adjusts point positions** so they **don't overlap**.
  - Points are arranged like a **swarm of bees** 🐝 → hence the name.
- Use *stripplot* **with** *jitter=True* for big datasets (faster).
- Use swarmplot for smaller datasets when you want a clean, non-overlapping visualization.



1. Spread of Salaries (Range)

- In **Sci&Tech** and **Comm&Mgmt**, salaries go very high (close to 1 million).
- For **Others**, salaries are low and tightly clustered (no wide spread).

2. Median Salary (white dot)

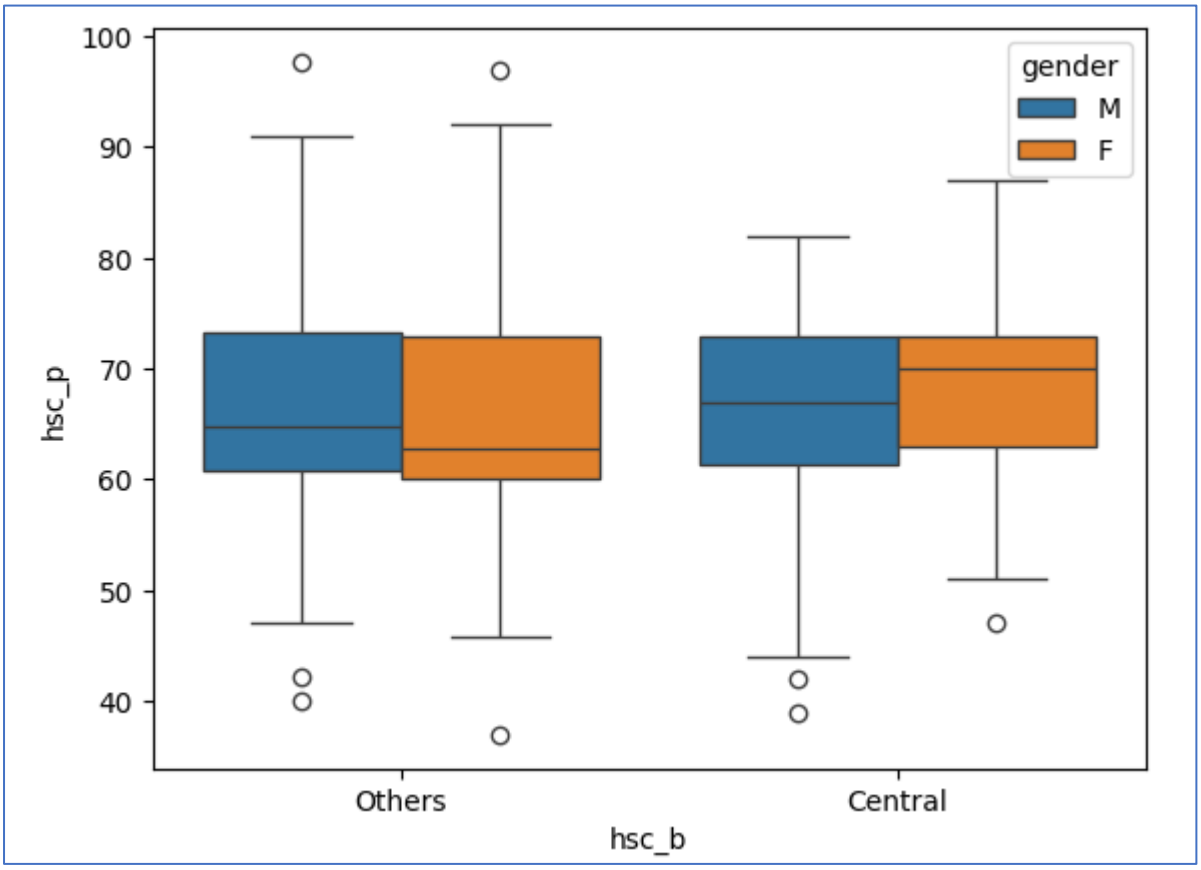
- Males (blue) in **Sci&Tech** and **Comm&Mgmt** have higher median salaries than females.
- In **Others**, male and female salaries are nearly the same (flat distribution).

3. **Distribution Shape**

- The violins are **fatter at the bottom** (many people earn in the lower salary range).
- The top of the violins is **narrower** → fewer people earn very high salaries.
- This shows a **right-skewed distribution** (most salaries are low–medium, few are very high).

4. **Variation by Degree**

- **Sci&Tech** and **Comm&Mgmt** degrees have **more salary opportunities** (wider distribution).
- **Others** degree group has **limited salary growth** (narrow and short violin).



1. **Median Performance**

- For both *Others* and *Central* boards, **females have slightly higher median scores** compared to males.
- Male and female medians are not very far apart (both around 60–70%).

2. **Spread (Interquartile Range)**

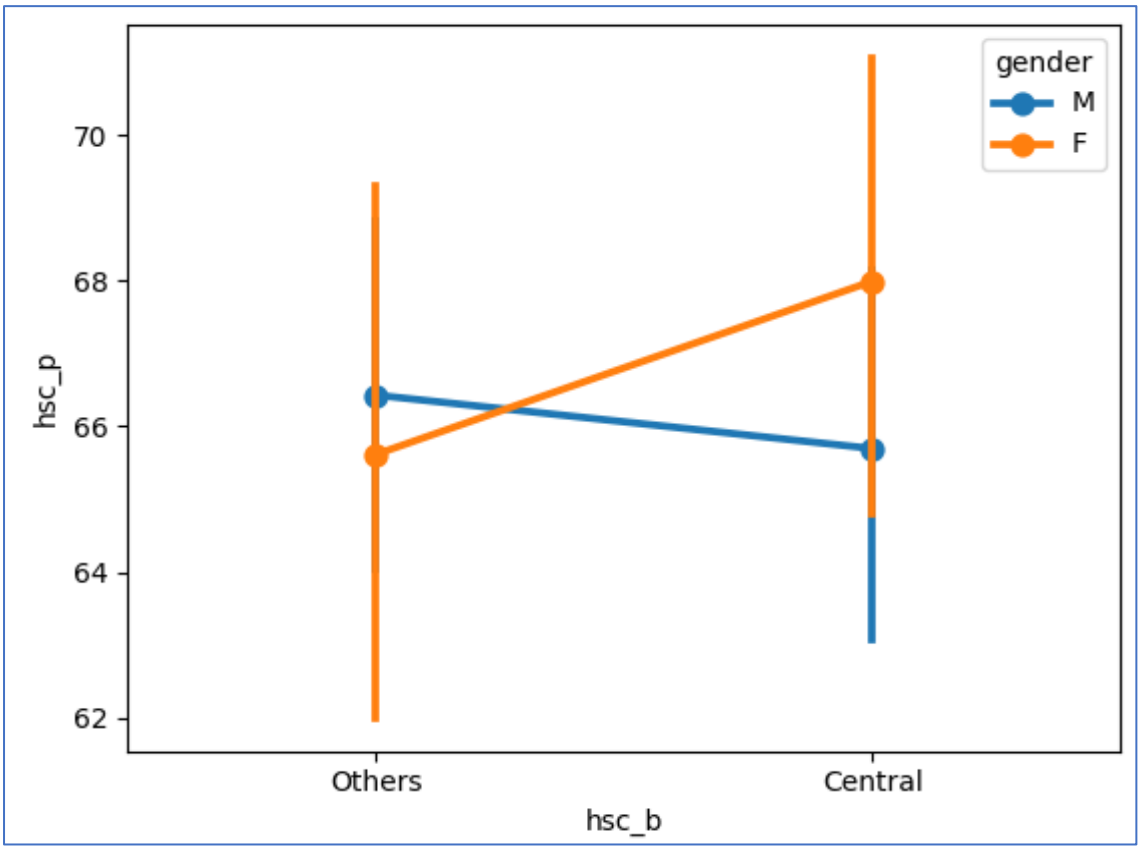
- Both genders across boards have a wide spread of scores (~45 to ~90).
- Central board students (both genders) show **slightly higher upper quartiles** compared to Others.

4. **Outliers**

- A few outliers are visible at the **lower end (~40%)** and **upper end (~100%)**, more in the "Others" board group.

5. **Comparison Between Boards**

- Overall performance distribution of *Others* vs *Central* is quite similar.
- Central board females seem to have a slight edge (higher concentration in the 70+ range).

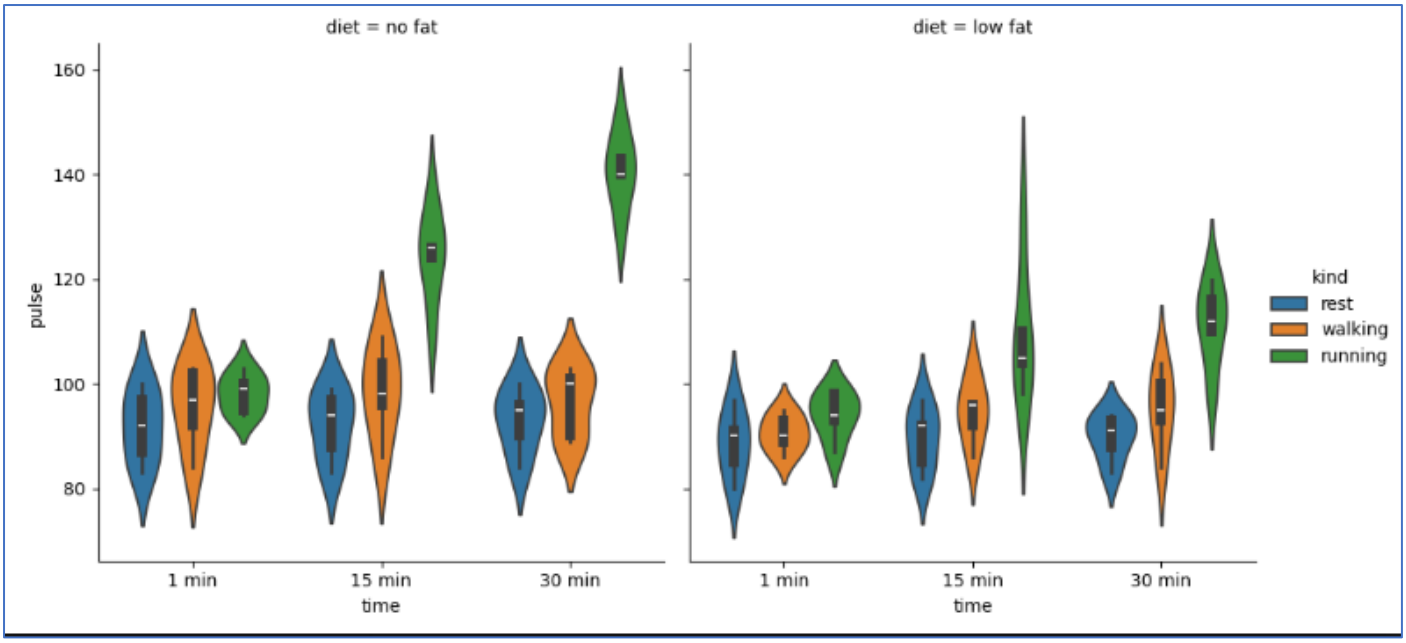


- **X-axis (hsc_b)** → Higher secondary school board (Others / Central).
- **Y-axis (hsc_p)** → Average percentage score.
- **Hue (gender)** → Male (blue) vs Female (orange).

🔎 **Interpretation**:

- For **"Others" board**:
  - Males have slightly higher average % than females.
- For **"Central" board**:
  - Females score higher than males.
- The vertical lines = **confidence intervals** (they show uncertainty or spread in the data).

So, females seem to do better in Central boards, while males do slightly better in Others boards.



What this violin plot shows

- **X-axis (time)** → 1 min, 15 min, 30 min after exercise.
- **Y-axis (pulse)** → Heart rate (pulse).
- **Colors (kind)** → Rest (blue), Walking (orange), Running (green).
- **Columns (diet)** → "no fat" diet (left) vs "low fat" diet (right).

The violin shape = shows the **distribution of pulse values** (wide parts = many people, narrow parts = fewer people).
Inside each violin you also see a **small boxplot** (median + spread).

Simple summary:

1. **Rest (blue):** Pulse is low and stable for both diets.
2. **Walking (orange):** Pulse increases a bit, but stays moderate.
3. **Running (green):** Pulse is much higher, especially after 30 min.
4. **Diet effect:**
    - People on **no fat diet** (left) tend to have slightly higher pulses during running compared to **low fat diet** (right).

👉 So, **exercise type affects pulse more strongly than diet**.

Plot Type	What it Shows	Best For	Example Question it Answers
Swarm plot	Each individual data point (like dots), spread out so they don't overlap.	Seeing **all raw data points** with categories.	"What are the actual pulse values of each person during rest, walking, and running?"

Plot Type	What it Shows	Best For	Example Question it Answers
Box plot	Summary of data: min, Q1, median, Q3, max (+ outliers).	Quick **summary of distribution** and detecting outliers.	"What is the typical range of student marks, and are there any outliers?"
Point plot	Average value for each category with error bars.	Comparing **means (averages)** across categories.	"Do boys or girls have higher average HSC % in different boards?"
Violin plot	Full distribution (like smoothed density) + boxplot inside.	Comparing **distribution shapes** between groups.	"How does pulse distribution change after exercise for different diets?"