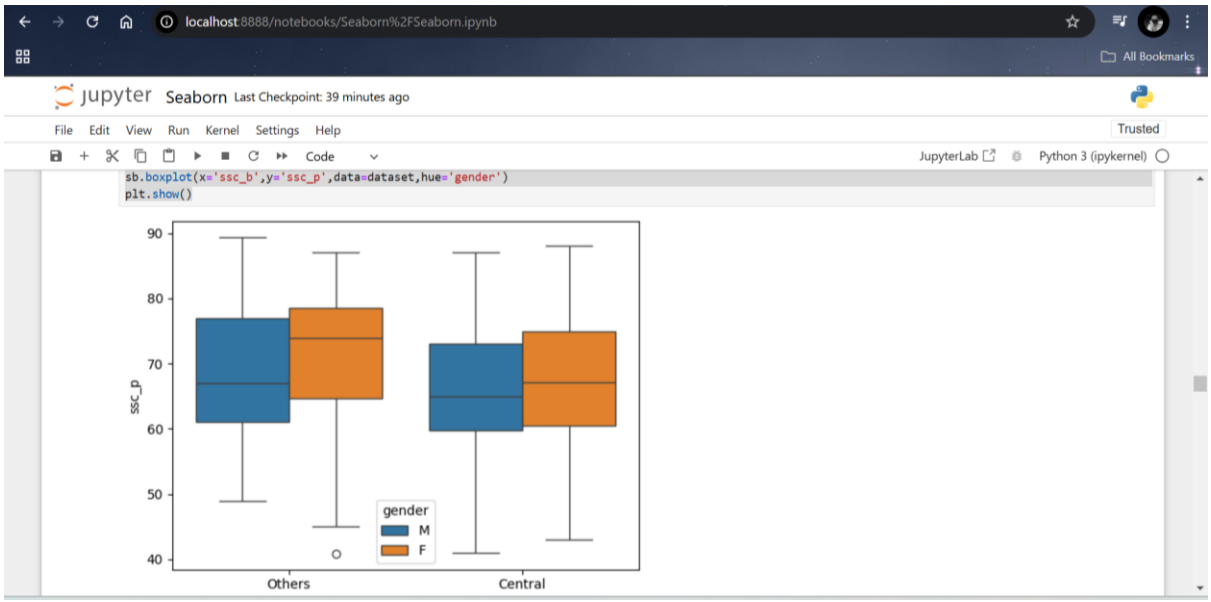
S. RAYEN SUBHIKSHA

# DATA VISUALISATION

## Seaborn:

### 1. Box Plot

The box plot in the Seaborn library is a method for visualizing the distribution of data across categories using quartiles. It provides a visual summary of the minimum, first quartile (Q1), median (Q2), third quartile (Q3), and maximum, and can highlight outliers.




It creates a box plot using the Seaborn library to analyze the relationship between students' secondary school board (ssc_b) and their percentage scores (ssc_p), further categorized by gender.

For each school board (ssc_b):

- Two box plots (one for each gender) are displayed side by side, showing the distribution of ssc_p for that board.
- The boxes span from the first quartile (Q1) to the third quartile (Q3).
- A line inside the box indicates the median score.
- Whiskers extend to data points within 1.5×IQR (Interquartile Range).
- Outliers, if any, appear as individual points outside the whiskers.

In Central board,

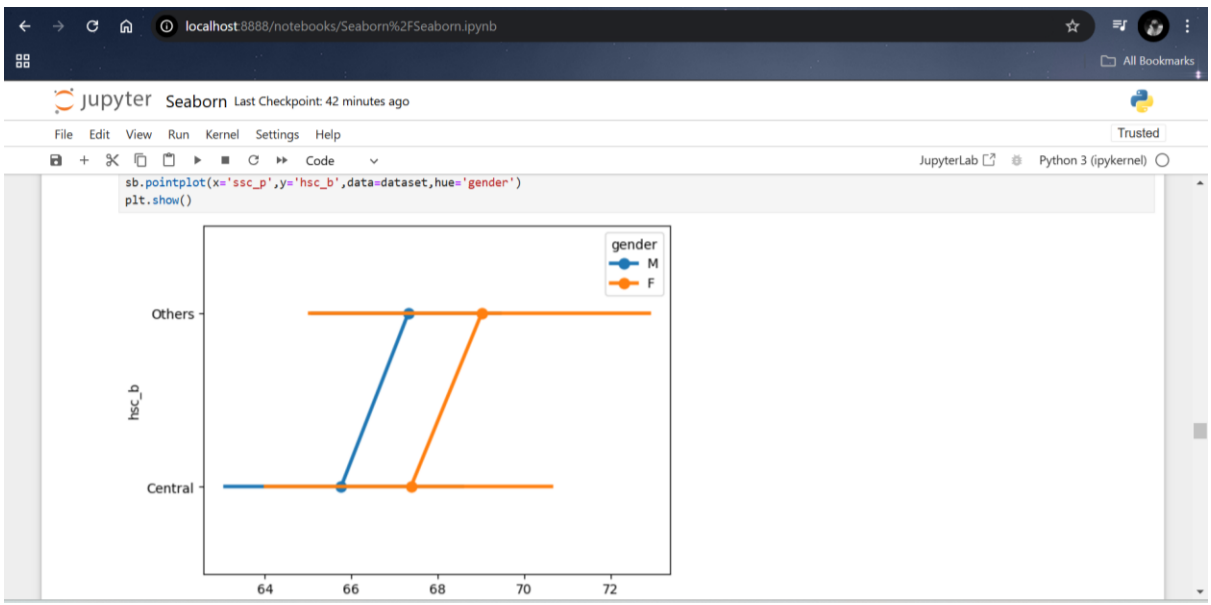
- Lowest score (40) in the entire class is scored by male students.
- Highest score (90) in the entire class is scored by female students.
- 60 marks is the median score scored by male students.
- 70 marks is the median score scored by female students.
- The overall performance of female students is better than male students in the class.

In other boards,

- Lowest mark (45) in the entire class is scored by female students.
- Highest mark (90) in the entire class is scored by male students.
- 65 marks is the median score of male students.
- 75 marks is the median score of female students.
- The overall performance of male students is better than female students in the class

## 2. Point Plot

The point plot in the Seaborn library is used to represent statistical estimates (like mean) and their confidence intervals as points with optional error bars. It is particularly useful for visualizing categorical data and making comparisons across different groups or conditions. One axis represents the categorical variable, and the other shows the numerical summary.



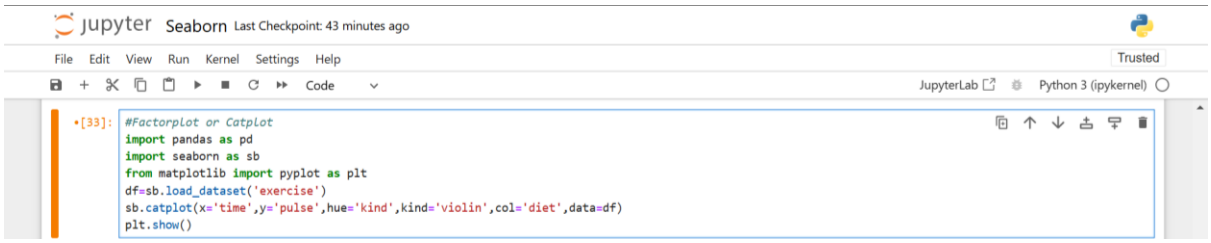
This code snippet creates a point plot using the Seaborn library to visualize the relationship between secondary school percentages (ssc_p) and higher secondary school boards (hsc_b), further categorized by gender.

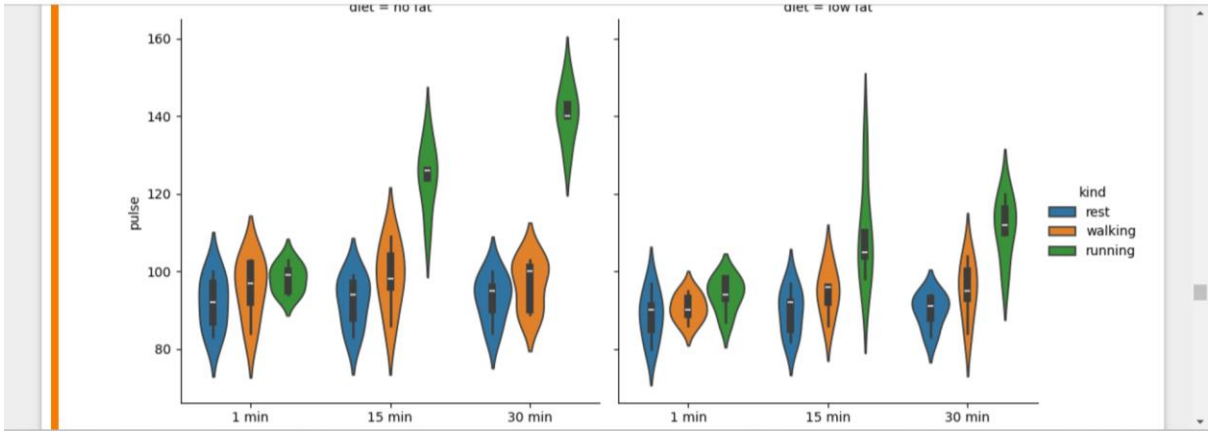
For each combination of hsc_b (higher secondary board) and gender, Seaborn calculates the mean of ssc_p and plots it as a point.

**Lines**: Points for different genders are connected by lines to indicate trends or comparisons across groups.

## 3. Factor plot/Catplot in latest version of Seaborn library

The Catplot function in the Seaborn library is a powerful tool for creating categorical plots. It is especially useful for creating faceted plots, which split data into subplots based on additional variables.




It creates a faceted violin plot using the Seaborn library to visualize the distribution of pulse rates across different time points, exercise types, and diets.

The exercise dataset contains information about pulse rates (pulse) measured under different conditions:

  - **time**: The time of exercise session (1 min, 15 min, etc.).
  - **pulse**: The pulse rate.
  - **kind**: The type of exercise (e.g., rest, walking, running).
  - **diet**: The type of diet (e.g., low fat, no fat).

- **x='time'**: The x-axis represents the time of the exercise session.
- **y='pulse'**: The y-axis represents the pulse rate.
- **hue='kind'**: Adds a grouping based on the type of exercise (rest, walking, running), with each group differentiated by color.
- **kind='violin'**: Specifies the plot type as a violin plot, which shows the distribution of pulse rates using a kernel density estimation.
- **col='diet'**: Creates a **facet** for each type of diet (e.g., low fat, no fat). Each diet gets its own subplot.
- **data=df**: Passes the loaded dataset.

**Low Fat Diet**:

- For 1 min of exercise, pulse rates are distributed differently for rest, walking, and running.
- The pulse rate of rest for the time intervals of 1 min, 15 mins and 30 mins shows similar violin plots i.e) distribution of data.
- Walking of 30 mins had done by many as it shows wider violin plot compared to other 2 time periods.
- Running of above 30 mins shows the highest peak pulse rate.

**No Fat Diet**:

- The violin plots for rest, walking and running in no fat diet shows less pulse rate compared to low fat.
- The time of resting is higher in all the 3 times like 1 min, 15 mins and 30 mins respectively.
- Time of 1 min running shows higher pulse rate when compared to 15 mins and 30 mins of running.
- Similarly, time of 1 min walking shows higher pulse rate when compared to 15 mins and 30 mins of walking.
- During the time interval between 15 mins and 30 mins, Pulse rate of running shows its peak.