Seaborn Basics – Histograms, Boxplots, Scatterplots

We'll use **histograms, boxplots, and scatterplots** to explore the sleep dataset.

## 🔍 Difference Between Matplotlib & Seaborn

Feature	Matplotlib	Seaborn
Level	Low-level → gives you full control of every element	High-level → built on Matplotlib, easier defaults
Syntax	More code needed for styling	Cleaner, simpler syntax
Default Style	Basic, sometimes plain-looking	Modern, attractive styles out-of-the-box
Best For	Custom, complex plots where you need full control	Quick, good-looking plots for data exploration

1️⃣ Histogram – Sleep Duration Distribution
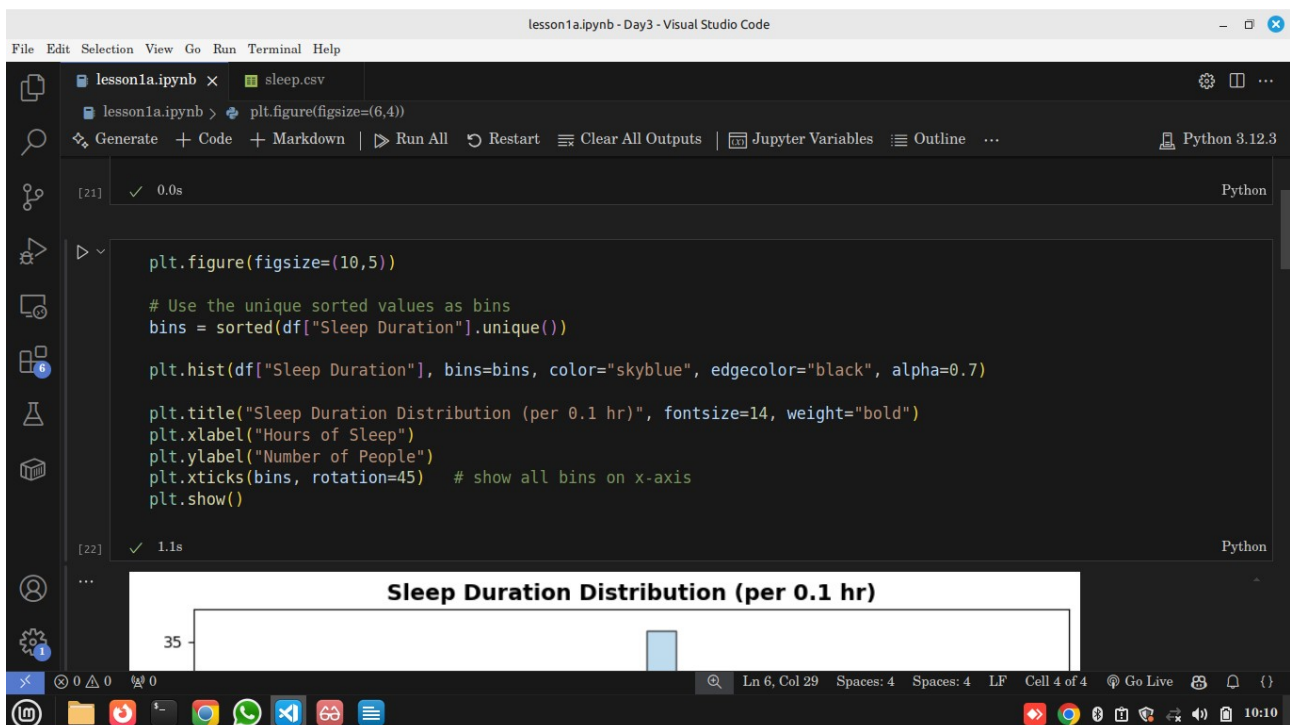
Output
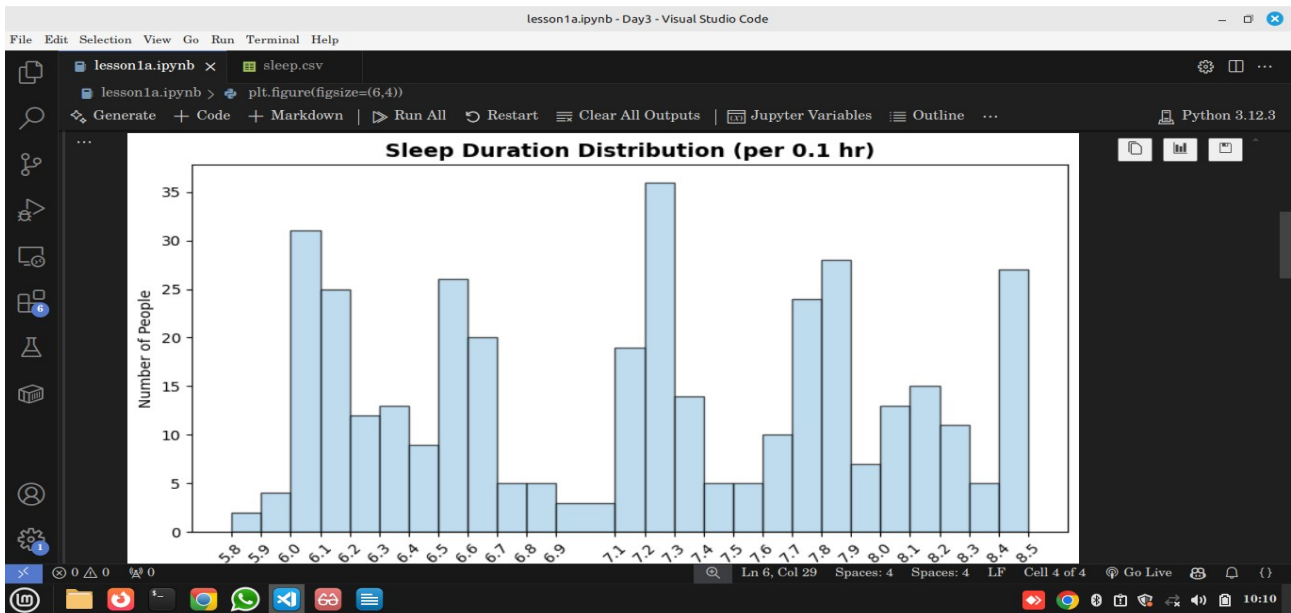
📖 **Explanation**

- `plt.hist(..., bins=bins)` → histogram grouped by **exact sleep duration values**.

- `xticks(bins)` → every sleep hour (6.1, 6.2 …) is shown on the x-axis.

- `alpha=0.7` → makes bars slightly transparent for clarity.

💡 **Insights**

- Most people sleep between **6.0 and 7.2 hours**.

- Few people are below **6 hours** → possible sleep-deprived cases.

- The data is not perfectly smooth, it has **spikes** at specific sleep values.
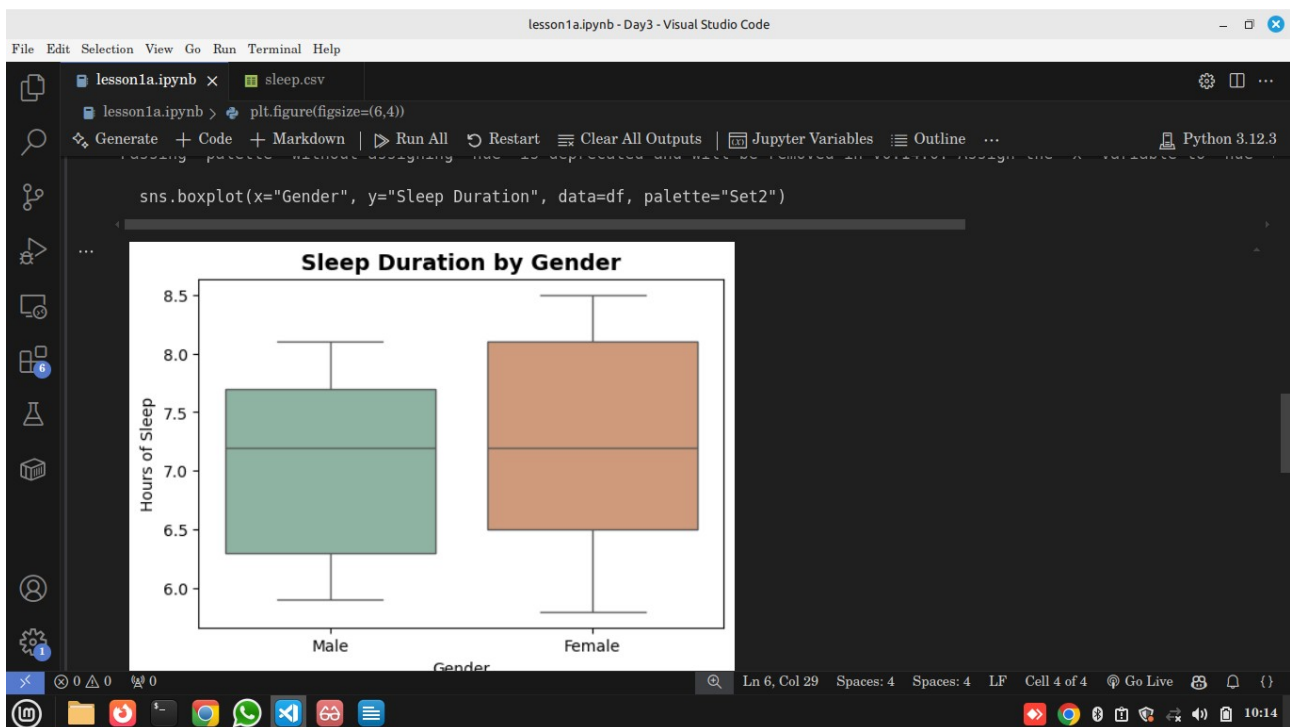
2️⃣ Boxplot – Sleep Duration by Gender

Practical


```python
plt.figure(figsize=(6,4))
sns.boxplot(x="Gender", y="Sleep Duration", data=df, palette="Set2")

plt.title("Sleep Duration by Gender", fontsize=14, weight="bold")
plt.xlabel("Gender")
plt.ylabel("Hours of Sleep")
plt.show()
```

Output

`How to read the boxplot:`

- **The box itself** → shows the middle 50% of the data (called the interquartile range, IQR).

Wha is Interquartile Range (IQR)

The **box** in a boxplot represents the **middle 50% of the data**.

To get it, we calculate **quartiles** (like slicing your data into 4 chunks):

Q1 (25th percentile) → 25% of the data falls below this value.

Q2 (50th percentile = median) → 50% of the data falls below this value.

Q3 (75th percentile) → 75% of the data falls below this value.

👉 The **IQR = Q3 – Q1** → this range covers the "middle bulk" of your data.

- **The line inside the box** → the **median** (middle value of sleep duration).
- **The whiskers (the vertical lines)** → show the range of typical values (not outliers).
- **Dots outside whiskers (if any)** → outliers (people who sleep way less or more).

📖 Explanation

- `sns.boxplot()` → shows **median, quartiles, and outliers**.
- Each box represents the **spread** of sleep hours for males and females.

💡 Insights

On The Sleep Duration by Gender chart

- For **Males**:
  - Bottom of the box ≈ **6.3 hrs** (Q1).
  - Top of the box ≈ **7.6 hrs** (Q3).
  - Middle line inside box ≈ **7.1 hrs** (median).
    → This means 50% of males sleep **between 6.3 and 7.6 hrs**.
- For **Females**:
  - Bottom of the box ≈ **6.1 hrs** (Q1).
  - Top of the box ≈ **8.1 hrs** (Q3).
  - Middle line inside box ≈ **7.2 hrs** (median).
    → This means 50% of females sleep **between 6.1 and 8.1 hrs**.

✨ Why it's useful

- The **box** ignores extreme values (outliers).

- It gives a **clear picture of spread and consistency**:

  - A **short box** = data is tightly clustered.

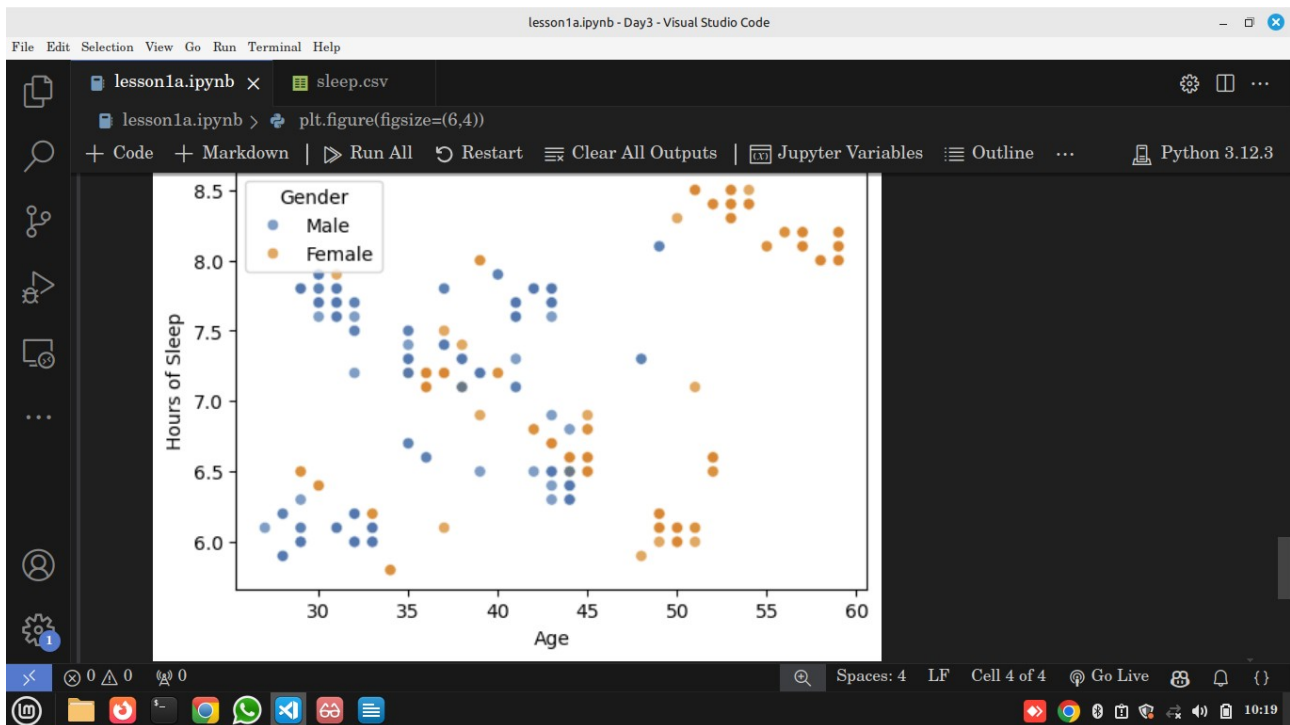 - A **tall box** = data is spread out (more variability).

So in this case:

- **Females have a taller box** → their sleep hours are more variable.

- **Males have a tighter box** → more consistent sleep hours.

-

3️⃣ Scatterplot – Age vs Sleep Duration

Practical
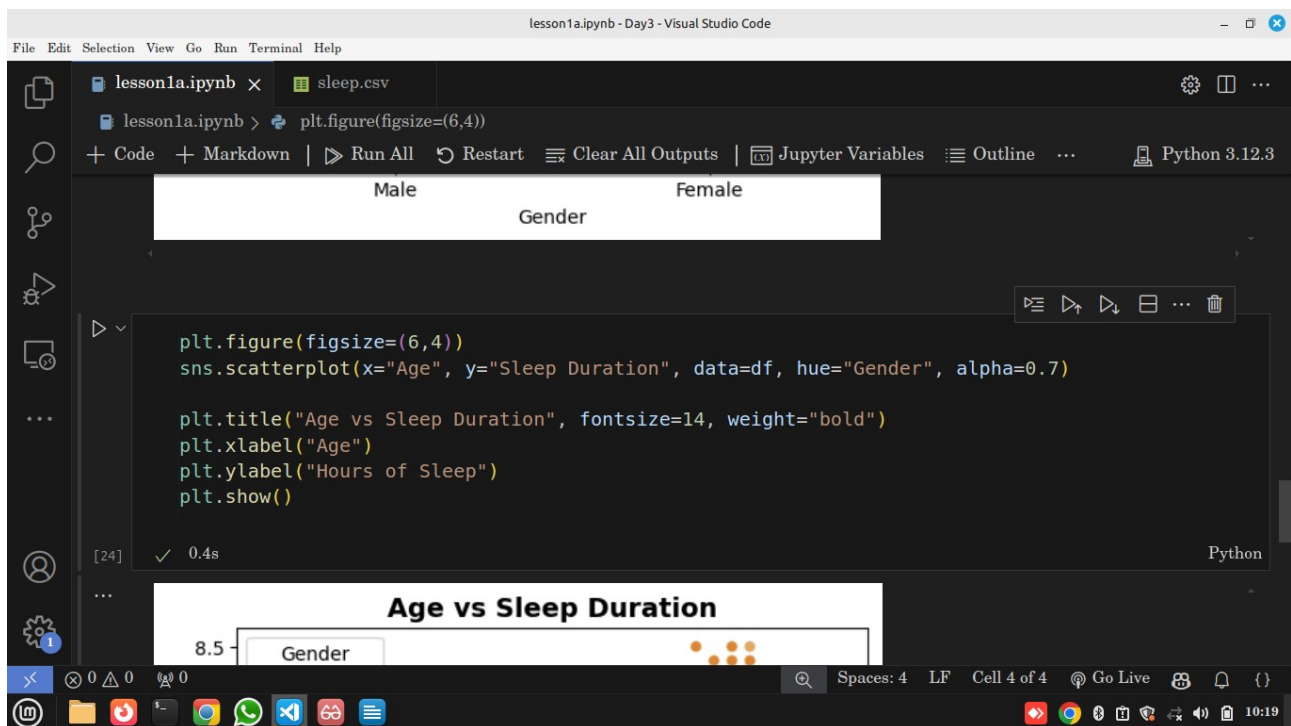
Output

📖 Explanation

- `sns.scatterplot()` → each point = one person.

- `hue="Gender"` → separates males and females with different colors.

💡 Insights

- Sleep duration does not vary much with age in this sample.

- Both males and females cluster around **6–8 hours**, showing consistent sleep behavior.