

Learn / Courses / Introduction to Data Visualiza...

← Course Outline →

Daily XP 150

Exercise

Omitting outliers

Now let's use the `student_data` dataset to compare the distribution of final grades ("G3") between students who have internet access at home and those who don't. To do this, we'll use the "internet" variable, which is a binary (yes/no) indicator of whether the student has internet access at home.

Since internet may be less accessible in rural areas, we'll add subgroups based on where the student lives. For this, we can use the "location" variable, which is an indicator of whether a student lives in an urban ("Urban") or rural ("Rural") location.

Seaborn has already been imported as `sns` and `matplotlib.pyplot` has been imported as `plt`. As a reminder, you can omit outliers in box plots by setting the `sym` parameter equal to an empty string ("").

Instructions

100 XP

- Use `sns.catplot()` to create a box plot with the `student_data` DataFrame, putting "internet" on the x-axis and "G3" on the y-axis.
- Add subgroups so each box plot is colored based on "location".
- Do not display the outliers.

Take Hint (-30 XP)

script.py

Light Mode

```
1 # Create a box plot with subgroups and omit the
2 outliers
3
4
5
6
7
8 # Show plot
9 plt.show()
```

↺ Loading... Loading...

IPython Shell

Slides

In [1]:

Final Document: Omitting Outliers in a Box Plot (Corrected Code)

Use `sns.catplot()` to create a box plot with the `student_data` DataFrame, putting 'internet' on the x-axis and 'G3' on the y-axis. Add subgroups so each box plot is colored based on 'location'. Do not display the outliers.

Correct Answer

To omit outliers in a box plot and add subgroups based on 'location', use `sns.catplot()` with 'kind' set to 'box' and set the 'sym' parameter to an empty string. Below is the working code:

```
import seaborn as sns
import matplotlib.pyplot as plt

# Create a box plot with subgroups and omit the outliers
sns.catplot(x='internet', y='G3',
            data=student_data,
            kind='box',
            hue='location',
            sym='')

# Show plot
plt.show()
```

Explanation

1. Import seaborn and matplotlib.pyplot for creating visualizations.
2. Use `sns.catplot()` to create a box plot with:
 - 'x' set to 'internet' to display whether students have internet access.
 - 'y' set to 'G3' to display final grades.
 - 'kind' set to 'box' to create a box plot.
 - 'hue' set to 'location' to color the box plots based on whether the location is Urban or Rural.
 - 'sym' set to an empty string (") to omit the outliers.
 - 'data' set to `student_data`, the DataFrame containing the data.
3. Use `plt.show()` to render and display the plot.