**Exercises**

**Creating subplots with col and row**

We've seen in prior exercises that students with more absences ("absences") tend to have lower final grades ("G3"). Does this relationship hold regardless of how much time students study each week?

To answer this, we'll look at the relationship between the number of absences that a student has in school and their final grade in the course, creating separate subplots based on each student's weekly study time ("study\_time").

Seaborn has been imported as sns and matplotlib.pyplot has been imported as plt.

**Instructions 3/3**

**30 XP**

Modify the code to use relplot() instead of scatterplot().

Modify the code to create one scatter plot for each level of the variable "study\_time", arranged in columns.

Adapt your code to create one scatter plot for each level of a student's weekly study time, this time arranged in rows.

**script.py**

# Change this scatter plot to arrange the plots in rows instead of columns

sns.relplot(x="absences", y="G3",

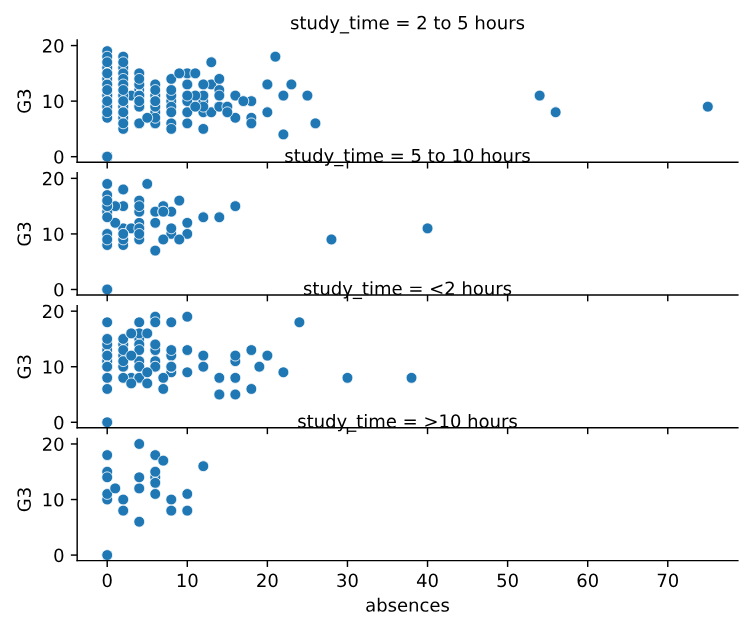
            data=student\_data,

            kind="scatter",

            row="study\_time")

# Show plot

plt.show()



\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Creating two-factor subplots**

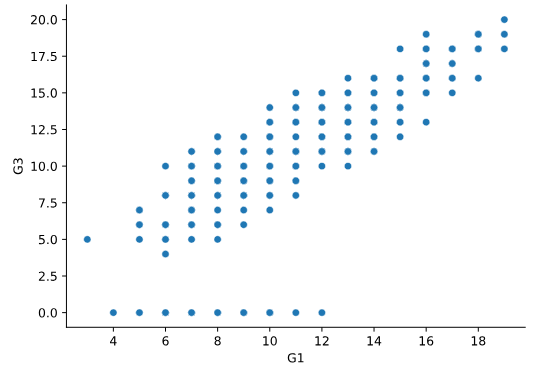
Let's continue looking at the student\_data dataset of students in secondary school. Here, we want to answer the following question: does a student's first semester grade ("G1") tend to correlate with their final grade ("G3")?

There are many aspects of a student's life that could result in a higher or lower final grade in the class. For example, some students receive extra educational support from their school ("schoolsup") or from their family ("famsup"), which could result in higher grades. Let's try to control for these two factors by creating subplots based on whether the student received extra educational support from their school or family.

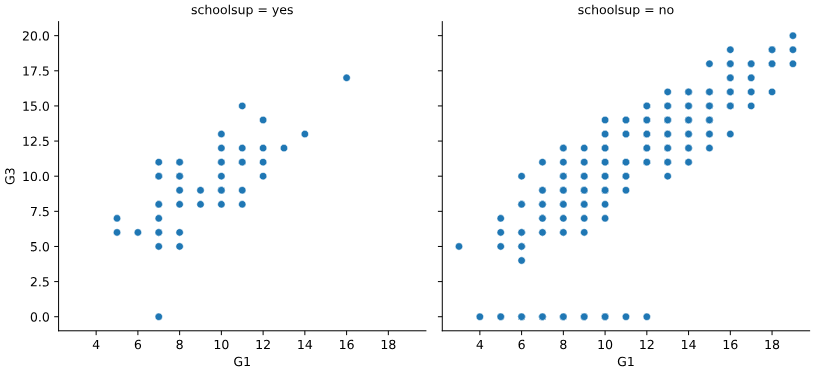
Seaborn has been imported as sns and matplotlib.pyplot has been imported as plt.

**Instructions 1/3**

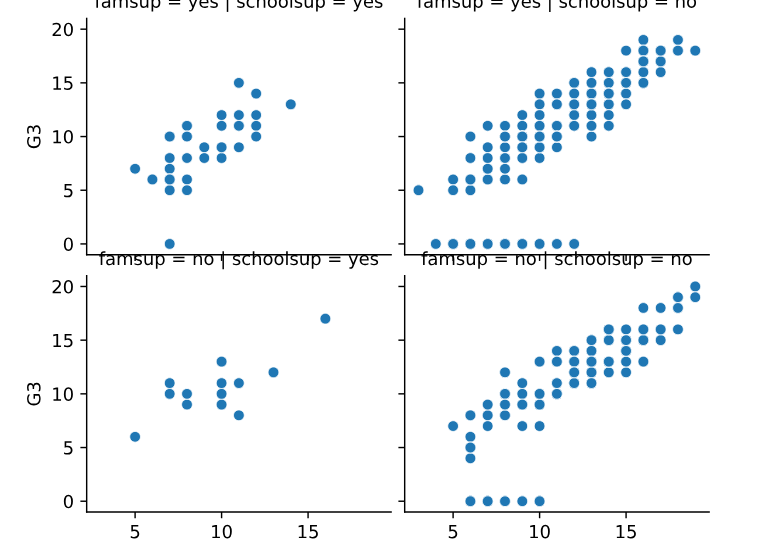
* Use relplot() to create a scatter plot with "G1" on the x-axis and "G3" on the y-axis, using the student\_data DataFrame.



* Create **column** subplots based on whether the student received support from the school ("schoolsup"), ordered so that "yes" comes before "no".



* Add **row** subplots based on whether the student received support from the family ("famsup"), ordered so that "yes" comes before "no". This will result in subplots based on two factors.



\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Changing the size of scatter plot points**

In this exercise, we'll explore Seaborn's mpg dataset, which contains one row per car model and includes information such as the year the car was made, the number of miles per gallon ("M.P.G.") it achieves, the power of its engine (measured in "horsepower"), and its country of origin.

What is the relationship between the power of a car's engine ("horsepower") and its fuel efficiency ("mpg")? And how does this relationship vary by the number of cylinders ("cylinders") the car has? Let's find out.

Let's continue to use relplot() instead of scatterplot() since it offers more flexibility.

**Instructions 1/2**

* Use relplot() and the mpg DataFrame to create a scatter plot with "horsepower" on the x-axis and "mpg" on the y-axis. Vary the size of the points by the number of cylinders in the car ("cylinders").

**script.py**

# Import Matplotlib and Seaborn

import matplotlib.pyplot as plt

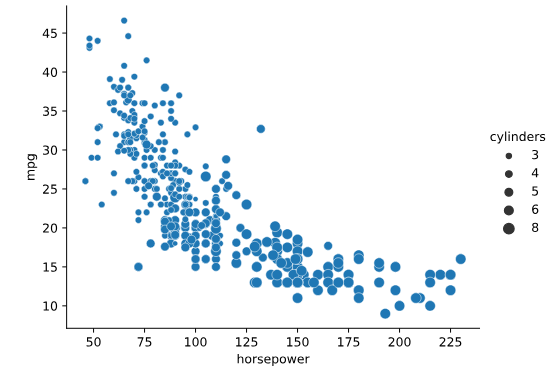
import seaborn as sns

# Create scatter plot of horsepower vs. mpg

sns.relplot(x="horsepower", y="mpg", data=mpg, kind="scatter", size="cylinders")

# Show plot

plt.show()



* To make this plot easier to read, use hue to vary the color of the points by the number of cylinders in the car ("cylinders").

**script.py**

# Import Matplotlib and Seaborn

import matplotlib.pyplot as plt

import seaborn as sns

# Create scatter plot of horsepower vs. mpg

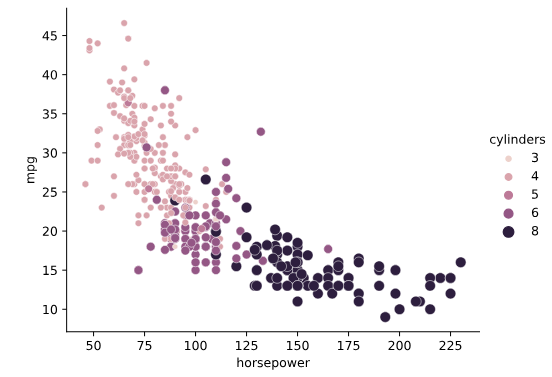
sns.relplot(x="horsepower", y="mpg",

            data=mpg, kind="scatter",

            size="cylinders", hue="cylinders")

# Show plot

plt.show()



\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Changing the style of scatter plot points**

Let's continue exploring Seaborn's mpg dataset by looking at the relationship between how fast a car can accelerate ("acceleration") and its fuel efficiency ("mpg"). Do these properties vary by country of origin ("origin")?

Note that the "acceleration" variable is the time to accelerate from 0 to 60 miles per hour, in seconds. Higher values indicate slower acceleration.

**Instructions**

* Use relplot() and the mpg DataFrame to create a scatter plot with "acceleration" on the x-axis and "mpg" on the y-axis. Vary the style and color of the plot points by country of origin ("origin").

**script.py**

# Import Matplotlib and Seaborn

import matplotlib.pyplot as plt

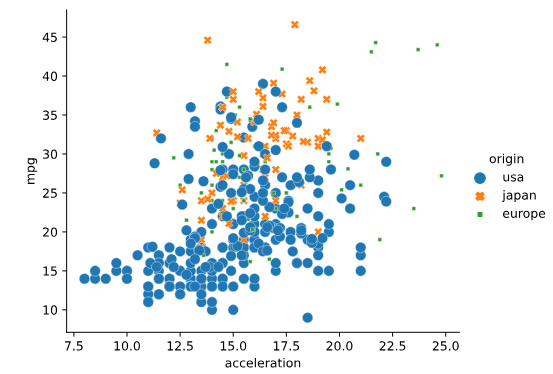
import seaborn as sns

# Create a scatter plot of acceleration vs. mpg

sns.relplot(x="acceleration", y="mpg", data=mpg, kind="scatter", size="origin", hue="origin", style="origin")

# Show plot

plt.show()



\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Interpreting line plots**

In this exercise, we'll continue to explore Seaborn's mpg dataset, which contains one row per car model and includes information such as the year the car was made, its fuel efficiency (measured in "miles per gallon" or "M.P.G"), and its country of origin (USA, Europe, or Japan).

How has the average miles per gallon achieved by these cars changed over time? Let's use line plots to find out!

**Instructions 1/2**

* Use relplot() and the mpg DataFrame to create a line plot with "model\_year" on the x-axis and "mpg" on the y-axis.

**script.py**

# Import Matplotlib and Seaborn

import matplotlib.pyplot as plt

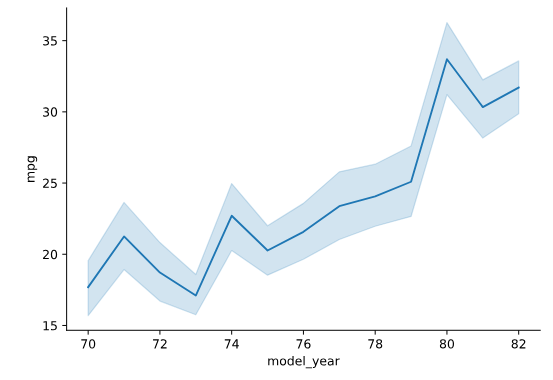
import seaborn as sns

# Create line plot

sns.relplot(x="model\_year", y="mpg", data=mpg, kind="line")

# Show plot

plt.show()



**Question**

Which of the following is NOT a correct interpretation of this line plot?

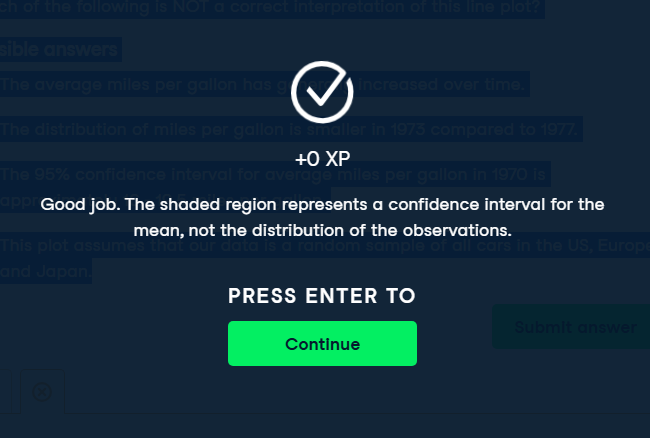
**Possible answers**

The average miles per gallon has generally increased over time.

The distribution of miles per gallon is smaller in 1973 compared to 1977.

The 95% confidence interval for average miles per gallon in 1970 is approximately 16 - 19.5 miles per gallon.

This plot assumes that our data is a random sample of all cars in the US, Europe, and Japan.



\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_