Introduction to Data Visualization with Seaborn

👋 Welcome to your workspace! Here, you can write and run Python code and add text in Markdown. Below, we've imported the datasets from the course *Introduction to Data Visualization with Seaborn* as DataFrames as well as the packages used in the course. This is your sandbox environment: analyze the course datasets further, take notes, or experiment with code!

Don't know where to start?

Try completing these tasks:

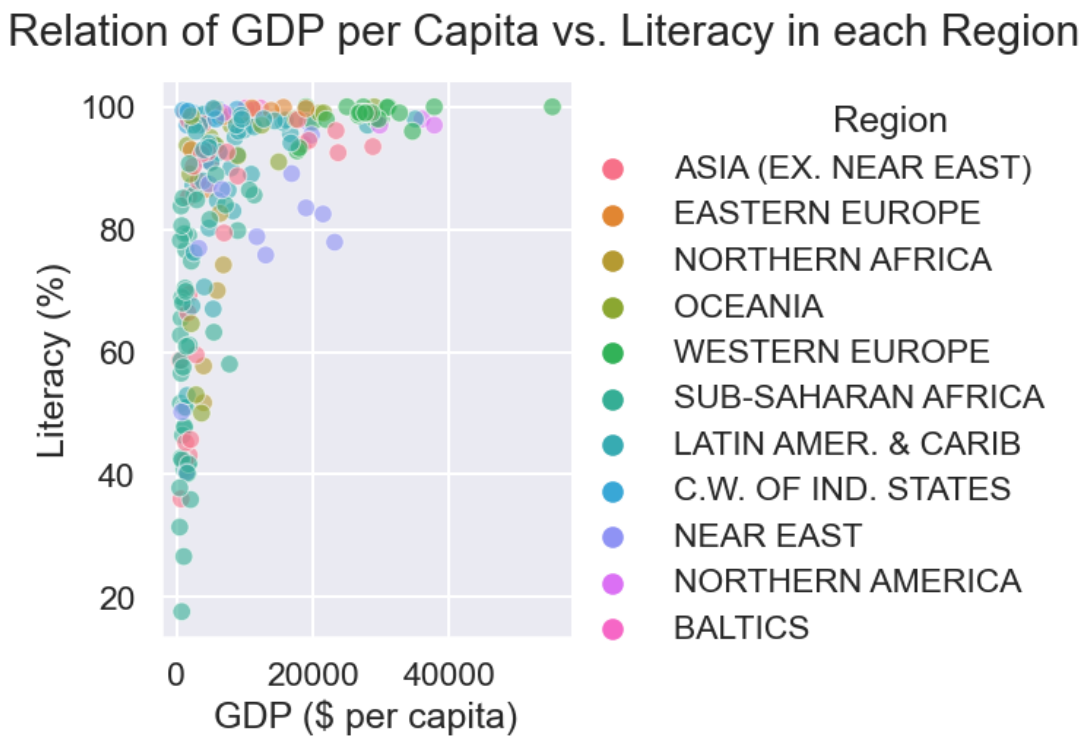
- From `country_data`, create a scatter plot to look at the relationship between GDP and Literacy. Use color to segment the data points by region.
- Use `mpg` to create a line plot with `model_year` on the x-axis and `weight` on the y-axis. Create differentiating lines for each country of origin (`origin`).
- Create a box plot from `student_data` to explore the relationship between the number of failures (`failures`) and the average final grade (`G3`).
- Create a bar plot from `survey` to compare how `Loneliness` differs across values for `Internet usage`. Format it to have two subplots for gender.
- Make sure to add titles and labels to your plots and adjust their format for readability!

In [ ]:
```
# Importing course packages; you can add more too!
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Importing course datasets as DataFrames
country_data = pd.read_csv('datasets/countries-of-the-world.csv', decimal=",")
mpg = pd.read_csv('datasets/mpg.csv')
student_data = pd.read_csv('datasets/student-alcohol-consumption.csv', index_col=0)
survey = pd.read_csv('datasets/young-people-survey-responses.csv', index_col=0)

sns.set_context('talk')

survey.head() # Display the first five rows of this DataFrame
```

Out[ ]:

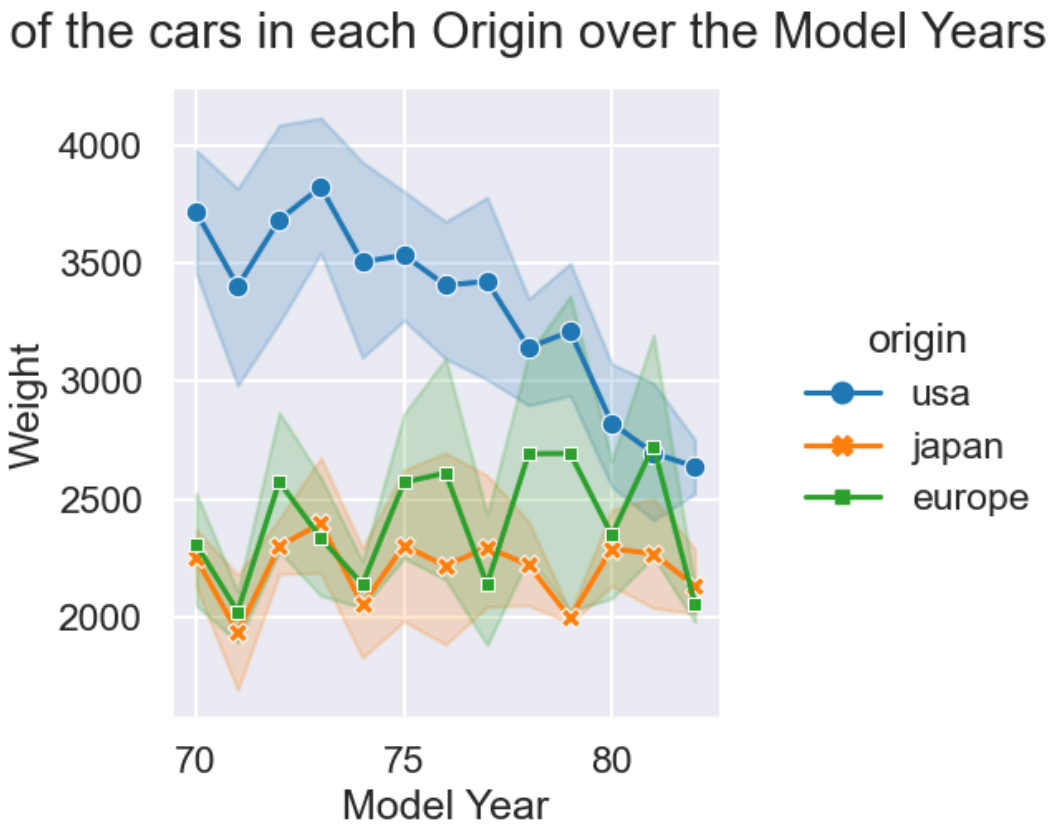
	Music	Techno	Movies	History	Mathematics	Pets	Spiders	Loneliness	Parents' advice	Internet usage	Fina
0	5.0	1.0	5.0	1.0	3.0	4.0	1.0	3.0	4.0	few hours a day	
1	4.0	1.0	5.0	1.0	5.0	5.0	1.0	2.0	2.0	few hours a day	
2	5.0	1.0	5.0	1.0	5.0	5.0	1.0	5.0	3.0	few hours a day	
3	5.0	2.0	5.0	4.0	4.0	1.0	5.0	5.0	2.0	most of the day	
4	5.0	2.0	5.0	3.0	2.0	1.0	1.0	3.0	3.0	few hours a day	

In [ ]:
```
sns.set_style('darkgrid')
g = sns.relplot(x='GDP ($ per capita)' , y='Literacy (%)', data=country_data,  kind='s
g.fig.suptitle("Relation of GDP per Capita vs. Literacy in each Region", y=1.04)
plt.show()
```



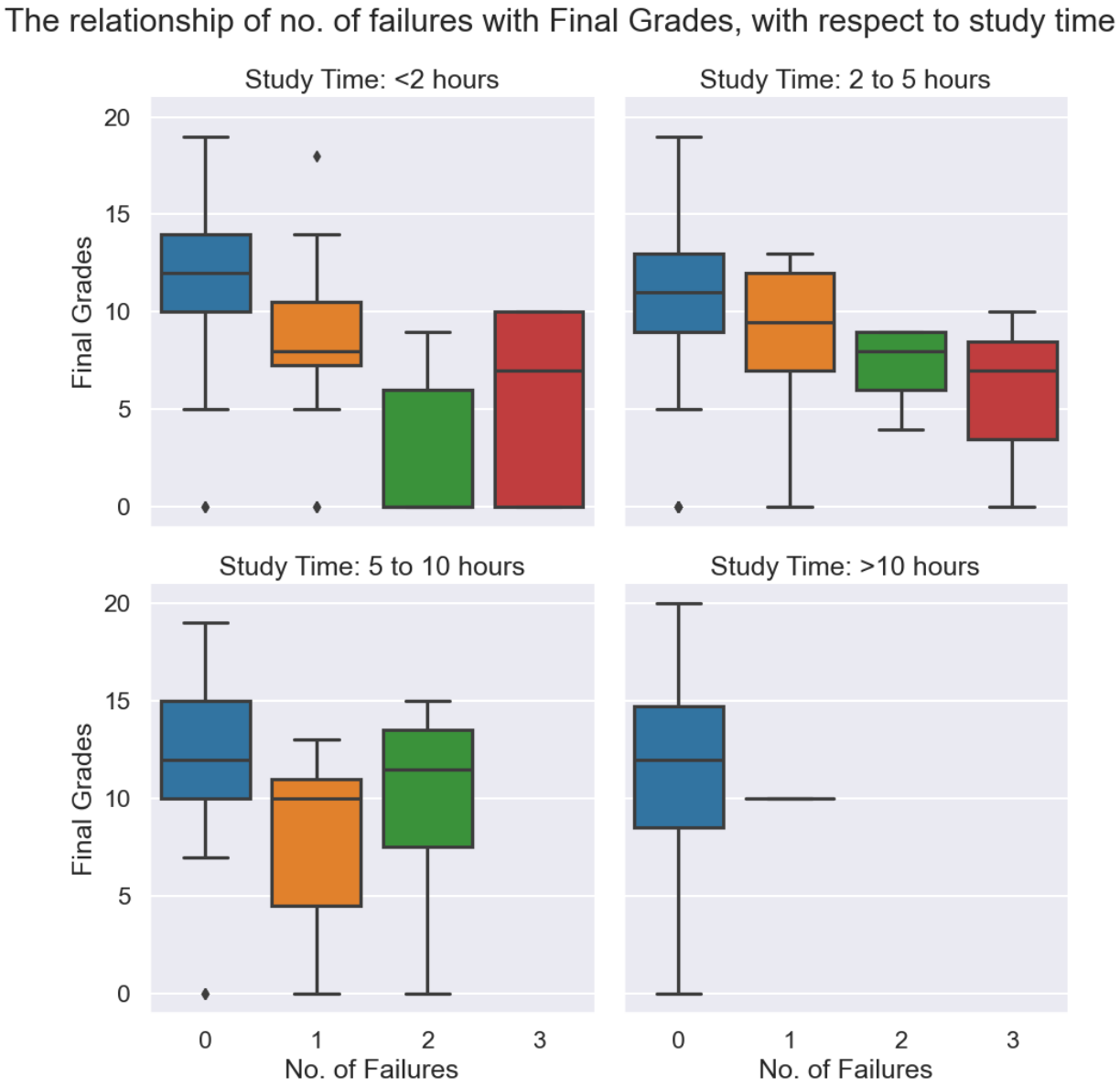
In [ ]:
```
plt.clf()
f = sns.relplot(y='weight', x='model_year', data=mpg, kind='line', hue='origin', style
f.fig.suptitle("Weight of the cars in each Origin over the Model Years", y=1.04)
f.set(xlabel='Model Year', ylabel='Weight')
plt.show()
```

```
<Figure size 640x480 with 0 Axes>
```



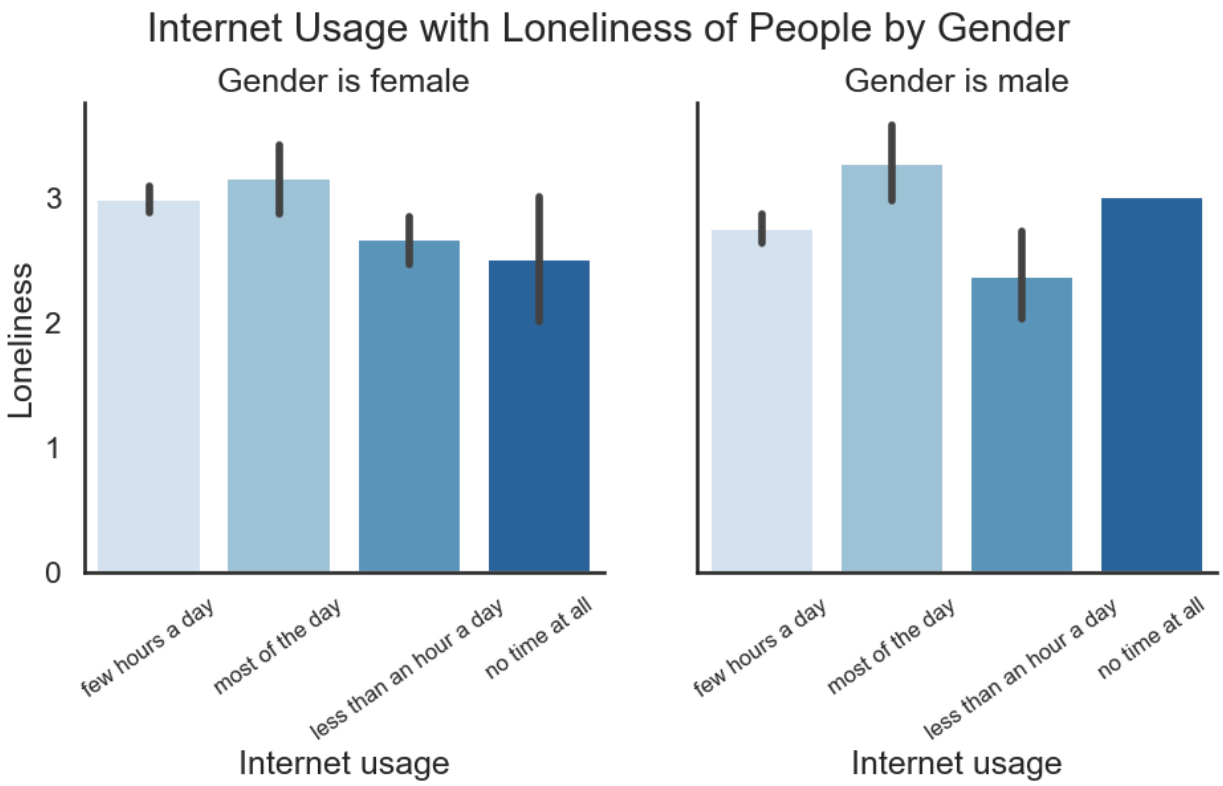
In [ ]:
```
plt.clf()
col_order = ['<2 hours', '2 to 5 hours', '5 to 10 hours', '>10 hours']
s = sns.catplot(x='failures', y='G3', kind='box', data=student_data, col='study_time'
s.fig.suptitle('The relationship of no. of failures with Final Grades, with respect to
s.set_titles('Study Time: {col_name}')
s.set(xlabel='No. of Failures', ylabel='Final Grades')
plt.show()
```

```
<Figure size 640x480 with 0 Axes>
```



In [ ]:
```
plt.clf()
sns.set_style('white')
fig = sns.catplot(y='Loneliness', x='Internet usage', data=survey, kind='bar', col='Ge
fig.set_xticklabels(rotation=35, fontsize=12)
fig.fig.suptitle('Internet Usage with Loneliness of People by Gender', y=1.03)
fig.set_titles('Gender is {col_name}')
plt.show()
```

```
<Figure size 640x480 with 0 Axes>
```


Internet Usage with Loneliness of People by Gender
Gender is female
Gender is male
Loneliness
0
1
2
3
few hours a day
most of the day
less than an hour a day
no time at all
Internet usage