# Exercise1

# Making a scatter plot with lists

In this exercise, we'll use a dataset that contains information about 227 countries. This dataset has lots of interesting information on each country, such as the country's birth rates, death rates, and its gross domestic product (GDP). GDP is the value of all the goods and services produced in a year, expressed as dollars per person.

We've created three lists of data from this dataset to get you started. gdp is a list that contains the value of GDP per country, expressed as dollars per person. phones is a list of the number of mobile phones per 1,000 people in that country. Finally, percent_literate is a list that contains the percent of each country's population that can read and write.

## Instructions

- Import Matplotlib and Seaborn using the standard naming convention.
- Create a scatter plot of GDP (gdp) vs. number of phones per 1000 people (phones).
- Display the plot.
- Change the scatter plot so it displays the percent of the population that can read and write (percent_literate) on the y-axis.

In [ ]:

```python
# Import Matplotlib and Seaborn

# Create scatter plot with GDP on the x-axis and number of phones on the y-axis

# Change this scatter plot to have percent literate on the y-axis

# Show plot

#_____#
#Solutions

# Import Matplotlib and Seaborn
import matplotlib.pyplot as plt
import seaborn as sns

# Create scatter plot with GDP on the x-axis and number of phones on the y-axis
sns.scatterplot(x=gdp,y=phones)

# Show plot
plt.show()

# Change this scatter plot to have percent literate on the y-axis
sns.scatterplot(x=gdp, y=percent_literate)

# Show plot
plt.show()
```

## Exercise2

# Making a count plot with a list

In the last exercise, we explored a dataset that contains information about 227 countries. Let's do more exploration of this data - specifically, how many countries are in each region of the world?

To do this, we'll need to use a count plot. Count plots take in a categorical list and return bars that represent the number of list entries per category. You can create one here using a list of regions for each country, which is a variable named region.

## Instructions

- Import Matplotlib and Seaborn using the standard naming conventions.
- Use Seaborn to create a count plot with region on the y-axis.
- Display the plot.

In [ ]:

```python
# Import Matplotlib and Seaborn

# Create count plot with region on the y-axis
____.____(y=____)

# Show plot


#_____#
#Solutions

# Import Matplotlib and Seaborn
import matplotlib.pyplot as plt
import seaborn as sns


# Create count plot with region on the y-axis
sns.countplot(y=region)

# Show plot
plt.show()
```

## Exercise3

# "Tidy" vs. "untidy" data

Here, we have a sample dataset from a survey of children about their favorite animals. But can we use this dataset as-is with Seaborn? Let's use pandas to import the csv file with the data collected from the survey and determine whether it is tidy, which is essential to having it work well with Seaborn.

To get you started, the filepath to the csv file has been assigned to the variable csv_filepath.

Note that because csv_filepath is a Python variable, you will not need to put quotation marks around it when you read the csv.

## Instructions

- Read the csv file located at csv_filepath into a DataFrame named df.
- Print the head of df to show the first five rows.

In [ ]:

```python
# Import pandas
import pandas as pd

# Create a DataFrame from csv file
____ = pd.____(csv_filepath)

# Print the head of df
print(df.____)

#_____#
#Solutions

# Import pandas
import pandas as pd

# Create a DataFrame from csv file
df = pd.read_csv(csv_filepath)

# Print the head of df
print(df.head())
```

## Exercise4

# Making a count plot with a DataFrame

In this exercise, we'll look at the responses to a survey sent out to young people. Our primary question here is: how many young people surveyed report being scared of spiders? Survey participants were asked to agree or disagree with the statement "I am afraid of spiders". Responses vary from 1 to 5, where 1 is "Strongly disagree" and 5 is "Strongly agree".

To get you started, the filepath to the csv file with the survey data has been assigned to the variable csv_filepath.

Note that because csv_filepath is a Python variable, you will not need to put quotation marks around it when you read the csv.

## Instructions

- Import Matplotlib, pandas, and Seaborn using the standard names.
- Create a DataFrame named df from the csv file located at csv_filepath.
- Use the countplot() function with the x= and data= arguments to create a count plot with the "Spiders" column values on the x-axis.
- Display the plot.

In [ ]:

```python
# Import Matplotlib, pandas, and Seaborn

# Create a DataFrame from csv file

# Create a count plot with "Spiders" on the x-axis

# Display the plot

#_____#
#Solutions

# Import Matplotlib, pandas, and Seaborn
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns


# Create a DataFrame from csv file
df = pd.read_csv(csv_filepath)

# Create a count plot with "Spiders" on the x-axis
sns.countplot(x='Spiders',data=df)

# Display the plot
plt.show()
```

# Exercise5

# Hue and scatter plots

In the prior video, we learned how hue allows us to easily make subgroups within Seaborn plots. Let's try it out by exploring data from students in secondary school. We have a lot of information about each student like their age, where they live, their study habits and their extracurricular activities.

For now, we'll look at the relationship between the number of absences they have in school and their final grade in the course, segmented by where the student lives (rural vs. urban area).

## Instructions

- Create a scatter plot with "absences" on the x-axis and final grade ("G3") on the y-axis using the DataFrame student_data. Color the plot points based on "location" (urban vs. rural)
- Make "Rural" appear before "Urban" in the plot legend.

In [ ]:

```python
# Import Matplotlib and Seaborn

# Create a scatter plot of absences vs. final grade

# Change the legend order in the scatter plot

# Show plot

#_____#
#Solutions

# Import Matplotlib and Seaborn
import matplotlib.pyplot as plt
import seaborn as sns

# Create a scatter plot of absences vs. final grade
sns.scatterplot(x='absences', y='G3', data=student_data,hue='location')

# Change the legend order in the scatter plot
sns.scatterplot(x="absences", y="G3",
                data=student_data,
                hue="location", hue_order=['Rural','Urban'])

# Show plot
plt.show()
```

## Exercise6

# Hue and count plots

Let's continue exploring our dataset from students in secondary school by looking at a new variable. The "school" column indicates the initials of which school the student attended - either "GP" or "MS".

In the last exercise, we created a scatter plot where the plot points were colored based on whether the student lived in an urban or rural area. How many students live in urban vs. rural areas, and does this vary based on what school the student attends? Let's make a count plot with subgroups to find out.

## Instructions

- Fill in the palette_colors dictionary to map the "Rural" location value to the color "green" and the "Urban" location value to the color "blue".
- Create a count plot with "school" on the x-axis using the student_data DataFrame.
- Add subgroups to the plot using "location" variable and use the palette_colors dictionary to make the location subgroups green and blue.

In [ ]:

```python
# Import Matplotlib and Seaborn

# Create a dictionary mapping subgroup values to colors
palette_colors = {____: "green", ____: "blue"}

# Create a count plot of school with location subgroups

# Display plot

#_____#
#Solutions

# Import Matplotlib and Seaborn
import matplotlib.pyplot as plt
import seaborn as sns

# Create a dictionary mapping subgroup values to colors
palette_colors = {'Rural': "green", 'Urban': "blue"}

# Create a count plot of school with location subgroups
sns.countplot(x='school',hue='location',data=student_data,palette=palette_colors)

# Display plot
plt.show()
```