

Assignment 1.2 Python Refresher

```
In [1]: # importing Libraries
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [2]: # importing warning to surpress the future warnings
import warnings
warnings.simplefilter(action='ignore', category=FutureWarning)
```

1. Write a summary of your data and identify at least two questions to explore visually with your data.

For this Python refresher activity I have chosen to use a dataset that I had wanted to use for a previous course but did not have the opportunity. The dataset contains different features of Pokemon. Pokemon are fictitious creatures best known from the Nintendo video games.

The data was sourced from:

Barradas, Alberto. (2016). *Pokemon with Stats*[Dataset]. Kaggle.
<https://www.kaggle.com/datasets/abcsds/pokemon>

Questions:

1. How many pokemon are of the various types?
2. Is there a correlation between pokemon height and weight?

```
In [3]: # reading dataset
df = pd.read_csv('Datasets/Pokemon.csv')
```

```
In [4]: # reviewing first 5 rows
df.head(5)
```

```
Out[4]:
```

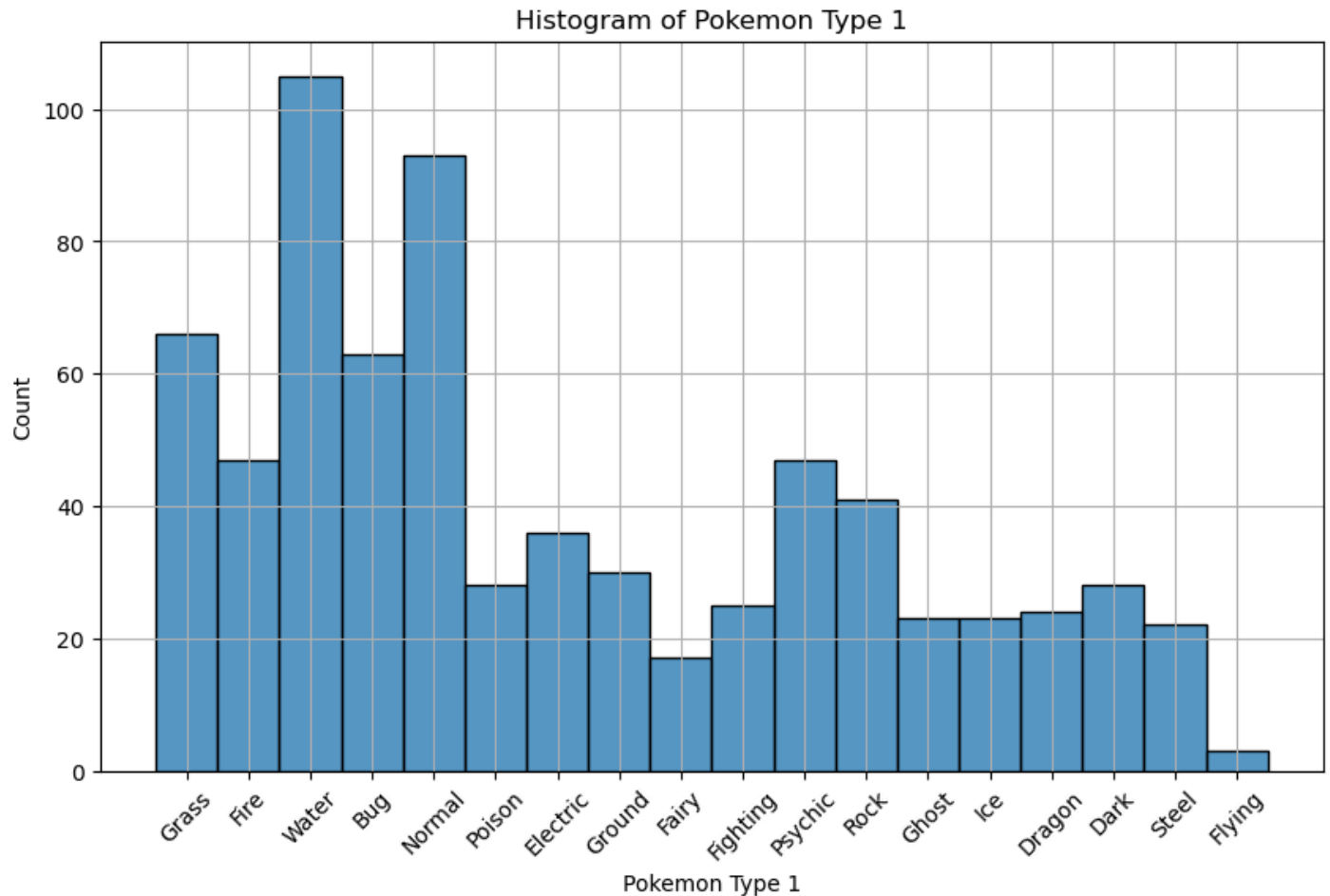
	Number	Name	Type_1	Type_2	Total	HP	Attack	Defense	Sp_Atk	Sp_Def	...	Color	hasGer
0	1	Bulbasaur	Grass	Poison	318	45	49	49	65	65	...	Green	
1	2	Ivysaur	Grass	Poison	405	60	62	63	80	80	...	Green	
2	3	Venusaur	Grass	Poison	525	80	82	83	100	100	...	Green	
3	4	Charmander	Fire	NaN	309	39	52	43	60	50	...	Red	
4	5	Charmeleon	Fire	NaN	405	58	64	58	80	65	...	Red	

5 rows × 23 columns



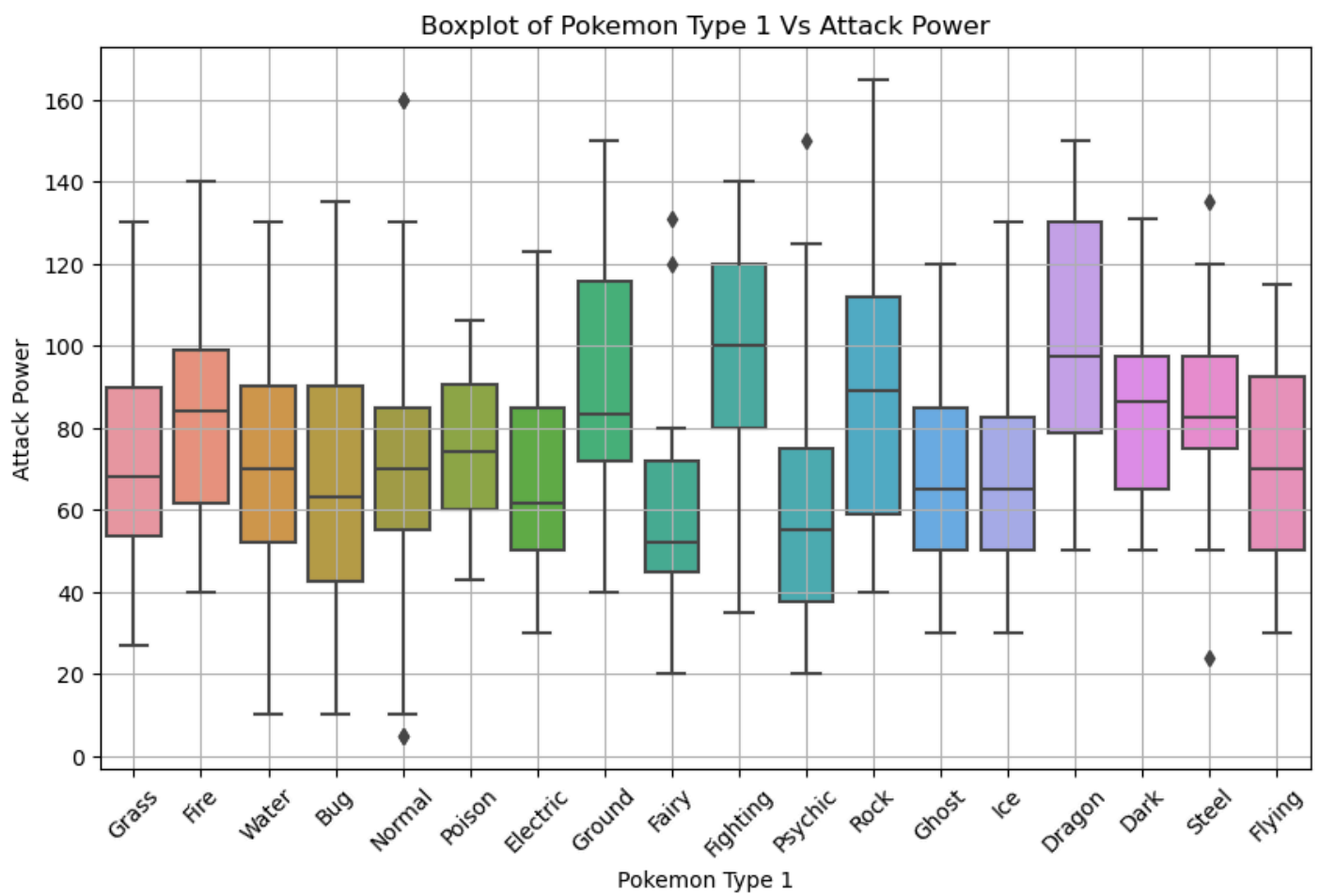
2. Create a histogram or bar graph from your data.

```
In [5]: # creating a histogram of pokemon types
plt.figure(figsize=(10, 6))
sns.histplot(data=df, x='Type_1')
plt.xlabel('Pokemon Type 1')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.title('Histogram of Pokemon Type 1')
plt.grid(True)
plt.show()
```



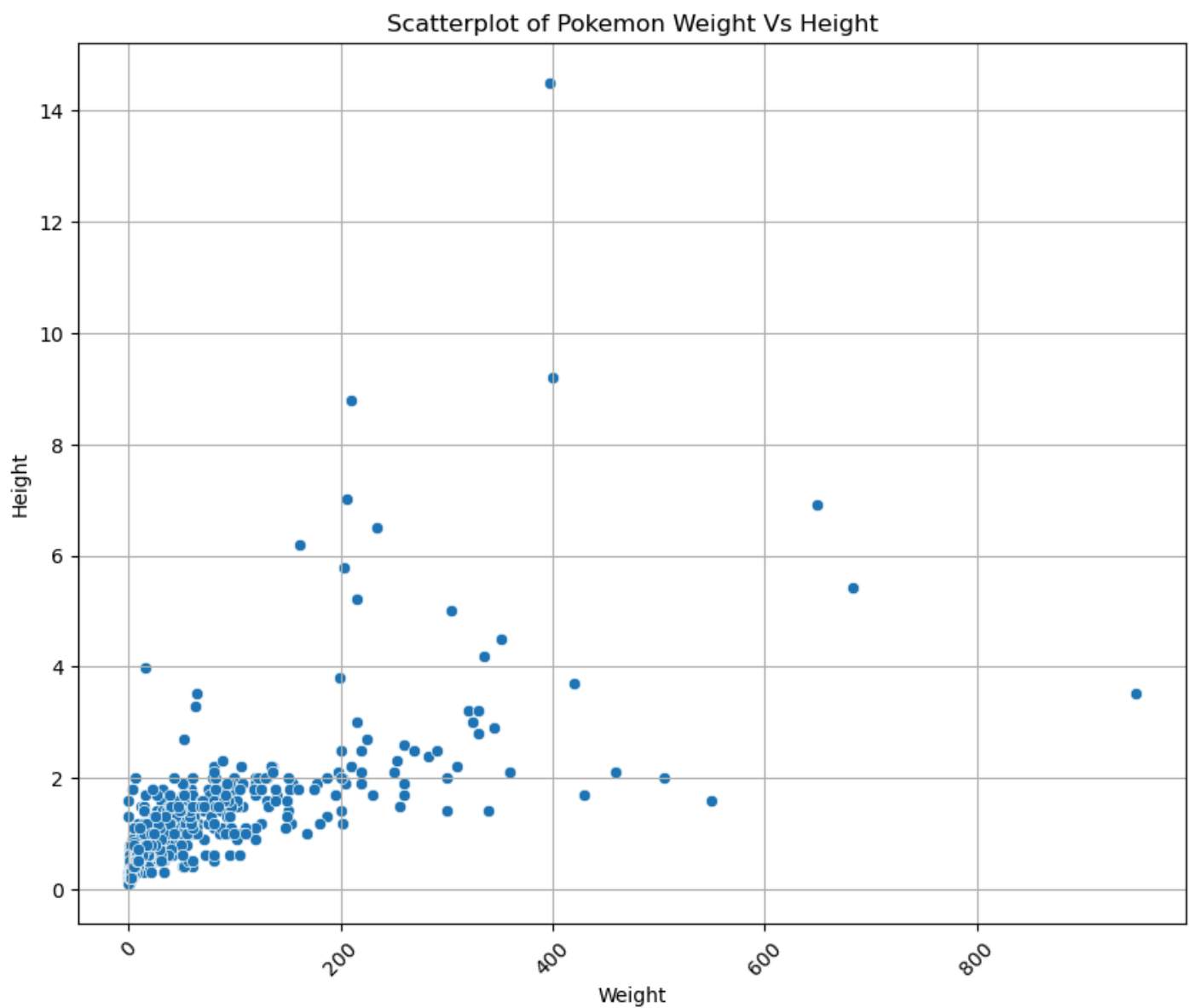
3. Create a boxplot from your data.

```
In [6]: # creating a boxplot of pokemon types and attack power by type
plt.figure(figsize = (10, 6))
sns.boxplot(x = 'Type_1', y = 'Attack', data = df)
plt.xlabel('Pokemon Type 1')
plt.ylabel('Attack Power')
plt.xticks(rotation = 45)
plt.title('Boxplot of Pokemon Type 1 Vs Attack Power')
plt.grid(True)
plt.show()
```



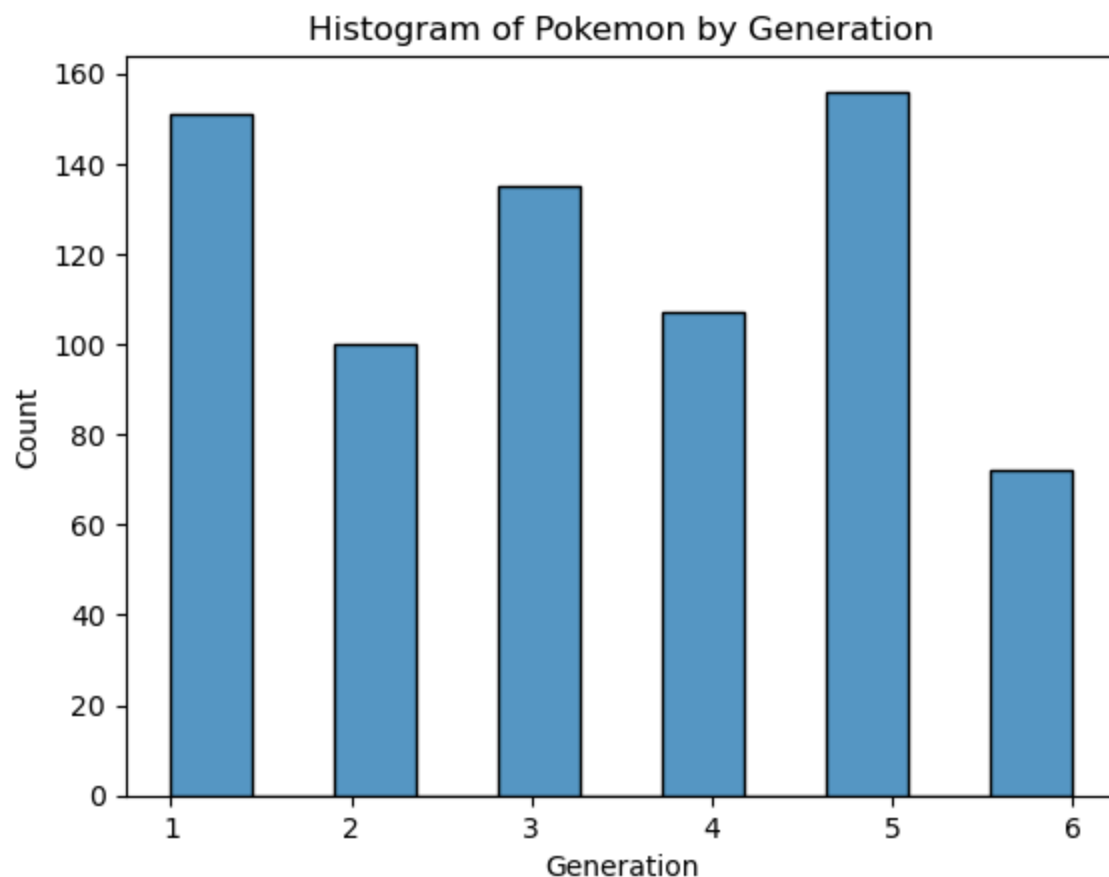
4. Create a bivariate plot from your data.

```
In [7]: # creating a scatterplot of weight vs height for pokemon
plt.figure(figsize = (10, 8))
sns.scatterplot(x = 'Weight_kg', y = 'Height_m', data = df)
plt.xlabel('Weight')
plt.ylabel('Height')
plt.xticks(rotation = 45)
plt.title('Scatterplot of Pokemon Weight Vs Height')
plt.grid(True)
plt.show()
```



5. Create any additional visualizations that will help to answer the question(s) you want to answer.

```
In [8]: # creating a histogram of pokemon by generation
sns.histplot(data=df, x='Generation')
plt.xlabel('Generation')
plt.ylabel('Count')
plt.title('Histogram of Pokemon by Generation')
plt.grid(False)
plt.show()
```



6. Summarize your results and make a conclusion. Explain how you arrived at this conclusion and how your visualizations support your conclusion.

In the Histogram of Pokemon Types 1, we can see the number of total pokemon by type, this visualization gives an immediate answer to the first question of "How many pokemon are of the various types?". It is also easy to see which Pokemon type is the most and least prevalent, water type being the most and flying being the least.

The second question, "Is there a correlation between pokemon height and weight?", is a little more challenging to answer simply based on visualizations. However, utilizing a bivariate visualization like a scatterplot can show correlation between the two plotted variables. In the scatterplot of Pokemon Weight Vs Height, we can see what appears to be a weak positive correlation between the two variables.

The visualizations are key to supporting the conclusions made. Both the Histogram and Scatterplot offer compelling visual evidence to make conclusions for both questions.