

# 20230328\_Data\_Bases\_Data\_Mining\_MV4

March 28, 2023

```
[1]: # Case Study Pokemon Dataset
```

```
[2]: # Packages und Libraries installieren und importieren
```

```
[3]: # Pandas brauchen wir um mit Datasets umzugehen
```

```
[4]: !pip install pandas
```

Requirement already satisfied: pandas in /Users/h4/anaconda3/lib/python3.9/site-packages (1.4.2)

Requirement already satisfied: python-dateutil>=2.8.1 in /Users/h4/anaconda3/lib/python3.9/site-packages (from pandas) (2.8.2)

Requirement already satisfied: pytz>=2020.1 in /Users/h4/anaconda3/lib/python3.9/site-packages (from pandas) (2021.3)

Requirement already satisfied: numpy>=1.20.0 in /Users/h4/anaconda3/lib/python3.9/site-packages (from pandas) (1.23.2)

Requirement already satisfied: six>=1.5 in /Users/h4/anaconda3/lib/python3.9/site-packages (from python-dateutil>=2.8.1->pandas) (1.16.0)

```
[5]: import pandas as pd
```

```
[6]: # Matplotlib brauchen wir um graphische Darstellungen zu ermöglichen
```

```
[7]: !pip install matplotlib
```

Requirement already satisfied: matplotlib in /Users/h4/anaconda3/lib/python3.9/site-packages (3.5.1)

Requirement already satisfied: kiwisolver>=1.0.1 in /Users/h4/anaconda3/lib/python3.9/site-packages (from matplotlib) (1.4.2)

Requirement already satisfied: numpy>=1.17 in /Users/h4/anaconda3/lib/python3.9/site-packages (from matplotlib) (1.23.2)

Requirement already satisfied: pyparsing>=2.2.1 in /Users/h4/anaconda3/lib/python3.9/site-packages (from matplotlib) (3.0.4)

Requirement already satisfied: fonttools>=4.22.0 in /Users/h4/anaconda3/lib/python3.9/site-packages (from matplotlib) (4.25.0)

Requirement already satisfied: cyycler>=0.10 in /Users/h4/anaconda3/lib/python3.9/site-packages (from matplotlib) (0.11.0)

Requirement already satisfied: pillow>=6.2.0 in  
/Users/h4/anaconda3/lib/python3.9/site-packages (from matplotlib) (9.0.1)  
Requirement already satisfied: python-dateutil>=2.7 in  
/Users/h4/anaconda3/lib/python3.9/site-packages (from matplotlib) (2.8.2)  
Requirement already satisfied: packaging>=20.0 in  
/Users/h4/anaconda3/lib/python3.9/site-packages (from matplotlib) (21.3)  
Requirement already satisfied: six>=1.5 in  
/Users/h4/anaconda3/lib/python3.9/site-packages (from python-  
dateutil>=2.7->matplotlib) (1.16.0)

[8]: *# von matplotlib brauchen wir nur pyplot*

```
from matplotlib import pyplot as plt
```

[9]: *# seaborn ermöglicht deutlich bessere graphische darstellungen als matplotlib*

```
!pip install seaborn
```

Requirement already satisfied: seaborn in  
/Users/h4/anaconda3/lib/python3.9/site-packages (0.11.2)  
Requirement already satisfied: matplotlib>=2.2 in  
/Users/h4/anaconda3/lib/python3.9/site-packages (from seaborn) (3.5.1)  
Requirement already satisfied: pandas>=0.23 in  
/Users/h4/anaconda3/lib/python3.9/site-packages (from seaborn) (1.4.2)  
Requirement already satisfied: scipy>=1.0 in  
/Users/h4/anaconda3/lib/python3.9/site-packages (from seaborn) (1.9.1)  
Requirement already satisfied: numpy>=1.15 in  
/Users/h4/anaconda3/lib/python3.9/site-packages (from seaborn) (1.23.2)  
Requirement already satisfied: kiwisolver>=1.0.1 in  
/Users/h4/anaconda3/lib/python3.9/site-packages (from matplotlib>=2.2->seaborn)  
(1.4.2)  
Requirement already satisfied: cycler>=0.10 in  
/Users/h4/anaconda3/lib/python3.9/site-packages (from matplotlib>=2.2->seaborn)  
(0.11.0)  
Requirement already satisfied: pyparsing>=2.2.1 in  
/Users/h4/anaconda3/lib/python3.9/site-packages (from matplotlib>=2.2->seaborn)  
(3.0.4)  
Requirement already satisfied: packaging>=20.0 in  
/Users/h4/anaconda3/lib/python3.9/site-packages (from matplotlib>=2.2->seaborn)  
(21.3)  
Requirement already satisfied: pillow>=6.2.0 in  
/Users/h4/anaconda3/lib/python3.9/site-packages (from matplotlib>=2.2->seaborn)  
(9.0.1)  
Requirement already satisfied: fonttools>=4.22.0 in  
/Users/h4/anaconda3/lib/python3.9/site-packages (from matplotlib>=2.2->seaborn)  
(4.25.0)  
Requirement already satisfied: python-dateutil>=2.7 in  
/Users/h4/anaconda3/lib/python3.9/site-packages (from matplotlib>=2.2->seaborn)

(2.8.2)

Requirement already satisfied: pytz>=2020.1 in  
/Users/h4/anaconda3/lib/python3.9/site-packages (from pandas>=0.23->seaborn)  
(2021.3)

Requirement already satisfied: six>=1.5 in  
/Users/h4/anaconda3/lib/python3.9/site-packages (from python-  
dateutil>=2.7->matplotlib>=2.2->seaborn) (1.16.0)

```
[10]: import seaborn as sns
```

```
[11]: # wir werden nun unser dataset hochladen
```

```
[12]: # wir generieren eine Variable "df" (dataframe) welches durch den Pfad den  
      ↪ Dataset hochlädt  
  
      # index_col = 0 bedeutet, dass die erste Spalte als Index benutzt wird  
  
      # encoding (nicht Prüfungsrelevant) bedeutet die "sprache" der Speicherung  
  
      df = pd.read_csv('/Users/h4/Desktop/20230328_Pokemon.csv', index_col = 0,  
      ↪ encoding = 'latin')
```

```
[14]: # df.head() beschreibt die ersten 5 Zeilen von einem Dataframe
```

```
df.head()
```

```
[14]:
```

	Name	Type 1	Type 2	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	\
#										
1	Bulbasaur	Grass	Poison	318	45	49	49	65	65	
2	Ivysaur	Grass	Poison	405	60	62	63	80	80	
3	Venusaur	Grass	Poison	525	80	82	83	100	100	
4	Charmander	Fire	NaN	309	39	52	43	60	50	
5	Charmeleon	Fire	NaN	405	58	64	58	80	65	

	Speed	Stage	Legendary
#			
1	45	1	False
2	60	2	False
3	80	3	False
4	65	1	False
5	80	2	False

```
[16]: df.tail() # die letzten 5 Zeilen
```

```
[16]:
```

	Name	Type 1	Type 2	Total	HP	Attack	Defense	Sp. Atk	\
#									
147	Dratini	Dragon	NaN	300	41	64	45	50	

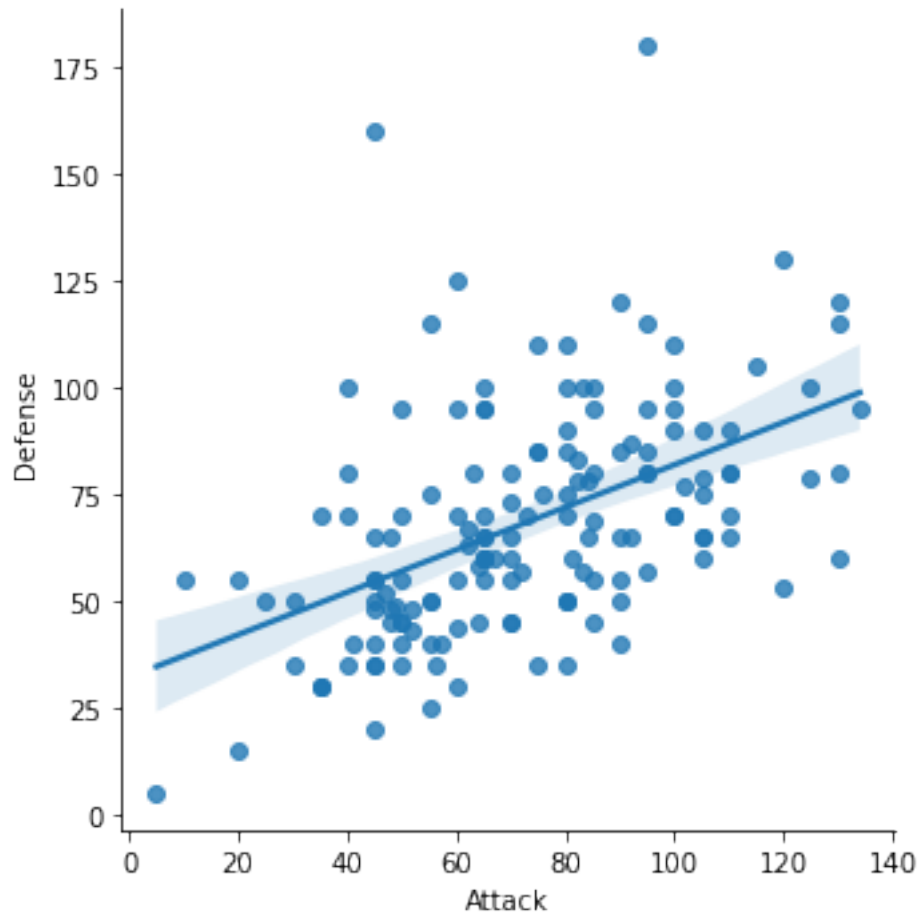
148	Dragonair	Dragon	NaN	420	61	84	65	70
149	Dragonite	Dragon	Flying	600	91	134	95	100
150	Mewtwo	Psychic	NaN	680	106	110	90	154
151	Mew	Psychic	NaN	600	100	100	100	100

	Sp.	Def	Speed	Stage	Legendary
#					
147		50	50	1	False
148		70	70	2	False
149		100	80	3	False
150		90	130	1	True
151		100	100	1	False

```
[17]: #scatterplot mit "Attack vs. Defense"
```

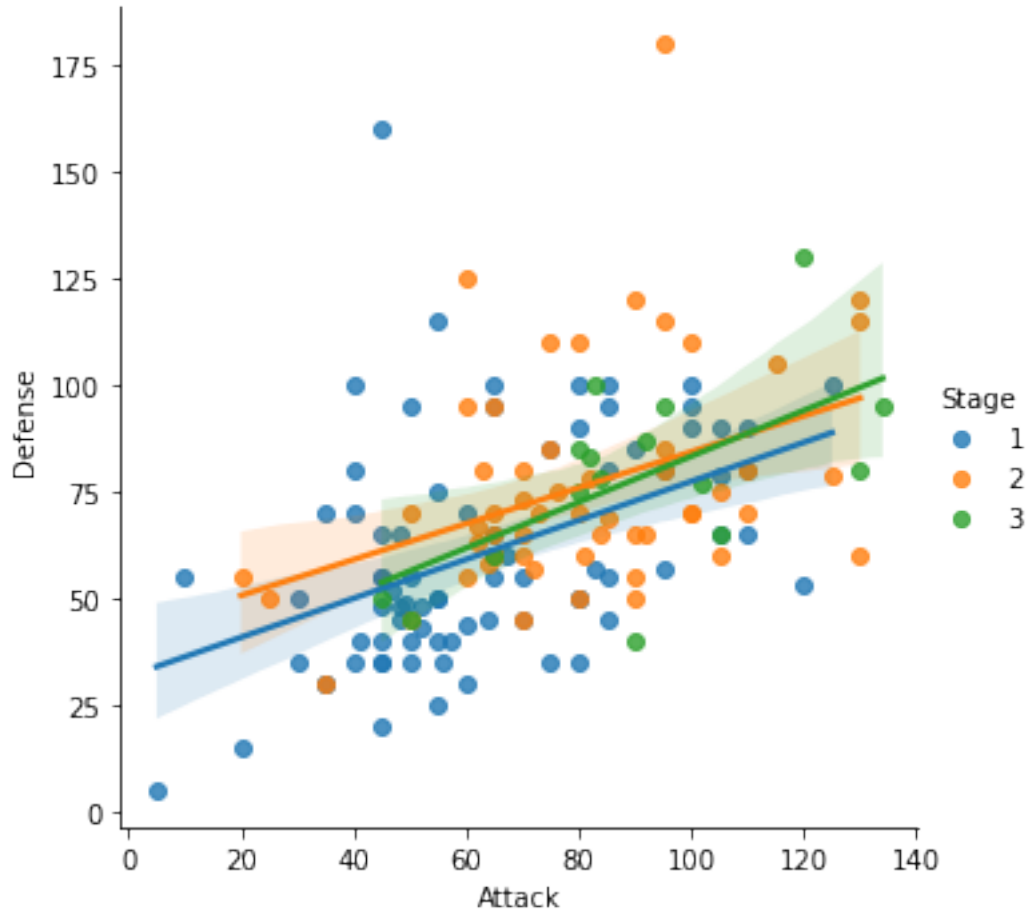
```
[18]: sns.lmplot(x='Attack', y='Defense', data=df)
```

```
[18]: <seaborn.axisgrid.FacetGrid at 0x1224e3700>
```



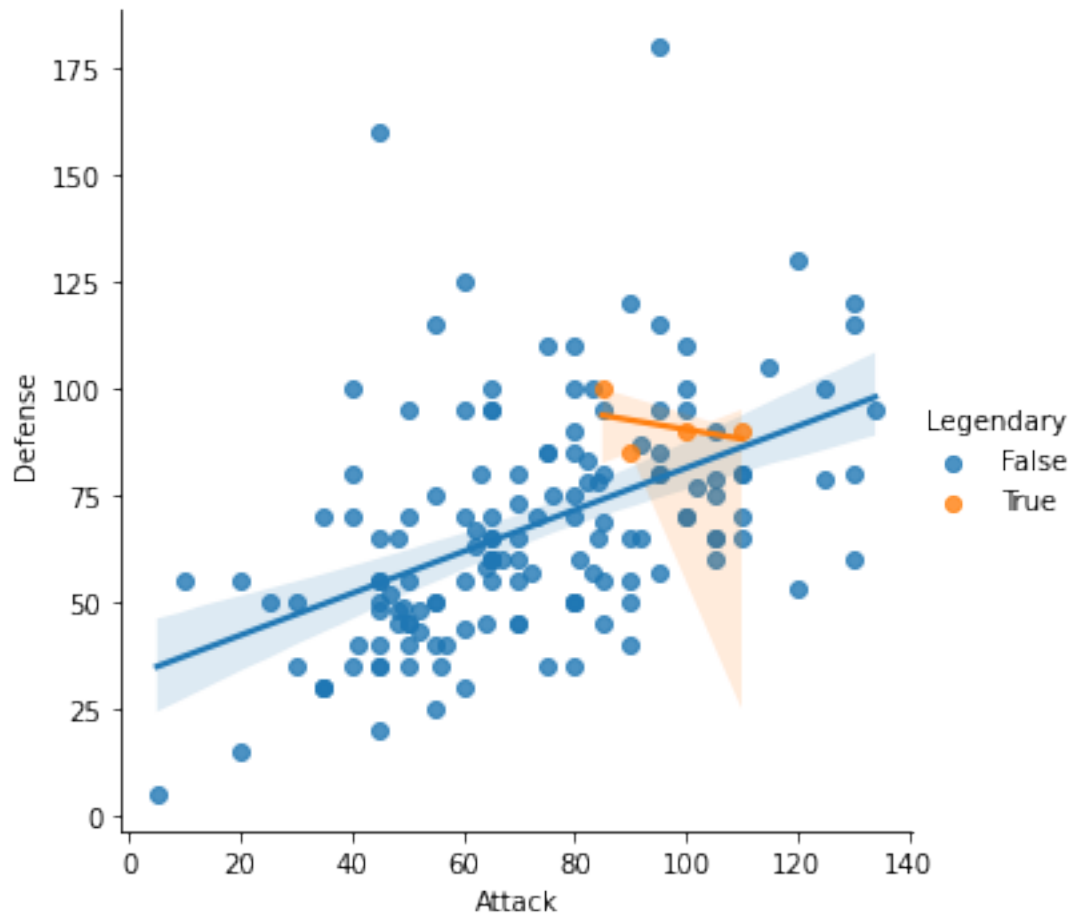
```
[19]: # Klassen innerhabl von einem Plot werden in Seaborn mit "hue" ermöglicht
```

```
[22]: sns.lmplot(x='Attack', y='Defense', data=df, hue='Stage')  
plt.show()
```

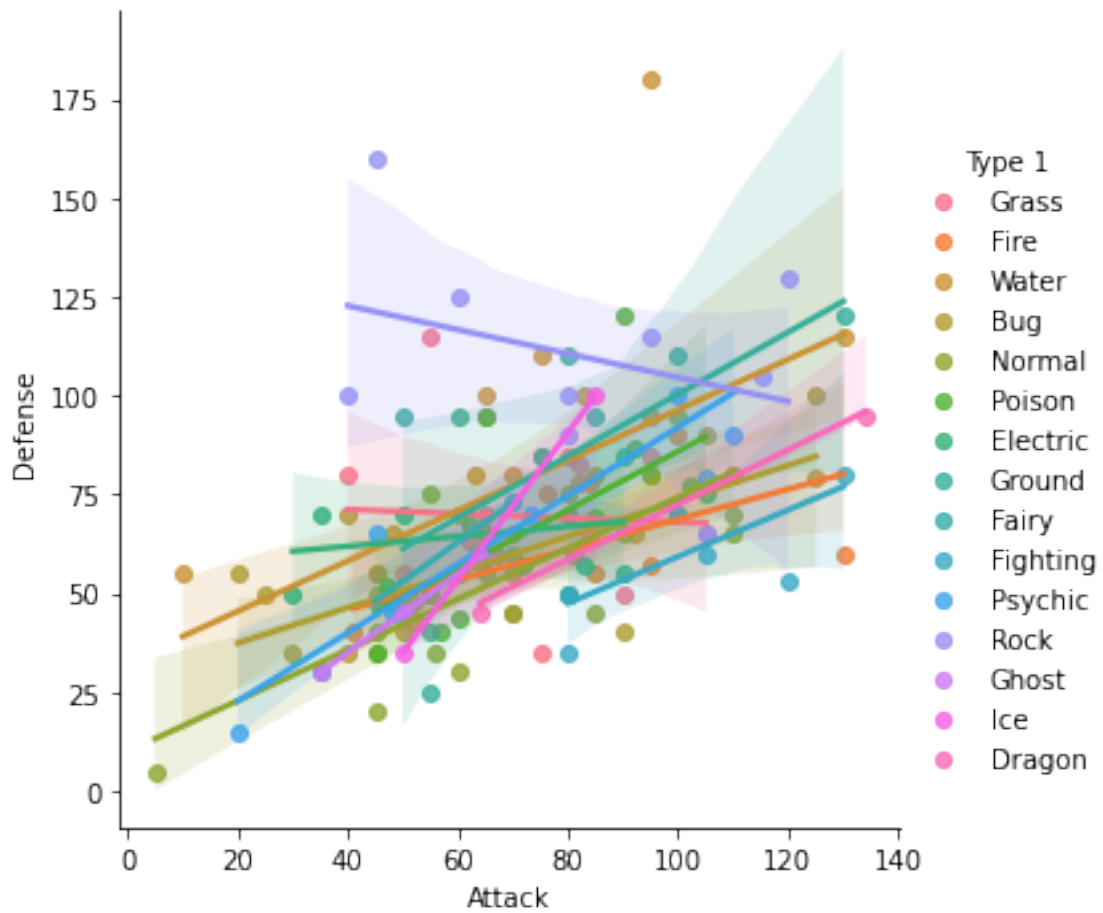


```
[23]: # Übung. (ähnlich wie in die Prüfung)  
# Bitte erstellen Sie einen Scatterplot mit x-Axes "Attack", y-Axes "Defense"  
# und hue=Legendary
```

```
[24]: sns.lmplot(x='Attack', y='Defense', data=df, hue='Legendary')  
plt.show()
```

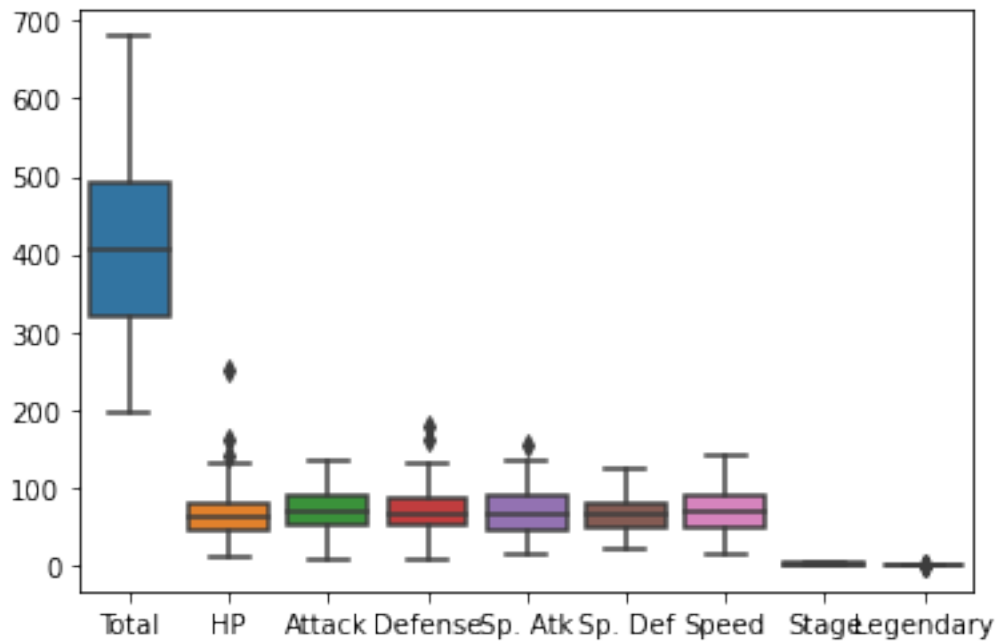


```
[25]: sns.lmplot(x='Attack', y='Defense', data=df, hue='Type 1')  
plt.show()
```



```
[26]: # box plot Darstellung
```

```
[29]: sns.boxplot(data=df)
plt.show()
```



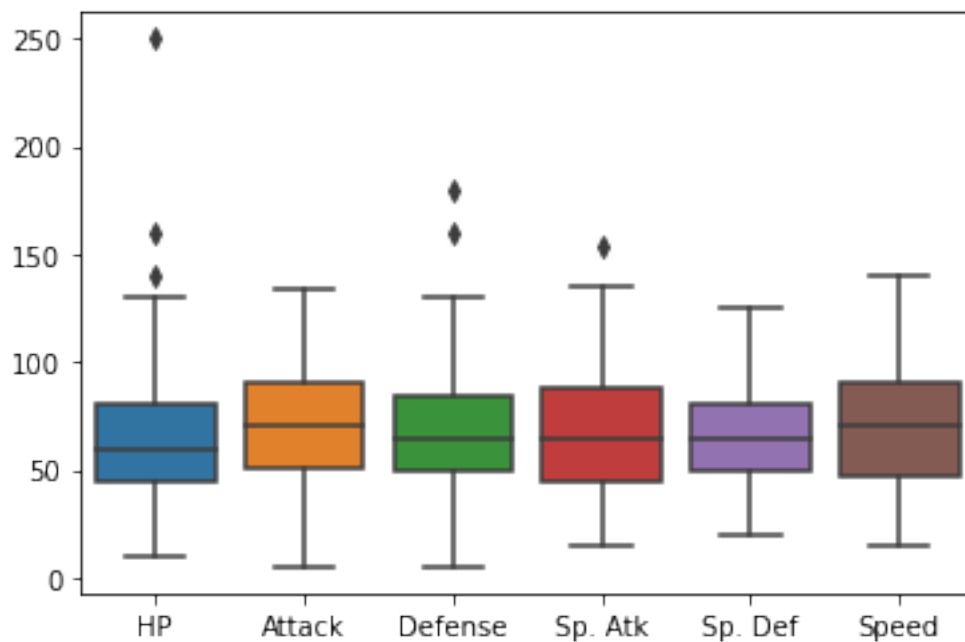
```
[30]: # wir erzeugen eine homogenisierte Darstellung der Daten,
      # dadurch dass wir "Total" "Stage" und "Legendary" ausfallen lassen.
```

```
[33]: stats_df = df.drop(['Total', 'Stage', 'Legendary'], axis=1)

      sns.boxplot(data=stats_df)
```

```
[33]: <AxesSubplot:>
```

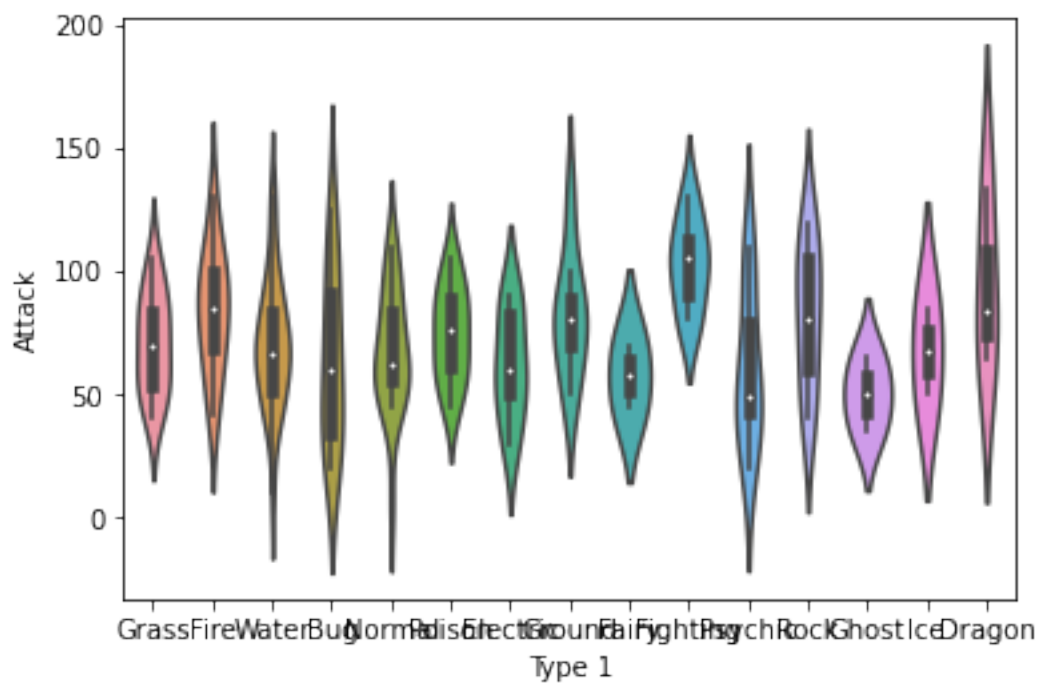




```
[34]: # violin plot
```

```
[35]: sns.violinplot(x='Type 1', y='Attack', data=df)
```

```
[35]: <AxesSubplot:xlabel='Type 1', ylabel='Attack'>
```

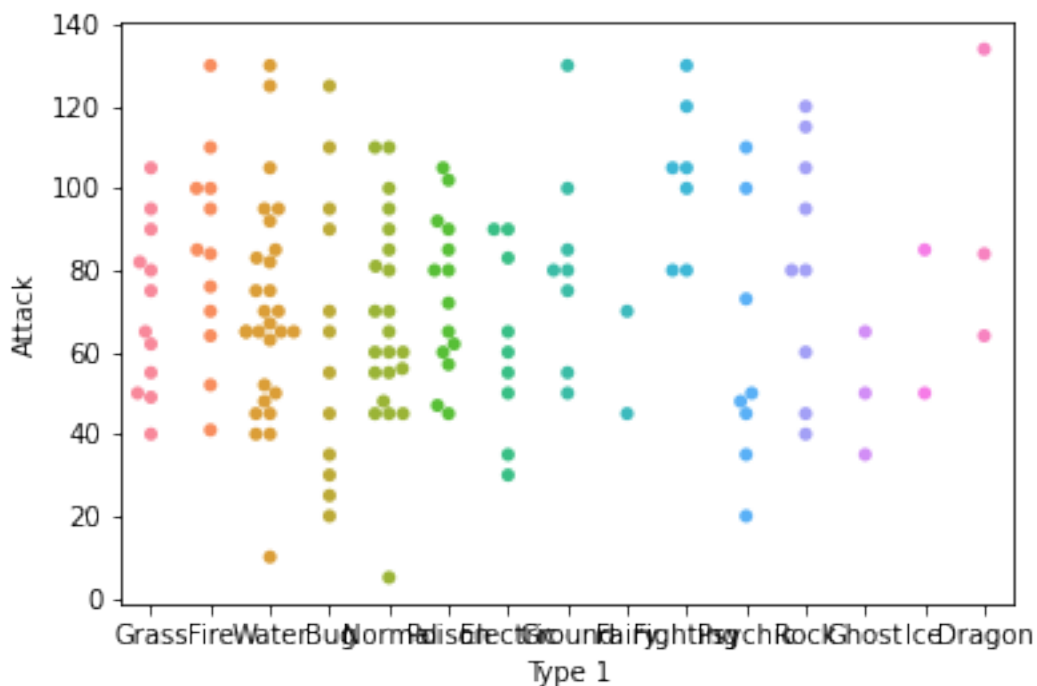


```
[36]: # swarmplot
```

```
[38]: sns.swarmplot(x='Type 1', y='Attack', data=df)
```

```
/Users/h4/anaconda3/lib/python3.9/site-packages/seaborn/categorical.py:1296:  
UserWarning: 10.7% of the points cannot be placed; you may want to decrease the  
size of the markers or use stripplot.  
warnings.warn(msg, UserWarning)
```

```
[38]: <AxesSubplot:xlabel='Type 1', ylabel='Attack'>
```



```
[39]: # plot overlay
```

```
[42]: # Set figure size with matplotlib  
plt.figure(figsize=(11,6))  
  
# Create plot  
sns.violinplot(x='Type 1',  
              y='Attack',  
              data=df, alpha=0.4)  
  
sns.swarmplot(x='Type 1',
```

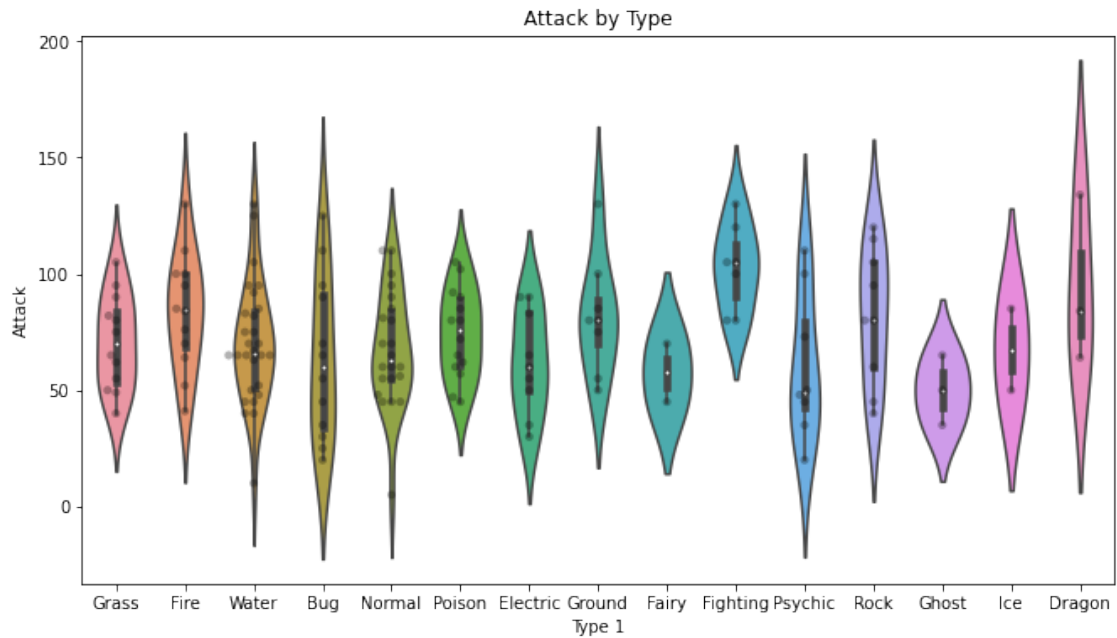
```

y='Attack',
data=df,
color='k', # Make points black
alpha=0.3) # and slightly transparent

# Set title with matplotlib
plt.title('Attack by Type')

```

```
[42]: Text(0.5, 1.0, 'Attack by Type')
```



```
[ ]:
```