

Portfolio assignment 12

30 min: Perform a bivariate analysis on at least 3 combinations of a numerical column with a categorical column in the dataset that you chose in portfolio assignment 4. Use `.groupby('columnname').mean()` to calculate the means. Is there a difference between categories? Then use seaborn barplots to check if there is a statistically significant difference.

In [1]:

```
import pandas as pd
import seaborn as sns
```

In [2]:

```
pokemons = pd.read_csv('../Pokemon.csv')
pokemons.drop(['#', 'Total', 'Legendary', 'Generation'], axis=1).groupby('Type 1').mean()
```

Out[2]:

	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed
Type 1						
Bug	56.884058	70.971014	70.724638	53.869565	64.797101	61.681159
Dark	66.806452	88.387097	70.225806	74.645161	69.516129	76.161290
Dragon	83.312500	112.125000	86.375000	96.843750	88.843750	83.031250
Electric	59.795455	69.090909	66.295455	90.022727	73.704545	84.500000
Fairy	74.117647	61.529412	65.705882	78.529412	84.705882	48.588235
Fighting	69.851852	96.777778	65.925926	53.111111	64.703704	66.074074
Fire	69.903846	84.769231	67.769231	88.980769	72.211538	74.442308
Flying	70.750000	78.750000	66.250000	94.250000	72.500000	102.500000
Ghost	64.437500	73.781250	81.187500	79.343750	76.468750	64.343750
Grass	67.271429	73.214286	70.800000	77.500000	70.428571	61.928571
Ground	73.781250	95.750000	84.843750	56.468750	62.750000	63.906250
Ice	72.000000	72.750000	71.416667	77.541667	76.291667	63.458333
Normal	77.275510	73.469388	59.846939	55.816327	63.724490	71.551020
Poison	67.250000	74.678571	68.821429	60.428571	64.392857	63.571429
Psychic	70.631579	71.456140	67.684211	98.403509	86.280702	81.491228
Rock	65.363636	92.863636	100.795455	63.340909	75.477273	55.909091
Steel	65.222222	92.703704	126.370370	67.518519	80.629630	55.259259
Water	72.062500	74.151786	72.946429	74.812500	70.517857	65.964286

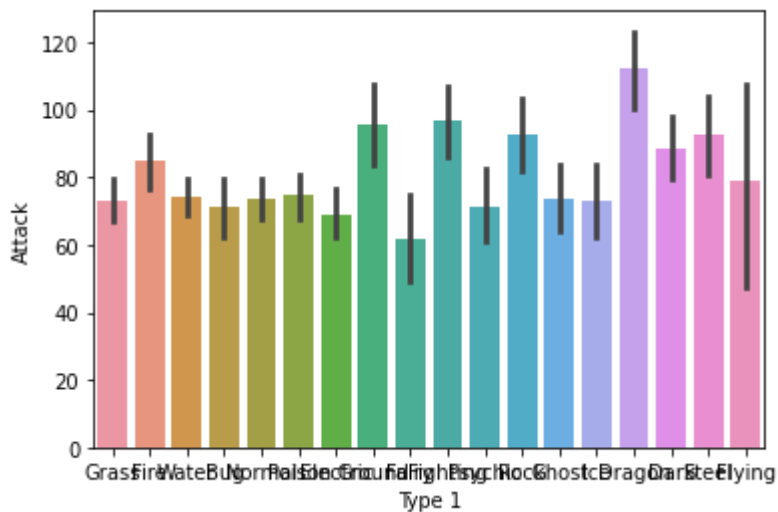
There is definitely a difference between the types of pokemon, but here is no way to determine if there's a reason for it or totally random.

In [3]:

```
sns.barplot(y="Attack", x="Type 1", data=pokemons)
```

Out[3]:

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



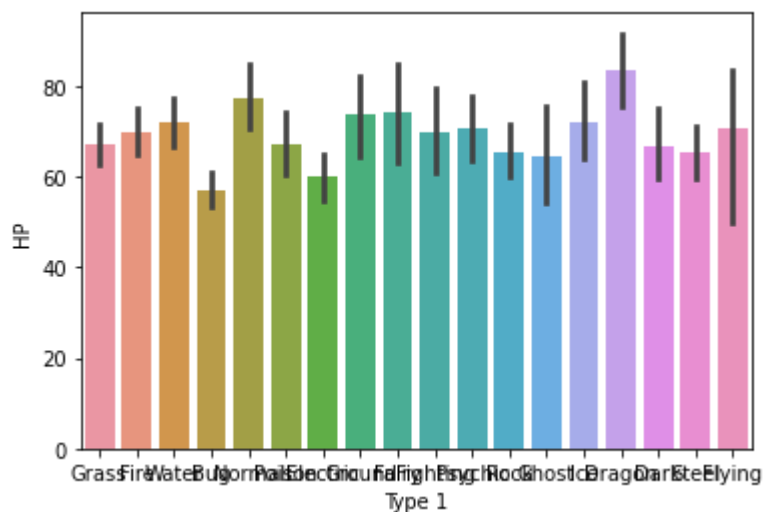
Dragon seems to be the type that had the highest Attack Stats, the Fighting, Ground and Steel followed after that.

In [4]:

```
sns.barplot(y="HP", x="Type 1", data=pokemons)
```

Out[4]:

```
<AxesSubplot:xlabel='Type 1', ylabel='HP'>
```



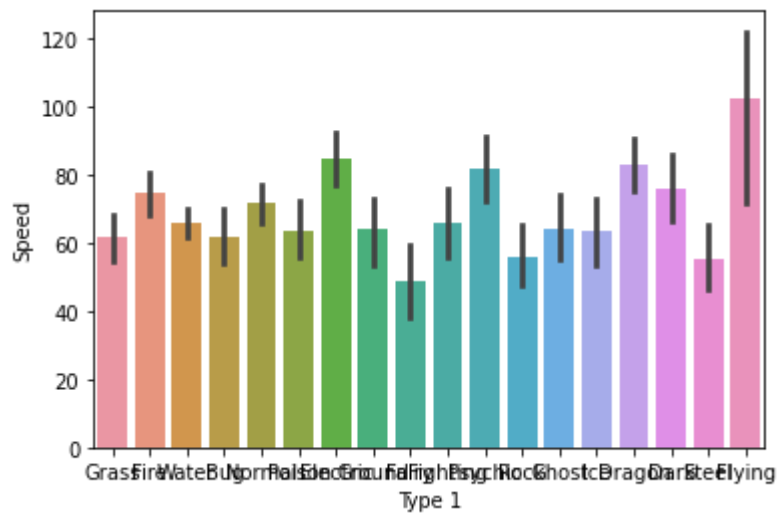
The HP on all Types seems to be more balanced than other stats in the dataset.

In [5]:

```
sns.barplot(y="Speed", x="Type 1", data=pokemons)
```

Out[5]:

<AxesSubplot:xlabel='Type 1', ylabel='Speed'>



The flying type has the highest speed stats, but there are other types that can reach the same amount.