2016 Rio Summer Olympics Analysis

Yasamin Yazdani, yy8429

Importing the Dataset

In [1]:
```
# First, I will need to import the pandas package.
import pandas as pd
```

In [18]:
```
# Now, I will import the dataset and take a look at it.
olympics = pd.read_csv('2016olympic.csv')
```

In [19]:
```
olympics.head()
```

Out[19]:

	country	gold2016	silver2016	bronze2016	nshemisphere
0	United States	46	37	38	northern
1	United Kingdom	27	23	17	northern
2	China	26	18	26	northern
3	Russia	19	18	19	northern
4	Germany	17	10	15	northern

This data comes from a website called "insidethegames" that includes world reports of data for different sports events. I used the data from the 2016 summer Olympics in Rio that includes the number of gold, silver, and bronze medals won by each country. There are 87 observations/countries. The variable "country" signifies the name of the country, "gold/silver/bronze.2016" signify the number of the different types of medals per country, and "ns.hemisphere" signifies whether the country was located in the northern or southern hemisphere (or both at the same time). I used a map to determine where the countries were located for the last variable.

Performing EDA

In [20]:
```
# First, I will need to import the seaborn package to make some visualizations a
import seaborn as sns
import numpy as np
```

In [21]:
```
# First, lets see the mean number of gold medals for the 2016 olympics
np.mean(olympics.gold2016)
```

Out[21]:
```
3.5172413793103448
```

The average number of gold medals per country was around 3-4 gold medals.

In [22]:
```
# Now, I will see the frequency of the number of countries per hemisphere.
olympics['nshemisphere'].value_counts()
```

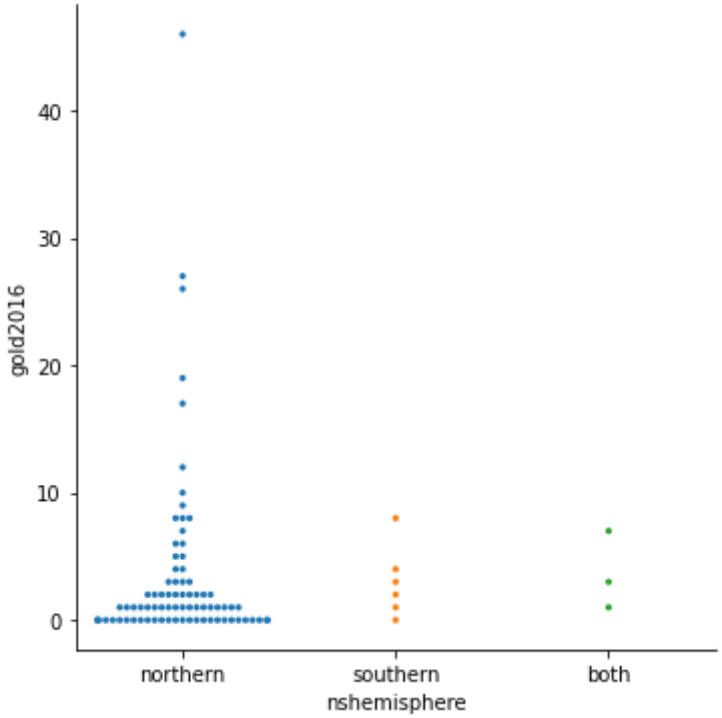
Out[22]:
```
northern    78
southern     6
```

```
both          3
Name: nshemisphere, dtype: int64
```

There are 78 countries from the northern hemisphere, 6 countries from the southern hemisphere, and 3 countries located in both hemispheres.

In [33]:
```
# Now, I will make a 'catplot' of the two variables to see the trend between the
sns.catplot(data=olympics, kind="swarm", x="nshemisphere", y="gold2016", s = 3.2
```

Out[33]:
```
<seaborn.axisgrid.FacetGrid at 0x7fb37e6946a0>
```



Most of the medals came from the northern hemisphere, but it is interesting to see how there are a few countries with extremely high gold medal counts and a significant amount with less than 5 gold medals.

In [ ]: