Data Frequency Analysis Notebook

This notebook will be used to investigate how usage of the word `data` has changed over the last few hundred years. First, I will need to impot my tools and data.

In [1]:
```python
# Import tools
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```python
# Import engram data
df = pd.read_csv('s3://bigdatabucket91621/eng_data_1gram.csv')

# Confirm import successful
df.head(5)
```
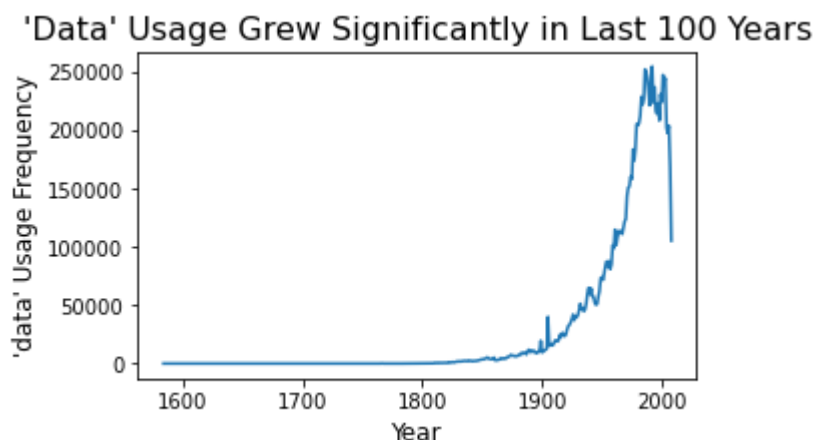
Out[2]:

	token	year	frequency	pages	books
0	data	1584	16	14	1
1	data	1614	3	2	1
2	data	1627	1	1	1
3	data	1631	22	18	1
4	data	1637	1	1	1

All imports appear to have been successful. I will now plot the usage frequency of the word `data` as a function of time (years).

In [3]:
```python
plt.figure(figsize=(5,3))
plt.plot(df['year'],df['frequency'])
plt.xlabel('Year', size=12)
plt.ylabel("'data' Usage Frequency", size=12)
plt.title("'Data' Usage Grew Significantly in Last 100 Years", size=16)
plt.show()
```


The usage of the word `data` was extremely low until about 1800, at which point it started to pick

up. From there, it increased slowly until about 1900, wherein usage of `data` increased exponentially until a few years before 2000. The word usage appears to be on a downturn since then.