Population Census In Ghana - Verification of Zipf's law

```
Work BY Appiah Boadu Prince Kwabena
```

Data from World Population Ranking

In [5]:

```
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline

df = pd.read_csv('realTime.csv')
data = pd.DataFrame(df)
data.index = data['Country Name']
data.drop('Country Name', axis=1, inplace=True)
data.drop('Indicator Name', axis=1, inplace=True)
data.drop('Indicator Code', axis=1, inplace=True)
data.drop('Country Code', axis=1, inplace=True)
data
```

Out[5]:

	1960	1961	1962	1963	1964	1965	1966	196
Country Name								
Aruba	54211.0	55438.0	56225.0	56695.0	57032.0	57360.0	57715.0	58055.
Afghanistan	8996973.0	9169410.0	9351441.0	9543205.0	9744781.0	9956320.0	10174836.0	10399926.
Angola	5454933.0	5531472.0	5608539.0	5679458.0	5735044.0	5770570.0	5781214.0	5774243.
Albania	1608800.0	1659800.0	1711319.0	1762621.0	1814135.0	1864791.0	1914573.0	1965598.
Andorra	13411.0	14375.0	15370.0	16412.0	17469.0	18549.0	19647.0	20758.
...	...	...	...	...	...	...	...	.
Kosovo	947000.0	966000.0	994000.0	1022000.0	1050000.0	1078000.0	1106000.0	1135000.
Yemen, Rep.	5315355.0	5393036.0	5473671.0	5556766.0	5641597.0	5727751.0	5816247.0	5907874.
South Africa	17099840.0	17524533.0	17965725.0	18423161.0	18896307.0	19384841.0	19888250.0	20406864.
Zambia	3070776.0	3164329.0	3260650.0	3360104.0	3463213.0	3570464.0	3681955.0	3797873.
Zimbabwe	3776681.0	3905034.0	4039201.0	4178726.0	4322861.0	4471177.0	4623351.0	4779827.

264 rows × 60 columns

Imported dataFrame with slight edits for Original Documents

File will be uploaded to my github repository for further integrations

This is the population growth per annum in Ghana

In [6]:
```python
data.loc['Ghana']
```

Out[6]:
```
1960             6635230.0
1961             6848295.0
1962             7071971.0
1963             7300116.0
1964             7524472.0
1965             7739473.0
1966             7941412.0
1967             8132803.0
1968             8321770.0
1969             8520015.0
1970             8735495.0
1971             8973244.0
1972             9229631.0
1973             9493556.0
1974             9749104.0
1975             9985946.0
1976            10199165.0
1977            10395452.0
1978            10590261.0
1979            10805314.0
1980            11056116.0
1981            11348289.0
1982            11676823.0
1983            12033575.0
1984            12405660.0
1985            12783613.0
1986            13164837.0
1987            13552021.0
1988            13947042.0
1989            14353410.0
1990            14773277.0
1991            15207367.0
1992            15653336.0
1993            16106765.0
1994            16561674.0
1995            17014057.0
1996            17462496.0
1997            17908985.0
1998            18357156.0
1999            18812359.0
2000            19278856.0
2001            19756928.0
2002            20246381.0
2003            20750299.0
2004            21272323.0
2005            21814642.0
2006            22379055.0
2007            22963946.0
2008            23563825.0
2009            24170940.0
2010            24779619.0
2011            25387710.0
2012            25996449.0
2013            26607642.0
2014            27224472.0
2015            27849205.0
2016            28481946.0
2017            29121471.0
2018            29767108.0
Unnamed: 63            NaN
Name: Ghana, dtype: float64
```

Zipf's law then predicts that out of a population of *N* elements, the normalized frequency of elements of rank *k*, *f*(*k*;*s*,*N*), is:

$$f(k;s,N) = \frac{1/k^s}{\sum_{n=1}^{N}(1/n^s)}$$

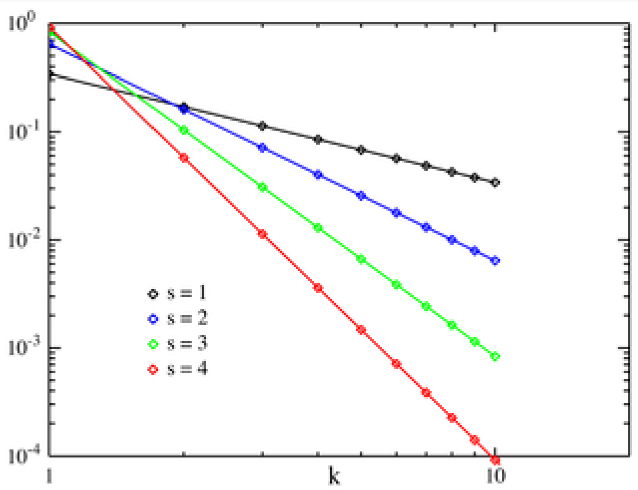
Zipf's law holds if the number of elements with a given frequency is a random variable with power law distribution $p(f) = \alpha f^{-1-1/s}$.[10]

This can be used in population growth and probabilistic analysis

With an output gragh similar to this below.

Zipf's law

Probability mass function

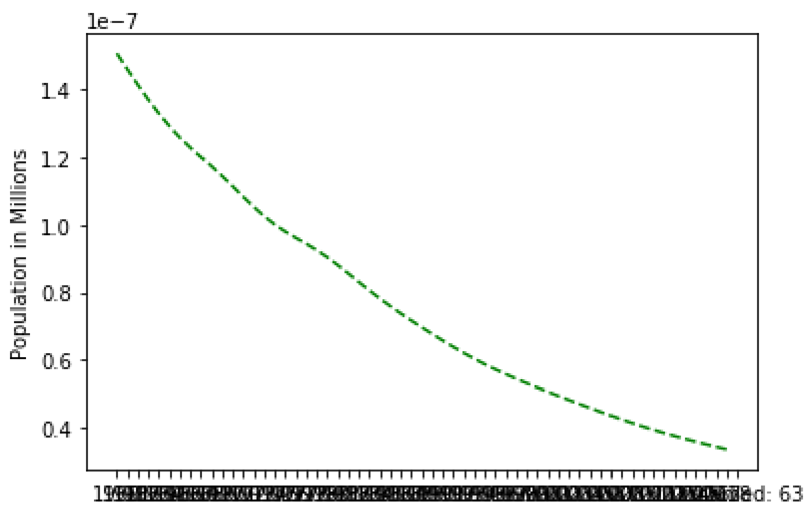


Zipf PMF for *N* = 10 on a log–log scale. The horizontal axis is the index *k* . (Note that the function is only defined at integer values of *k*. The connecting lines do not indicate continuity.)

GHANAIAN populis under the scope of the Zipf Law

In [7]:
```
zipf = 1 / data.loc['Ghana']
plt.plot(zipf, 'g--')
plt.ylabel("Population in Millions")
# size is inversely proportional to its rank
```

Out[7]: `Text(0, 0.5, 'Population in Millions')`



Original plot of Population against Years

In [11]:
```
plt.plot(data.loc['Ghana'])
plt.xlabel("Years")
plt.ylabel("Population in Millions")
```

Out[11]:
```
Text(0, 0.5, 'Population in Millions')
```