Assignment 9 NLP

Ubaid Ullah Tariq & Nafisa Ahmed

AIT - 580

Importing Essential Libraries:

In [1]:
```
import matplotlib as mpl
import matplotlib.pyplot as plt
import os
import nltk
from nltk.corpus import stopwords
nltk.download('stopwords')
stopwords = stopwords.words('english')
```

In [2]:
```
import nltk
from nltk.corpus import stopwords
nltk.download('stopwords')
stopwords = stopwords.words('english')
```

```
[nltk_data] Downloading package stopwords to C:\Users\ubaid.LAPTOP-60A
[nltk_data]     EGHFJ\AppData\Roaming\nltk_data...
[nltk_data]   Unzipping corpora\stopwords.zip.
```

In [3]:
```
os.getcwd()
os.chdir('C:/Users/ubaid.LAPTOP-60AEGHFJ/Desktop/Ait580')
os.getcwd()
```

Out[3]:
```
'C:\\Users\\ubaid.LAPTOP-60AEGHFJ\\Desktop\\Ait580'
```

Opening, Reading and Displaying the ABC News Input File:

In [2]:
```
# open, read, and display the ABC News input file
ABCNews_textfile = open('FDACovidVaccineKids_ABCNEWS.txt', mode='r', encoding='utf-8')
ABCNews_CovidVaccine = ABCNews_textfile.read()
print(ABCNews_CovidVaccine)
```

Output:

```
FDA authorizes COVID-19 vaccine for kids 5-11
If the CDC signs off, kids' vaccinations could begin next week.

FDA authorizes COVID-19 vaccine for children 5 to 11

Following Food and Drug Administration authorization, if the Centers for Disease Contro...Read More
Shawn Rocco/Duke University via Reuters
Another 28 million Americans are one step closer to getting vaccinated against COVID-19 after the Food and Drug Administrati
on on Friday authorized the Pfizer shot for 5- to 11-year-olds.

Children will be one of the last groups in the U.S. to become eligible for the vaccine. Protecting them against COVID-19 is
a major step in getting the country back on the path to normalcy after an unexpected late-summer surge that disproportionate
ly impacted unvaccinated Americans and filled hospitals to the brim.

Advertisement
```

Opening, Reading and Displaying the New York Times Input File:

```
In [3]:
# open, read, and display the New York Times input file
NYTimes_textfile = open('CovidShotsForChildren_NYTimes.txt', mode='r', encoding='utf-8')
NYTimes_CovidShots = NYTimes_textfile.read()
print(NYTimes_CovidShots)
```

Output:

```
Covid Shots Are a Go for Children, but Parents Are Reluctant to Consent
Vaccinating 5- to 11-year-olds could be a big step toward returning to normal life in the U.S., but even parents who got
the shot are worried about how it might affect their kids.



The Food and Drug Administration's authorization of a Covid-19 vaccine for ages 5 to 11 on Friday makes 28 million unvacc
inated children in the United States suddenly eligible for the shot and offers the country an opportunity to make big inr
oads in its efforts to achieve broad immunity against the coronavirus.

But in a nation that has already struggled mightily with Covid vaccine hesitancy, getting shots into those little arms ma
y present health authorities with the toughest vaccination challenge yet.

Even many parents who are themselves vaccinated and approved the shot for their teenagers are churning over whether to gi
ve consent for their younger children, questioning if the risk of the unknowns of a brand-new vaccine is worth it when mo
st coronavirus cases in youngsters are mild.
```

Extracting the Words and Converting all To Lower Case for ABC News File:

```
In [5]:
# extract the words, convert all to lower case for ABCNews_textfile(ABCNews_CovidVaccine)
from nltk.tokenize import RegexpTokenizer
tokenizer = RegexpTokenizer(r'\w+')
tokens = tokenizer.tokenize(ABCNews_CovidVaccine.lower())
print(tokens)
```

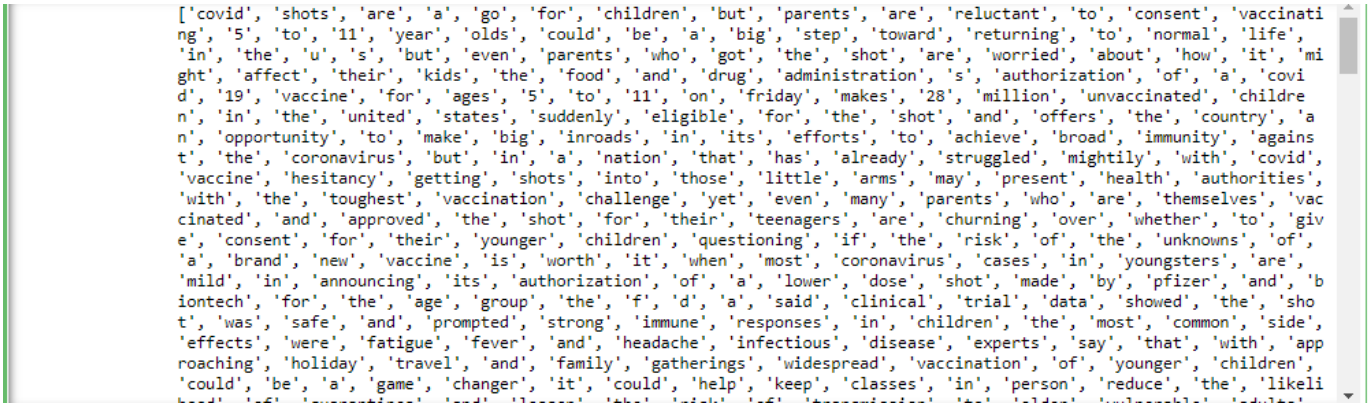
Output:

```
['fda', 'authorizes', 'covid', '19', 'vaccine', 'for', 'kids', '5', '11', 'if', 'the', 'cdc', 'signs', 'off', 'kids', 'vacci
nations', 'could', 'begin', 'next', 'week', 'fda', 'authorizes', 'covid', '19', 'vaccine', 'for', 'children', '5', 'to', '1
1', 'following', 'food', 'and', 'drug', 'administration', 'authorization', 'if', 'the', 'centers', 'for', 'disease', 'contr
o', 'read', 'more', 'shawn', 'rocco', 'duke', 'university', 'via', 'reuters', 'another', '28', 'million', 'americans', 'ar
e', 'one', 'step', 'closer', 'to', 'getting', 'vaccinated', 'against', 'covid', '19', 'after', 'the', 'food', 'and', 'drug',
'administration', 'on', 'friday', 'authorized', 'the', 'pfizer', 'shot', 'for', '5', 'to', '11', 'year', 'olds', 'children',
'will', 'be', 'one', 'of', 'the', 'last', 'groups', 'in', 'the', 'u', 's', 'to', 'become', 'eligible', 'for', 'the', 'vaccin
e', 'protecting', 'them', 'against', 'covid', '19', 'is', 'a', 'major', 'step', 'in', 'getting', 'the', 'country', 'back',
'on', 'the', 'path', 'to', 'normalcy', 'after', 'an', 'unexpected', 'late', 'summer', 'surge', 'that', 'disproportionately',
'impacted', 'unvaccinated', 'americans', 'and', 'filled', 'hospitals', 'to', 'the', 'brim', 'advertisement', 'the', 'rationa
le', 'here', 'is', 'protect', 'your', 'children', 'so', 'that', 'they', 'can', 'get', 'back', 'towards', 'normal', 'lives',
'said', 'dr', 'peter', 'marks', 'the', 'fda', 's', 'vaccine', 'chief', 'in', 'a', 'press', 'conference', 'after', 'the', 'au
thorization', 'was', 'announced', 'the', 'tremendous', 'cost', 'of', 'this', 'pandemic', 'has', 'not', 'just', 'been', 'in',
'physical', 'illness', 'it', 's', 'been', 'in', 'the', 'psychological', 'the', 'social', 'development', 'of', 'children', 't
he', 'process', 'now', 'heads', 'to', 'the', 'centers', 'for', 'disease', 'control', 'and', 'prevention', 'an', 'advisory',
'committee', 'for', 'the', 'cdc', 'will', 'meet', 'on', 'tuesday', 'to', 'discuss', 'the', 'pediatric', 'vaccine', 'safety',
'and', 'efficacy', 'data', 'as', 'an', 'advisory', 'fda', 'panel', 'did', 'this', 'past', 'week', 'and', 'then', 'cdc', 'dir
ector', 'rochelle', 'walensky', 'is', 'expected', 'to', 'give', 'the', 'final', 'signoff', 'soon', 'afterward', 'that', 'mea
ns', 'kids', 'could', 'begin', 'getting', 'shots', 'at', 'some', 'point', 'next', 'week', 'and', 'become', 'fully', 'vaccina
ted', 'by', 'december', 'in', 'anticipation', 'the', 'white', 'house', 'planned', 'to', 'unleash', 'millions', 'of', 'vaccin
e', 'shipments', 'across', 'the', 'nation', 'as', 'soon', 'as', 'fda', 'authorization', 'was', 'announced', 'vaccine', 'site
s', 'will', 'have', 'to', 'wait', 'for', 'the', 'cdc', 's', 'word', 'to', 'begin', 'administering', 'the', 'vaccine', 'but',
'stock', 'will', 'be', 'on', 'hand', 'more', 'fda', 'panel', 'greenlights', 'vaccines', 'for', 'kids', 'paving', 'the', 'wa
y', 'for', 'authorization', 'the', 'bottom', 'line', 'is', 'that', 'we', 'will', 'be', 'ready', 'immediately', 'following',
```

Extracting the Words and Converting all To Lower Case for New York Times File:

In [6]:
```
# extract the words, convert all to lower case for NYTimes_textfile(NYTimes_CovidShots)
from nltk.tokenize import RegexpTokenizer
tokenizer = RegexpTokenizer(r'\w+')
NYTimes_tokens2 = tokenizer.tokenize(NYTimes_CovidShots.lower())
print(NYTimes_tokens2)
```

Output:



```
['covid', 'shots', 'are', 'a', 'go', 'for', 'children', 'but', 'parents', 'are', 'reluctant', 'to', 'consent', 'vaccinati
ng', '5', 'to', '11', 'year', 'olds', 'could', 'be', 'a', 'big', 'step', 'toward', 'returning', 'to', 'normal', 'life',
'in', 'the', 'u', 's', 'but', 'even', 'parents', 'who', 'got', 'the', 'shot', 'are', 'worried', 'about', 'how', 'it', 'mi
ght', 'affect', 'their', 'kids', 'the', 'food', 'and', 'drug', 'administration', 's', 'authorization', 'of', 'a', 'covi
d', '19', 'vaccine', 'for', 'ages', '5', 'to', '11', 'on', 'friday', 'makes', '28', 'million', 'unvaccinated', 'childre
n', 'in', 'the', 'united', 'states', 'suddenly', 'eligible', 'for', 'the', 'shot', 'and', 'offers', 'the', 'country', 'a
n', 'opportunity', 'to', 'make', 'big', 'inroads', 'in', 'its', 'efforts', 'to', 'achieve', 'broad', 'immunity', 'agains
t', 'the', 'coronavirus', 'but', 'in', 'a', 'nation', 'that', 'has', 'already', 'struggled', 'mightily', 'with', 'covid',
'vaccine', 'hesitancy', 'getting', 'shots', 'into', 'those', 'little', 'arms', 'may', 'present', 'health', 'authorities',
'with', 'the', 'toughest', 'vaccination', 'challenge', 'yet', 'even', 'many', 'parents', 'who', 'are', 'themselves', 'vac
cinated', 'and', 'approved', 'the', 'shot', 'for', 'their', 'teenagers', 'are', 'churning', 'over', 'whether', 'to', 'giv
e', 'consent', 'for', 'their', 'younger', 'children', 'questioning', 'if', 'the', 'risk', 'of', 'the', 'unknowns', 'of',
'a', 'brand', 'new', 'vaccine', 'is', 'worth', 'it', 'when', 'most', 'coronavirus', 'cases', 'in', 'youngsters', 'are',
'mild', 'in', 'announcing', 'its', 'authorization', 'of', 'a', 'lower', 'dose', 'shot', 'made', 'by', 'pfizer', 'and', 'b
iontech', 'for', 'the', 'age', 'group', 'the', 'f', 'd', 'a', 'said', 'clinical', 'trial', 'data', 'showed', 'the', 'sho
t', 'was', 'safe', 'and', 'prompted', 'strong', 'immune', 'responses', 'in', 'children', 'the', 'most', 'common', 'side',
'effects', 'were', 'fatigue', 'fever', 'and', 'headache', 'infectious', 'disease', 'experts', 'say', 'that', 'with', 'app
roaching', 'holiday', 'travel', 'and', 'family', 'gatherings', 'widespread', 'vaccination', 'of', 'younger', 'children',
'could', 'be', 'a', 'game', 'changer', 'it', 'could', 'help', 'keep', 'classes', 'in', 'person', 'reduce', 'the', 'likeli
```

Recreating Token List Without Stop Words for ABC News Text File:

In [10]:
```
# recreate token list without stopwords for ABCNews_textfile(ABCNews_CovidVaccine)
from nltk.corpus import stopwords
token = [token for token in token if token not in stopwords.words('english')]
print(token)
```

Output:

```
['fda', 'authorizes', 'covid', '19', 'vaccine', 'kids', '5', '11', 'cdc', 'signs', 'kids', 'vaccinations', 'could', 'begin',
'next', 'week', 'fda', 'authorizes', 'covid', '19', 'vaccine', 'children', '5', '11', 'following', 'food', 'drug', 'administ
ration', 'authorization', 'centers', 'disease', 'contro', 'read', 'shawn', 'rocco', 'duke', 'university', 'via', 'reuters',
'another', '28', 'million', 'americans', 'one', 'step', 'closer', 'getting', 'vaccinated', 'covid', '19', 'food', 'drug', 'a
dministration', 'friday', 'authorized', 'pfizer', 'shot', '5', '11', 'year', 'olds', 'children', 'one', 'last', 'groups',
'u', 'become', 'eligible', 'vaccine', 'protecting', 'covid', '19', 'major', 'step', 'getting', 'country', 'back', 'path', 'n
ormalcy', 'unexpected', 'late', 'summer', 'surge', 'disproportionately', 'impacted', 'unvaccinated', 'americans', 'filled',
'hospitals', 'brim', 'advertisement', 'rationale', 'protect', 'children', 'get', 'back', 'towards', 'normal', 'lives', 'sai
d', 'dr', 'peter', 'marks', 'fda', 'vaccine', 'chief', 'press', 'conference', 'authorization', 'announced', 'tremendous', 'c
ost', 'pandemic', 'physical', 'illness', 'psychological', 'social', 'development', 'children', 'process', 'heads', 'center
s', 'disease', 'control', 'prevention', 'advisory', 'committee', 'cdc', 'meet', 'tuesday', 'discuss', 'pediatric', 'vaccin
e', 'safety', 'efficacy', 'data', 'advisory', 'fda', 'panel', 'past', 'week', 'cdc', 'director', 'rochelle', 'walensky', 'ex
pected', 'give', 'final', 'signoff', 'soon', 'afterward', 'means', 'kids', 'could', 'begin', 'getting', 'shots', 'point', 'n
ext', 'week', 'become', 'fully', 'vaccinated', 'december', 'anticipation', 'white', 'house', 'planned', 'unleash', 'million
s', 'vaccine', 'shipments', 'across', 'nation', 'soon', 'fda', 'authorization', 'announced', 'vaccine', 'sites', 'wait', 'cd
c', 'word', 'begin', 'administering', 'vaccine', 'stock', 'hand', 'fda', 'panel', 'greenlights', 'vaccines', 'kids', 'pavin
g', 'way', 'authorization', 'bottom', 'line', 'ready', 'immediately', 'following', 'fda', 'cdc', 'decisions', 'parents', 'ge
t', 'kids', 'vaccinated', 'quickly', 'easily', 'conveniently', 'white', 'house', 'covid', '19', 'coordinator', 'jeff', 'zien
ts', 'said', 'briefing', 'reporters', 'thursday', 'white', 'house', 'purchased', 'enough', 'shots', '28', 'million', '5', '1
1', 'year', 'olds', 'thursday', 'announced', 'plans', 'purchase', 'another', '50', 'million', 'shots', 'april', '30', '202
2', 'could', 'also', 'used', 'children', '5', 'authorization', 'age', 'group', 'vaccines', 'available', 'pharmacies', 'pedia
```

Recreating Token List Without Stop Words for New York Times Text File:

```
In [3]:
# recreate token list without stopwords for NYTimes_textfile(NYTimes_CovidShots)
from nltk.corpus import stopwords
NYTimes_tokens2 = [token for token in NYTimes_tokens2 if token not in stopwords.words('english')]
print(NYTimes_tokens2)
```

Output:

```
['covid', 'shots', 'are', 'a', 'go', 'for', 'children', 'but', 'parents', 'are', 'reluctant', 'to', 'consent', 'vaccinati
ng', '5', 'to', '11', 'year', 'olds', 'could', 'be', 'a', 'big', 'step', 'toward', 'returning', 'to', 'normal', 'life',
'in', 'the', 'u', 's', 'but', 'even', 'parents', 'who', 'got', 'the', 'shot', 'are', 'worried', 'about', 'how', 'it', 'mi
ght', 'affect', 'their', 'kids', 'the', 'food', 'and', 'drug', 'administration', 's', 'authorization', 'of', 'a', 'covi
d', '19', 'vaccine', 'for', 'ages', '5', 'to', '11', 'on', 'friday', 'makes', '28', 'million', 'unvaccinated', 'childre
n', 'in', 'the', 'united', 'states', 'suddenly', 'eligible', 'for', 'the', 'shot', 'and', 'offers', 'the', 'country', 'a
n', 'opportunity', 'to', 'make', 'big', 'inroads', 'in', 'its', 'efforts', 'to', 'achieve', 'broad', 'immunity', 'agains
t', 'the', 'coronavirus', 'but', 'in', 'a', 'nation', 'that', 'has', 'already', 'struggled', 'mightily', 'with', 'covid',
'vaccine', 'hesitancy', 'getting', 'shots', 'into', 'those', 'little', 'arms', 'may', 'present', 'health', 'authorities',
'with', 'the', 'toughest', 'vaccination', 'challenge', 'yet', 'even', 'many', 'parents', 'who', 'are', 'themselves', 'vac
cinated', 'and', 'approved', 'the', 'shot', 'for', 'their', 'teenagers', 'are', 'churning', 'over', 'whether', 'to', 'giv
e', 'consent', 'for', 'their', 'younger', 'children', 'questioning', 'if', 'the', 'risk', 'of', 'the', 'unknowns', 'of',
'a', 'brand', 'new', 'vaccine', 'is', 'worth', 'it', 'when', 'most', 'coronavirus', 'cases', 'in', 'youngsters', 'are',
'mild', 'in', 'announcing', 'its', 'authorization', 'of', 'a', 'lower', 'dose', 'shot', 'made', 'by', 'pfizer', 'and', 'b
iontech', 'for', 'the', 'age', 'group', 'the', 'f', 'd', 'a', 'said', 'clinical', 'trial', 'data', 'showed', 'the', 'sho
t', 'was', 'safe', 'and', 'prompted', 'strong', 'immune', 'responses', 'in', 'children', 'the', 'most', 'common', 'side',
'effects', 'were', 'fatigue', 'fever', 'and', 'headache', 'infectious', 'disease', 'experts', 'say', 'that', 'with', 'app
```

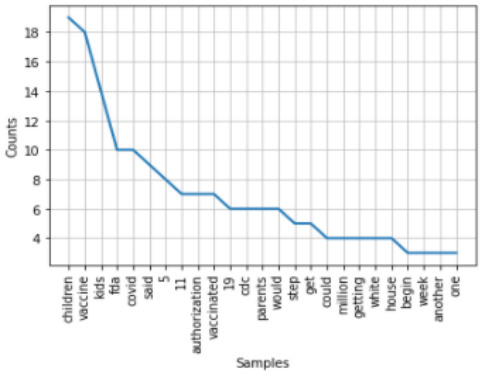
Displaying the Graph, Word Frequency and Few Specific Words for ABC News Text File:

```
In [12]:
# display and graph the word frequncies, plus a few specific words
freq_dist = nltk.FreqDist(tokens)
freq_dist
freq_dist['covid']
freq_dist['vaccine']
freq_dist['children']

print(freq_dist)
print(freq_dist.most_common(25))
freq_dist.plot(25)
```

Output:

```
<FreqDist with 365 samples and 600 outcomes>
[('children', 19), ('vaccine', 18), ('kids', 14), ('fda', 10), ('covid', 10), ('said', 9), ('5', 8), ('11', 7), ('authorizat
ion', 7), ('vaccinated', 7), ('19', 6), ('cdc', 6), ('parents', 6), ('would', 6), ('step', 5), ('get', 5), ('could', 4), ('m
illion', 4), ('getting', 4), ('white', 4), ('house', 4), ('begin', 3), ('week', 3), ('another', 3), ('one', 3)]
```




```
Out[12]: <AxesSubplot:xlabel='Samples', ylabel='Counts'>
```

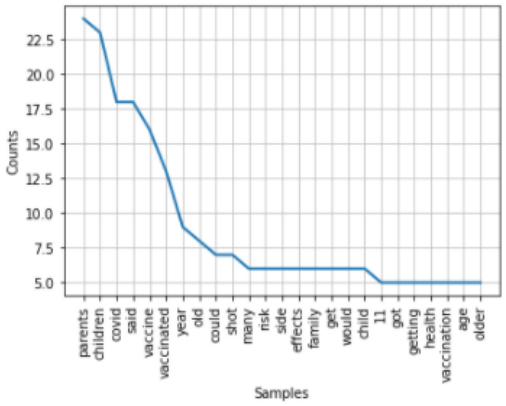
Displaying the Graph, Word Frequency and Few Specific Words for New York Times Text File:

```
In [13]:
# display and graph the word frequncies, plus a few specific words
freq_dist = nltk.FreqDist(NYTimes_tokens2)
freq_dist
freq_dist['covid']
freq_dist['vaccine']
freq_dist['children']

print(freq_dist)
print(freq_dist.most_common(25))
freq_dist.plot(25)
```

Output:

```
<FreqDist with 614 samples and 992 outcomes>
[('parents', 24), ('children', 23), ('covid', 18), ('said', 18), ('vaccine', 16), ('vaccinated', 13), ('year', 9), ('old',
8), ('could', 7), ('shot', 7), ('many', 6), ('risk', 6), ('side', 6), ('effects', 6), ('family', 6), ('get', 6), ('would',
6), ('child', 6), ('11', 5), ('got', 5), ('getting', 5), ('health', 5), ('vaccination', 5), ('age', 5), ('older', 5)]
```




```
Out[13]: <AxesSubplot:xlabel='Samples', ylabel='Counts'>
```

Difference Between Two Articles:

- Frequency count for **NY Times** text file is higher than the **ABC News** text file, as can be clearly seen in the word frequency count outcome of both text files.

- Specific words such as Covid, Vaccine, and Children has a higher word count in **NY Times** file as compared to the **ABC News** file.

References

Haslett, C. (2021, October 29). FDA authorizes COVID-19 vaccine for kids 5-11. Retrieved October 31, 2021, from https://abcnews.go.com/Politics/fda-authorizes-covid-19-vaccine-kids-11/story?id=80846188

Hoffman, J. (2021, October 30). Covid Shots Are a Go for Children, but Parents Are Reluctant to Consent. Retrieved October 31, 2021, from https://www.nytimes.com/2021/10/30/health/covid-vaccine-kids-parents.html