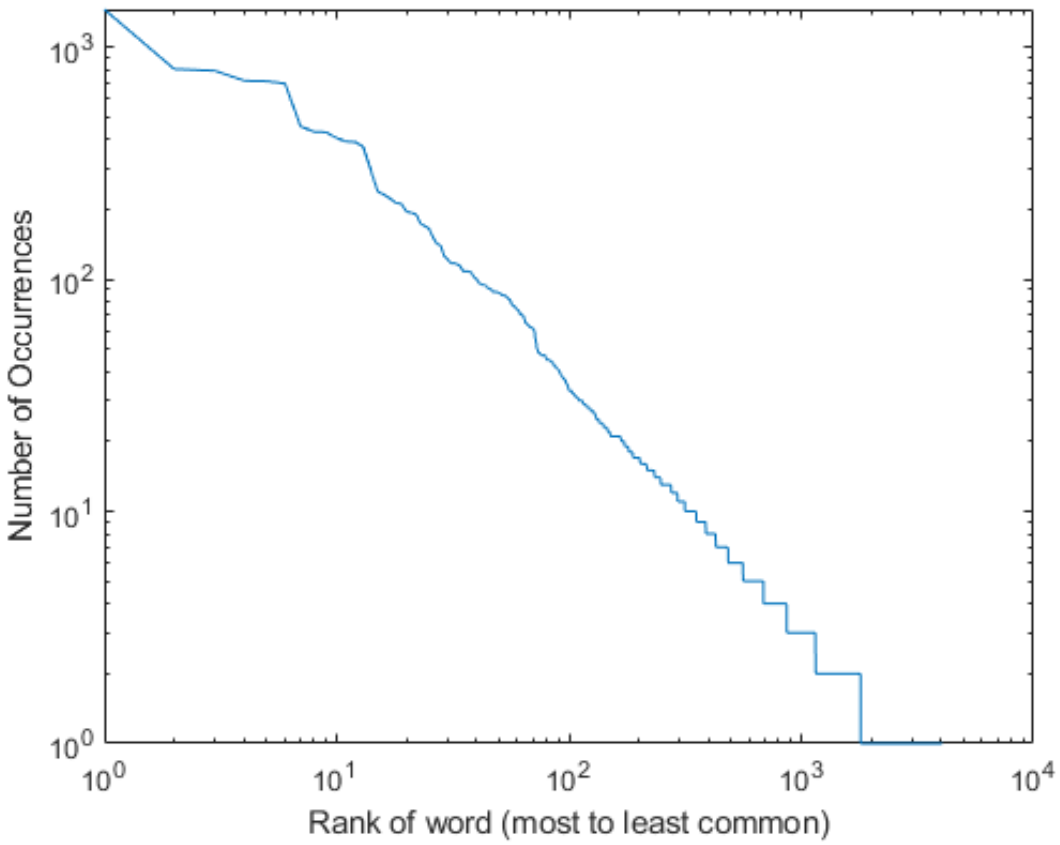
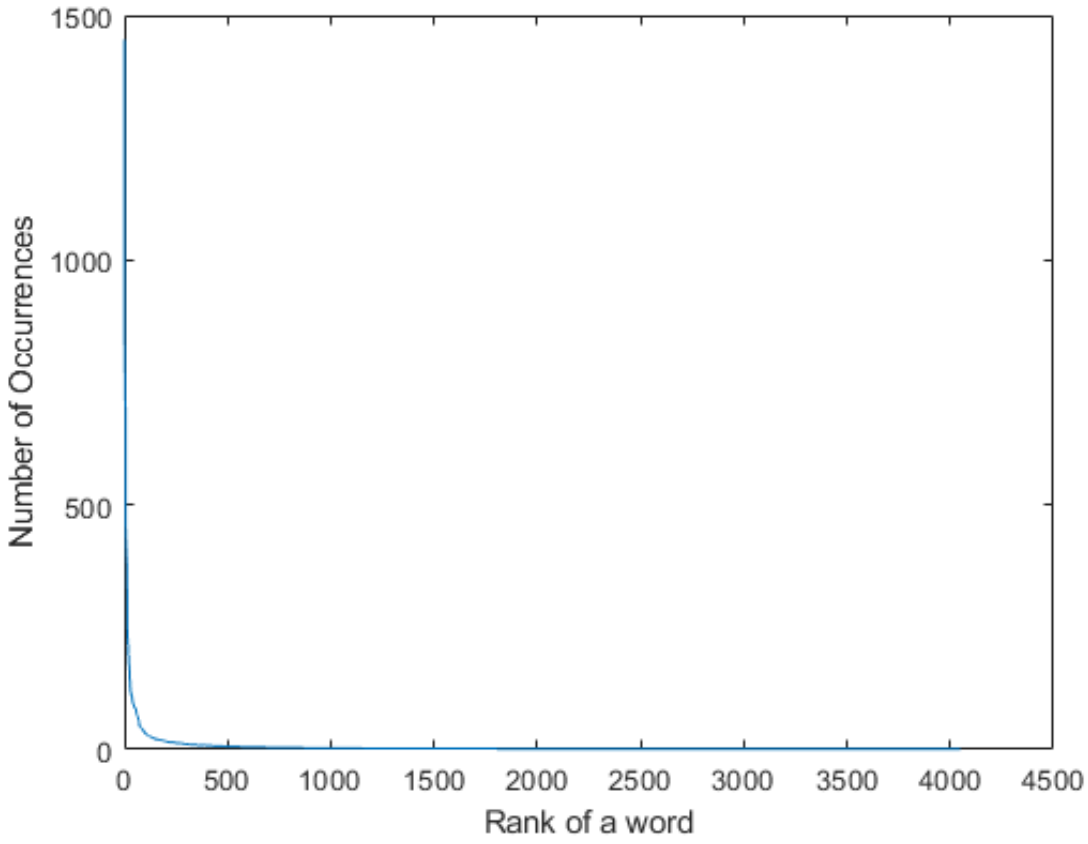
```
clc; clear; close all;

zipf();
```



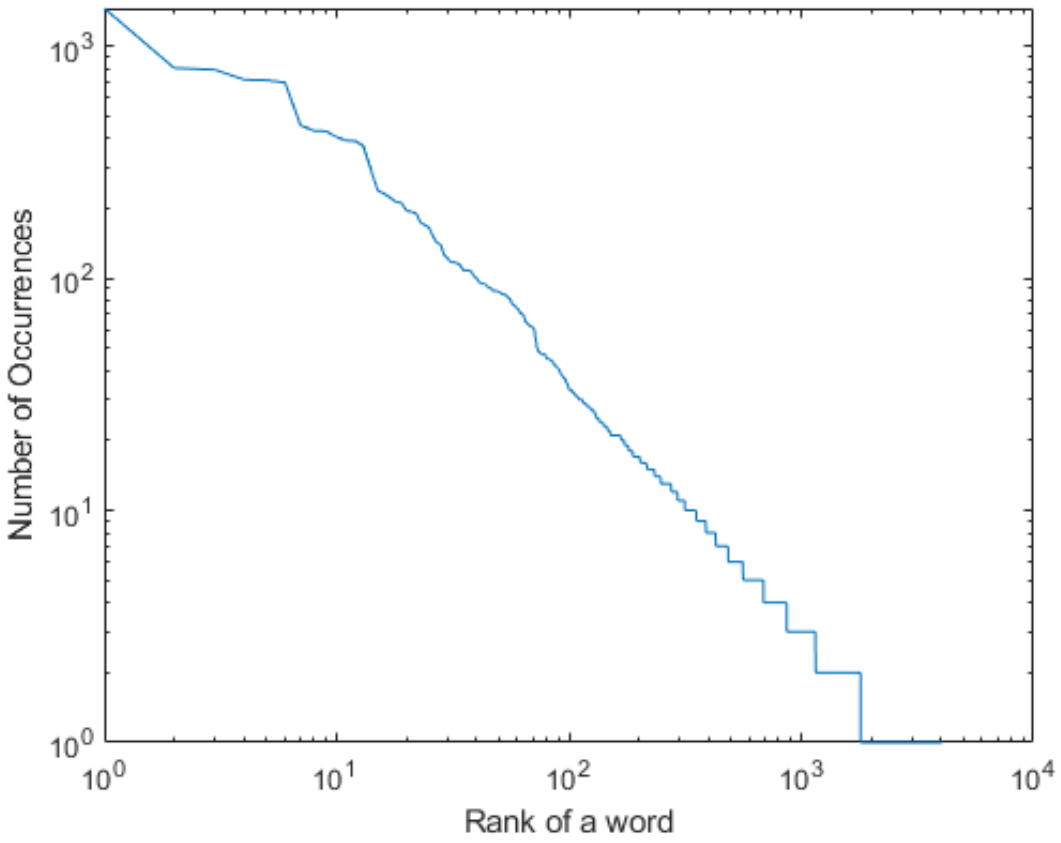
ans = 10×4 table

	Words	NumOccurren...	PercentOfText	CumulativePe...
1	"the"	1450	5.2846	5.2846
2	"and"	802	2.9230	8.2076
3	"i"	789	2.8756	11.0832
4	"to"	714	2.6022	13.6854
5	"of"	711	2.5913	16.2767
6	"a"	695	2.5330	18.8097
7	"it"	454	1.6546	20.4643
8	"in"	431	1.5708	22.0351
9	"that"	429	1.5635	23.5987
10	"he"	405	1.4761	25.0747

```
%5a)
plot(1:size(words), numOccurrences);
xlabel('Rank of a word');
ylabel('Number of Occurrences');
```




```
%Better Plotted on log graph
loglog(numOccurrences);
xlabel('Rank of a word');
ylabel('Number of Occurrences');
```




```
%5b)
%Top 10 words based on number of occurrences
words(rankIndex(1:10))
```

```
ans = 10×1 string array
    "the"
    "and"
    "i"
    "to"
    "of"
    "a"
    "it"
    "in"
    "that"
    "he"
```

```
%5c)
% frequency of most common word
fMostCommon = numOccurrences(1)
```

```
fMostCommon = 1450
```

```
% N = 5
f5CommonMeasured = numOccurrences(5)
```

```
f5CommonMeasured = 711
```

```
f5CommonCalculated = fMostCommon * 1/5
```

```
f5CommonCalculated = 290
```

```
% N = 10
f10CommonMeasured = numOccurrences(10)
```

```
f10CommonMeasured = 405
```

```
f10CommonCalculated = fMostCommon * 1/10
```

```
f10CommonCalculated = 145
```

```
% while the raw data may seem to differ by a significant amount, if the
% calculated and measured values are compared in relation to the log-log
% graphical representation, it can be seen that Zipf's Law holds such that
% the frequency of the nth most common word is approximately 1/n times the
% frequency of the most common word.
```