# Visualizing Textual Data



Matthew Renze
SOFTWARE CONSULTANT

@matthewrenze www.matthewrenze.com



# Overview



**Visualizing Textual Data** 

Demo



# Textual Data Analysis

Body of text

Extract new information

Patterns and trends

Unstructured data

To be, or not to be: that is the question:
Whether 'tis nobler in the mind to suffer
The slings and arrows of outrageous fortune,
Or to take arms against a sea of troubles,
And by opposing end them? To die: to sleep;
No more; and by a sleep to say we end
The heart-ache and the thousand natural shocks
That flesh is heir to, 'tis a consummation
Devoutly to be wish'd. To die, to sleep;
To sleep: perchance to dream: ay, there's the rub;
For in that sleep of death what dreams may come



To be, or not to be: that is the question:
Whether 'tis nobler in the mind to suffer
The slings and arrows of outrageous fortune,
Or to take arms against a sea of troubles,
And by opposing end them? To die: to sleep;
No more; and by a sleep to say we end
The heart-ache and the thousand natural shocks
That flesh is heir to, 'tis a consummation
Devoutly to be wish'd. To die, to sleep;
To sleep: perchance to dream: ay, there's the rub;
For in that sleep of death what dreams may come



#### Convert to lower case

to be, or not to be: that is the question: whether 'tis nobler in the mind to suffer the slings and arrows of outrageous fortune, or to take arms against a sea of troubles, and by opposing end them? to die: to sleep; no more; and by a sleep to say we end the heart-ache and the thousand natural shocks that flesh is heir to, 'tis a consummation devoutly to be wish'd. to die, to sleep; to sleep: perchance to dream: ay, there's the rub; for in that sleep of death what dreams may come.



# Convert to lower case Remove punctuation

to be or not to be that is the question whether 'tis nobler in the mind to suffer the slings and arrows of outrageous fortune or to take arms against a sea of troubles and by opposing end them to die to sleep no more and by a sleep to say we end the heartache and the thousand natural shocks that flesh is heir to 'tis a consummation devoutly to be wish'd to die to sleep to sleep perchance to dream ay there's the rub for in that sleep of death what dreams may come



Convert to lower case
Remove punctuation
Remove stopwords

question
whether nobler mind suffer
slings arrows outrageous fortune
take arms sea troubles
opposing end die sleep
sleep say end
heartache thousand natural shocks
flesh heir consummation
devoutly wish'd die sleep
sleep perchance dream rub
sleep death dreams may come



Convert to lower case
Remove punctuation
Remove stopwords
Reduce to stem words

whether noble mind suffer sling arrow outrag fortun take arm sea troubl oppos end die sleep sleep say end heartach thousand natur shock flesh heir consumm devout wish die sleep sleep perchanc dream rub sleep death dream may come



Convert to lower case
Remove punctuation
Remove stopwords
Reduce to stem words
Strip extra whitespace

question whether noble mind suffer sling arrow outrag fortun take arm sea troubl oppos end die sleep sleep say end heartach thousand natur shock flesh heir consumm devout wish die sleep sleep perchanc dream rub sleep death dream may come



Convert to lower case

Remove punctuation

Remove stopwords

Reduce to stem words

Strip extra whitespace

**Convert to frequencies** 

Term	Frequency
sleep	5
die	2
dream	2
end	2
arm	1
arrow	1
come	1



### Structured Text Data

Bag-of-words model
Frequency of terms
Additional variables

Term	Frequency
aaa	1
aang	1
aaron	2
aback	1
abagnale	1
abandon	1
abandoned	10



# Frequency Word Cloud

Visualization of terms
Sized by frequency
Random or ordered

#### Plot Keywords by Frequency



### Quantitative Word Cloud

# Sized by numeric variable Add a main chart title

#### Plot Keywords by Average Box Office Revenue

```
gadgetry succumbs
guerrilla shatters paraplegic shane bese showmanh frodo anakin aragorn frodo anakin aragorn panemminion reindeer obi transformers of voluntarily of o
                                                                                                                                                                                                                                                            peacekeepingmightiest
```



### Colored Word Cloud

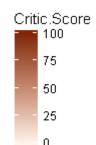
Sized by 1<sup>st</sup> variable

Colored by 2<sup>nd</sup> variable

Add title and color scale

#### Plot Keywords by Average Box Office Revenue and Critic Score

```
hermione modified mightiest enacting voldemort enacting voldemort enacting voldemort enacting peacekeeping dragged voluntarily mellark televised looked voluntarily enactions and enactions attices to be a second enaction of the control of the cont
                                                                                            attic shifty jurassic rebuilding is actional sattic shifty jurassic rebuilding is saruman ith witwicky object donkey object donkey is engard in emotions of a dormant shatters peeta agaze armor was panem districts skulls and all satticts in a doctor stark of addetor stark of a doctor stark of a docto
                                              containment
                                                                                                                                                          everdeengotham nazareth
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 joker frodo
                                                                                                                                   accuracyweaponizedenslaving
                                                                      decepticons gru install spacecraft sidious darth
                                                                                                    bane paraplegic reindeer snowman catwoman
```





### Textual Data Visualization

```
findslives stop first help smallbest life american life old seen around friends forced become falls come one city war u.s. begins can around friends forced become falls come one city war u.s. begins can around friends forced become falls come one city war u.s. begins can around friends forced become falls come one city war u.s. begins can around friends forced become together tries one city war u.s. begins can around friends forced become set last drugtime becomes manget town set based college manget town set by trying couple make living living logic whose son logets of group order follows meets school high story former discovers years
```

```
gadgetry succumbs

enslaving containment

kathiss spacecraft

grufanatical woody

gandarweren hordes of cybertronian gaze ogres

invite jurassic of countiess padme mellark

cybertronian gaze ogres

invite jurassic of cybertronian gaze ogres
```







- 1. What are the top keywords by frequency?
- 2. What are the top keywords by box office?
- 3. What are the top keywords by critic score?



# Load the Plots Data



# Create a Corpus



# Transform the Corpus



# Create a Frequency Word Cloud



# Load the Words Data



### Create Quantitative Word Cloud



# Create a Color Palette



# Create a Colored Word Cloud





# Summary



**Visualizing Textual Data** 

Demo

