



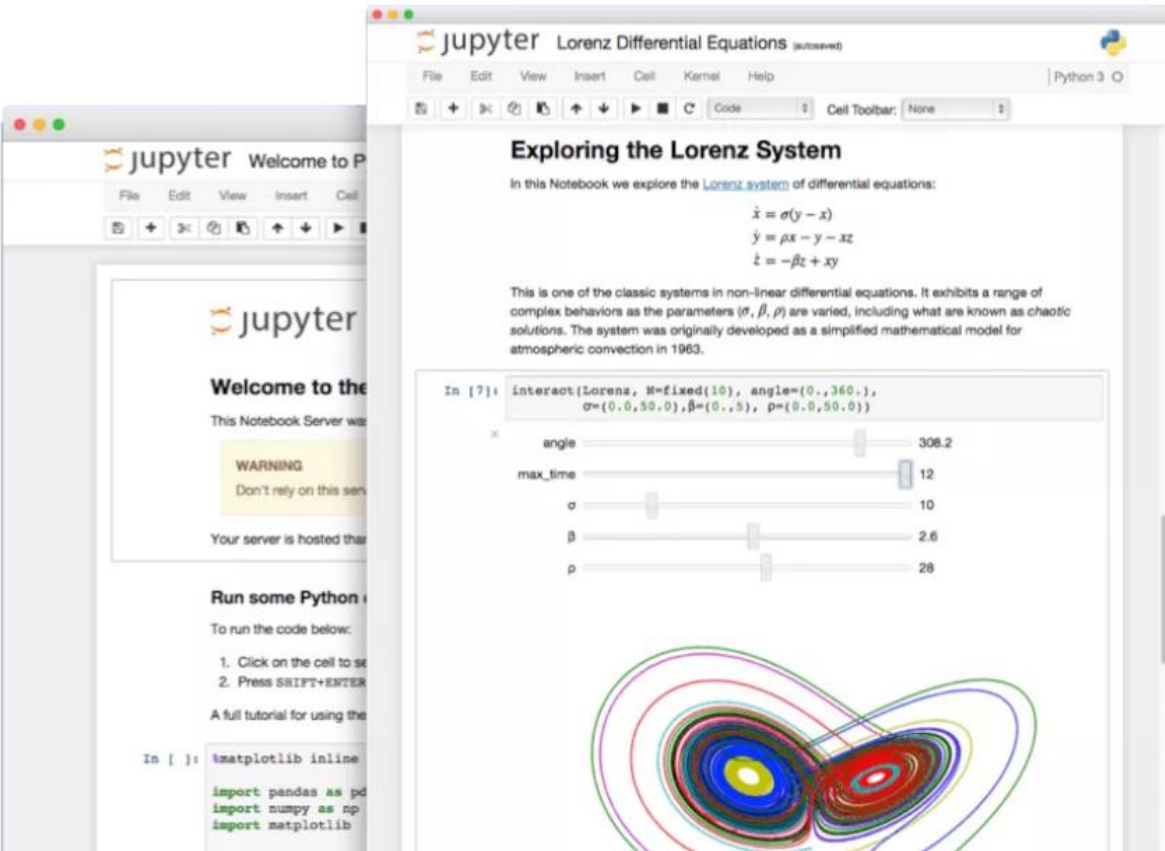
# Romeo and Juliet Word Cloud

ML Tlachac

Spring Teaching Summit '23

# Why Word Clouds?

- ◆ AA640: Data Visualization and Text Mining
  - ◆ Needed a transition between visualizations and text mining
- ◆ Text mining required use of Python programming
  - ◆ Some students coded in other classes
  - ◆ Some students had no prior coding experience
- ◆ Visualize raw text instead of numeric features
- ◆ Wanted text mining introduction to be fun and memorable



## Jupyter Notebook: The Classic Notebook Interface

The Jupyter Notebook is the original web application for creating and sharing computational documents. It offers a simple, streamlined, document-centric experience.

[Try it in your browser](#)

[Install the Notebook](#)

# Jupyter Notebooks for Python



Products ▼

Pricing

Solutions ▼

Resources ▼

Partners ▼

Blog

Company ▼

Contact Sales

Individual Edition is now

# ANACONDA DISTRIBUTION

The world's most popular open-source Python distribution platform

## Anaconda Distribution

Download 

For Windows

Python 3.9 • 64-Bit Graphical Installer • 621 MB

Get Additional Installers

<https://www.anaconda.com/products/distribution>

Google: Anaconda download





Home



Environments



Learning



Community

A full Python IDE directly

[Documentation](#)[Anaconda Blog](#)

Applications on

base (root) ▾

Channels



Notebook

 6.4.8

Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.

Launch



Powershell Prompt

0.0.1

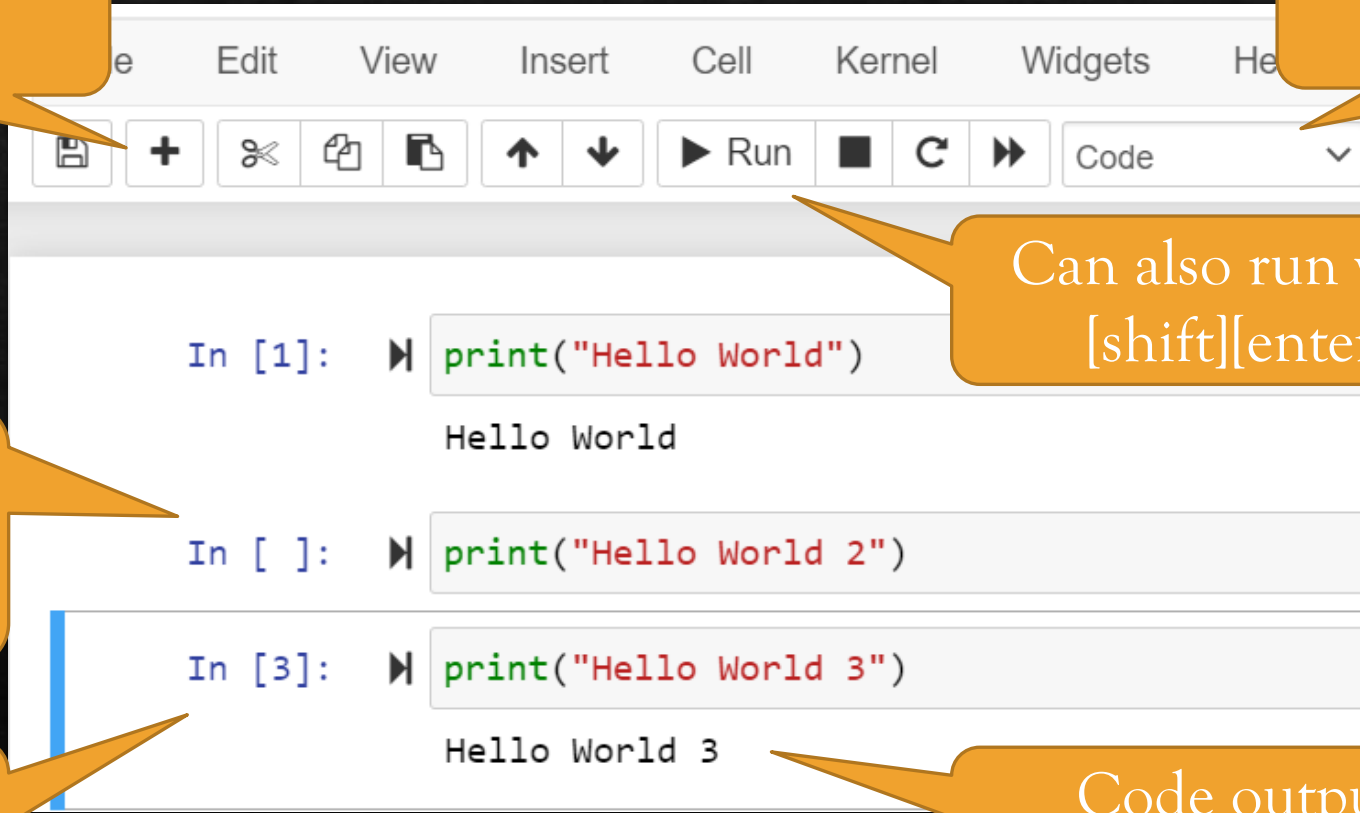
Run a Powershell terminal with your current environment from Navigator activated

Launch

# Coding in Notebooks

Add new cell below current cell

Indicates cell is for code



Can also run with  
[shift][enter]

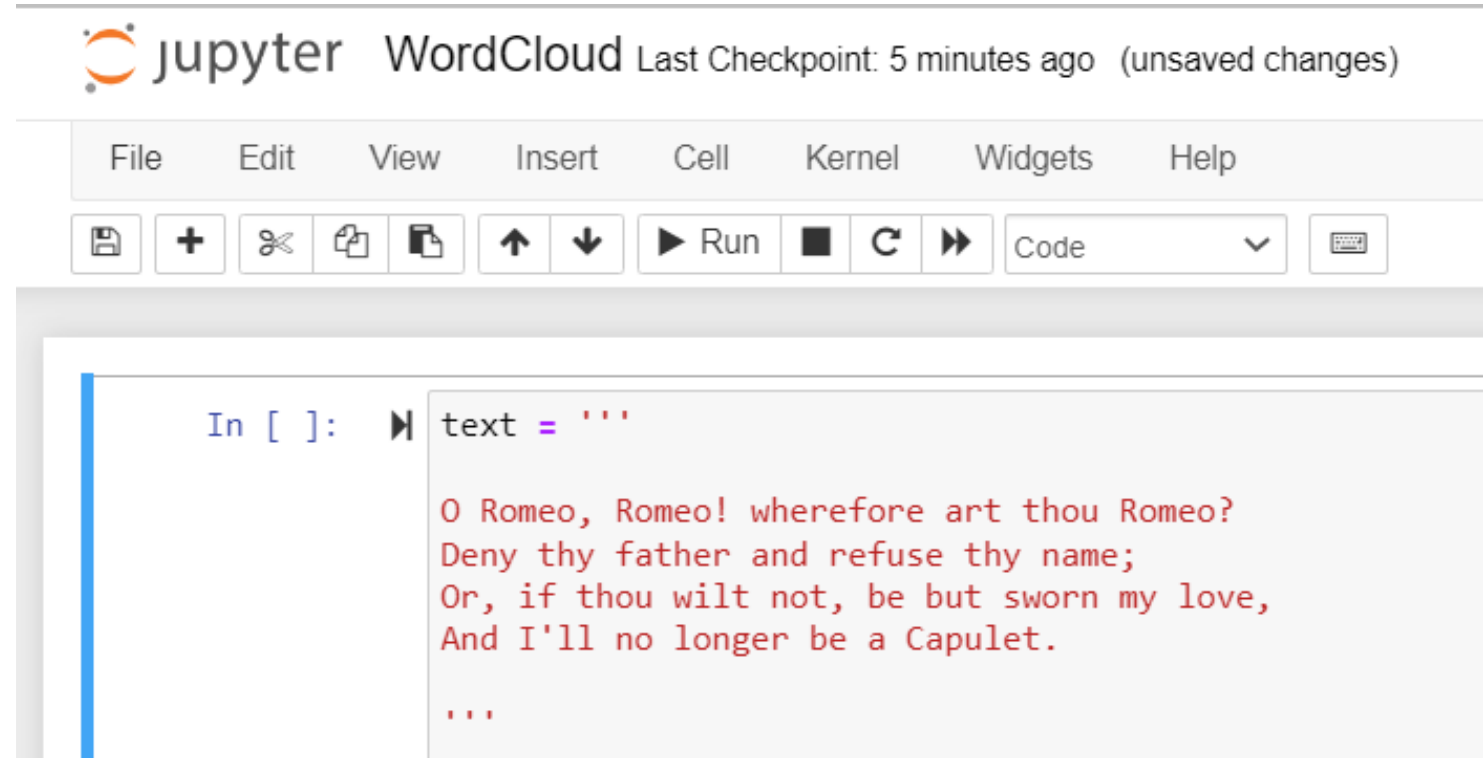
Blank because  
cell was not run  
(star if running)

Watch order, this  
cell was run third

Code output  
printed on screen

# What to get ready

1. Open Jupyter Notebook
2. Find the Wikipedia text you want to analyze
3. Copy and paste text into Jupyter Notebook



The screenshot shows the Jupyter Notebook interface. At the top, the Jupyter logo is followed by the text "WordCloud" and "Last Checkpoint: 5 minutes ago (unsaved changes)". Below this is a menu bar with "File", "Edit", "View", "Insert", "Cell", "Kernel", "Widgets", and "Help". Under the menu bar is a toolbar with icons for saving, adding a new notebook, undo, redo, copy, paste, up/down arrows, a "Run" button, a "Kernel" button, a "Help" button, and a dropdown menu currently set to "Code". The main area of the notebook shows a code cell with the prompt "In [ ]:" followed by a right-pointing arrow. The code in the cell is: `text = '''`  
`O Romeo, Romeo! wherefore art thou Romeo?`  
`Deny thy father and refuse thy name;`  
`Or, if thou wilt not, be but sworn my love,`  
`And I'll no longer be a Capulet.`  
`'''`

I added more text but shortened for slide so notation is visible

```
#install wordcloud library  
!pip install wordcloud
```

```
#import libraries  
from wordcloud import WordCloud  
import matplotlib.pyplot as plt
```



# Word Clouds do not require much code

```
#create basic word cloud
wordcloud = WordCloud().generate(text)

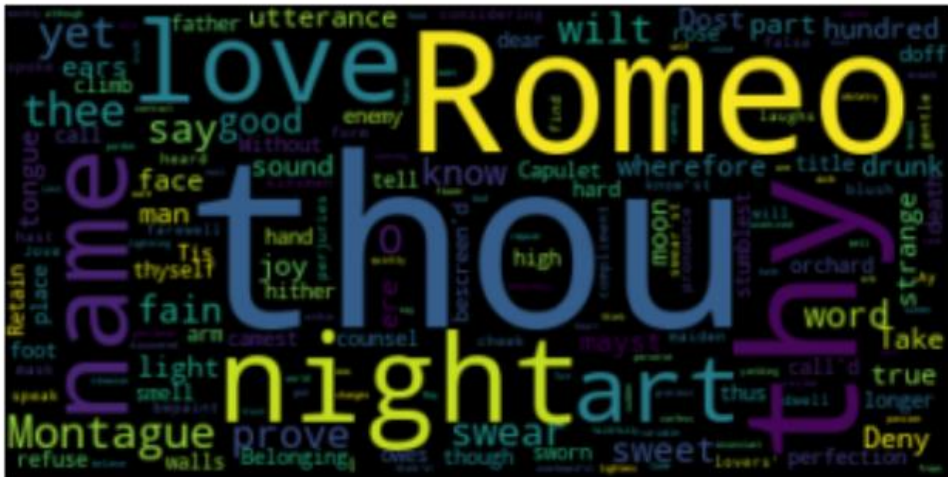
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```



# Does Juliet Talk About Herself?

```
#create basic word cloud
wordcloud = WordCloud().generate(text)

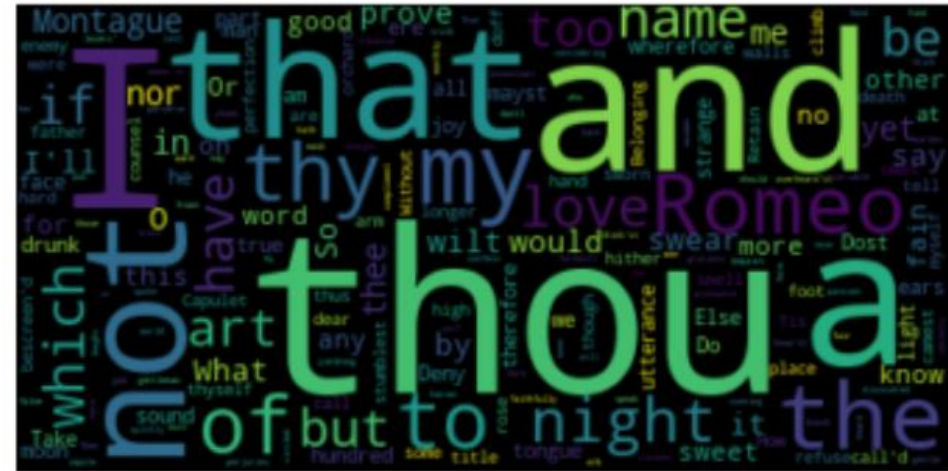
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```



## Parameter Changed

```
#create basic word cloud
wordcloud = WordCloud(stopwords = []).generate(text)

plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```





# Let's Change the Default Colors

```
#create basic word cloud with stop words
wordcloud = WordCloud(colormap = 'PiYG', background_color = 'white').generate(text)

plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```

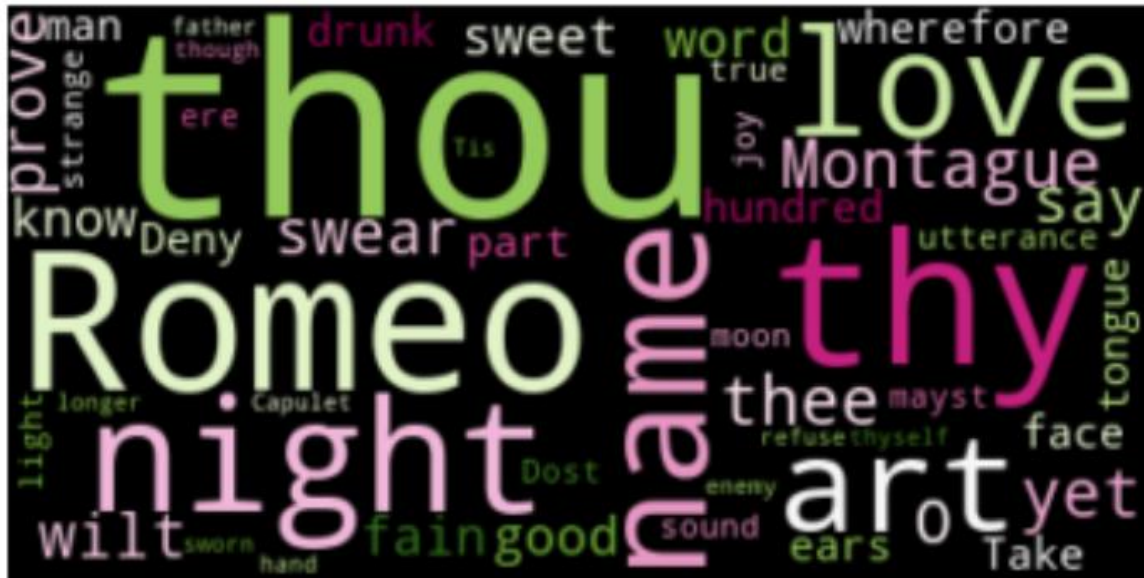
Parameters Changed



# Default is the Top 200 words

```
#create basic word cloud with stop words  
wordcloud = WordCloud(colormap = 'PiYG', max_words = 50).generate(text)  
  
plt.imshow(wordcloud, interpolation='bilinear')  
plt.axis("off")  
plt.show()
```

Parameter Changed



flower.jpg is a  
silhouette

```
#create basic word cloud with stop words - rose shaped
wordcloud = WordCloud(colormap = 'PiYG', mask = mask_img).generate(text)

plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```





```
#create basic word cloud with stop words - increase size and save
```

```
fig = plt.figure(figsize = (6,6))
```

```
wordcloud = WordCloud(colormap = 'PiYG', mask = mask_img, background_color = 'white').generate(text)
```

```
plt.imshow(wordcloud, interpolation='bilinear')
```

```
plt.axis("off")
```

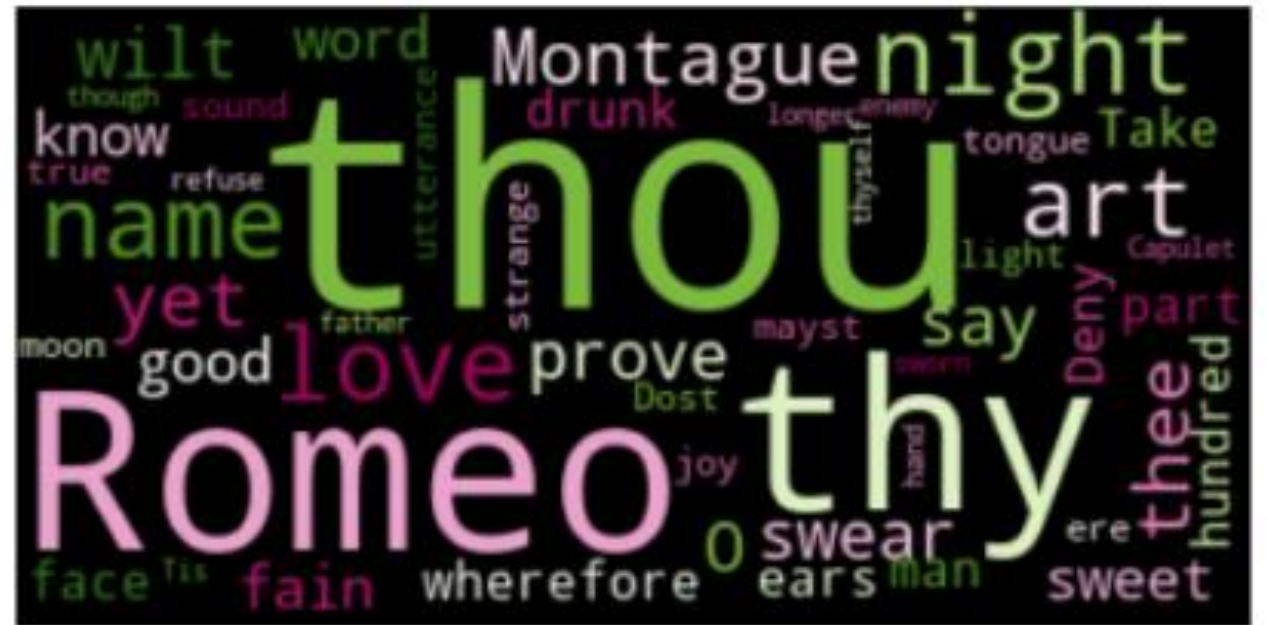
```
plt.savefig("juliet.png", bbox_inches = "tight")
```

```
plt.show()
```

```
plt.close()
```



# What Question do I have?







# In Class Activity

- ◇ Pick a topic or person
- ◇ Plot the popularity on Google n-grams
- ◇ Create a word cloud using the Wikipedia entry
- ◇ What can you conclude based on the plots?
- ◇ What questions do you have based on the plots?

Individual submission, include all plots

# Google Books Ngram Viewer

romeo and juliet,king lear

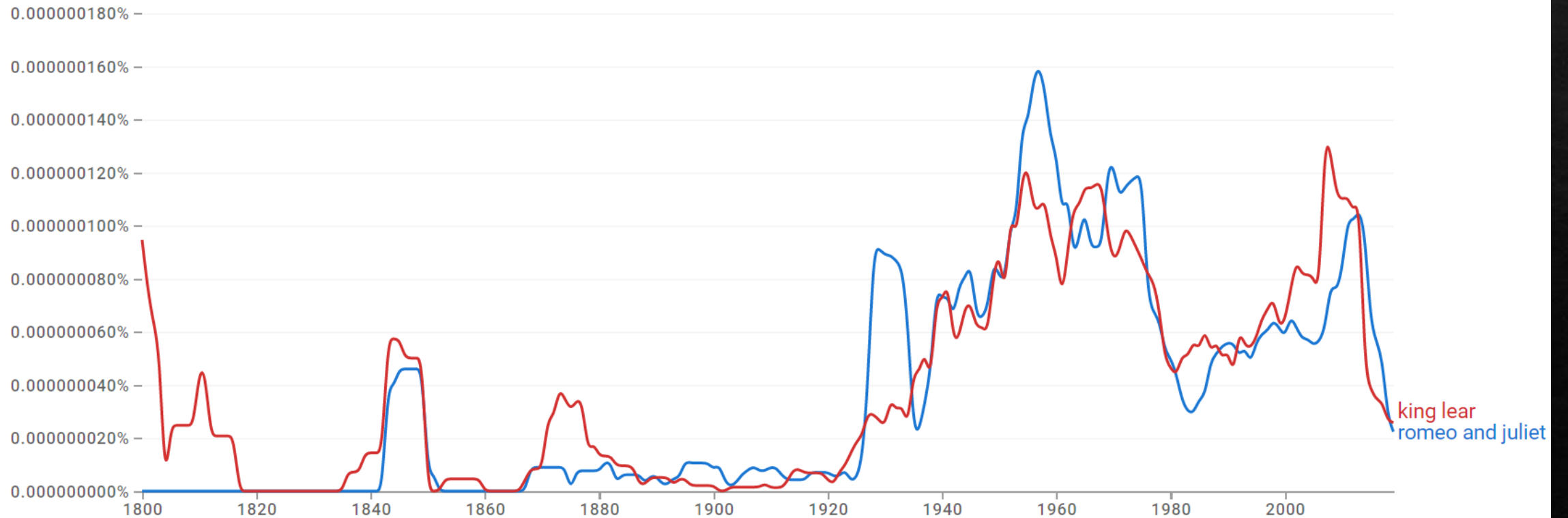


1800 - 2019

English (2019)

Case-Insensitive

Smoothing





Student Examples Redacted For Public Posting