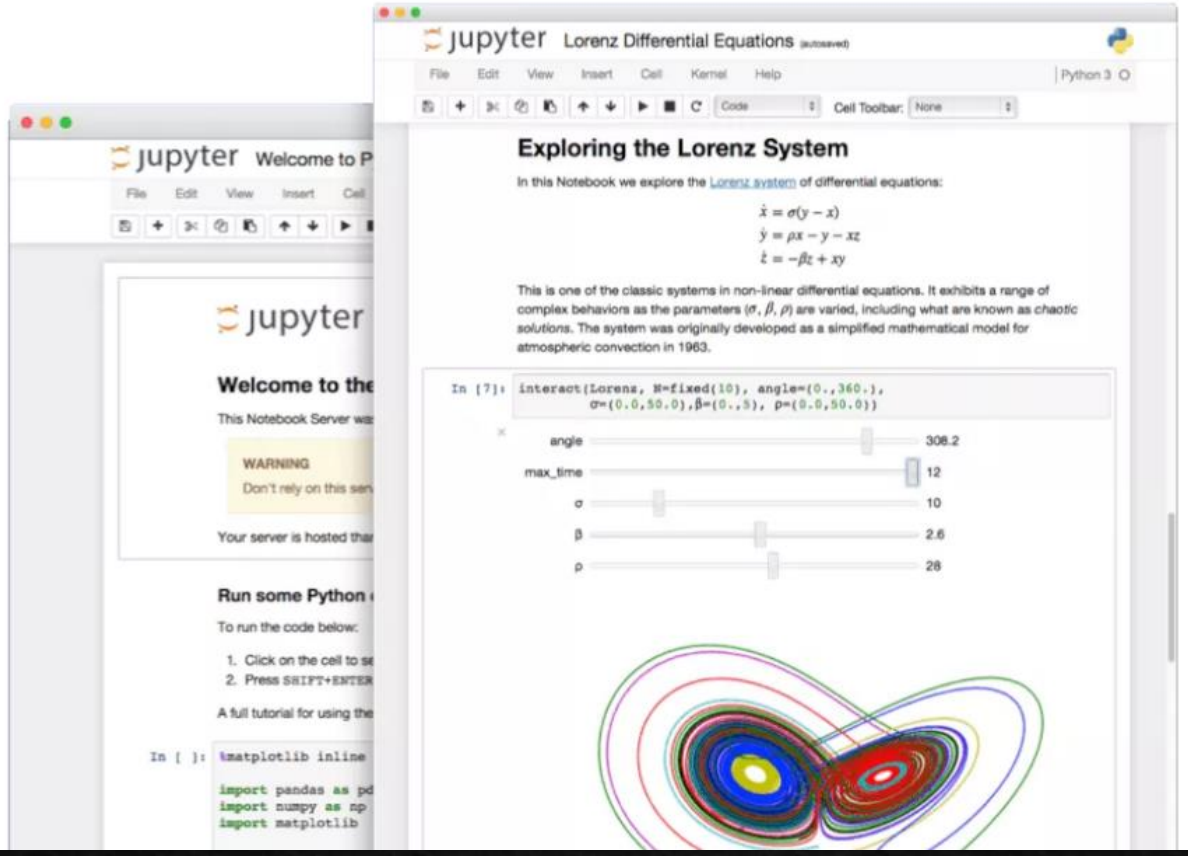
Romeo and Juliet Word Cloud

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

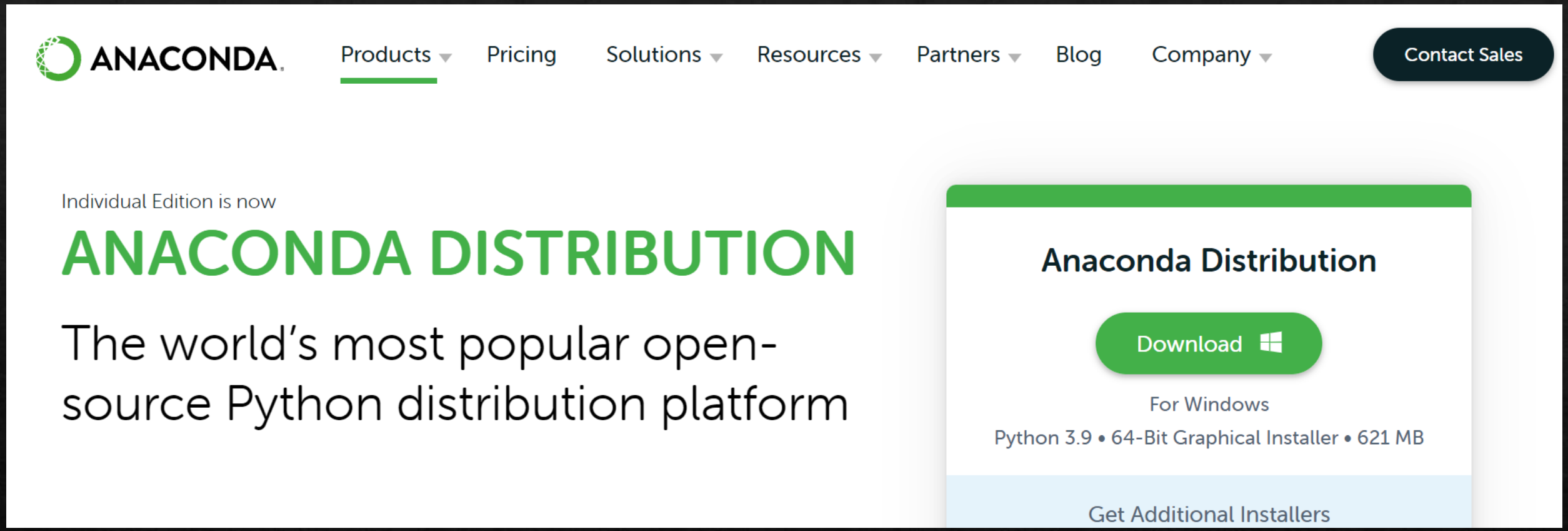


Jupyter Notebook: The Classic Notebook Interface

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

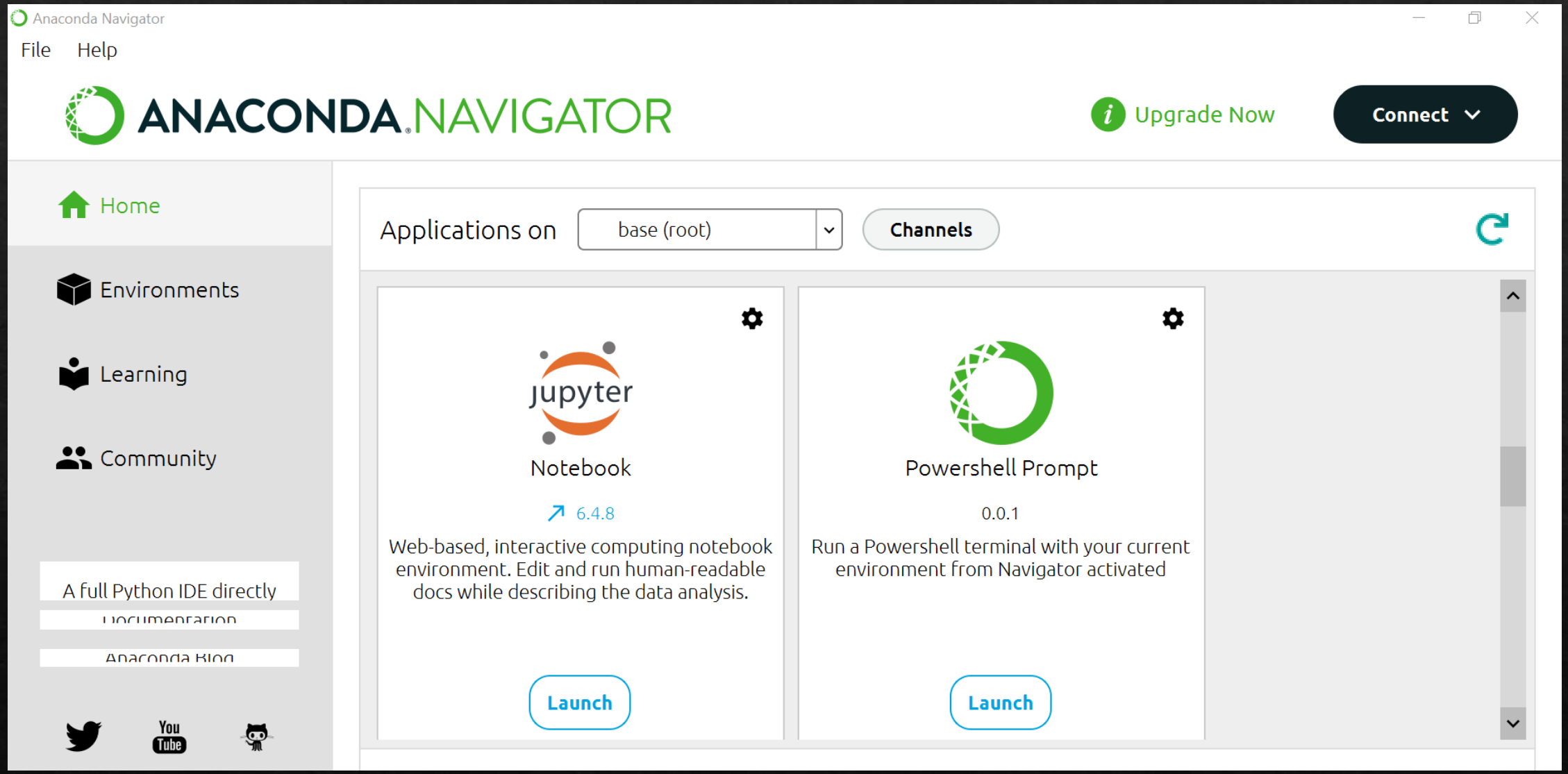
Try it in your browser | Install the Notebook

Jupyter Notebooks for Python



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

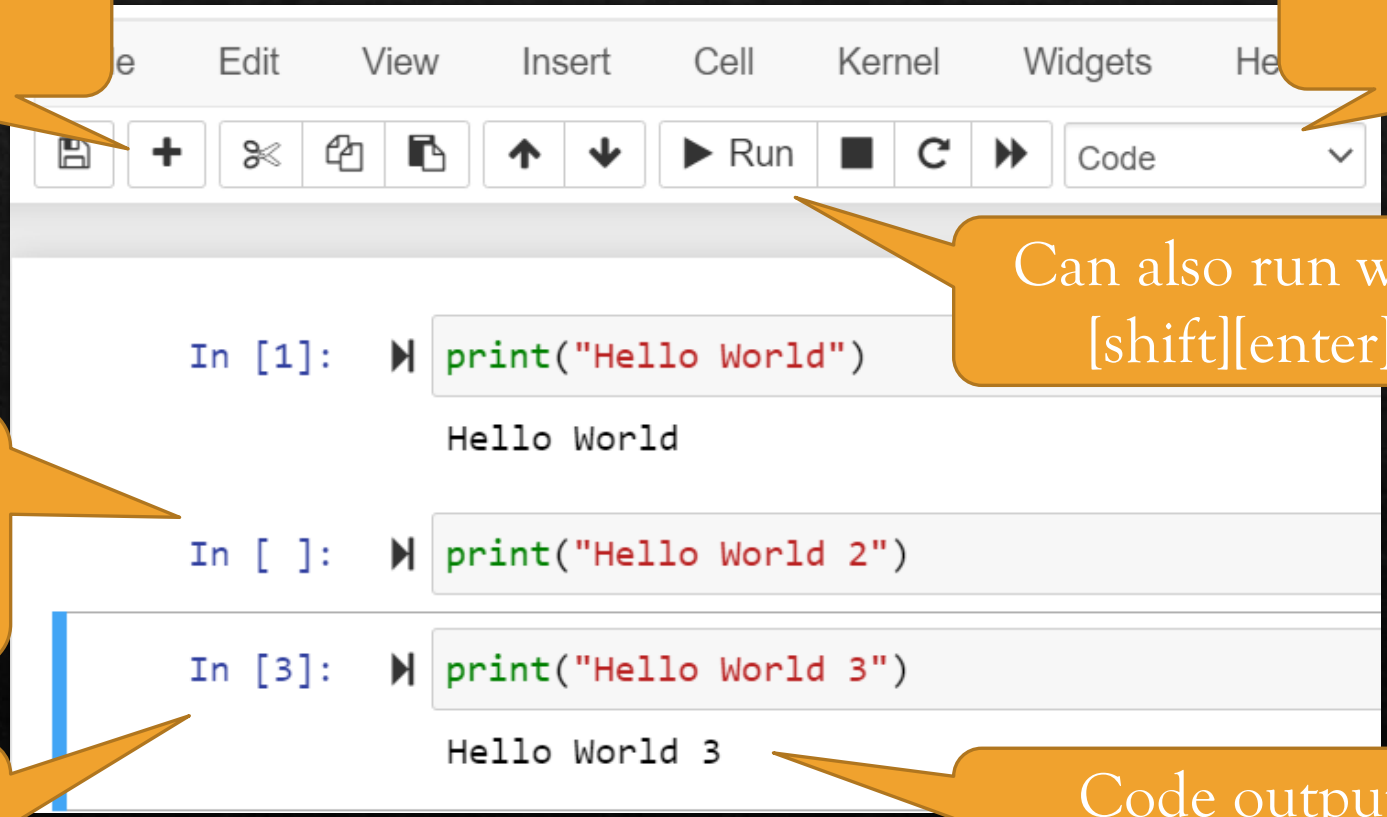
Google: Anaconda download


Anaconda Navigator
File
Help
ANACONDA.NAVIGATOR
Upgrade Now
Connect
Home
Environments
Learning
Community
A full Python IDE directly
Documentation
Anaconda Blog
Applications on
base (root)
Channels
jupyter
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]

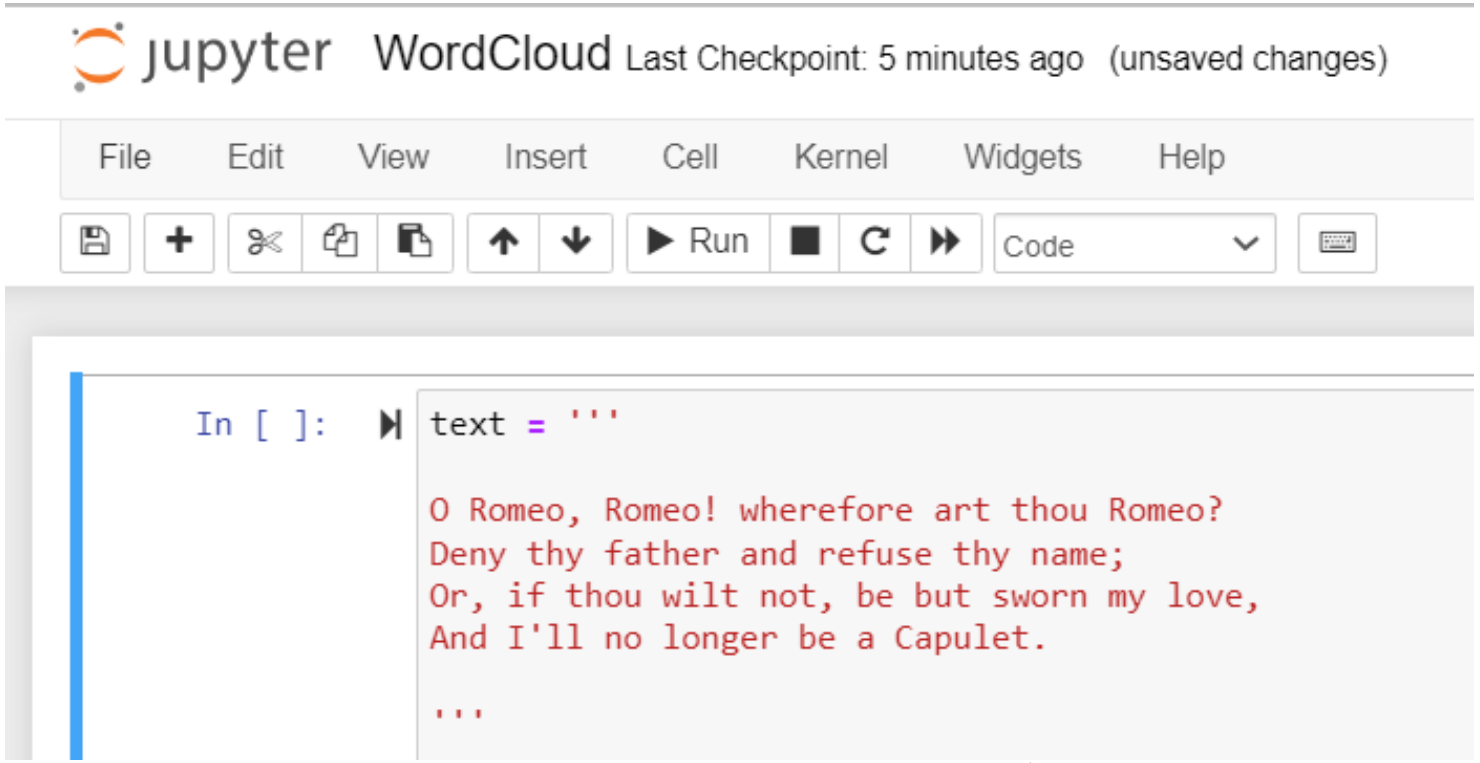
Blank because cell was not run (star if running)

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



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

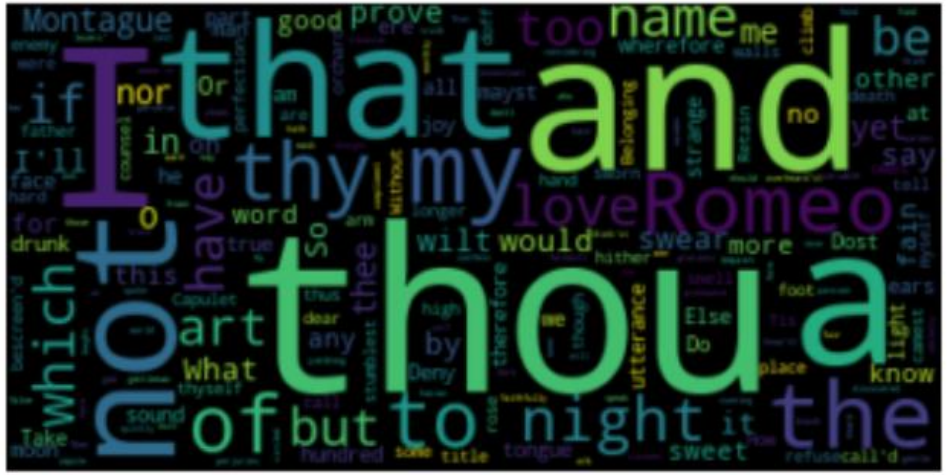
```
#create basic word cloud
wordcloud = WordCloud().generate(text)

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

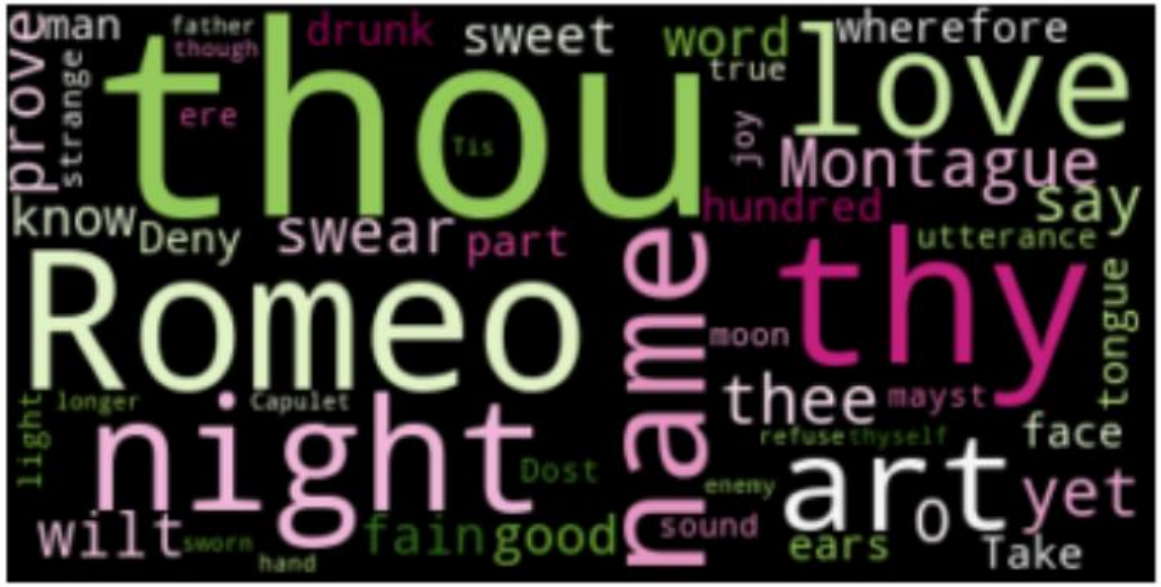
```
#create basic word cloud with stop words
wordcloud = WordCloud(colormap = 'PiYG', background_color = 'white').generate(text)

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



Changing the Cloud Shape

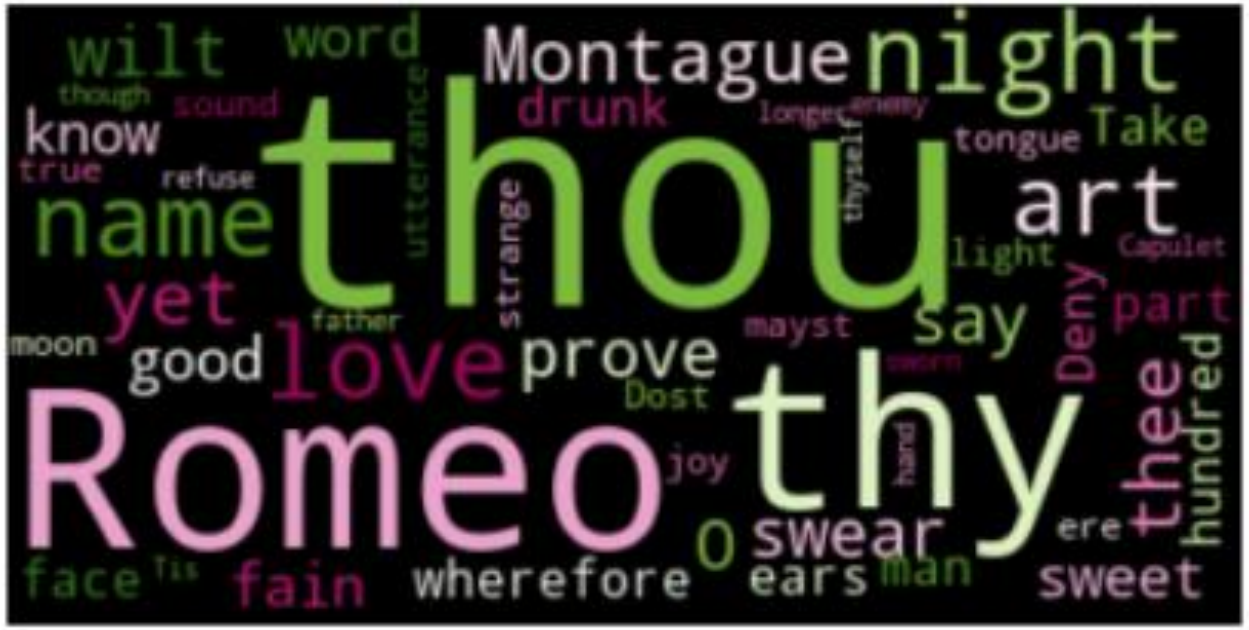
flower.jpg is a silhouette

```
from PIL import Image
import numpy as np

mask_img = np.array(Image.open("flower.jpg"))
```

```
#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()
```

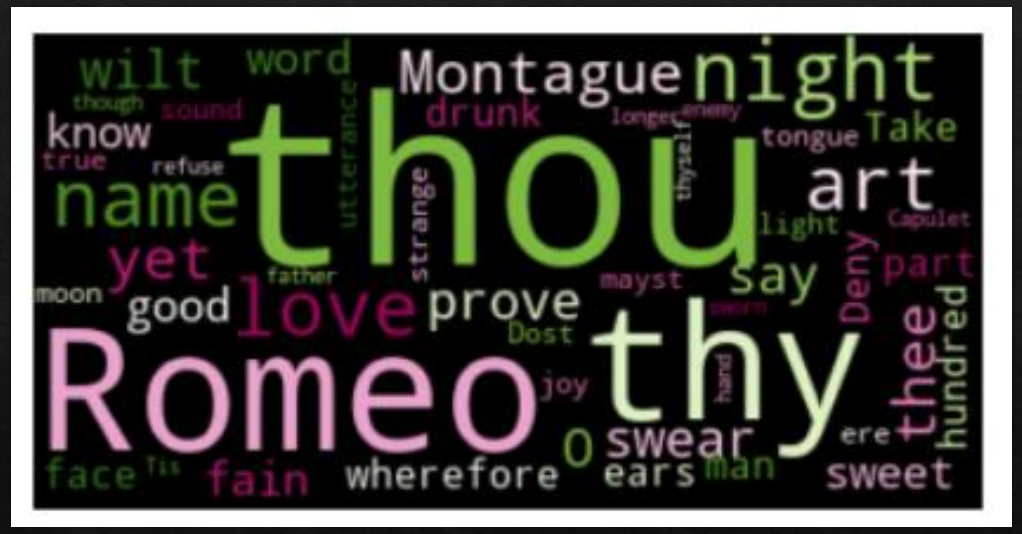
```
#create basic word cloud with stop words - increase size and save
fig = plt.figure(figsize = (6,6))
wordcloud = WordCloud(colormap = 'PiYG', mask = mask_img, background_color = 'white').generate(text)

plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.savefig("juliet.png", bbox_inches = "tight")
plt.show()
plt.close()
```

What Question do I have?



Romeo Doesn't Call Juliet by Name

Juliet



Romeo

In Class Activity

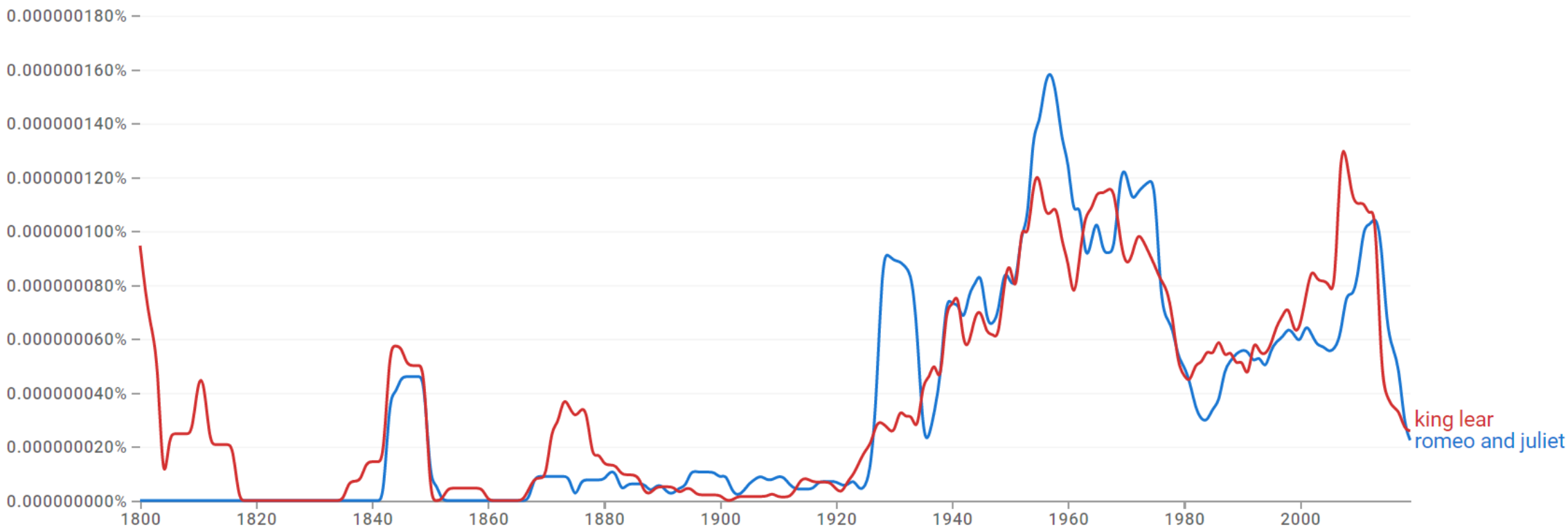
- Pick a topic or person
- Plot the popularity on Google n-grams
- Create a word cloud using the Wikipedia entry
- What can you conclude based on the plots?
- What questions do you have based on the plots?

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