## Names: Yuan Shi, Jingcong Hu

## GitHub Repository:<https://github.com/hujingc/SI-206-Final-Project>

## Old Goals

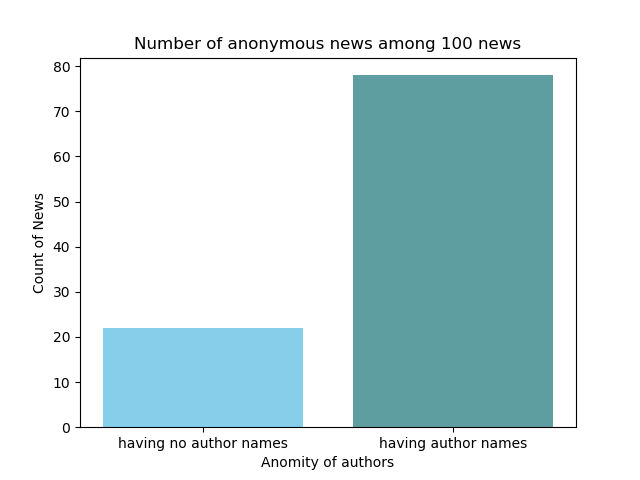
* Compare and contrast the Top BBC news with the news from New York Times

## Achieved Goals

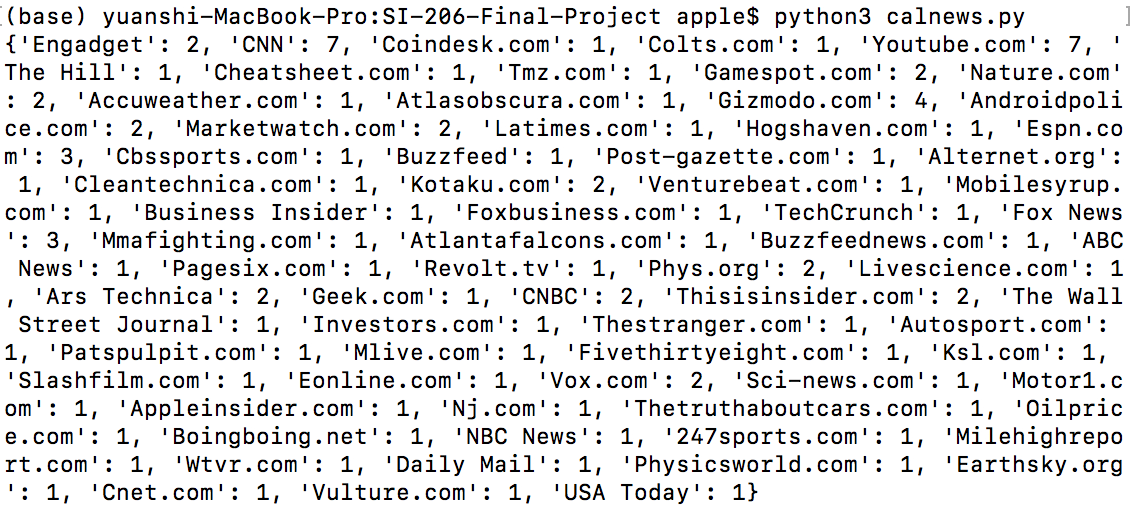
* Grabbed stories from News and New York Times website from 5 different sections
* Calculated number of unique words in the top titles in NYT

## calculation file

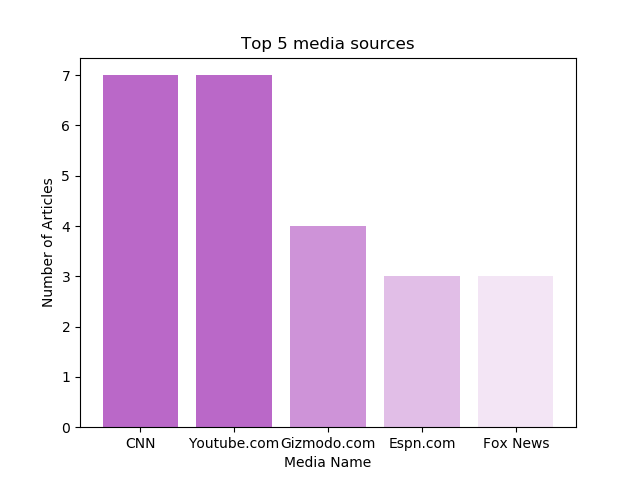
* Calnew.py
  + Calculate the number of articles that miss author’s name



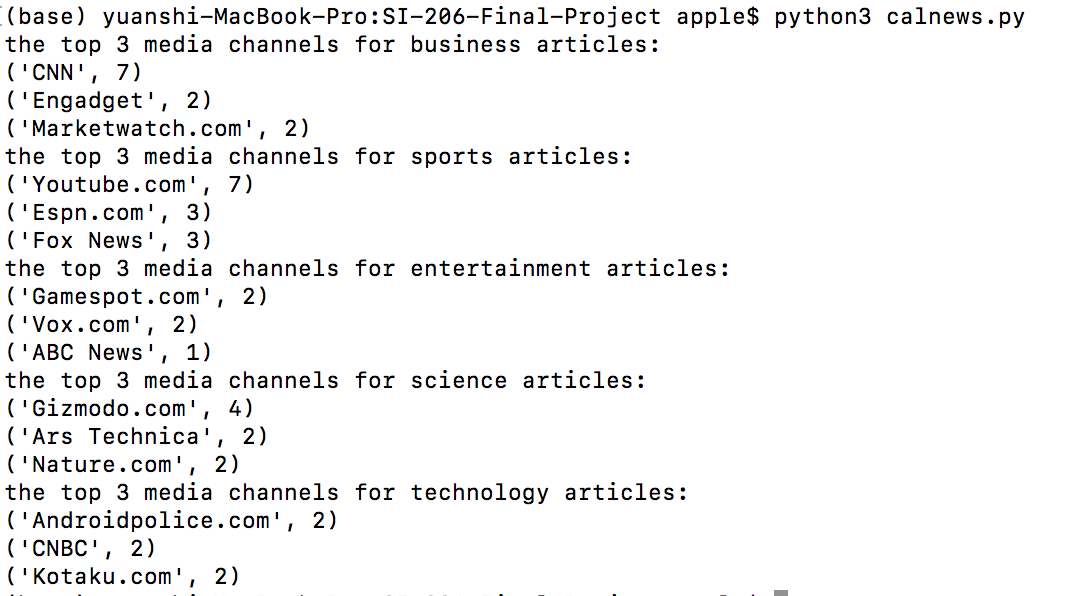
* + Created a dictionary with keys being the media channels (e.g. CNN, Youtube.com), and values being number of times



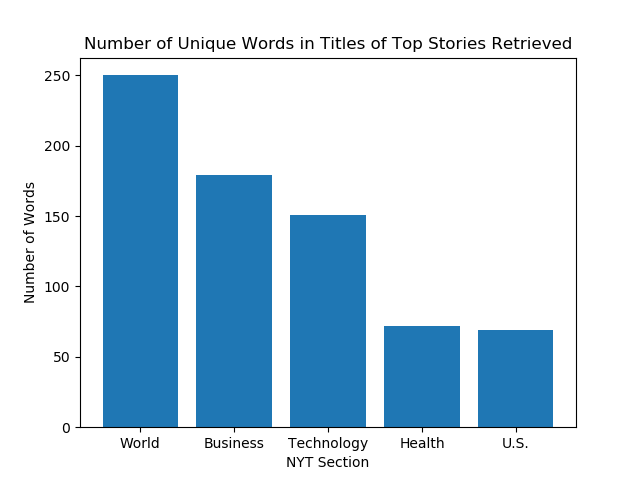
Draw a bar chart for the top 5 media channels



* + Counted top 3 media channels for all 5 sections



* NYTcalculations.py
  + For each section, found the non-stopwords that appeared in the titles of the articles
  + Created a dictionary with keys being the sections, and values being dictionaries with a key of number of times a word appeared and values being a list of the words that appeared ‘key’ number of times
  + See NYTcleaned.json for the calculation output
  + See NYTformatted.txt for a pretty formatted version of the calculation output
  + Visualization coded in NYTvisualization.py and saved in NYTbar.png
    - Shows that of the 5 sections and data gotten, World had the most number of unique words in the article titles.



## Instructions for running the code

* The following should be inside our zip file
  + setup.py
    - Pulls data from News API
  + NYT\_data\_getter.py
    - Pulls data from NYT Top Stories API
  + new.sqlite
    - Database used to do calculations
* The news.sqlite is complete with 2 tables of 100 rows each, and is the result of running setup.py 5 times and NYT\_data\_getter.py 5 times
  + If you wish to verify that we are correctly getting the data through our code, please change the name of the database in
    - line 15 for NYT\_data\_getter.py
    - line 8 for setup.py
  + from “news.sqlite” to something else like “other.sqlite” to use a new database
  + Then run NYT\_data\_getter.py and setup.py each 5 times to populate the new database.
  + NOTES
    - In the event that there are not enough news stories to run each file 5 times, running the files the first time will print out a statement saying as such, and populate the database with more titles from the other sections
    - Please wait at least 7 seconds after NYT\_data\_getter.py is ran each time, otherwise there will be KeyError: 'num\_results' due to NYT’s call limit

## Code Documentation

* Both of us did not write functions.

## Resource Documentation

* StackExchange for how to check if a title is already in the database
* NYT API documentation
* Various Google searches about how to get data from NYT API
* News API documentation