

MEDIA MODELING: JOURNALISM DATA COLLECTION AND ANALYSIS

MD SHUEY, 2019



PROJECT INCEPTION

- Former internship project
- E-mail newsletter summarizing relevant news across 7 major English-language sources
- Originally Excel analysis, later reconstructed via R

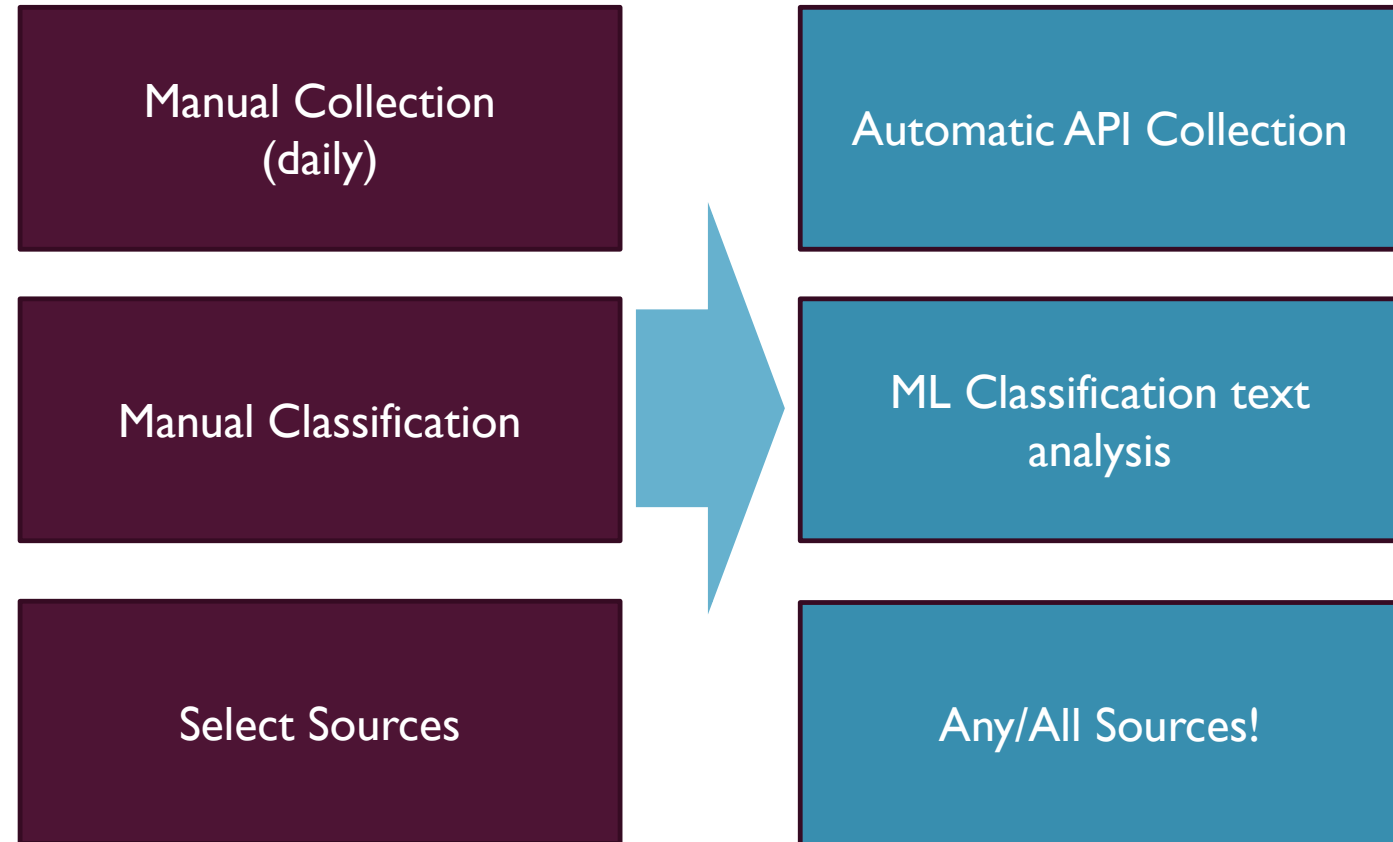
Manual Collection
(daily)

Manual Classification

Few, Select Sources

HOW TO RECREATE? HOW TO EXPAND?

- Automatic headline pulling
- Manual classification
- Sentiment analysis





News API is a simple and easy-to-use API that returns JSON metadata for headlines and articles live all over the web right now.



FREAKING FAST

Everything is asynchronously cached for a super-fast response.



FREE FOR DEVELOPERS

Jump right in if you're in development or open source.



EASY INTEGRATION

Get JSON results with simple GET HTTP requests

```

GET https://newsapi.org/v2/top-headlines?country=us&category=business&apiKey=AP
I_KEY

{
  status: "ok",
  totalResults: 69,
  - articles: [
    - {
      - source: {
        id: "the-wall-street-journal",
        name: "The Wall Street Journal"
      },
      author: "Maureen Farrell, Liz Hoffman, Eliot Brown, David Benoit",
      title: "Some WeWork Board Members Seek to Remove Adam Neumann as CEO - The Wall Street Journal",
      description: "SoftBank officials are among those expected to push for Neumann ouster",
      url: https://www.wsj.com/articles/some-wework-board-members-seek-to-remove-adam-neumann-as-ceo-11569171188,
      publishedAt: "2019-09-22T17:14:00Z",
      content: "A bloc of WeWork directors is planning to push \r\nAdam Neumann\r\nto step down as chief executive after a tumultuous week in which his eccentric behavior and drug use came to light and the startup delayed its much-anticipated stock-market listing.A group includi... [+4199 chars]"
    },
    - {
      - source: {
        id: null,
        name: "Bitcoinist.com"
      },
      author: null,
      title: "Sundav Digest: Bitcoin(BTC) Price Rangebound. Altcoins Start To Move -

```

	title	description
	FIFA 20 review: Volta brings flair to the familiar	FIFA 20 largely keeps
	Spain's Attempt to Aton for a 500-Year-Old Sin	The country is offerin
	15 video games that won't rot your kids' brains - CN...	Just because it's a vi
	LATAM unveils Star Wars "Galaxy's Edge" themed li...	LATAM Brazil recentl
	Climate Protesters and World Leaders: Same Planet,...	Friday's global clima
	Ex-Man Utd Star Anderson Officially Retires From Pr...	News from around th
World	'Millions' protest in youth-led global climate strike	Masses of children s
	A New Urgency at Marni	Francesco Risso's sh
	How big was the global climate strike? Here's the b...	It was likely the large
	Brazil: Rio favela residents rally after 8-year-old sho...	Hundreds gather in B

NEWSAPI R WRAPPER

- Github.com/MDshuey/NewsAPI
 - Credit to mkearney for 1.0
 - Possible future pull request for 2.0 updates
- Using R parameters, creates the HTML API call, and converts from JSON to data frame format
- daily_recaps loop function allows for more efficient pulling (from the last 31 days)
- Variables:
 - Source, author, title, URL, image URL, time published, content (first sentences of article)

USE CASES

- Customized news alerts
- Time series event detection
- Determine news cycle beginning, peak, end
 - From this, predict article popularity
- Sentiment analysis of text from difference sources
 - Compare popularity to positivity/negativity index
 - sentiment dictionary matrix (lexicon)

anger	anticipation	disgust	fear	joy	negative	positive	sadness	surprise	trust
			war		war				
	time				staring				
			pain		pain		pain		
				love		love			
				friend		friend		guess	friend
				found		found			found
				favorite		favorite			favorite
			die		die		die		
death	death	death	death		death		death	death	
					cry		cry		
						civil			
bad		bad	bad		bad		bad		
			afraid		afraid				

CONCLUSION



[GitHub.com/Mdshuey/Portfolio](https://github.com/Mdshuey/Portfolio)



LinkedIn: Mitchell D. Shuey

Open to career opportunities!