MEDIA MODELING:
JOURNALISM DATA COLLECTION AND ANALYSIS
MD SHUEY, 2019
Former internship project

E-mail newsletter summarizing relevant news across 7 major English-language sources

Originally Excel analysis, later reconstructed via R
HOW TO RECREATE? HOW TO EXPAND?

- Automatic headline pulling
- Manual classification
- Sentiment analysis

Manual Collection (daily) -> Automatic API Collection

Manual Classification

Select Sources

ML Classification text analysis

Any/All Sources!
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.
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)
CONCLUSION

GitHub.com/Mdshuey/Portfolio

LinkedIn: Mitchell D. Shuey

Open to career opportunities!