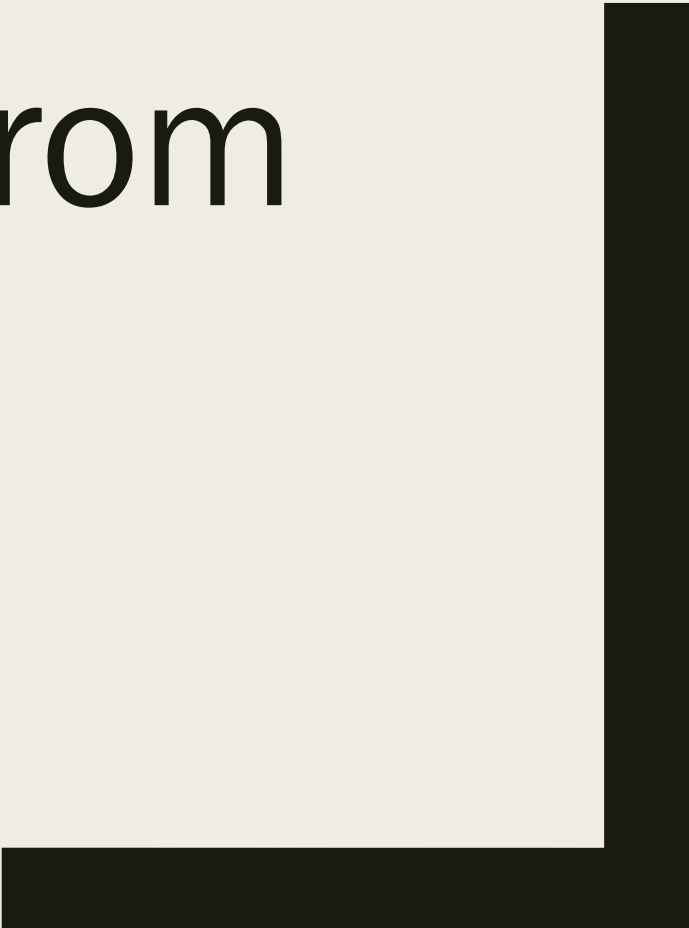# Deriving Alpha from News: GDELT

Ty Painter
Capstone Update #8
4/21/22

AWS EC2

- Script has been running for over a week

- Process time
  - *Roberta sentiment* → *70.4%*
  - *Navigating to URL* → *12.2%*
  - *Web scraping contents* → *8.1%*
  - *Tagging articles* → *4.0%*
  - *Extracting text from webpage* → *3.0%*

- Amount to 97.7%

Sample Data

- Return data
  - *Same & next day raw returns*
  - *S&P 500 average returns*
  - *Same & next day relative returns*
- Summarized data
  - *Scaled and averaged readability scores*
  - *Volume counts*
  - *Summarized sentiment & readability score per day for each ticker*
  - *Merged with Return data*
  - *How to handle look back for lag score?*
- Explore distribution of ticker tags

Microsoft Excel
Worksheet

Portfolio Construction

■ Long-short portfolio

    – *3,5,10 bins*

    – *Correlation between scores and return*

    – *How to order?*

        ■ Net Sentiment score

■ Quantile portfolio

■ Factor-weighted portfolio

Next Steps

- Gather additional months of data

- Run processes on January 2021 data

- Construct portfolio and evaluate results

- Finish final paper & presentation

- Organize GitHub