

Web Scraping: Above the Law

Logan Childers
NYC Data Science Academy
October 2018



Agenda

1. Website Overview
2. Scraping Process
3. Analysis

About Above the Law

Founded in 2006, Above the Law (ATL) is news website owned by Breaking Media.

ATL's five editors publish, share, argue, and defend takes on the news that matters to lawyers and the legal industry.

Combined with a network of 1000+ expert columnists sharing ground-level viewpoints, ATL publishes healthy stats:

- Approximately 20 original articles published per day
- Average of 10,000,000 page views per article
- 1,300,000 unique readers per month

The image shows a laptop screen with the Above the Law website loaded. The header features the "ABOVE THE LAW" logo. Below the header is a navigation bar with links: BIGLAW, SMALL LAW, LAW SCHOOLS, IN-HOUSE, LEGAL TECH, JUSTICE, GOVERNMENT, CAREER CENTER, and INNOVATION. A search bar is also present. The main content area includes a headline about a quick survey on litigation finance firms, followed by several news articles with images and brief descriptions. The articles include: "Overcriminalization Leads To Biased Outcomes" (by Tyler Broker), "Yale Law School Alumni Tell Paper They Aren't Surprised By Faculty Sexual Harassment Allegations" (by Gaston Kroub), "A Vote For IP Transparency" (by Kathryn Durino), and "The Law Firms With Appellate Practice (2018)" (by Scott Powers).



Overcriminalization Leads To Biased Outcomes



Above The Law's Supreme Court Power Index



A Vote For IP Transparency



The Law Firms With The Hottest Appellate Practice (2018)



SCOTUS Watch: 2 High-Profile LGBTQ Cases Before The High Court



BIGLAW SMALL LAW LAW SCHOOLS IN-HOUSE LEGAL TECH JUSTICE GOVERNMENT CAREER CENTER INNOVATION

Search ...



The Litigation Analytics Revolution – Read More Here



Sheets made for more than sleeping in

SHOP NOW



Biglaw



BIGLAW

The Law Firms With The Hottest Appellate Practice (2018)

By KATHRYN RUBINO

Congratulations to all the hard working lawyers recognized.



Project Objective

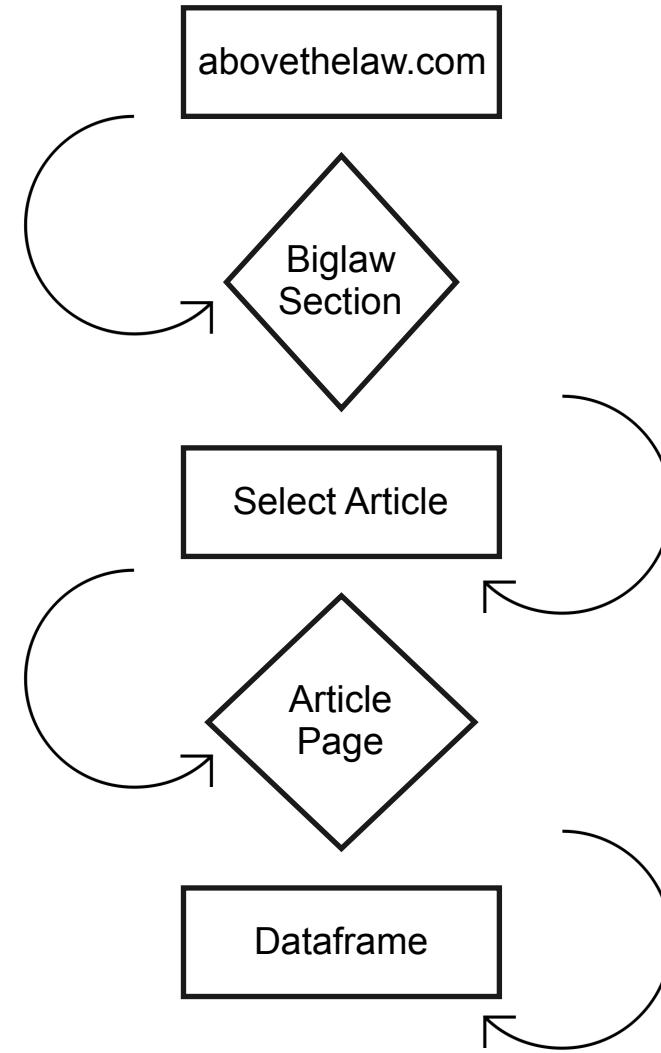
- Scrape subset of ATL articles
- Explore most-used tags & keywords
- Perform sentiment analysis per editor

Scraping Process



Process

- Start in the Biglaw section of ATL (greatest % of articles)
- Step through 20 articles per page (200 pages total)
- In the article page, collect the following info:
 - Title
 - Author
 - Publish date
 - Content (article summary)
 - Tags



Analysis



Overview of Scraped Data

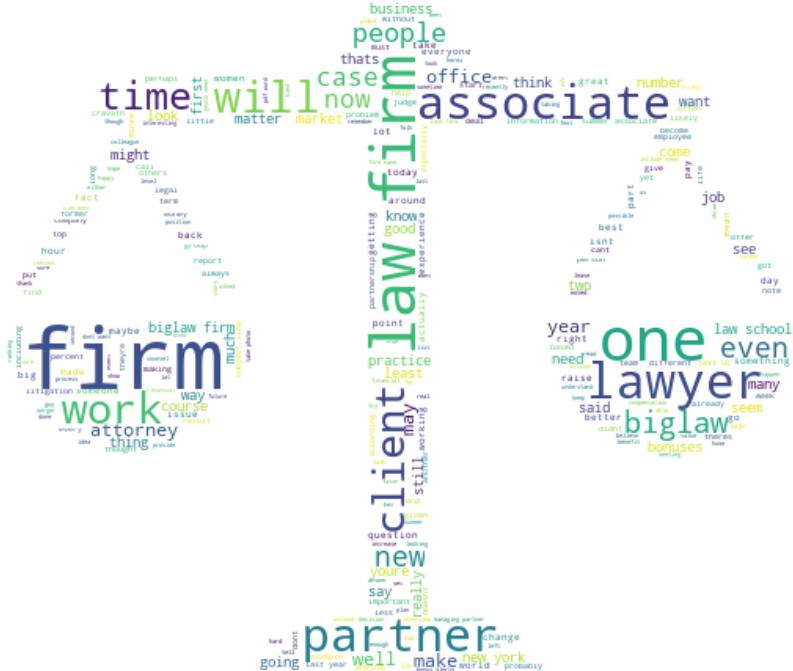
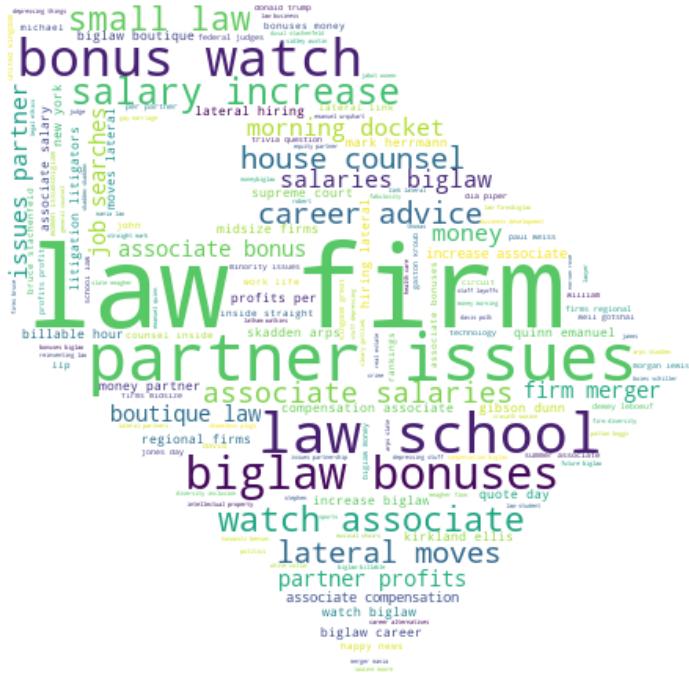
5 Years

3.5K Articles

84 Authors

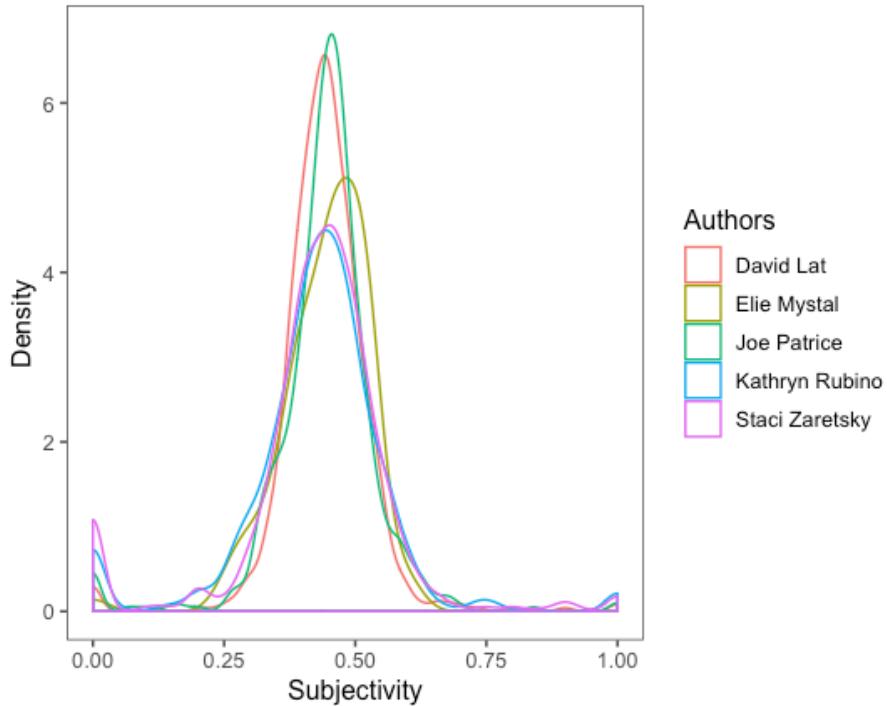
8K Topics

Word Clouds

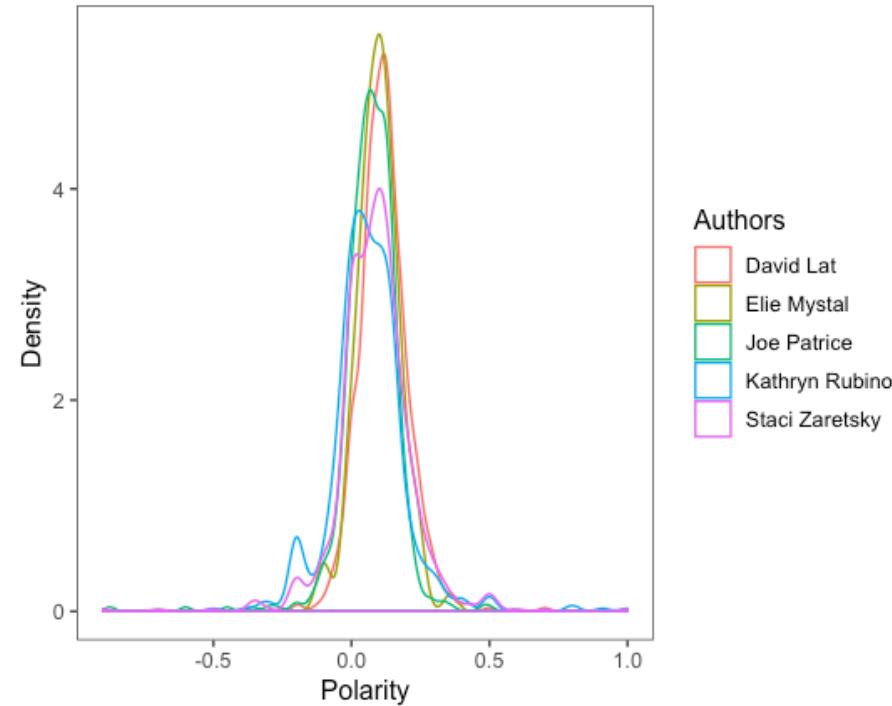


Sentiment Analysis

Subjectivity Analysis, 5 Permanent Editors



Polarity Analysis, 5 Permanent Editors



Questions/Comments?