Measuring Bias of News/Media Outlets 11/17/2021

Known News/Media Outlets with Known Bias

- Some media bias charts come from surveys in which people are polled about their News/Media preferences (almost always subjective)

- Some media bias charts come from studies done by journalists, researchers, experts in Journalism, etc. (can be somewhat subjective and not 100% objective)

Static Media Bias Charts:

https://www.allsides.com/media-bias/media-bias-chart

https://adfontesmedia.com/product/media-bias-chart-latest-edition/

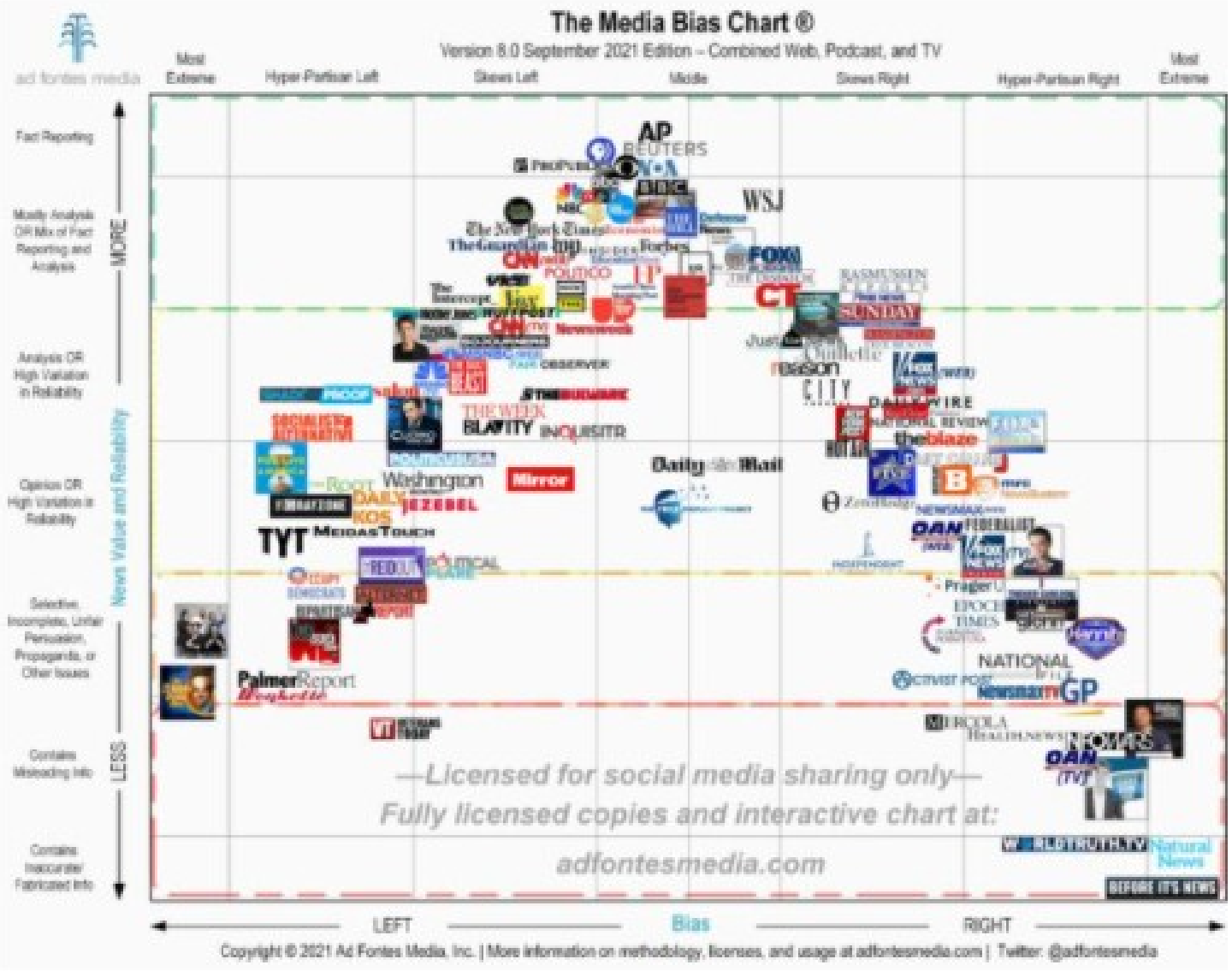
https://libguides.willamette.edu/c.php?g=697656&p=4958591

Interactive Comprehensive Media Bias Chart:

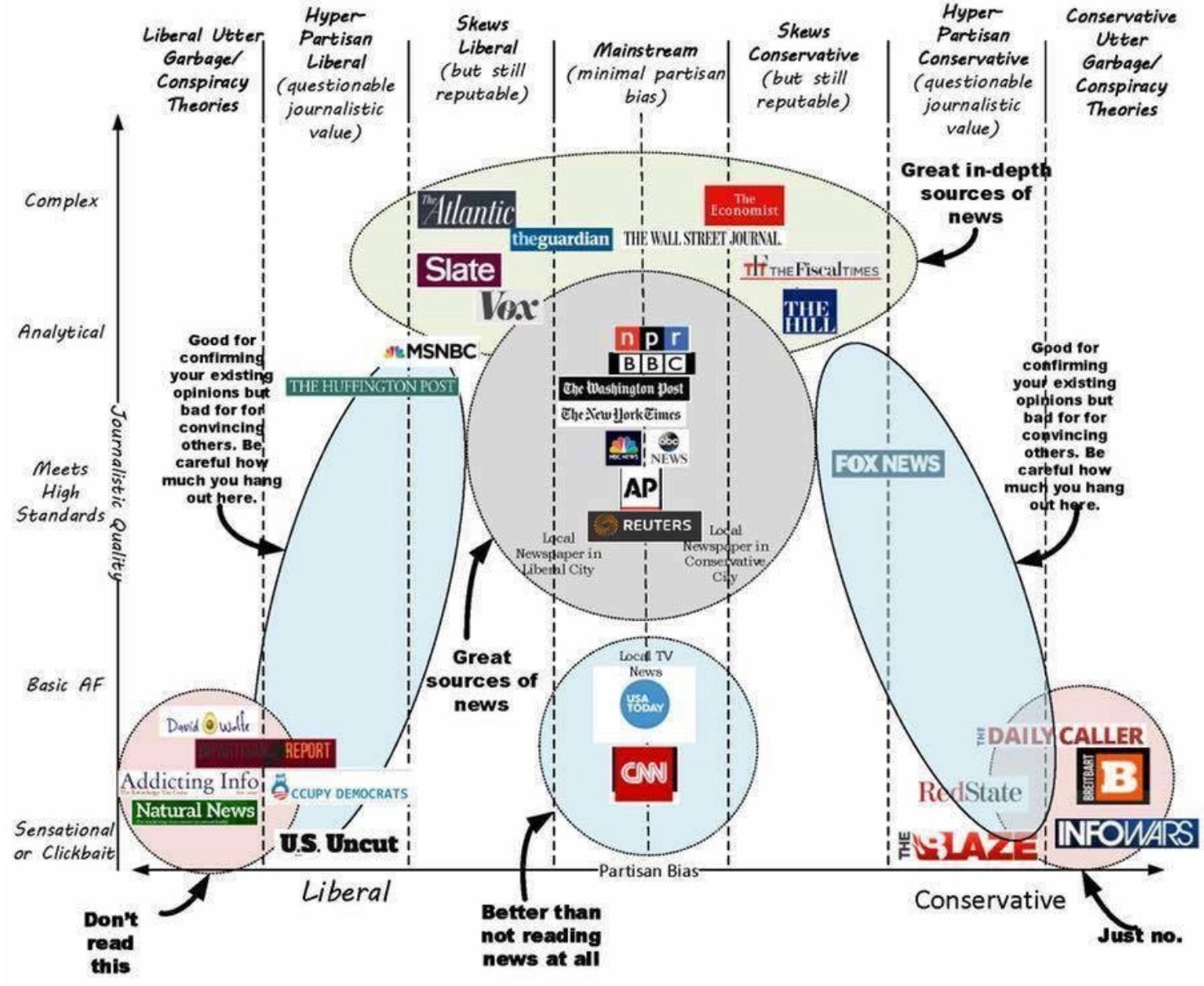
https://adfontesmedia.com/interactive-media-bias-chart/

https://www.allsides.com/media-bias/media-bias-chart

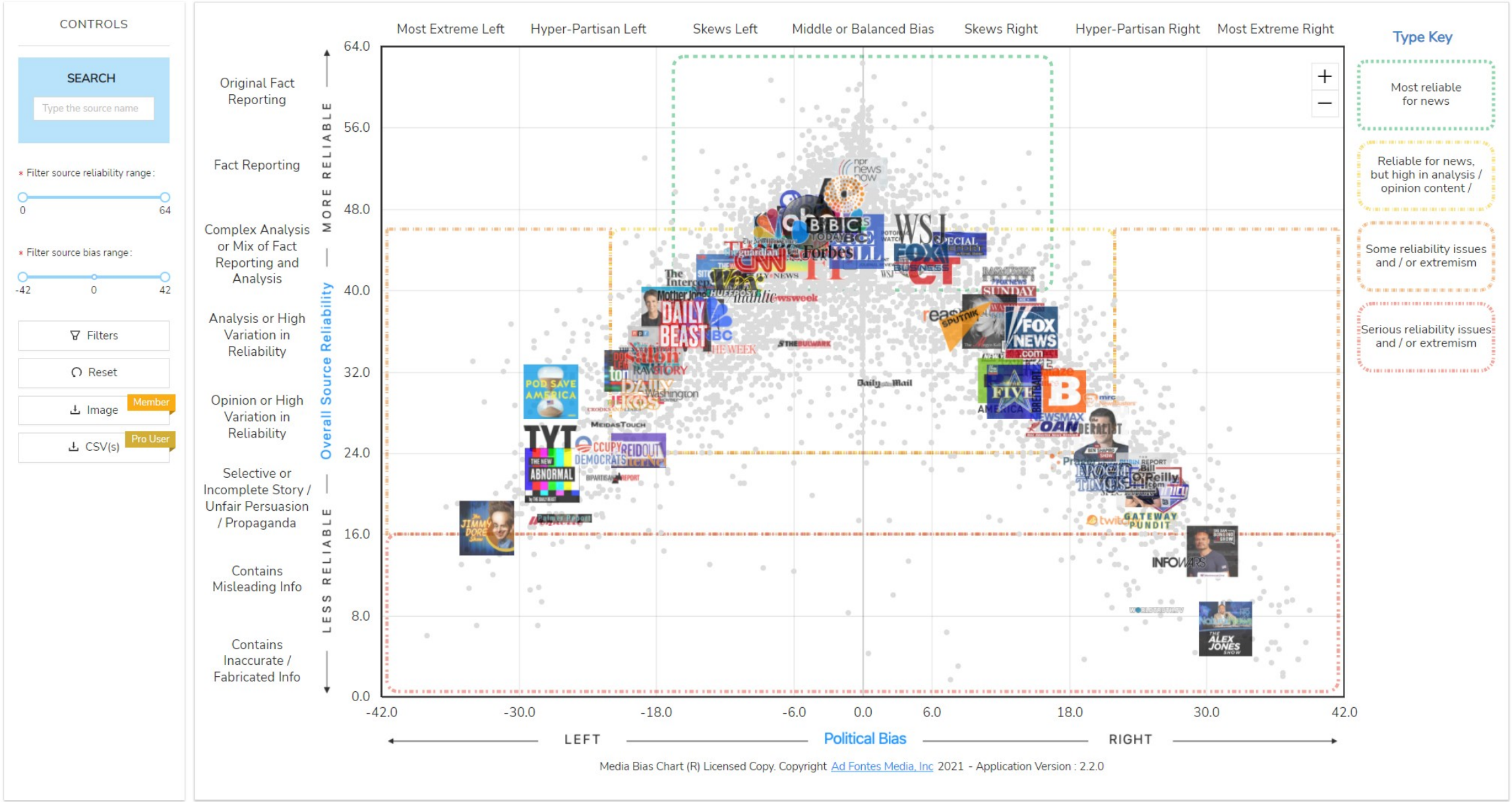
https://adfontesmedia.com/product/media-bias-chart-latest-edition/



https://libguides.willamette.edu/c.php?g=697656&p=4958591

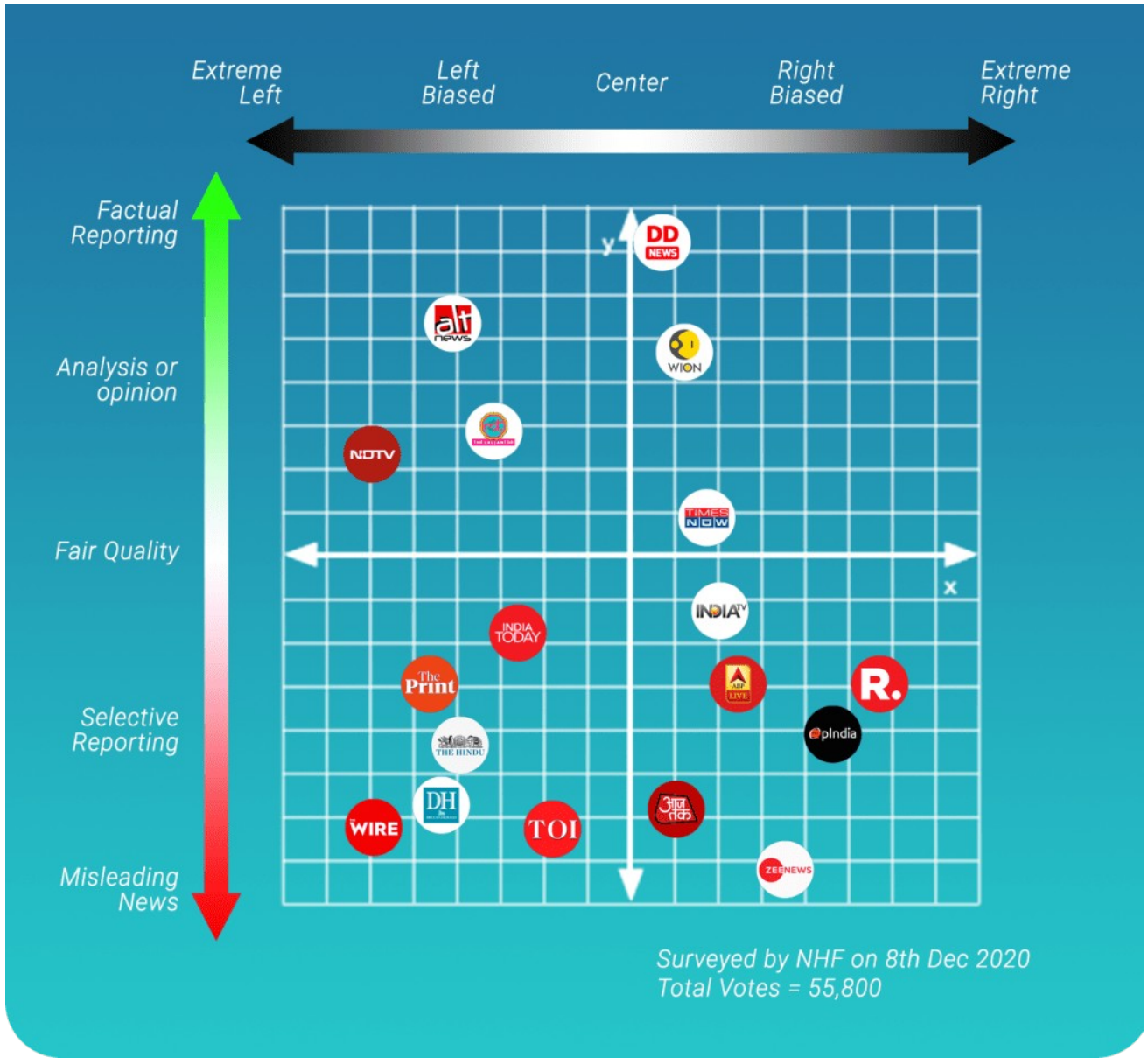


https://adfontesmedia.com/interactive-media-bias-chart/



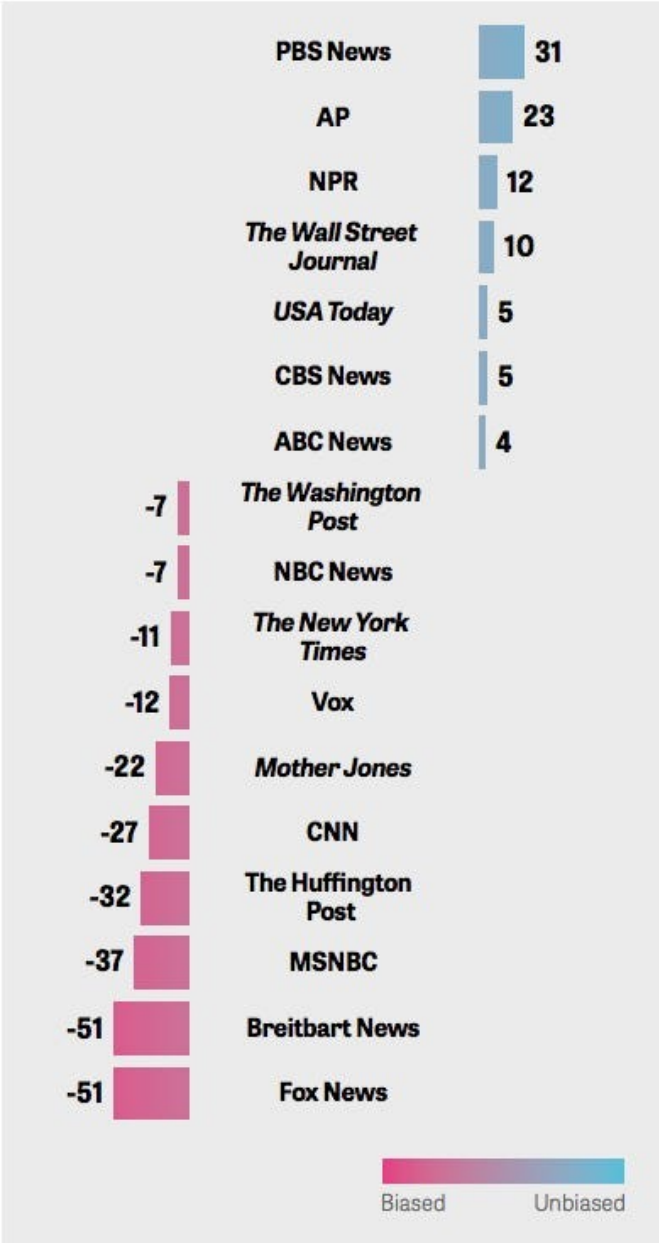
Countries other than US

https://newsheadlinesfixed.com/the-most-accurate-indian-media-bias-chart-2020/



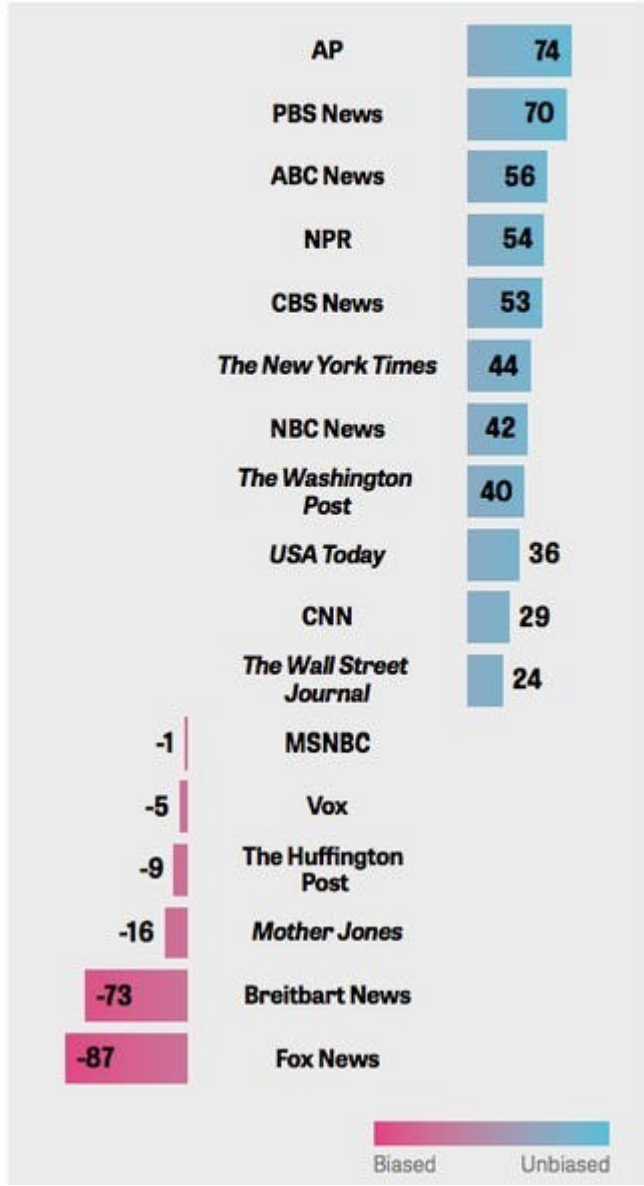
Net Bias Scores of News Organizations, U.S. Adults

Percentage rating each as unbiased minus percentage rating each as biased



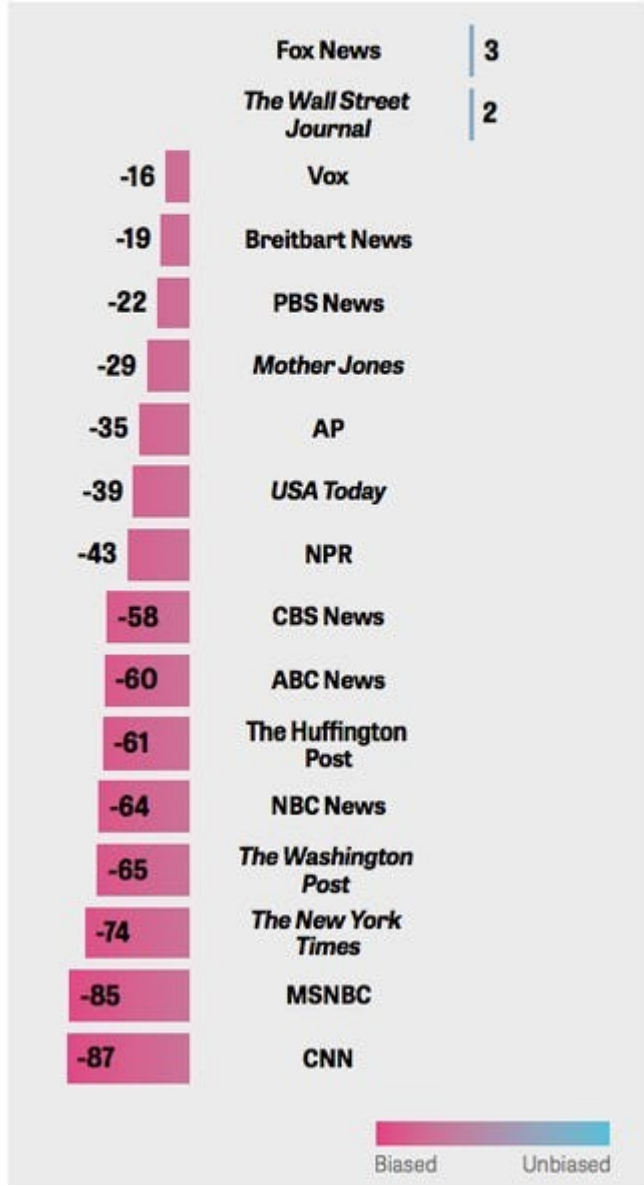
Net Bias Scores of News Organizations, Democrats/Democratic Leaners

Percentage rating each as unbiased minus percentage rating each as biased



Net Bias Scores of News Organizations, Republicans/Republican Leaners

Percentage rating each as unbiased minus percentage rating each as biased



Proposed research questions:

- Can we produce a Media Bias Chart for news websites/outlets that has not been surveyed based on the similarities of textual data with news outlets that has been surveyed or went under researchers' studies?
- Can we apply the same method to produce a Media Bias chart from Arabic news websites/outlets based on their similarities with English ones?
- Can we produce a mapping between English news outlets and Arabic ones?

Textual Data Sources:

- News Outlet’s Website:
  1. Site Map
  2. RSS feed
  3. Scrapping Homepage
- News Outlet’s accounts on social media (Twitter, etc)
- Archived news pages:
  1. Crawled Homepages
  2. Homepages pushed to an archive on the web regularly
- Archived social media posts

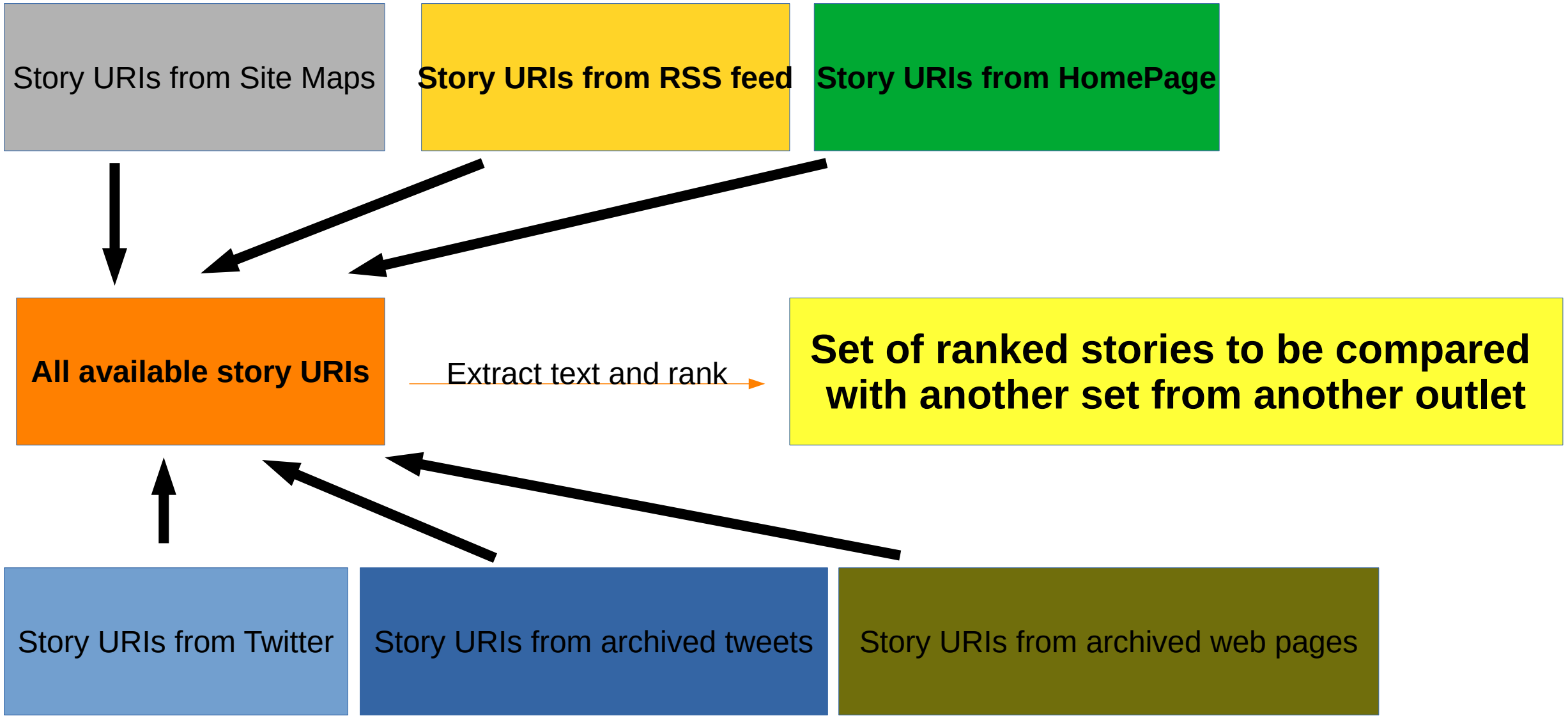
News Story Ranking Per News Outlet

- News stories collected from a news outlet’s websites can be ranked if it was scraped from the homepage.

- Links to news stories from news outlets’ social media accounts are difficult to rank because posts cannot be rearranged. Example: In Twitter, only one tweet can be pinned at any time.

- It is possible to collect the “pinned” status of a news story(s) but ranking stories that never been pinned is not possible (at least in Twitter).

Textual Data Sources (each outlet):



News Story Ranking on News Website Homepage

- A homepage is divided into areas based on their position.
- An integer between 1 and 10 represents the area ranking.
- News stories that belong to the same area have the same ranking.

Measures

- Similarities between stories within one news outlet and with other news outlets.
- Story and story position matching between news outlets (Find Similarities between news outlets)
- Combining all stories from all news outlets to measure coverage for each news outlet
- Story and story position matching between English and Arabic news outlets (Find Similarities between news outlets across both languages and create a mapping from one to the other based on story and position matching)

Arabic News Outlets Clusters 11/24/2021

Clustering Arabic News Outlets Based on:

- Similarities between stories
- Similarities between stories’ positions
- Similarities between stories’ duration in positions

K-Means Clustering

- Unsupervised machine learning algorithm
- Nondeterministic
- Number of clusters is preassigned
- Several runs of the algorithm is required to choose the best initialization of centroids by choosing the one with the lowest SSE (error)

Using BoW Model (Bag of Words)

- Collect stories from news outlet websites for a specific day.
- Extract named entities from each story and combine all named entities grouped by news outlet.
- Use TF-IDF to extract Features Matrix.
- It is possible to have cluster prediction for new data from a new media outlet.