

# Programm

1. My hypothesis
2. Data source
3. Data pre-processing and sentiment analysis
4. Observations
5. Conclusions
6. What would I change?

# My Hypothesis

1. Feminist porn is a movement that is more present in Germany than in other countries.
2. That movement is also perceived better in Germany than in other parts of the world.

# Data Source

I tried first to get my data from different APIs and libraries:

- <https://newsapi.org/>
- <https://webhose.io/>
- <https://pypi.org/project/GoogleNews/>

These methods seemed great but they were too limiting in term of price or archived data

**Scraping Google News was the answer!**

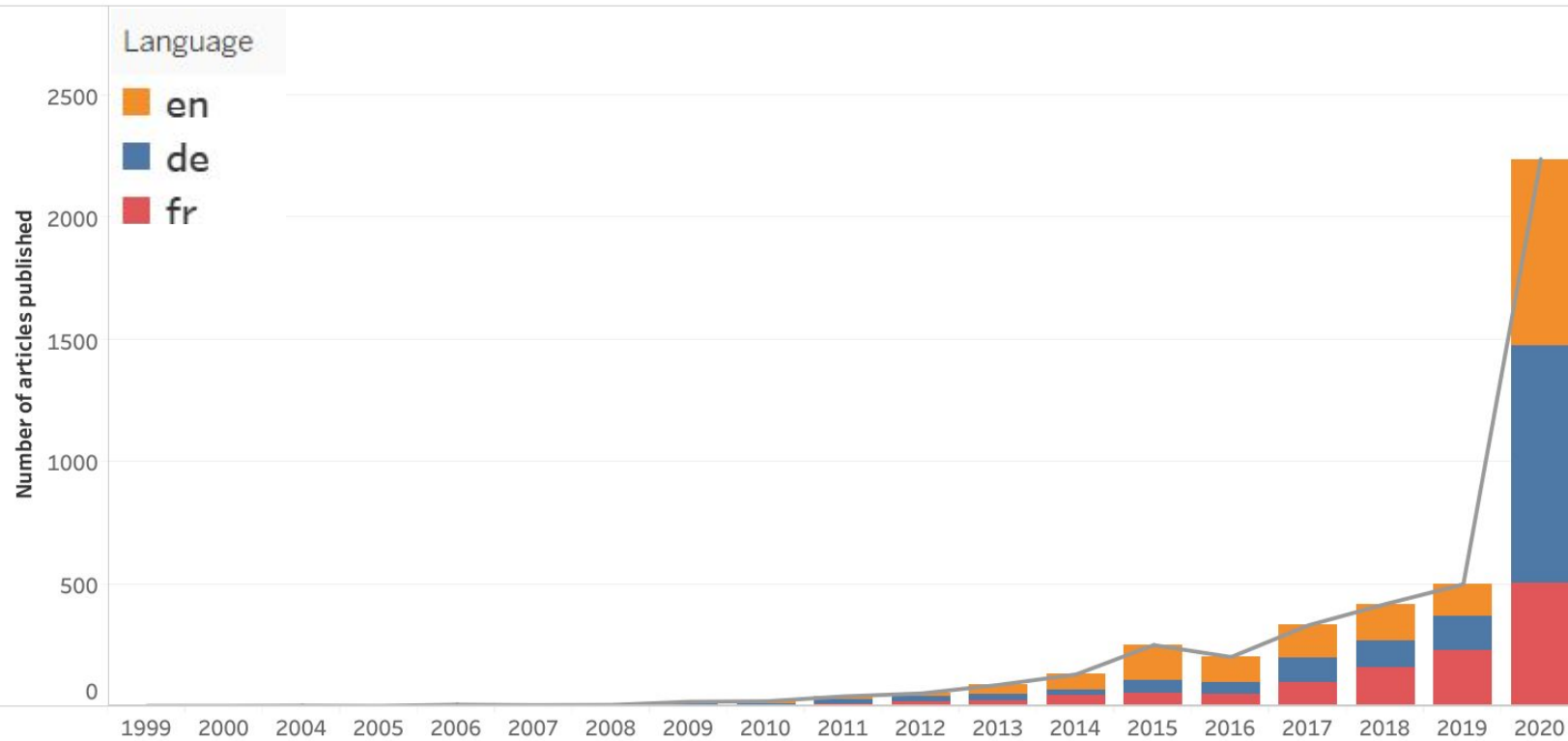
I could not directly scrape the country so I decided to focus on languages instead and scrape articles in: **EN, DE, FR**

# Data pre-processing and sentiment analysis

1. Scraping the data with BeautifulSoup using different keywords
2. Set up one dataframe for each language and clean them
3. Process the data pre sentiment analysis (different depending on the language)
4. Run a sentiment analysis on each dataframe
5. Merge all the data frames together

# Observations

Traction of feminist porn in the media over time

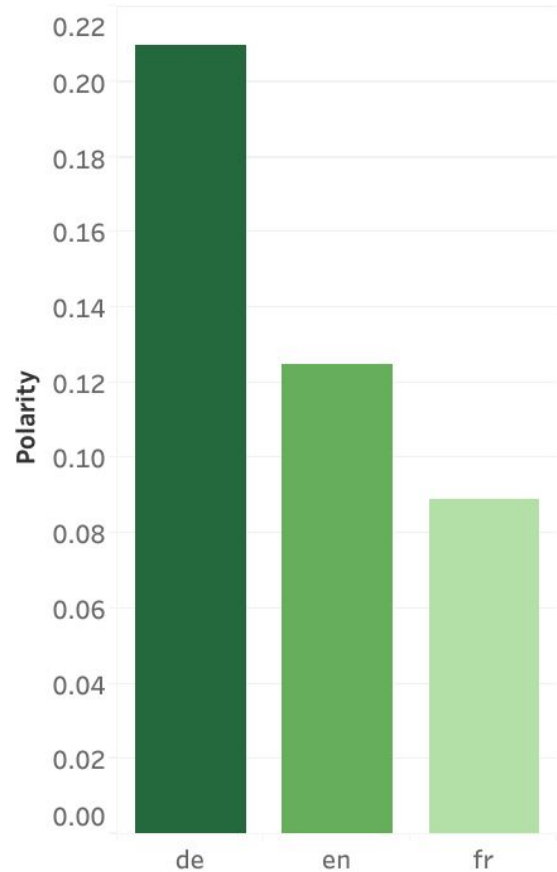


# Observations





## Perception of feminist porn in different languages



# Conclusions

1. A bit more articles on feminist porn in german does not prove a real difference in interest in german speaking countries.
2. There seems to be clearly a more positive view on feminist porn in german articles than in the other language. French being the least positive.
3. I would need to get more and more reliable data in order to be able to look more into details.  
(4304 articles)



# What would I change?

- The amount of articles scraped was not that great and the method to do it could have been more efficient.
- I would have loved to have data on the country the articles was written in.
- The calculation of the polarity was done with the same library (TextBlob), but with different extensions.