



# Topic modeling of job reviews

Andrew Hall

September 2022

Metis NLP short course

# **Why analyze anonymized job reviews?**

Job reviews on public sites (e.g., Glassdoor, LinkedIn) are increasingly prevalent

Companies need to know the signal they are sending to potential employees and whether they represent action items for current employees

Applying topic modeling to job reviews can help inform how much importance a company should put on interpreting online job reviews

# Project aims

- **Project Goal:** Implement topic modeling to derive business-relevant conclusions about a corpus of job reviews from an anonymous job review website
  - Are there detectable topics that can be extracted in the anonymous reviews that employees leave for their companies online?
  - How does this structure differ based on the topic modeling approach implemented?
  - Are there meaningful differences in the adjectives used to describe one's job when prompted to leave a positive (Pros) vs. negative (Cons) reviews?

# Data pipeline

## Data source



## Data manipulation



- Data extracted from Blind API<sup>1</sup>

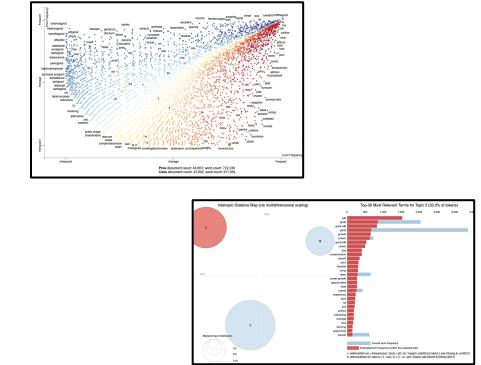
- Data cleaning
- Vectorizers (Count and TF-IDF)

## NLP/Dim reduction



- Document-term matrix construction
- Adjective, term identifier
- Dimensionality reduction using LDA, NMF, and LSA

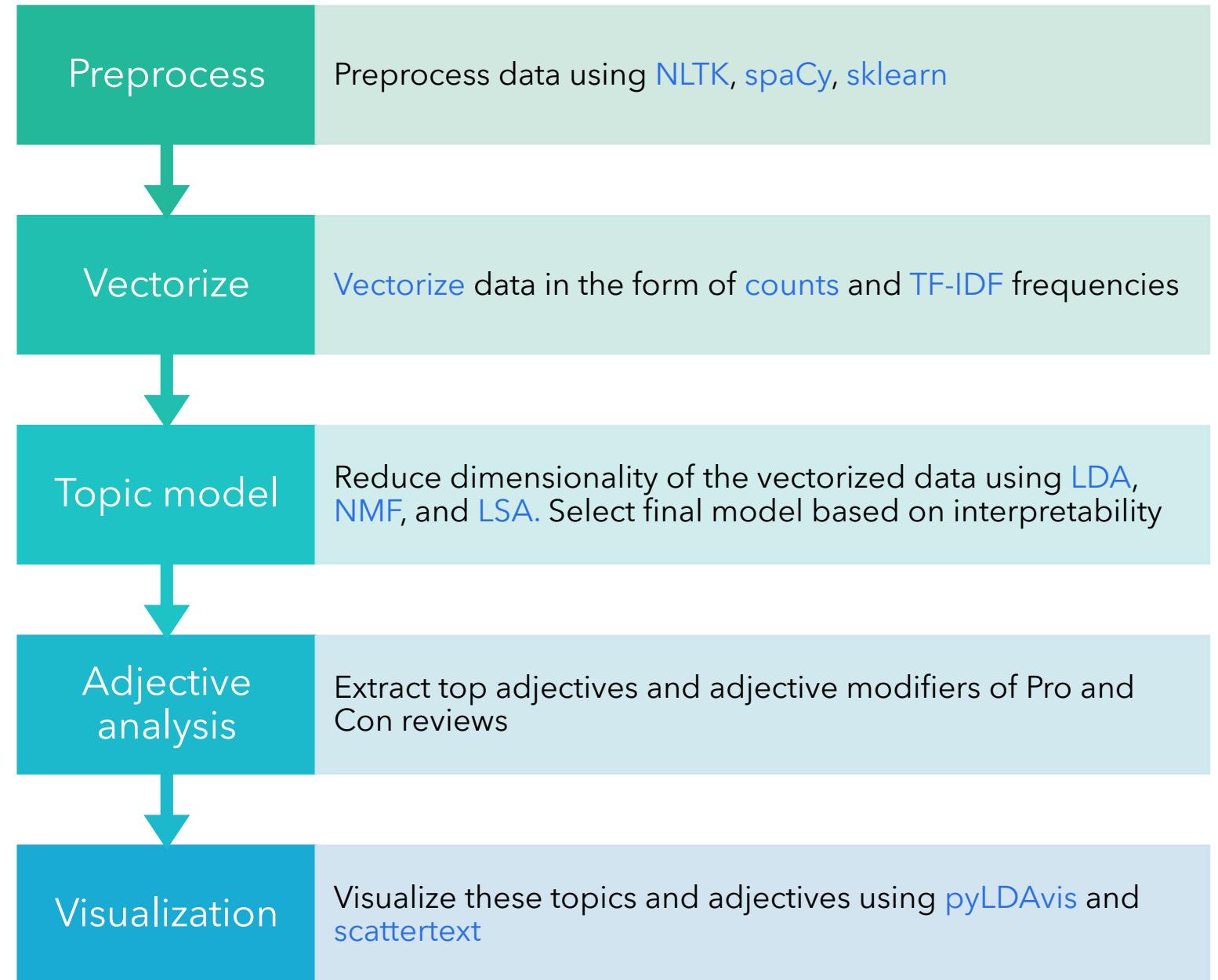
## Visualizations



- Topics visualized using pyLDAvis and scattertext

<sup>1</sup>Harsh Bardhan Mishra. (2022). *Blind App – Company Reviews* [Data set]. Kaggle. <https://doi.org/10.34740/KAGGLE/DSV/3603028>

# Approach to analysis



# Final model: LDA with 3 topics



## 1) Work life balance

Top words: *Life balance, work life, balance, life*



## 2) Culture and teaming

Top words: *Culture, team, growth, good*

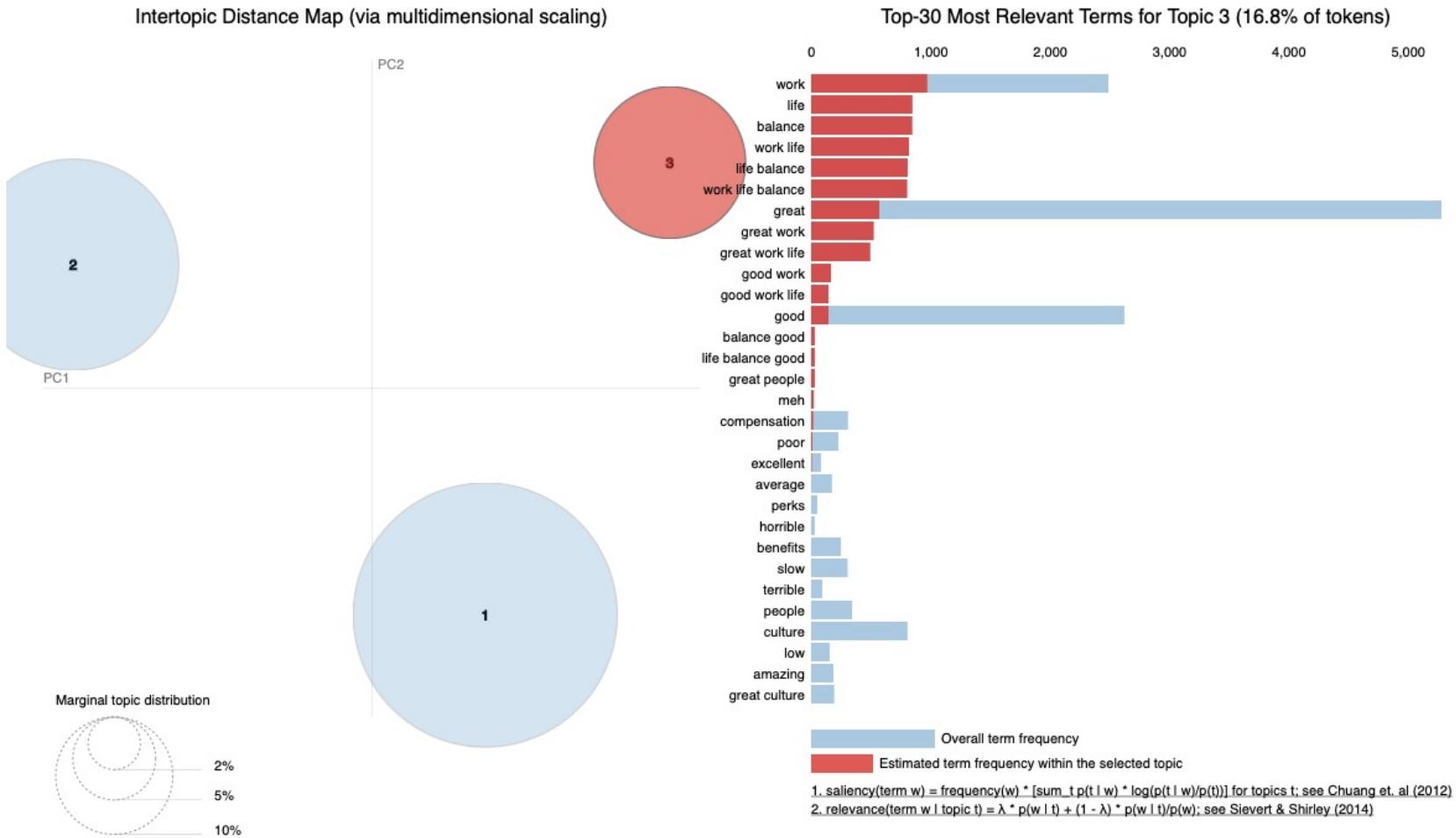


## 3) Company aspects

Top words: *Great, company, place, work*

Selected Topic: 3    Previous Topic    Next Topic    Clear Topic

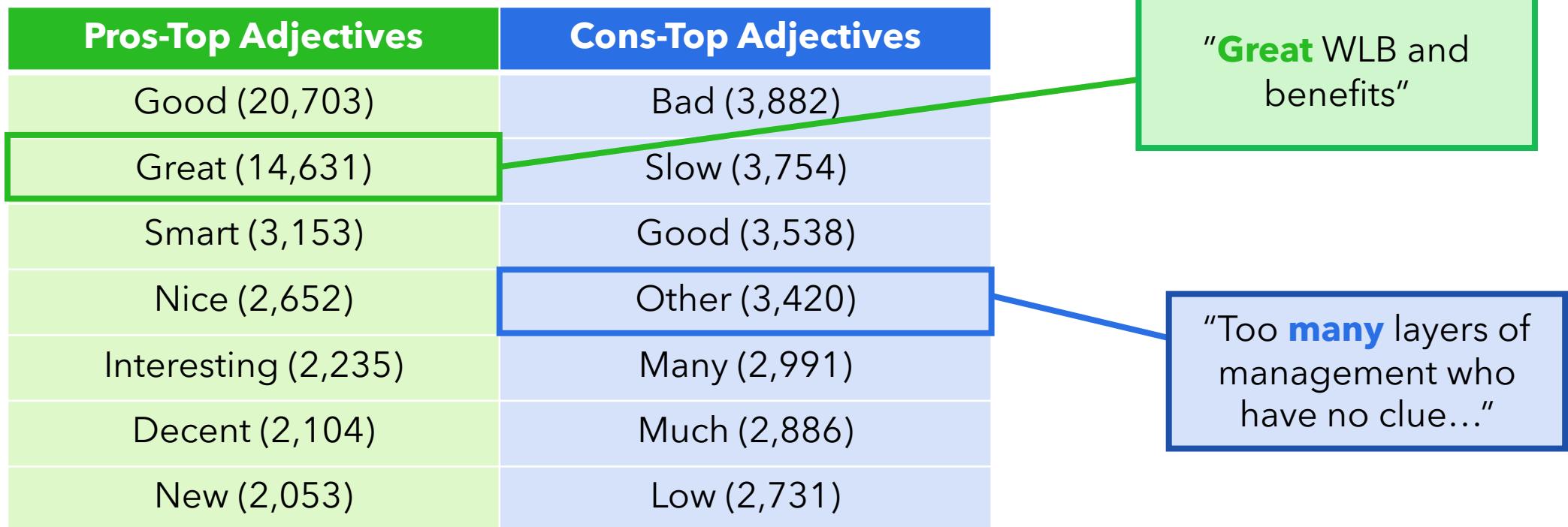
Slide to adjust relevance metric:(2)  
 $\lambda = 1$     0.0 0.2 0.4 0.6 0.8 1



Sample output from pyLDAvis:

- Topic 3 (**Work life balance**) has top terms that are clearly relevant to work life balance

# Comparing adjectives between pros and cons

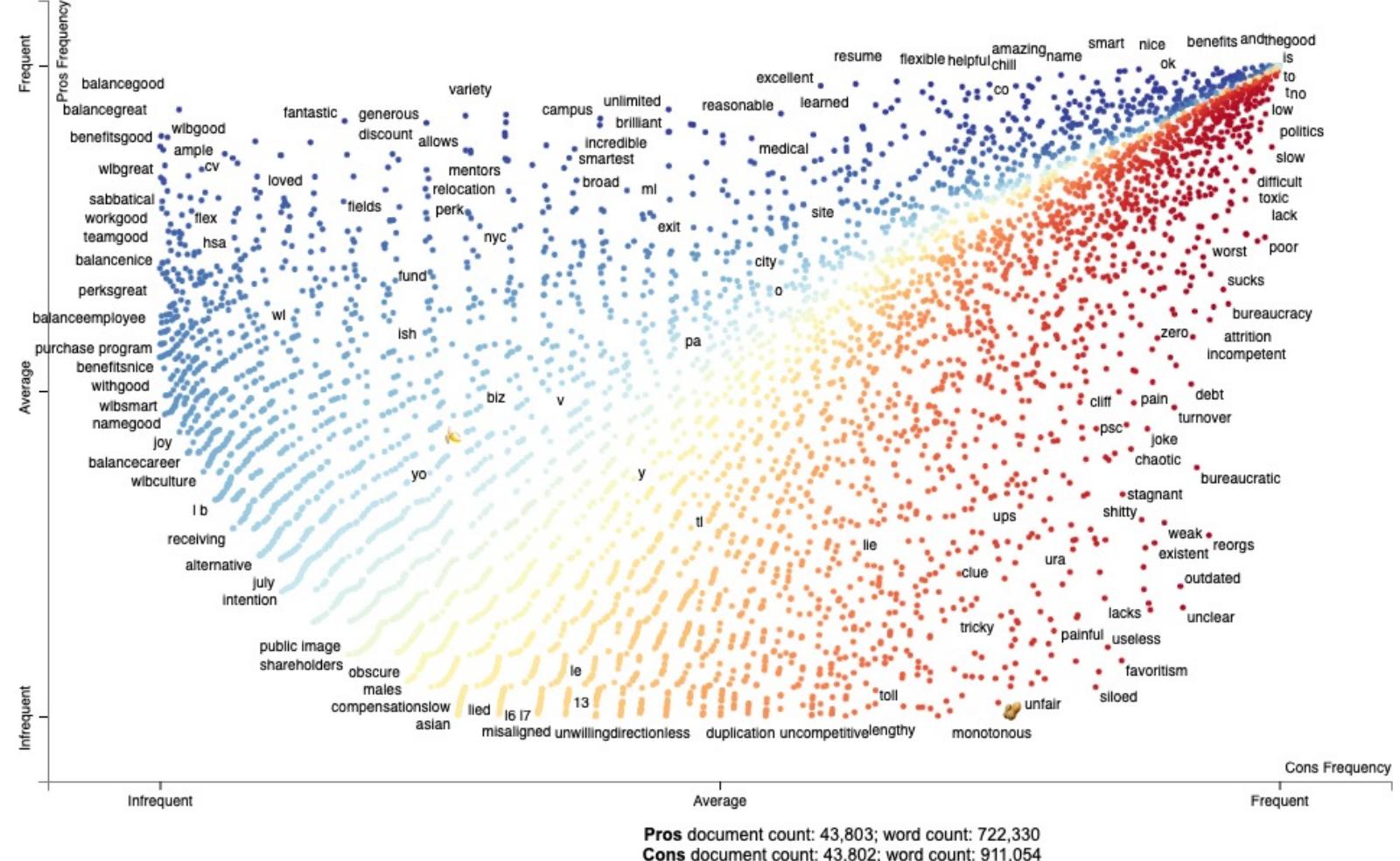


# Adjective modifiers of “Company”

When people refer to their “company” they are likely to modify it with an adjective. For example, a “great” company or a “bad” company. How do these adjective modifiers differ between pro and con reviews?

Pros-Modifiers of “Company”	Cons-Modifiers of “Company”
Great (298)	Big (327)
Good (273)	Large (200)
Big (189)	Other (88)
Large (105)	Huge (61)
Stable (86)	Tech (52)
Other (67)	Good (48)
Best (59)	Old (26)

- Pro reviews tend to focus on the positive qualities of the company (“good,” “great”)
- Con reviews focus on the size of the company, with “big” and “large” at the top



Top Pros	Overall top terms
Amazing	Wlb
Nice	Perks
Decent	Faang
Smart	Coworkers
Resume	Refreshers
Top Cons	Orgs
Poor	Covid
Slow	Oncall
Lack	AWS
Political	WFH
Terrible	Compensation

Scattertext shows top terms for "Pro" and "Con" reviews as well as overall

# THANK YOU!