Jeopardy Questions: Can They be Generalized?

Louisa Reilly
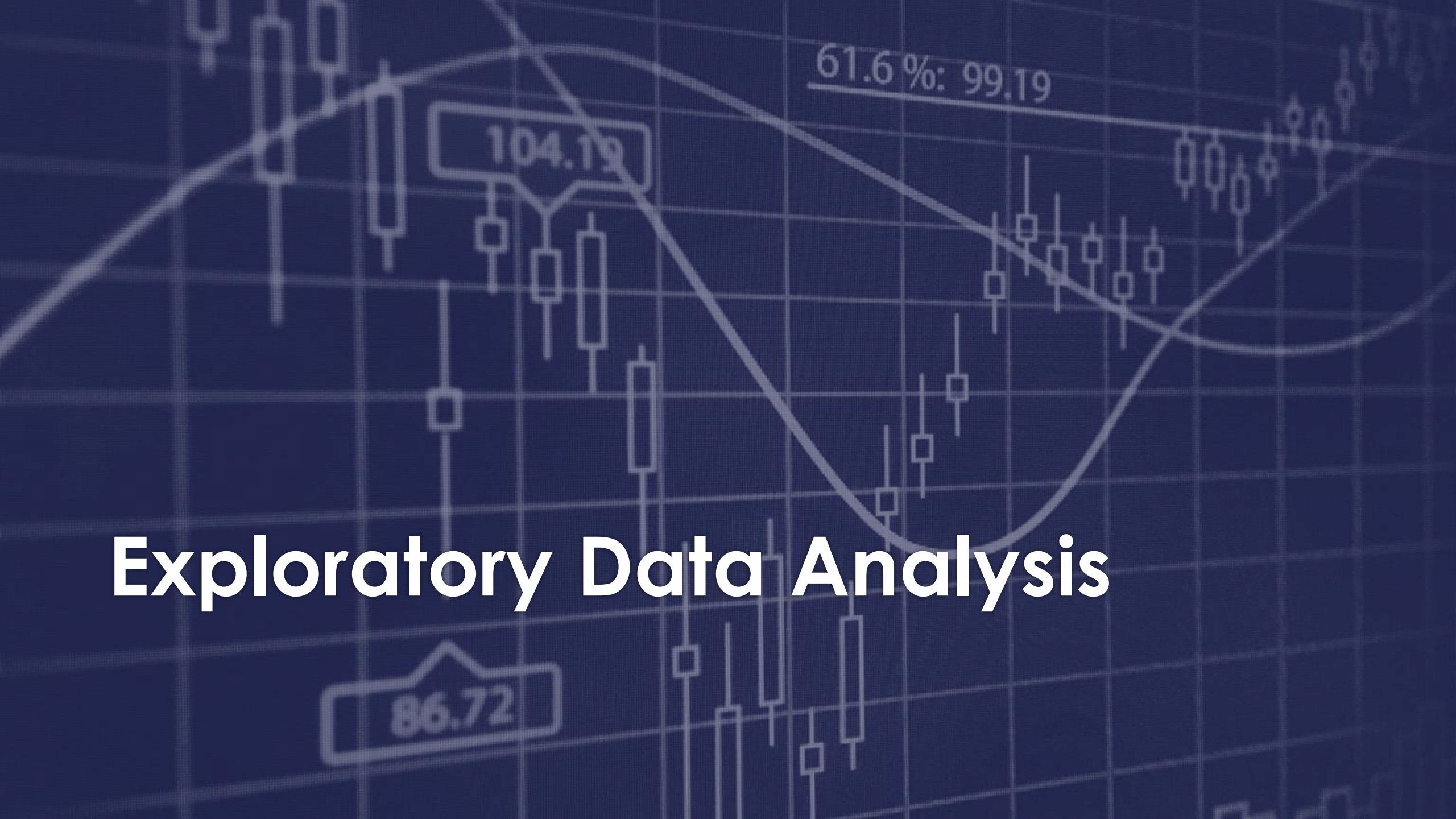
In case you have not seen the show…

- **Answer → Question**
- Contestants are given quiz clues in the form of answers
- They must respond in the form of a question.
- Audition:
  - Exam, IRL group audition, mock game, interviews
- No previous contestants
- Have not auditioned for at least the past 18 months

EDIBLE RHYME TIME	BOOKS IN GERMAN	3 "T"s	CHOP CHOP!	THEY SAID IT WOULDN'T LAST	THEY WERE RIGHT
$200	$200	$200	$200	$200	$200
$400	$400	$400	$400	$400	$400
$600	$600	$600	$600	$600	$600
$800	$800	$800	$800	$800	$800
$1000	$1000	$1000	$1000	$1000	$1000

Outline

- Dataset
  - 216k Jeopardy questions and answers
    - Spans from 1984 to 2012
    - Scrapped from J-archive by a Reddit User
- EDA
  - Tableau
- Topic Modeling
  - NLP tools: NLTK, sklearn
  - Processing tools: pandas, regex
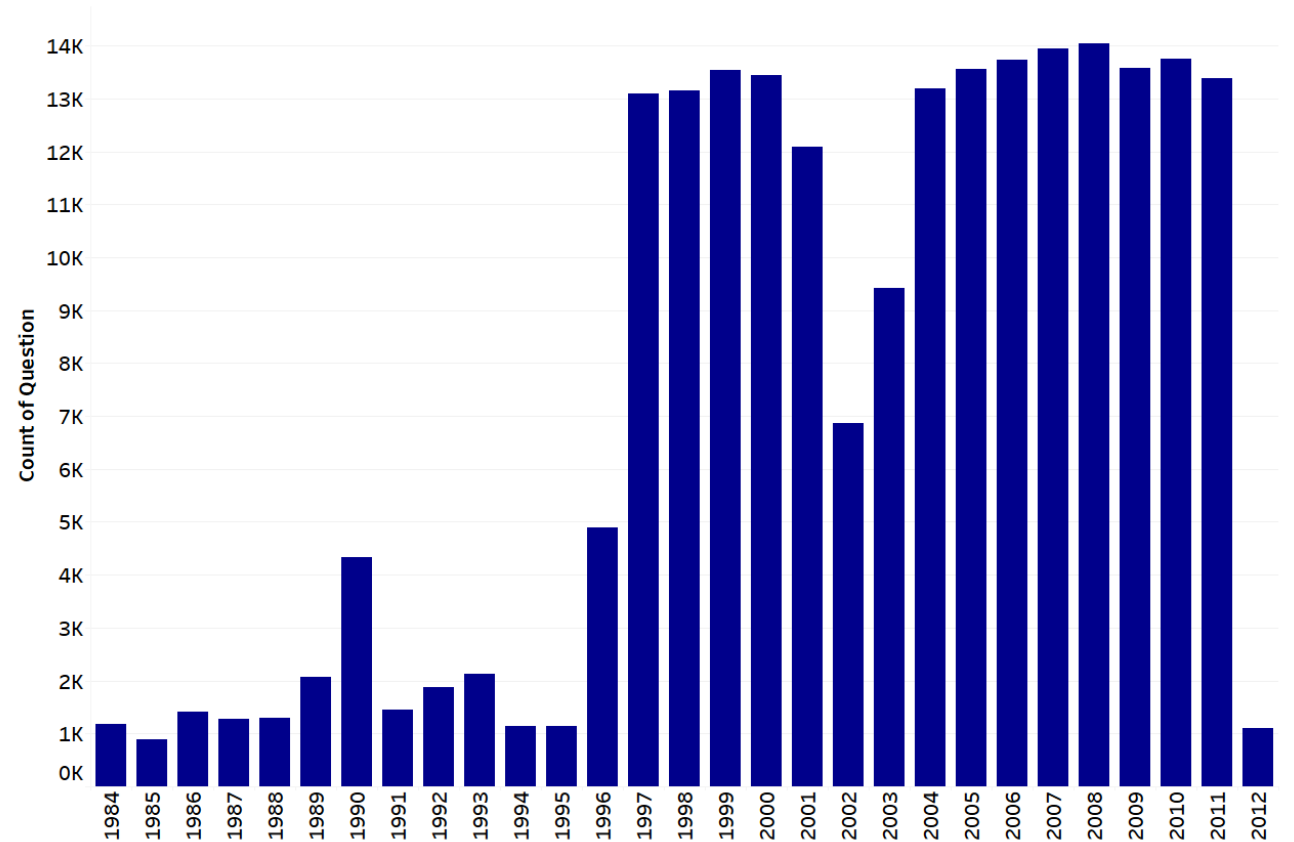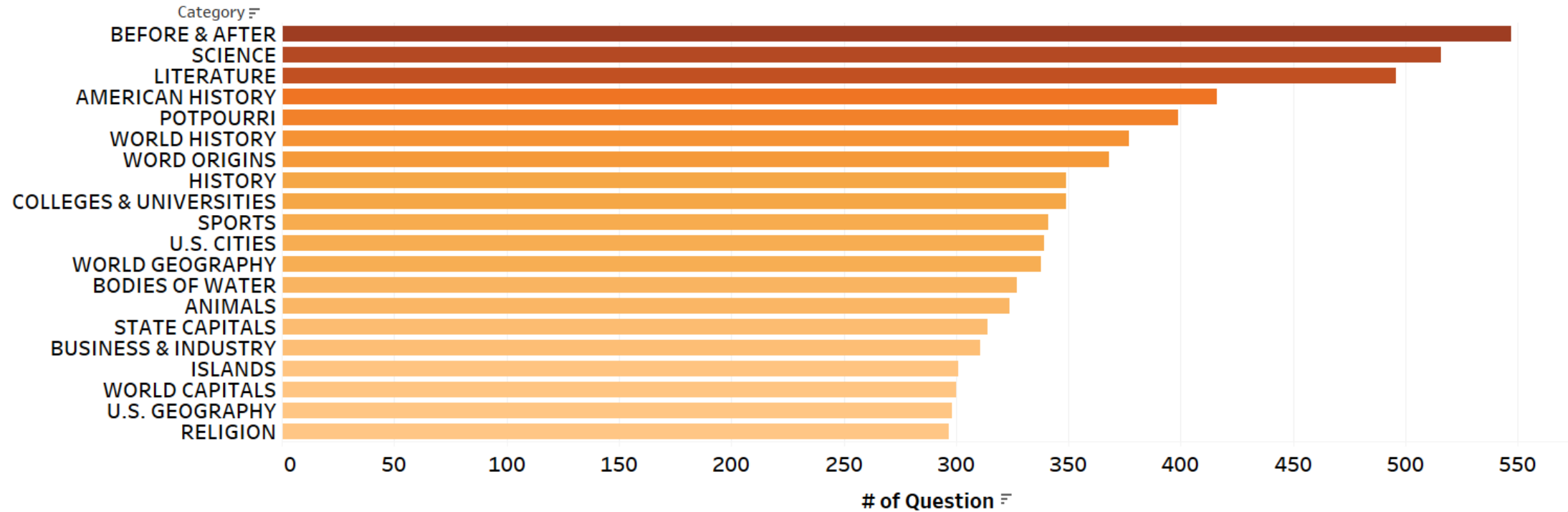 - Algo: NMF
  - Vectorizer: TFIDF

Exploratory Data Analysis

Majority of Q's in the 2000's decade
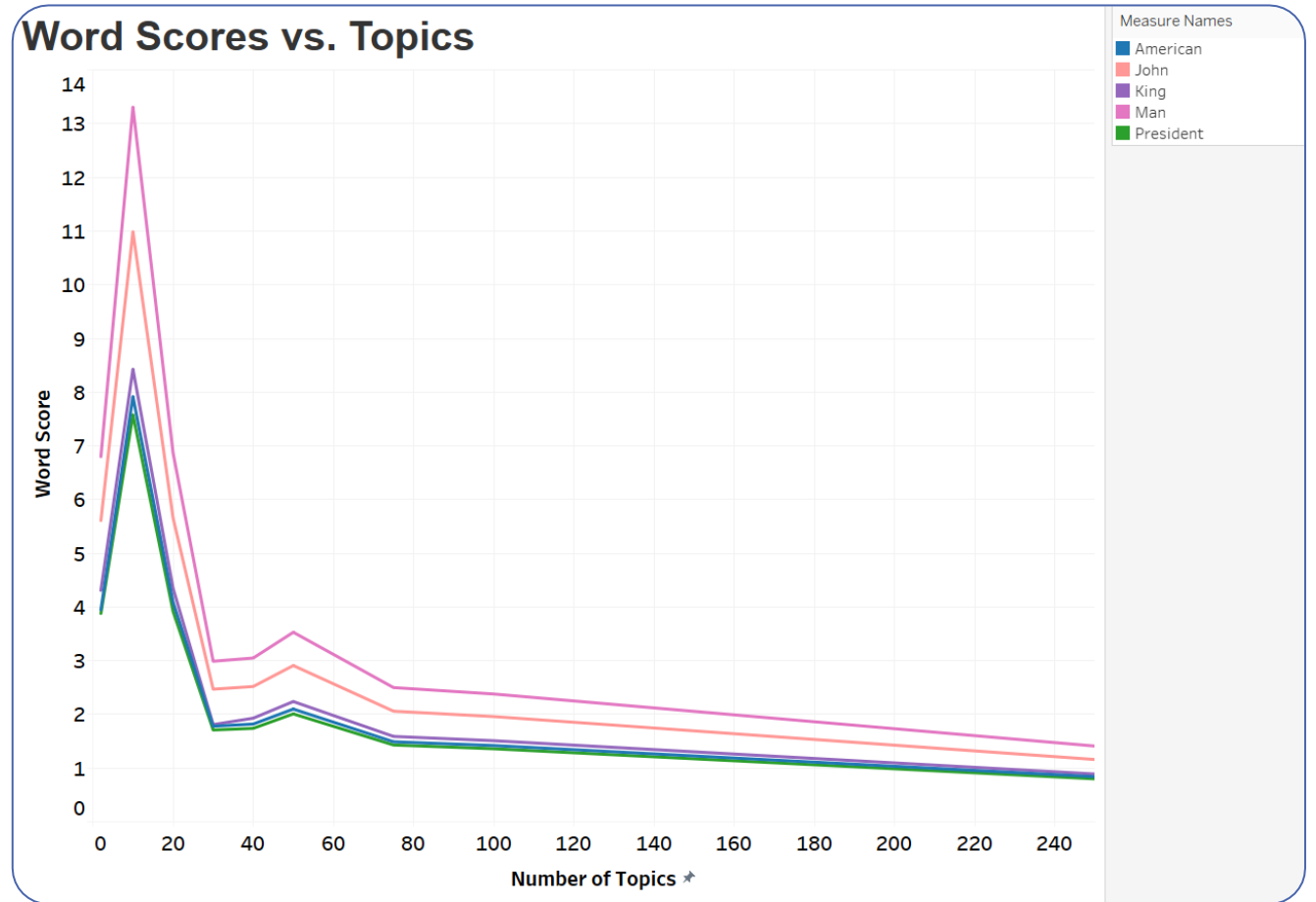
# of Questions per year
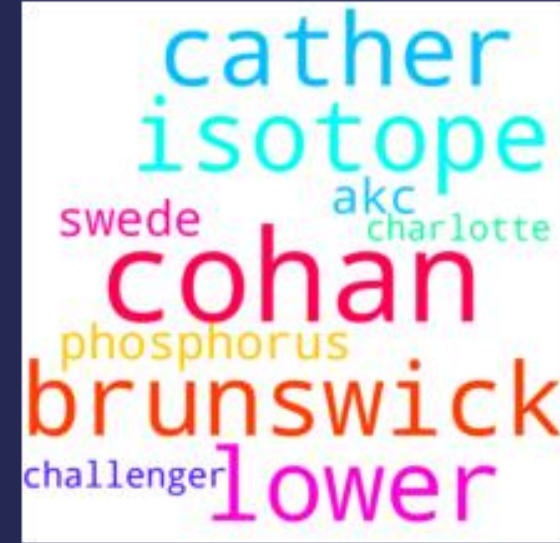
Top 20 Categories

- Emphasis on Geography and History
- Before & After = combo of 2 things to form an answer
- Potpourri = random unrelated questions

Topic Modeling

Increasing the number of topics

The topics: examples

- Science-ish
- General arts

Conclusions

Diverse dataset

Topic modeling needs more fine tuning

Filter for just nouns and adjectives

History

1964
Created by Merv Griffin
- Answer → Question
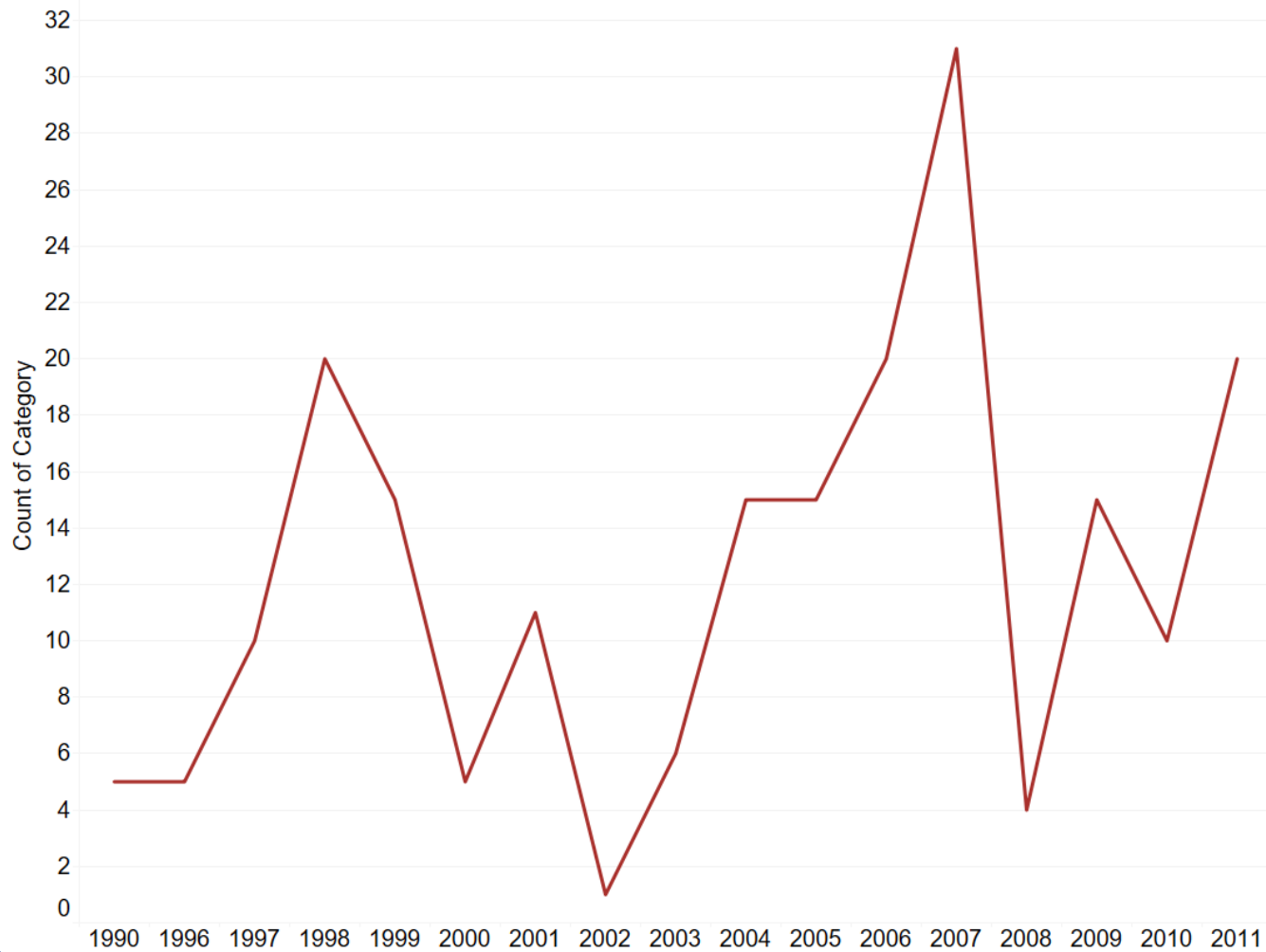
1964 - 1975
Aired on NBC

1984 - 2020
Revived by Sony and hosted by Alex Trebek

2020 - Present
Since Trebek's Death, different hosts but still going strong

Computer and Technology Categories

Tech Categories

- Did not start appearing until the 90s
- Increase overtime with a few dips
- Corresponding to economic dips?

Increasing the number of topics

Geography!

Word Scores vs. Topics