

The background is a dark blue, semi-transparent image of a Jeopardy! game board. It features a grid of 30 squares (5 rows by 6 columns). The top row contains category names: 'SELF-PORTRAIT OF THE ARTIST', '-OLOGIES', 'VISITING ITALY', 'WHOSE SONGS?', 'QUOTATIONS BY THE NUMBER', and 'THE SILENT LETTER IN...'. The subsequent rows contain monetary values: '\$400', '\$800', '\$1200', '\$1600', and '\$2000'. Each value is repeated across all six columns. The text 'Jeopardy Questions: Can They be Generalized?' is overlaid in large white font, and 'Louisa Reilly' is in a smaller white font at the bottom left.

Jeopardy Questions: Can They be Generalized?

Louisa Reilly

In case you have not seen the show...

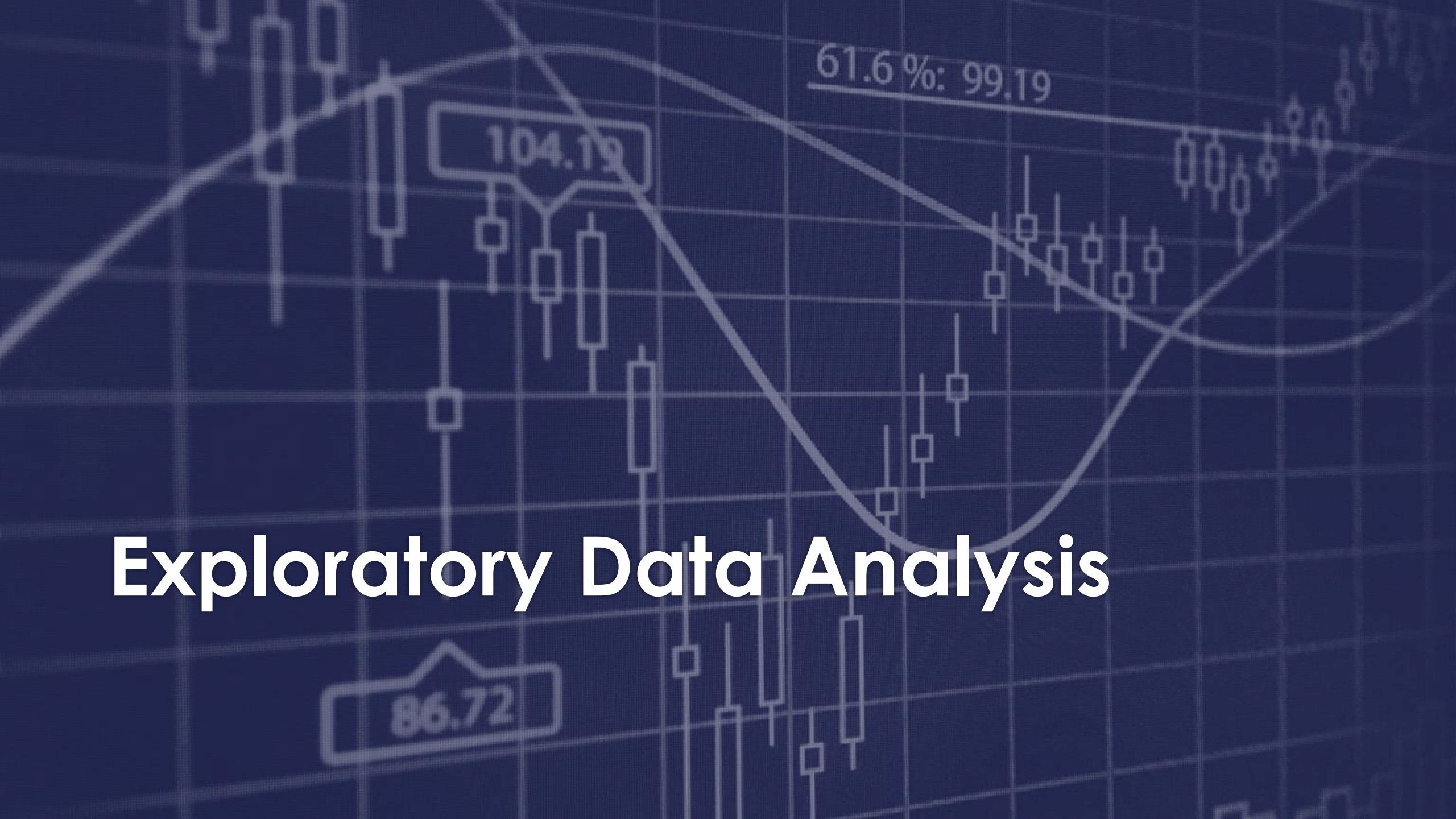
- Answer → Question
- Contestants are given quiz clues in the form of answers
- They must respond in the form of a question.
- Audition:
 - Exam, IRL group audition, mock game, interviews
- No previous contestants
- Have not auditioned for at least the past 18 months

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

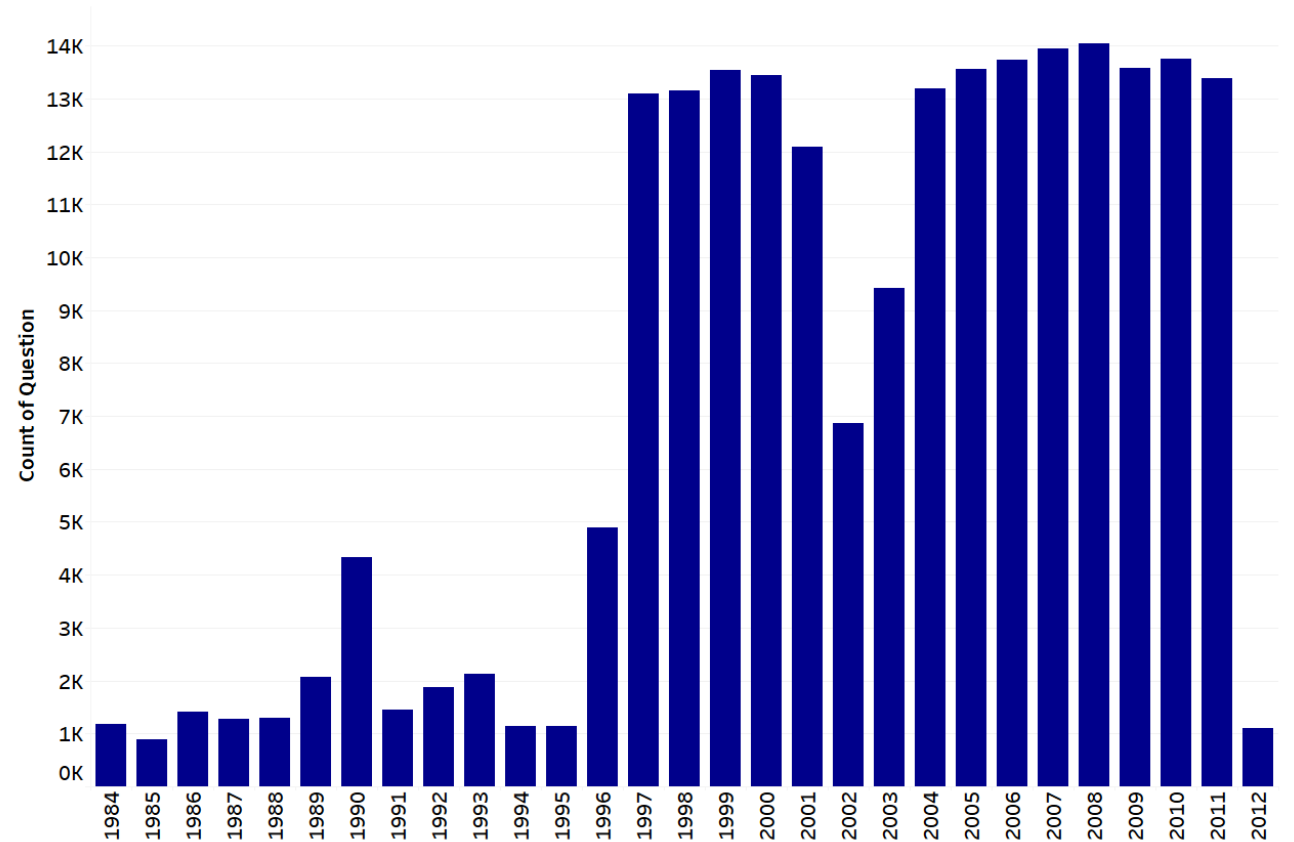




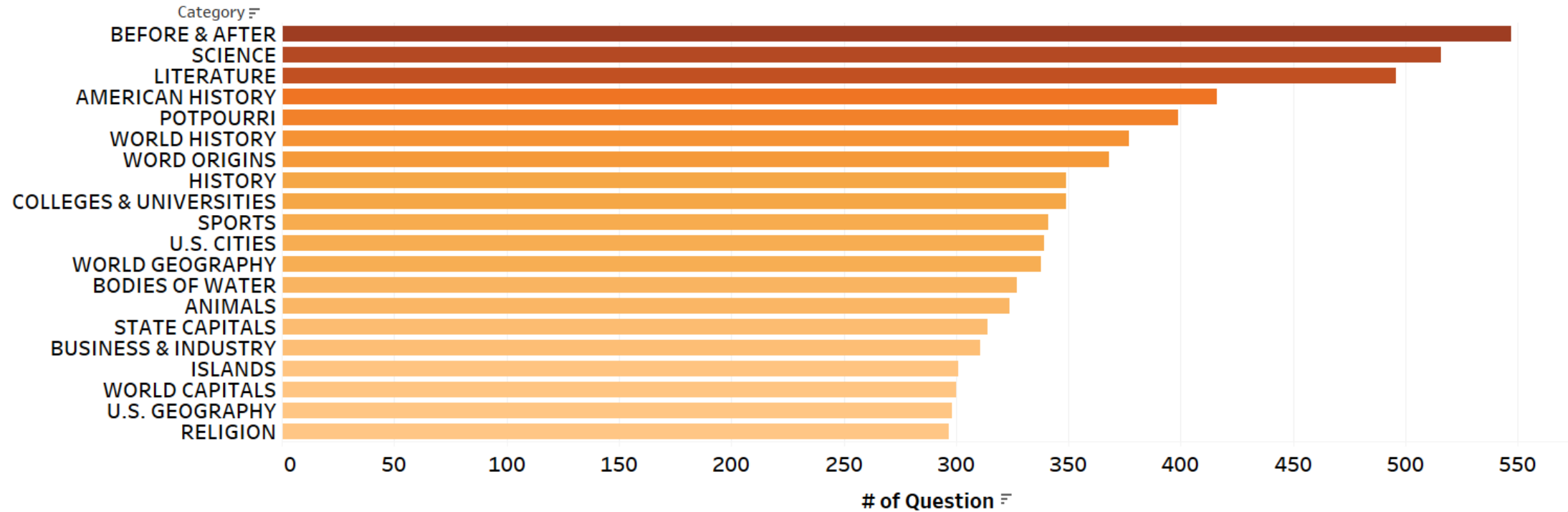
Exploratory Data Analysis

Majority of
Q's in the
2000's
decade

of Questions per year



Top 20 Categories



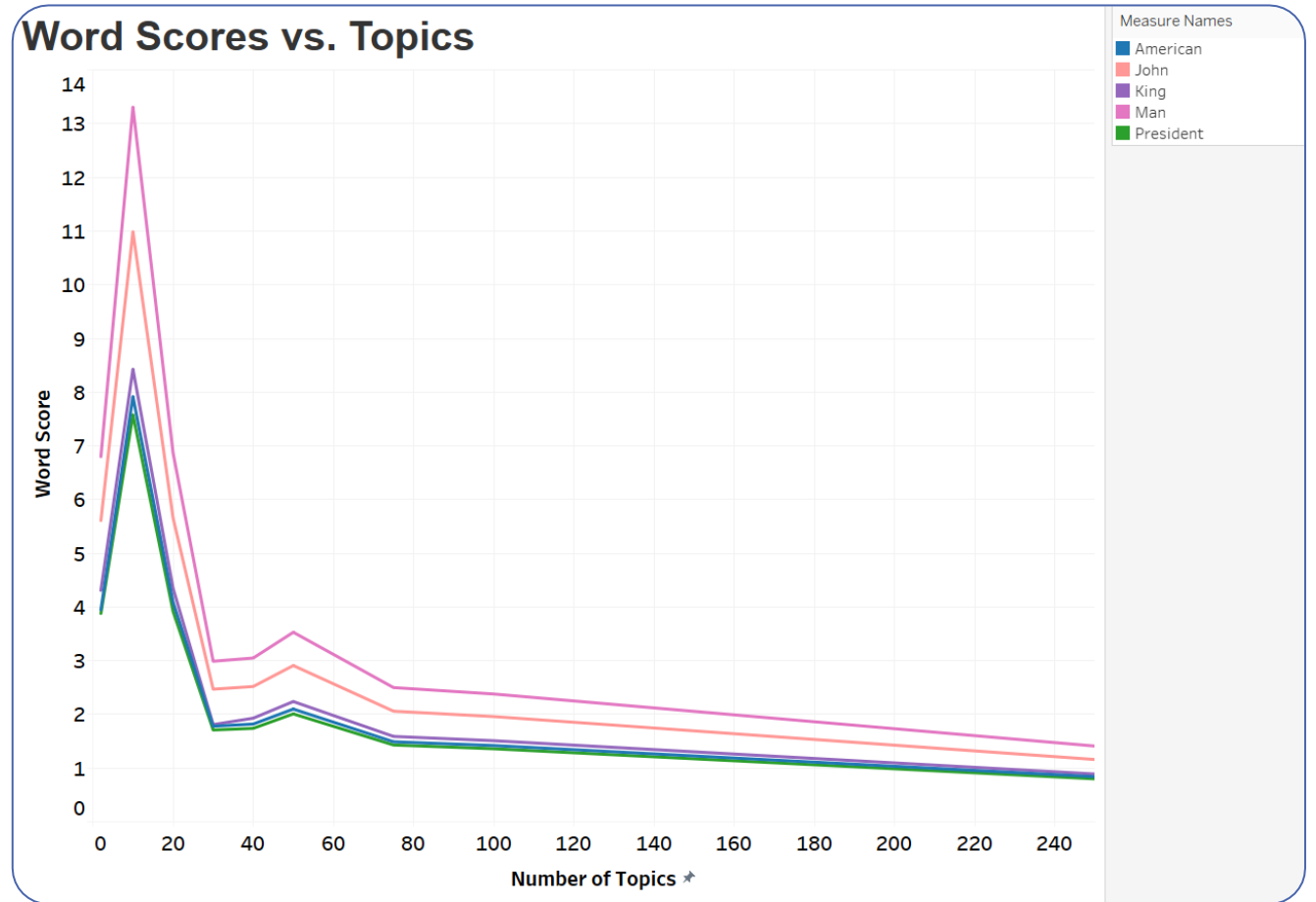
Top 20 Categories

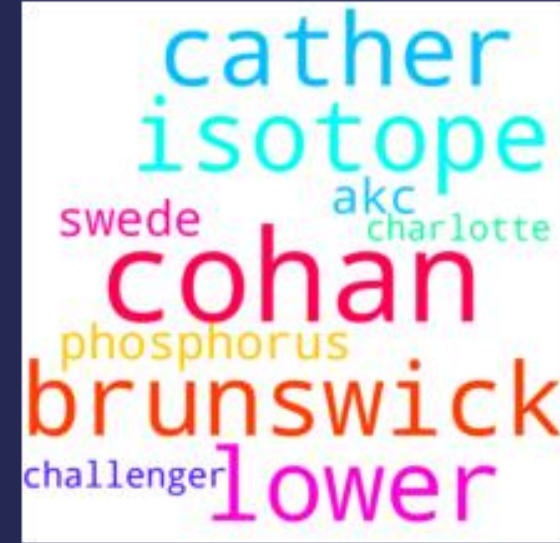
- Emphasis on Geography and History
- Before & After = combo of 2 things to form an answer
- Potpourri = random unrelated questions

Topic Modeling

The background of the slide is a dark blue gradient. Overlaid on this is a complex, abstract network of white lines and dots. The dots, representing nodes, are of varying sizes and are scattered across the right half of the image. They are interconnected by a dense web of thin white lines, creating a sense of connectivity and data flow. The overall effect is a high-tech, digital aesthetic.

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

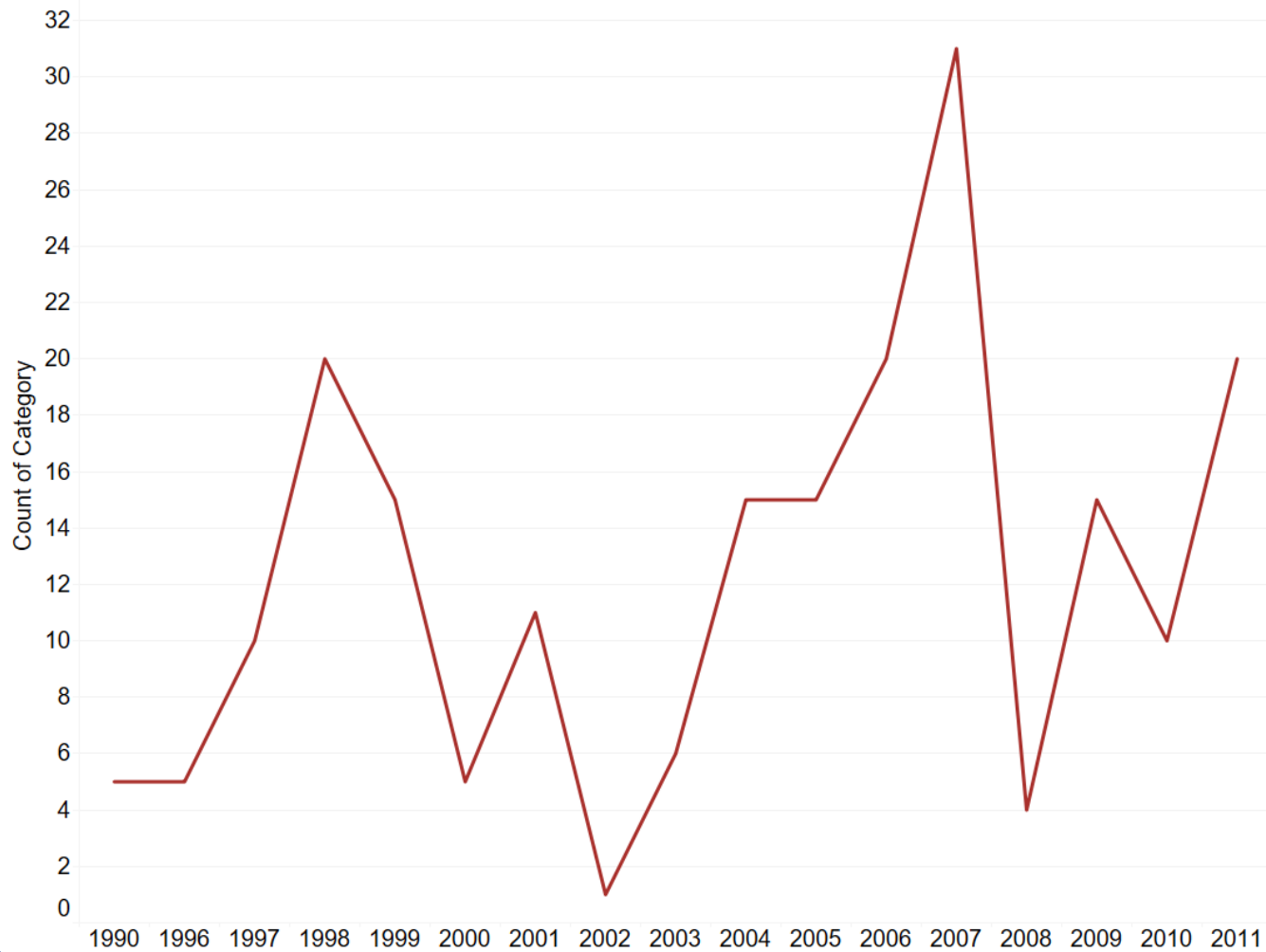
Revived by
Sony and
hosted by
Alex Trebek



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

