Your Worst Nightmare

childhood thoughts black red missing
college lock family follow forest officer friend real serial
never dead secrets mrs love society body man crazy
find far mind bad rules mother insane surrounded
old noise shadows horror night run live life legend late
room children dream scared death bloody
scary mr
sleep alone monster god dark dealer help woods killer final
demon urban hotel unknown
strange gone hell worse lady cursed nightmare
haunted creepy wish control dying future
ghost stuck dog wrong please mist

Brittney Hosac

Problem Statement

As members of Nightmare Publishing, a company that produces horror narratives, it is our mission to release stories that will leave our audiences not only capitaved, but absolutely terrified.

How can this be accomplished?
- We'll start off by asking, what is it that you're truly afraid of? With this knowledge, can we tap into our deepest fears and create a nightmarish novel.

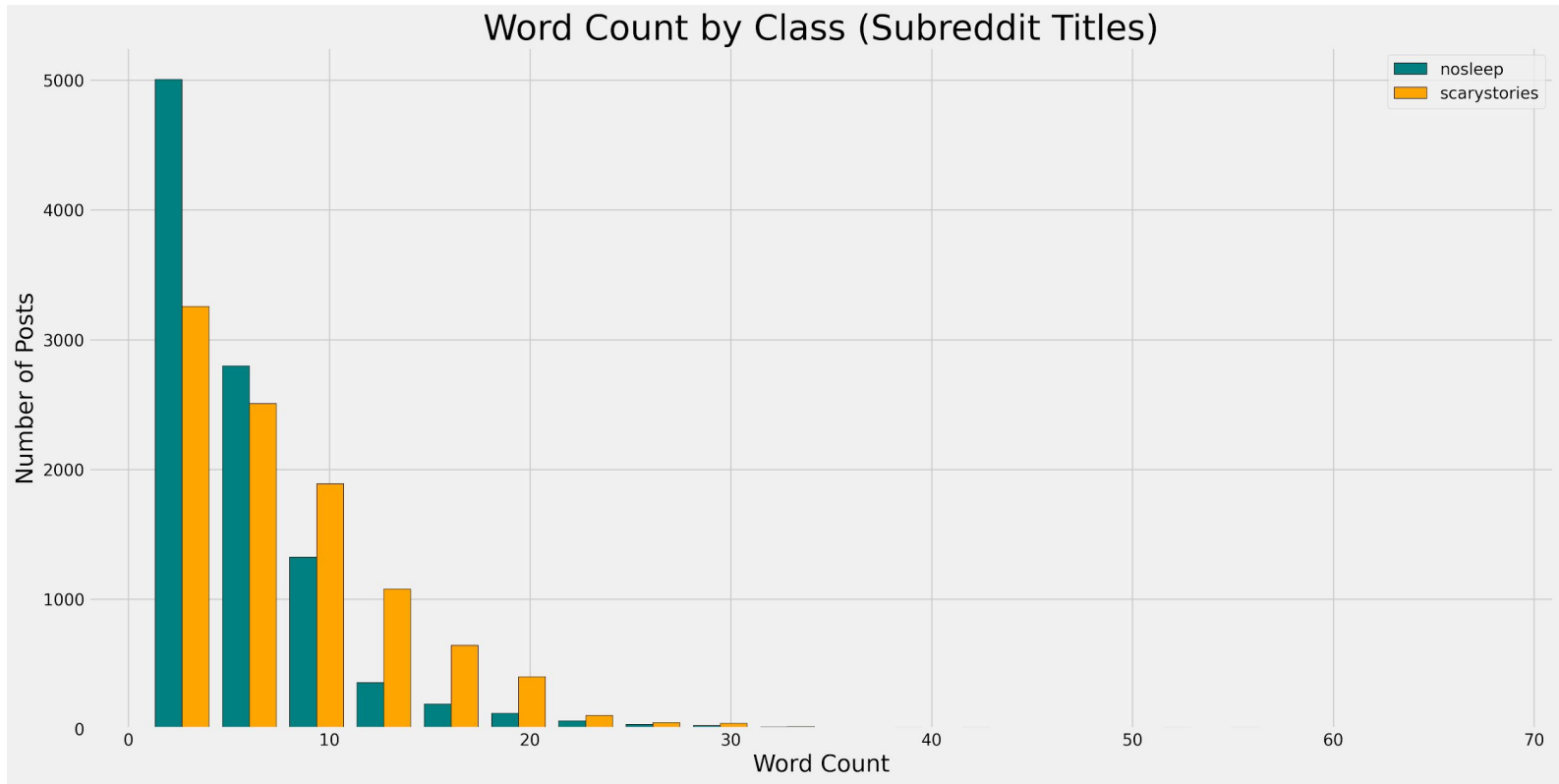# The Process

- ## Data wrangling
  - Focused on nosleep and scarystories subreddits
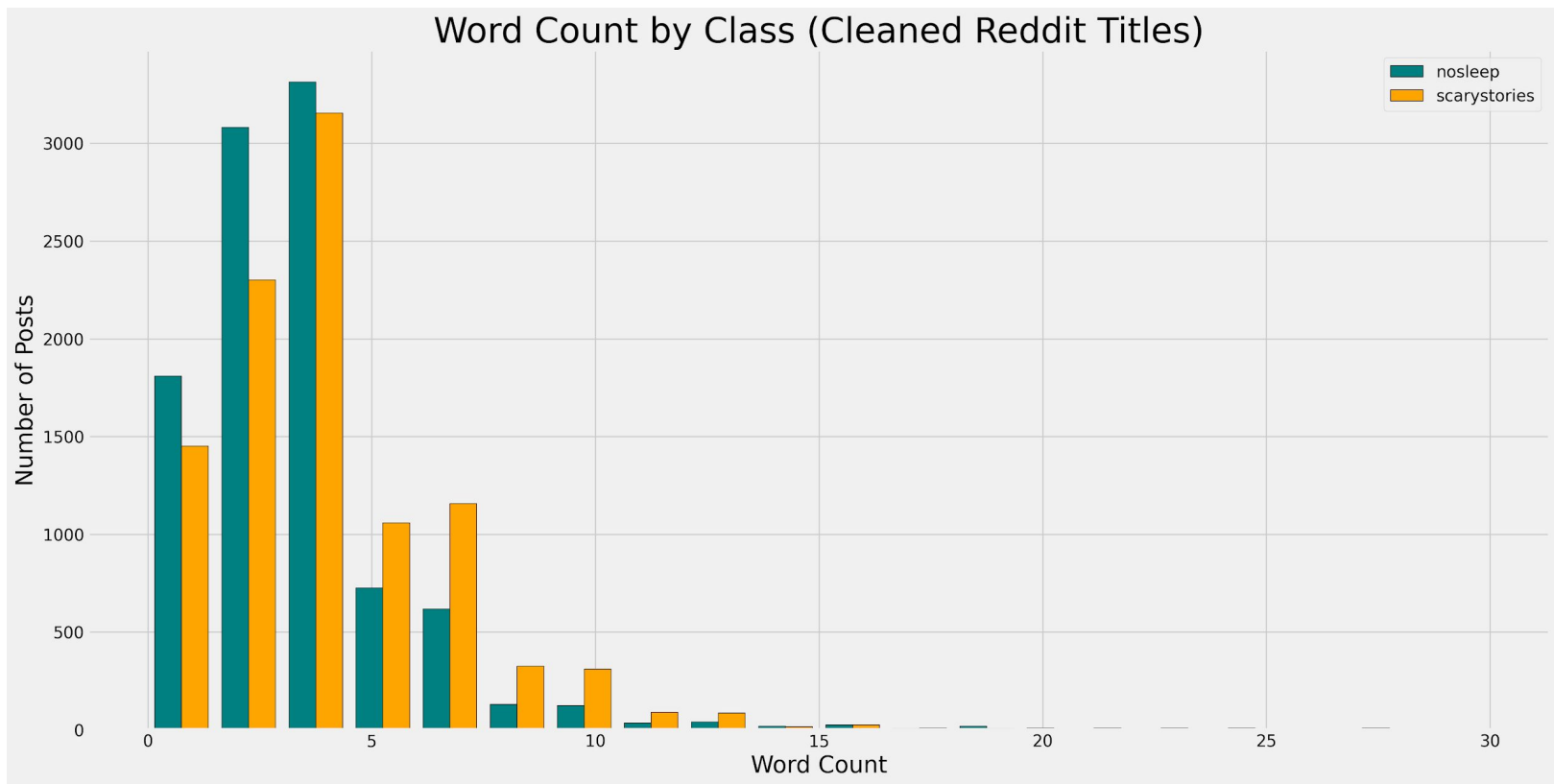- ## Natural Language Processing
  - Making our text data meaningful
- ## Classification Modeling
  - Multiclass Logistic Regression
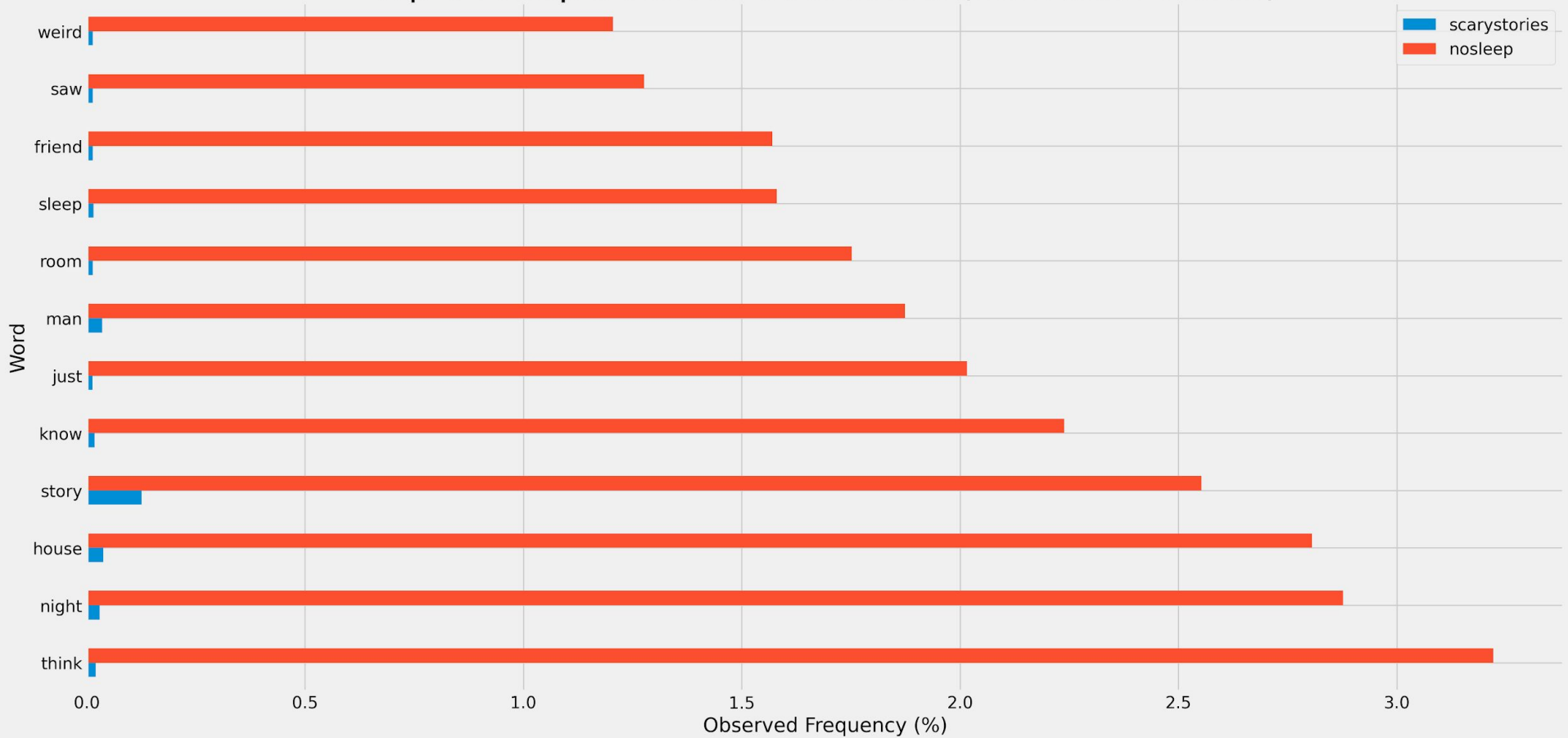
Word Count by Class (Subreddit Titles)

Raw title text: nosleep subreddit titles are longer by 2 words on average compared to scarystories subreddit titles.
(8 word titles versus 6 word titles)
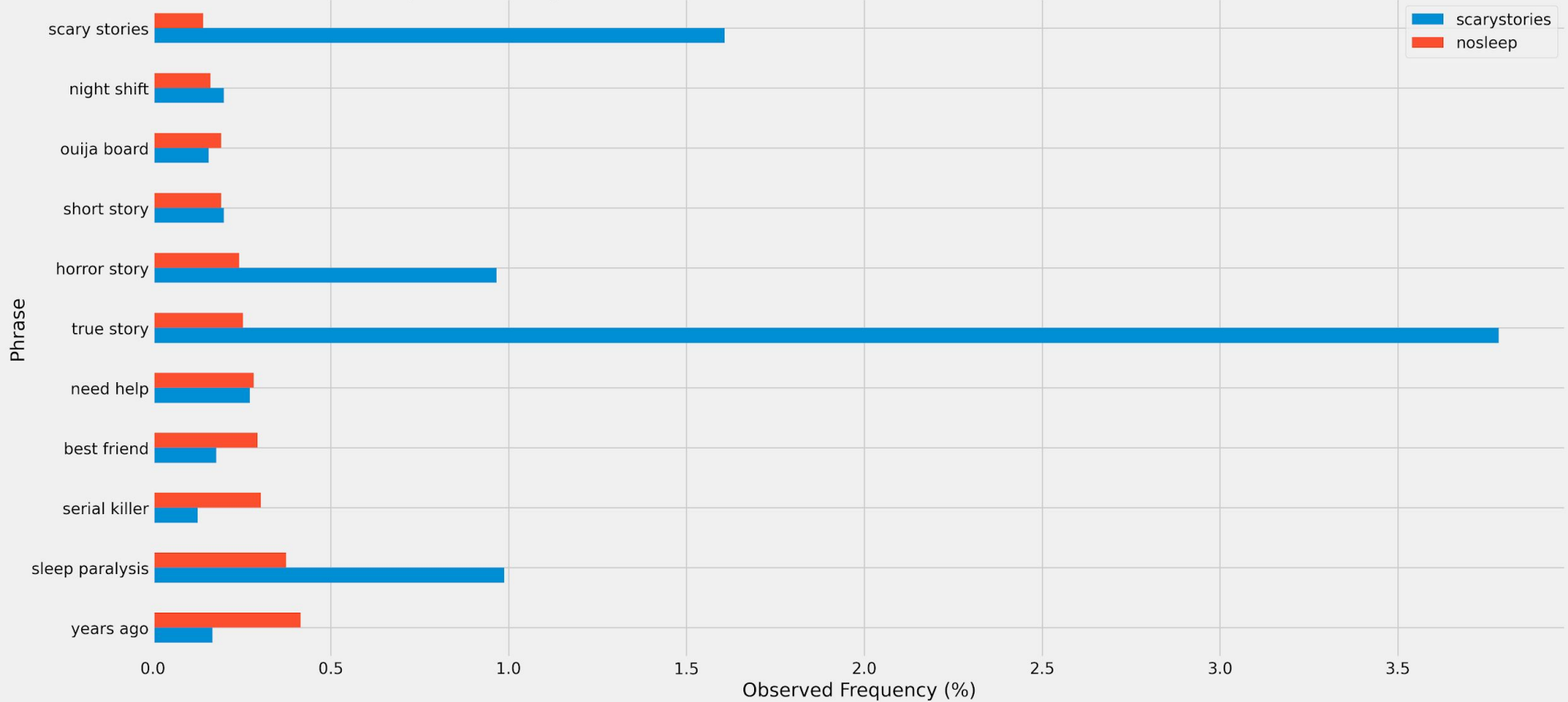
Word Count by Class (Cleaned Reddit Titles)

Clean meaningful title text: the average title word count dropped by 50% for both subreddits. How do we grab the attention of our audience in a more efficient way?
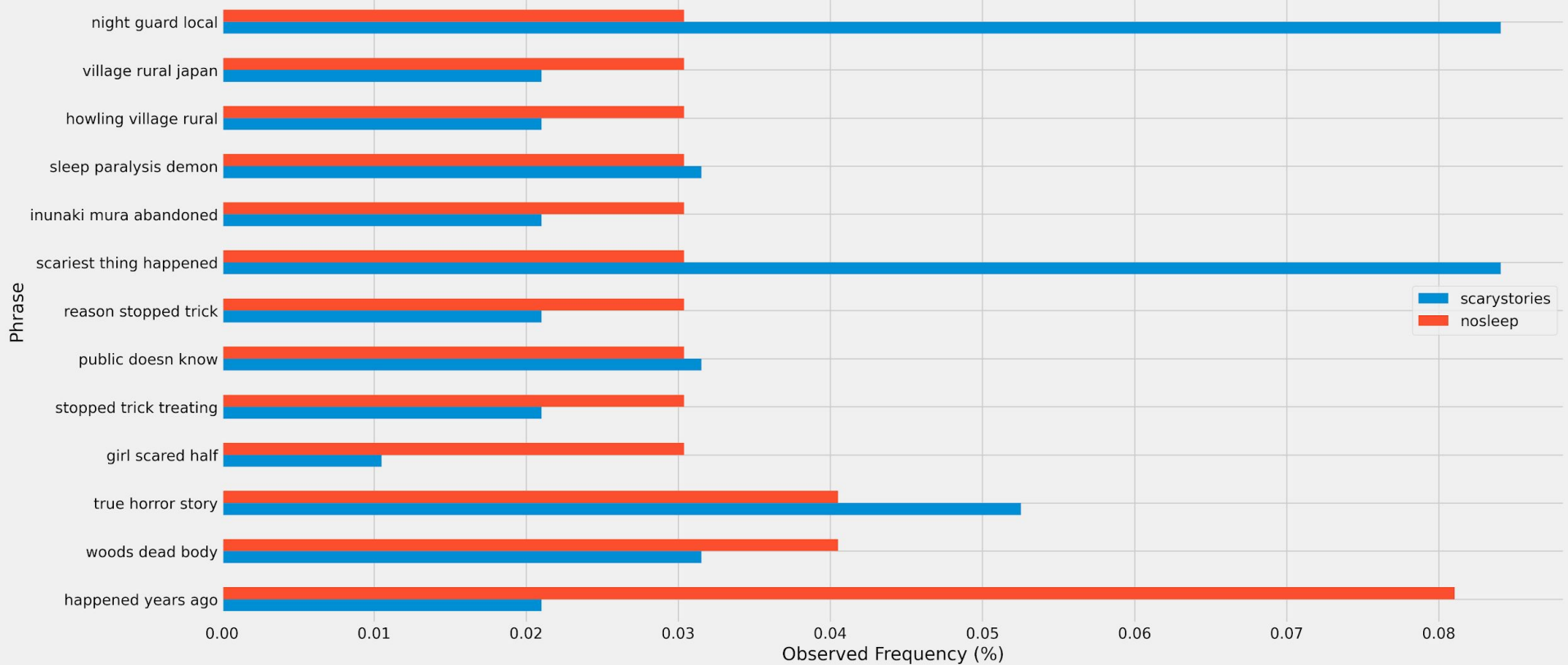
Top Overlap Between Subreddits (Individual Words)

Top Overlap Between Subreddit Titles (Two Word Phrases)

Top Overlap Between Subreddit Titles (Three Word Phrases)

Supporting Evidence

These high frequency words are routinely dark fiction themes

- Sleep Paralysis by Patrick Lacey
  - sleep paralysis demon
- Ring by Koji Suzuki
  - rural village japan
- Psycho by Robert Bloch
  - serial killer

Our Logistic Regression Model:
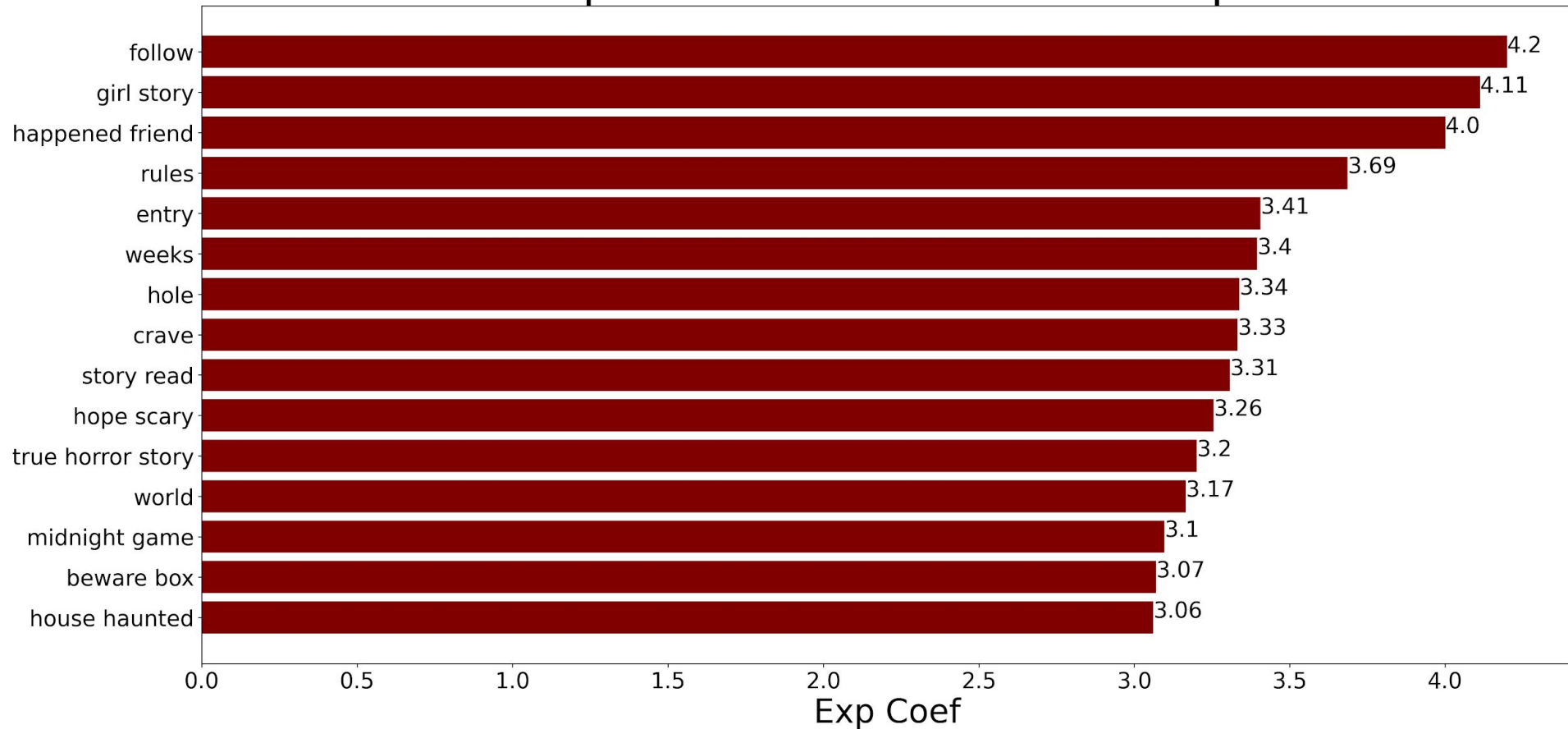
Predictive Target
- nosleep subreddit
- We can use the text that is predictive of nosleep to bring new ideas to creative writing

Statistics
- Accuracy: 72.6%
- Misclassification rate: 27.4%
- Overfit model
  - Training Accuracy: 93.63%.

Top 10 Title Features - nosleep

Feature	Exp Coef
follow	4.2
girl story	4.11
happened friend	4.0
rules	3.69
entry	3.41
weeks	3.4
hole	3.34
crave	3.33
story read	3.31
hope scary	3.26
true horror story	3.2
world	3.17
midnight game	3.1
beware box	3.07
house haunted	3.06

Coef. Interpretation: As there is a one word increase in our word 'follow', a reddit post is 4.2 times as likely to be from the nosleep subreddit.

Recommendations and Next Steps

Recommendations

Advertisement Shock Value:
- Decapitate text fluff and focus on buzz words for titles

Give the people what they want:
- ouija boards, serial killers, and sleep paralysis
- Kidnap this guide to build future narrative plots
- Abuse classification modeling to grab fresh ideas

Next Steps

Further exploration:
- Stalk these subreddits obsessively and branch out to other potential victims

In Conclusion...

Whatever you do, don't fall asleep.