HOW TO WIN AN OSCAR*

*FOR SCREENWRITING

Braden Baseley

BACKGROUND

- The Academy Awards, commonly known as The Oscars:
  Awards that honor the artistic and technical achievements in film each year.

- Given by the **Academy of Motion Picture Arts and Sciences**, an organization comprised of ~10,000 industry professionals spanning all major branches of filmmaking, such as actors, producers, writers, directors, musicians, makeup artists and hairstylists.

- *Are there significant differences between winning and losing films?*

DATA

- *Best Original Screenplay*: analyzing **46 screenplays** nominated over the past decade. There were 50 nominated films, but some of them (see: Quentin Tarantino) were unavailable. Data set comprised of **9 winning screenplays** and **37 losing screenplays**.

- *Best Original Screenplay* honors outstanding achievement in screenwriting for an original piece of work (as opposed to the *Best Adapted Screenplay* award, which is given to a screenplay that has been adapted from an existing body of work, such as a novel or a musical).

- Movie studios have "For Your Consideration" websites where they publish screenplays. Pulled scripts from those websites, otherwise used industry screenwriting sites (e.g. scriptslug.com)

METHODS

- *Are there significant differences between winning and losing films?*
  - First step: how to operationalize?

- Looking at differences in terms of:
  (i) Linguistic Complexity
  (ii) Sentiment

LINGUISTIC COMPLEXITY

- Many different measures, but ultimately settled on 3 different metrics:
  (i) Proportion of Unique Words (proxy for variation in word choice)
  (ii) Total Word Count (is the screenplay concise or detailed?)
  (iii) Average Syllables per Word (proxy for difficulty of words)

- Calculated these metrics using 3 different bodies of text:
  (i) Full
  (ii) Stemmed
  (iii) Lemmatized

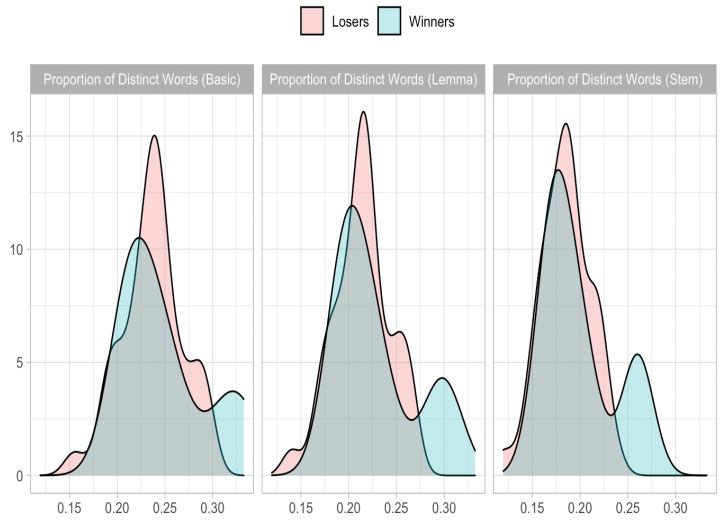
SENTIMENT

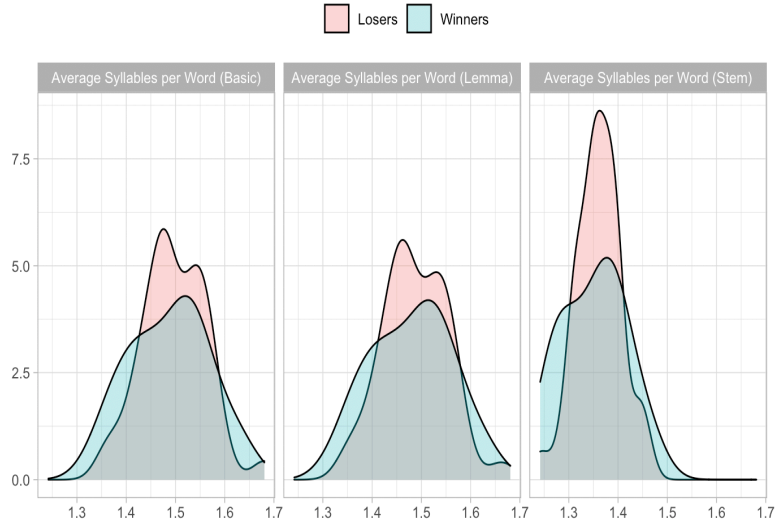
- Classified each word as either positive or negative using the list of positive and negative words from Hu and Bing (2004).
- Positive words assigned a value of +1, whereas negative words assigned a value of -1.
- Sentiment Score $= \frac{\text{Number of Pos. Words} - \text{Number of Neg.Words}}{\text{Number of Pos. Words} + \text{Number of Neg.Words}}$

RESULTS: LINGUISTICS PT. I

- Mean value of *proportion of unique words* higher for winners (true for full/stem/lemma).
- Mean value of *word count* higher for losers.
- Mean value of *average syllables per word higher* for losers (true for full/stem/lemma).
- *But are these observed differences in group means actually significant?*
  —**No**! After running a Welch Two Sample t-test for each variable, I failed to reject the null hypothesis that the true difference b/w group means is 0.
- Not an entirely surprising outcome…

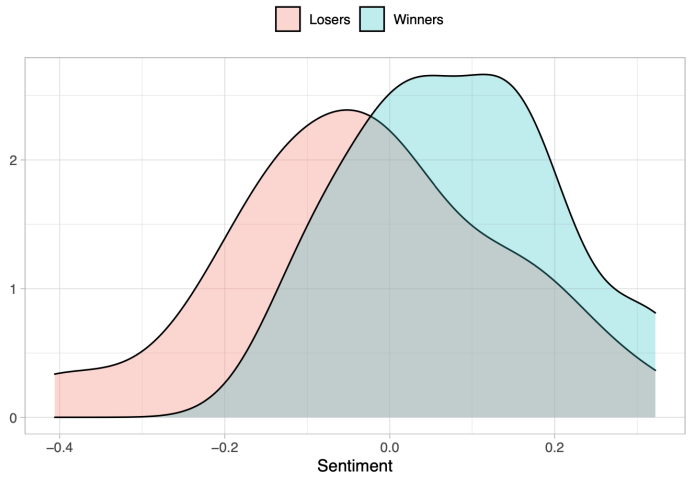
RESULTS: LINGUISTICS PT. II





RESULTS: SENTIMENT

- Distributions of sentiment for both winners/losers are roughly normal.
- Mean sentiment score is positive for winners, negative for losers.
- *But are these observed differences in group means actually significant?*
  --**Yes**! Welch Two Sample t-test had a p-value of 0.043, so there is a statistically significant difference between the two group means.
- Academy more partial to "positive" stories?



CONCLUSION

- In conclusion:
  - No statistically significant difference in linguistic complexity between winning/losing screenplays.
  - Statistically significant difference in sentiment between two groups—winners are more positive.
- Areas for improvement:
  - Measure of readability, perhaps Flesch-Kincaid score?
  - Use bigrams or other n-grams to get more sophisticated sentiment scores.
  - Increase sample size.