The Evolution of Language in Films

Jakob Svenningsson and Robert Injac

École Politechnique Fédérale de Lausanne

Introduction

Are films getting too simple; full of short sentences and predictable plots, catering to an attention-deficit audience? Is it true that contemporary films have a lot more profanity than the old ones? And finally, do people like happy or sad endings?

The Lowest Common Denominator

Many cinemaphiles claim that movies of today are "dumbed-down", and this phenomenon is generally known as the lowest common denominator. So are films getting less complex? Quantifying complexity of a film is a difficult task. We decided to use the Dale-Chall readability formula which measures how readable a text is. Larger score means that the text is less readable.

Mean readability score per decade

Mean readability score per genre
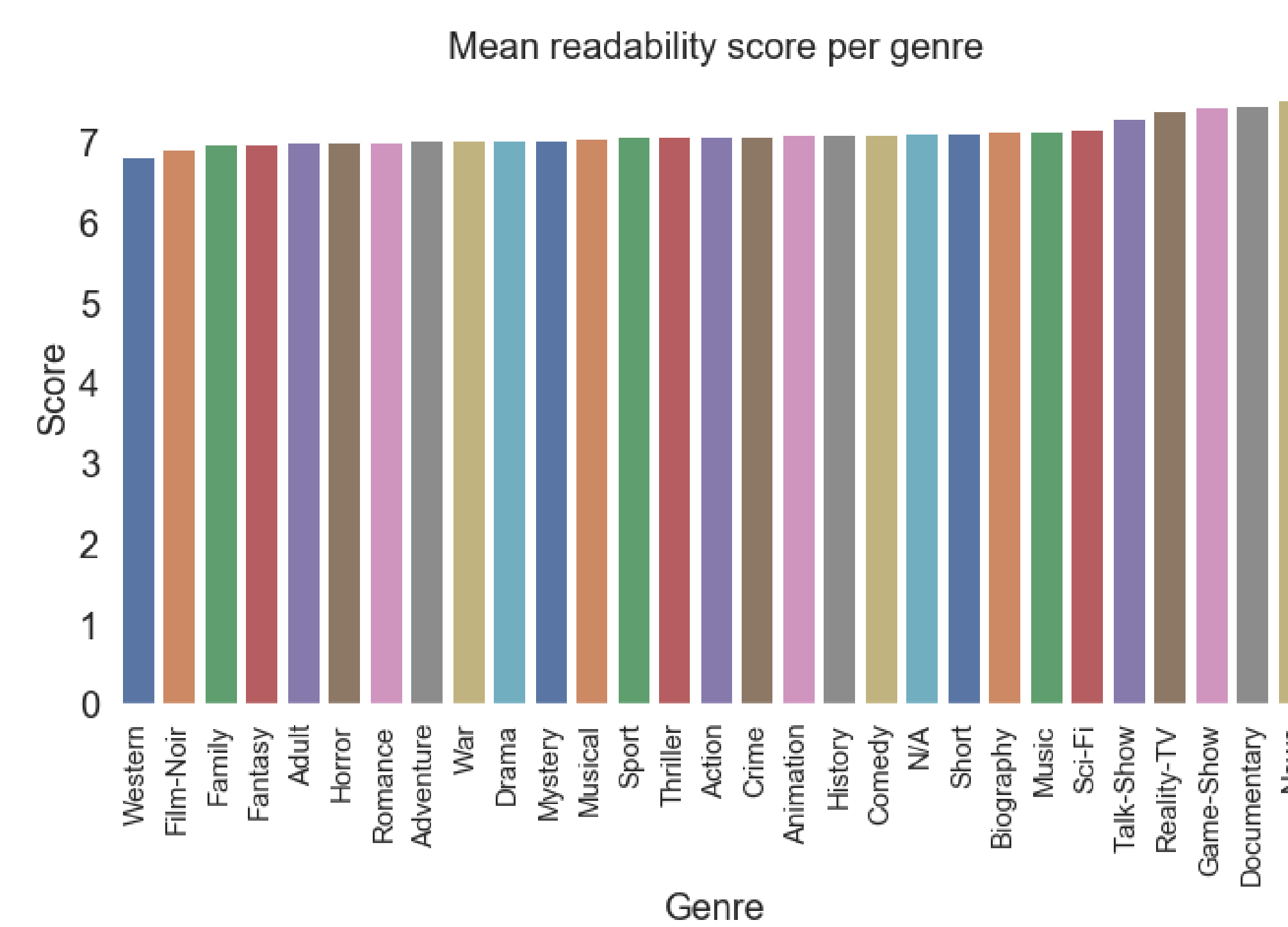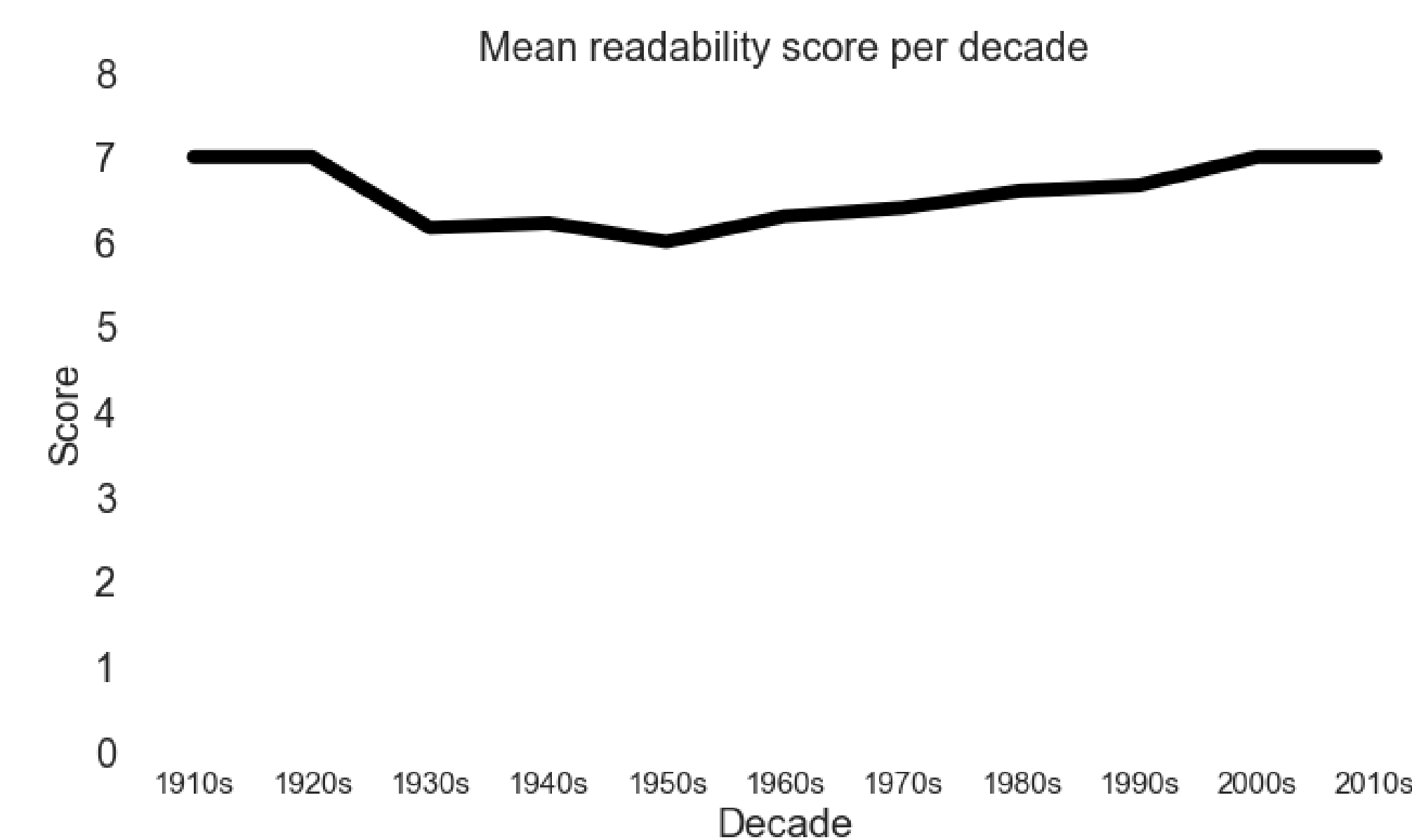
Looking at the plots above, it seems that **readability is pretty much constant in time and through genres**. However, we cannot make a definite conclusion: our chosen measure was designed for written text, and not for spoken language used in movies.

The F-bombs

Scorsese's *The Wolf of Wall Street* broke the all-time record for number of f-words in a movie: the word was used a whooping 506 times. That is more than 9 f-words used every minute! However, this is just one example. Are newer films in general containing more and more offensive language? Using our movie data and a dictionary of profanities, we will answer this question.

Mean proportion of profanities in movies per decade
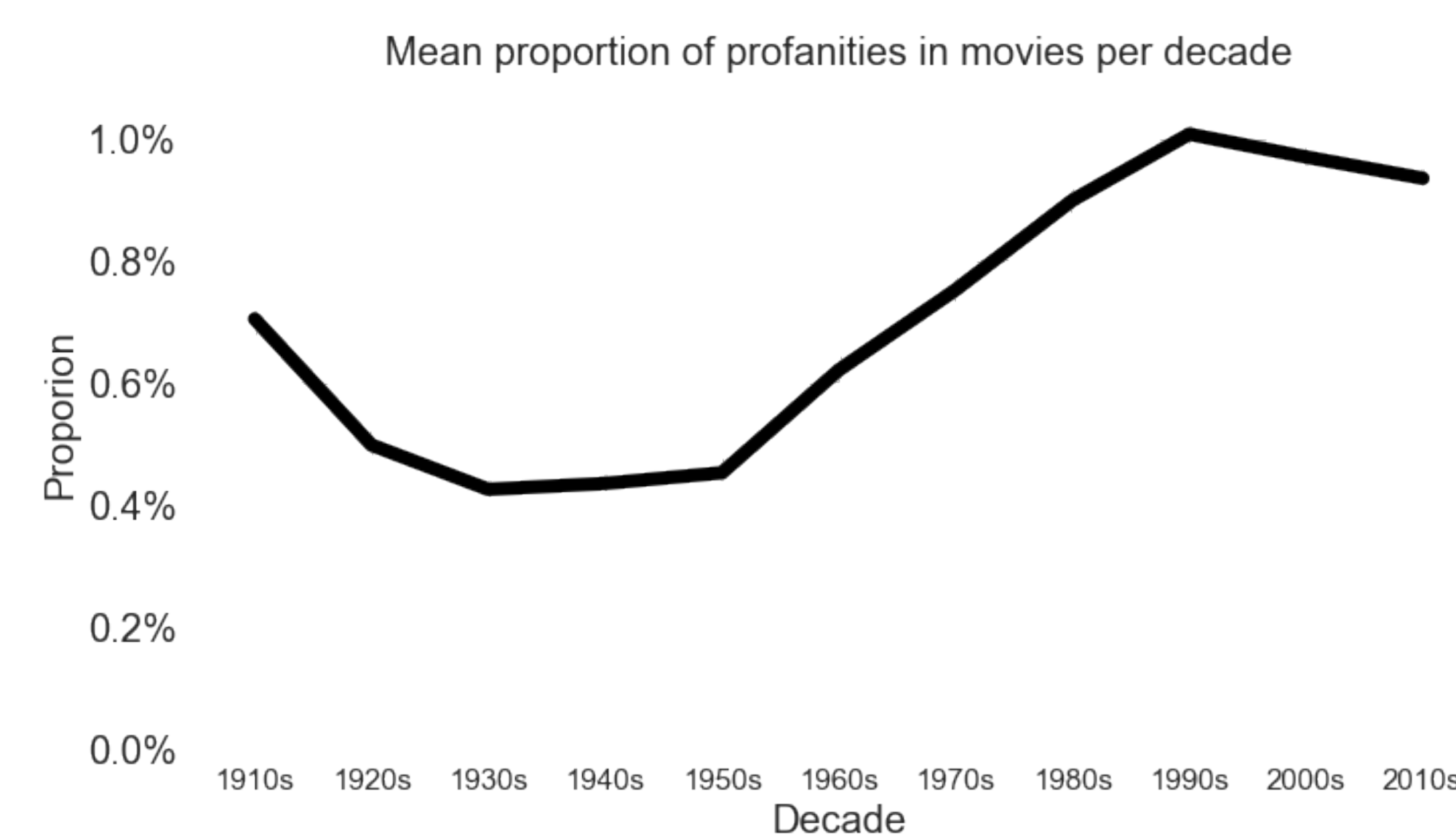
We can observe several interesting things in the above plot. From early 30s until the late 50s Hollywood movies were highly censored under the set of moral guidelines called the Hays code. The uncensored era before the code is known as pre-code Hollywood era.

Another thing we can see is that the proportion of profanities in an average new movie has more than doubled from the mid-20th century.

Mean proportion of profanities in movies per genre
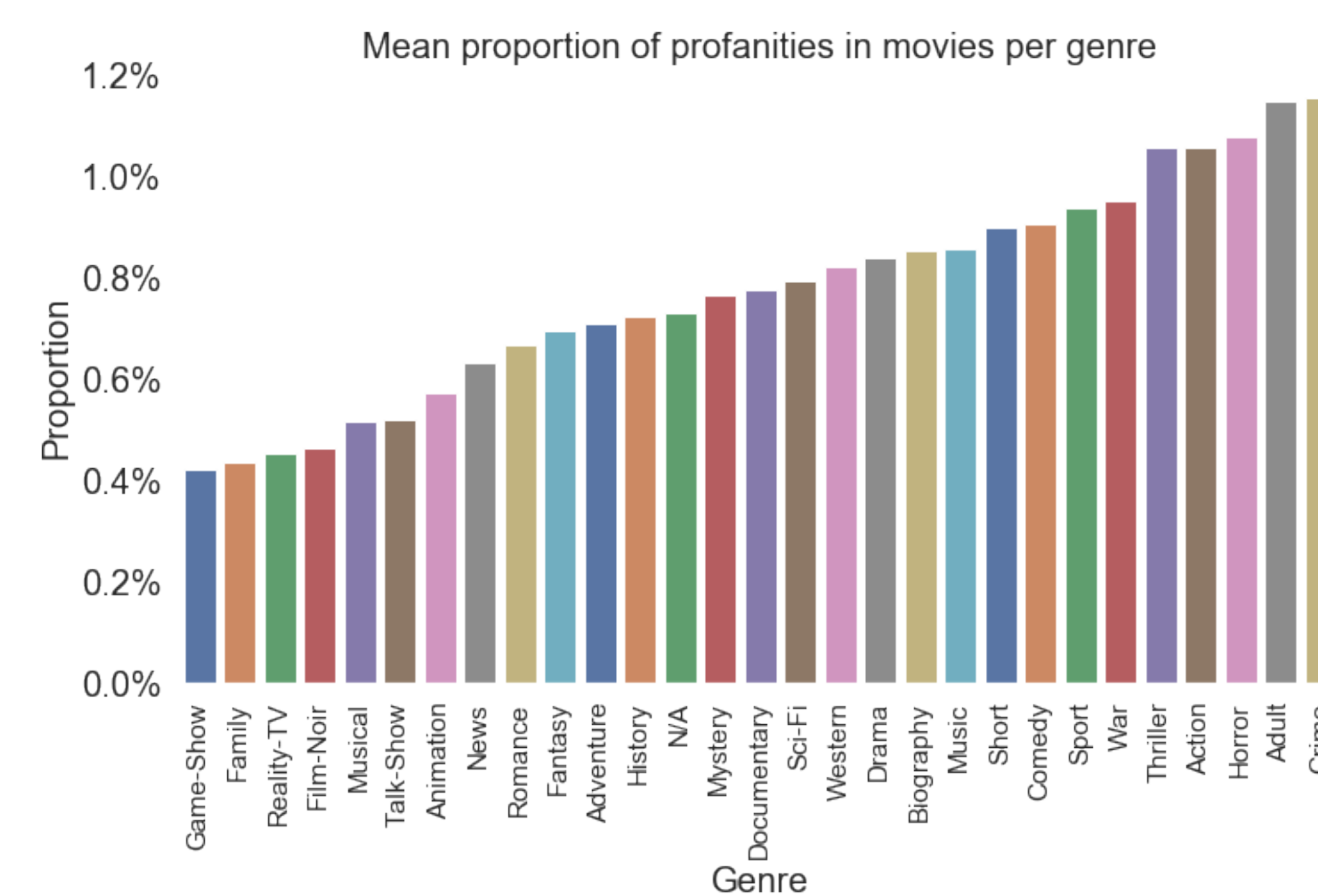
As expected, crime, adult, horror and action movies have the highest proportion of profanities. On the other hand, game-shows and family movies are more mild.

And They Lived Happily Ever After

Films in general often have a positive tone and a happy ending; after all, people have enough sadness in their real life. We will check if it is truly a fact that we have more happy movies today, and also find out which genres to avoid if you desire some realism.

Measuring the sentiment of a film is not an easy task. We used TextBlob and their implementation of sentiment analysis to calculate polarity scores for each of our movies. Positive score implies positive sentiment, while negative score implies negative.

Mean polarity score of films per decade
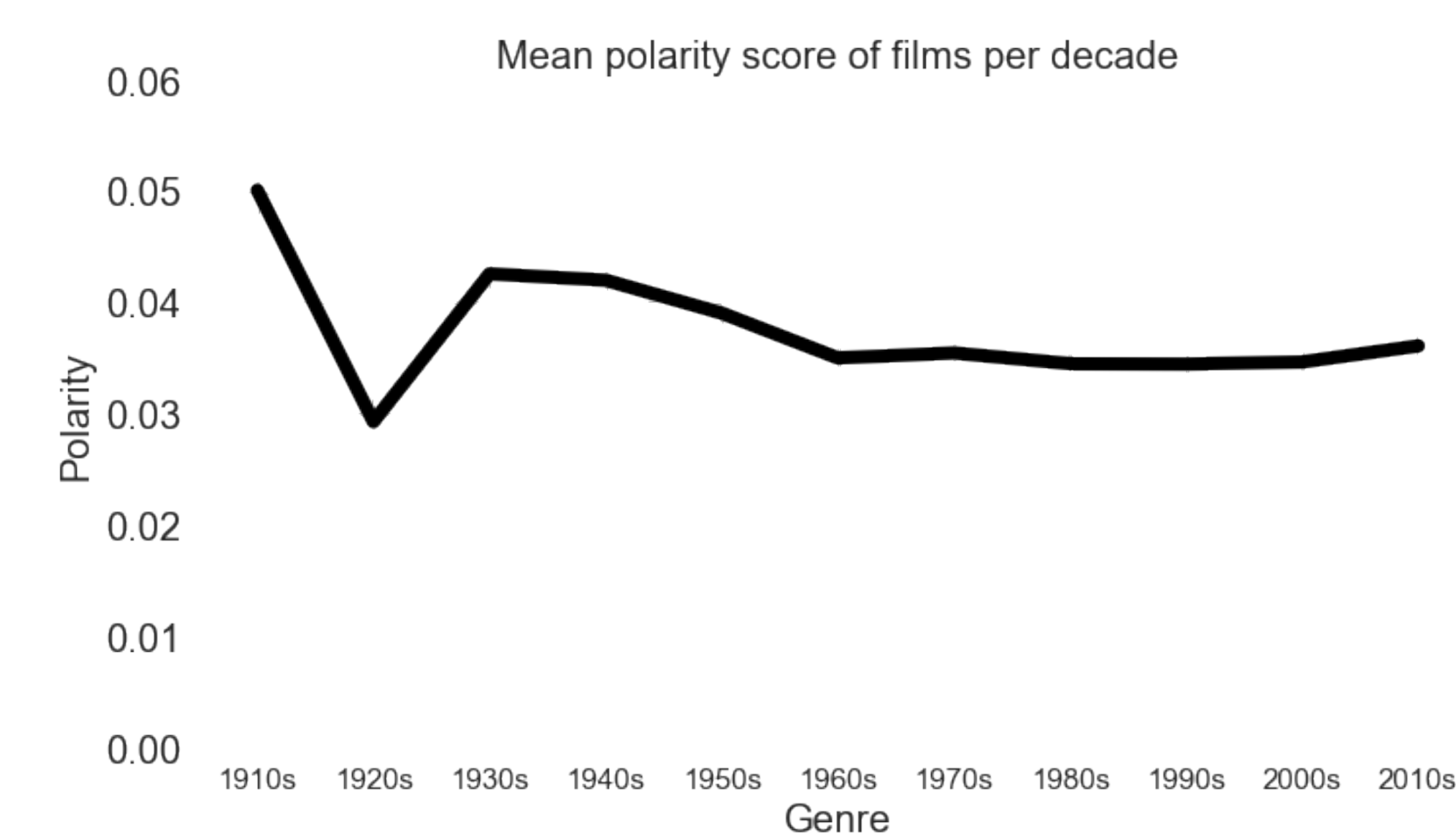
Except some slight variations in early 20th century (which can be explained by small sample sizes for those decades) the sentiment of films seems to be quite constant.
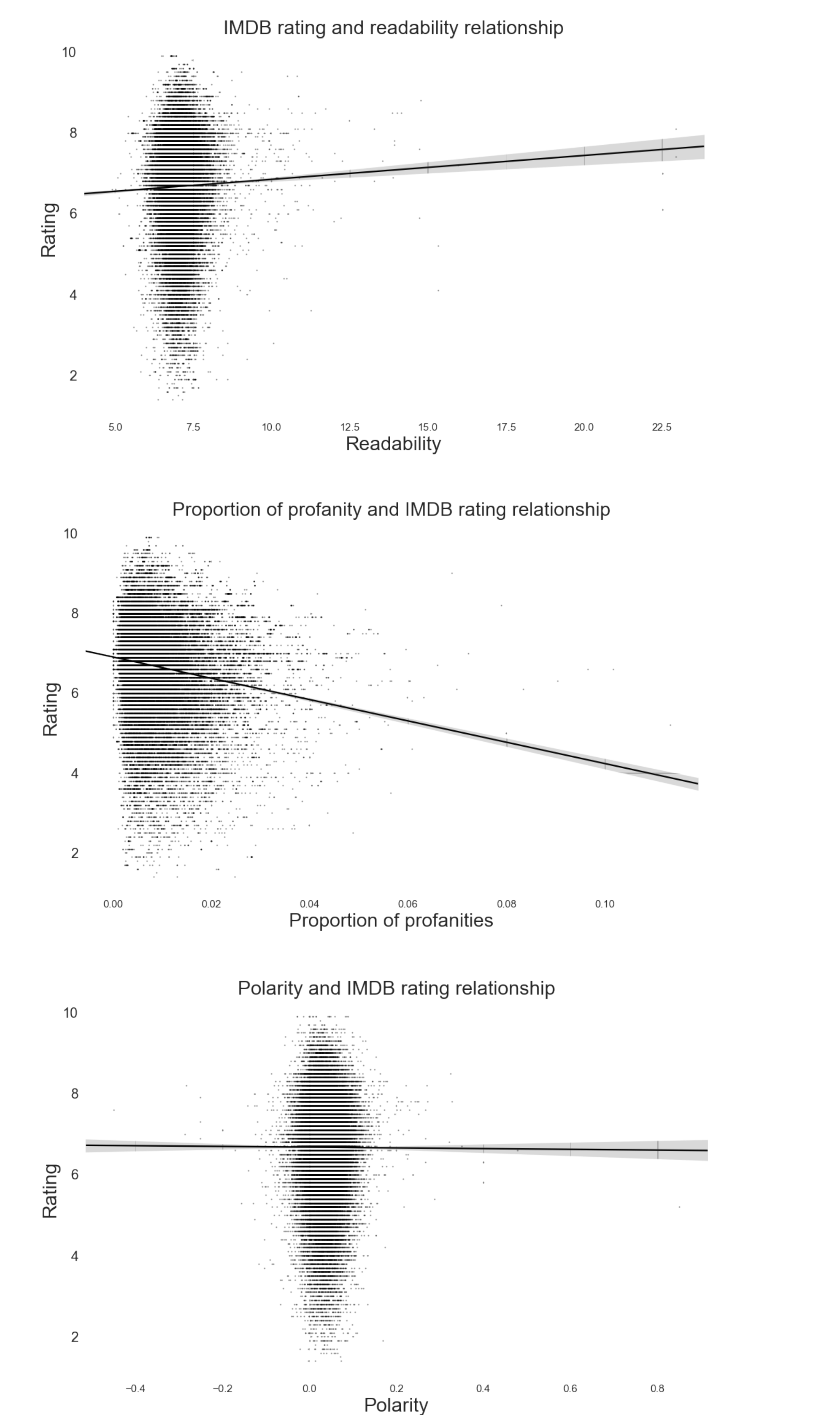

Mean polarity score per genre
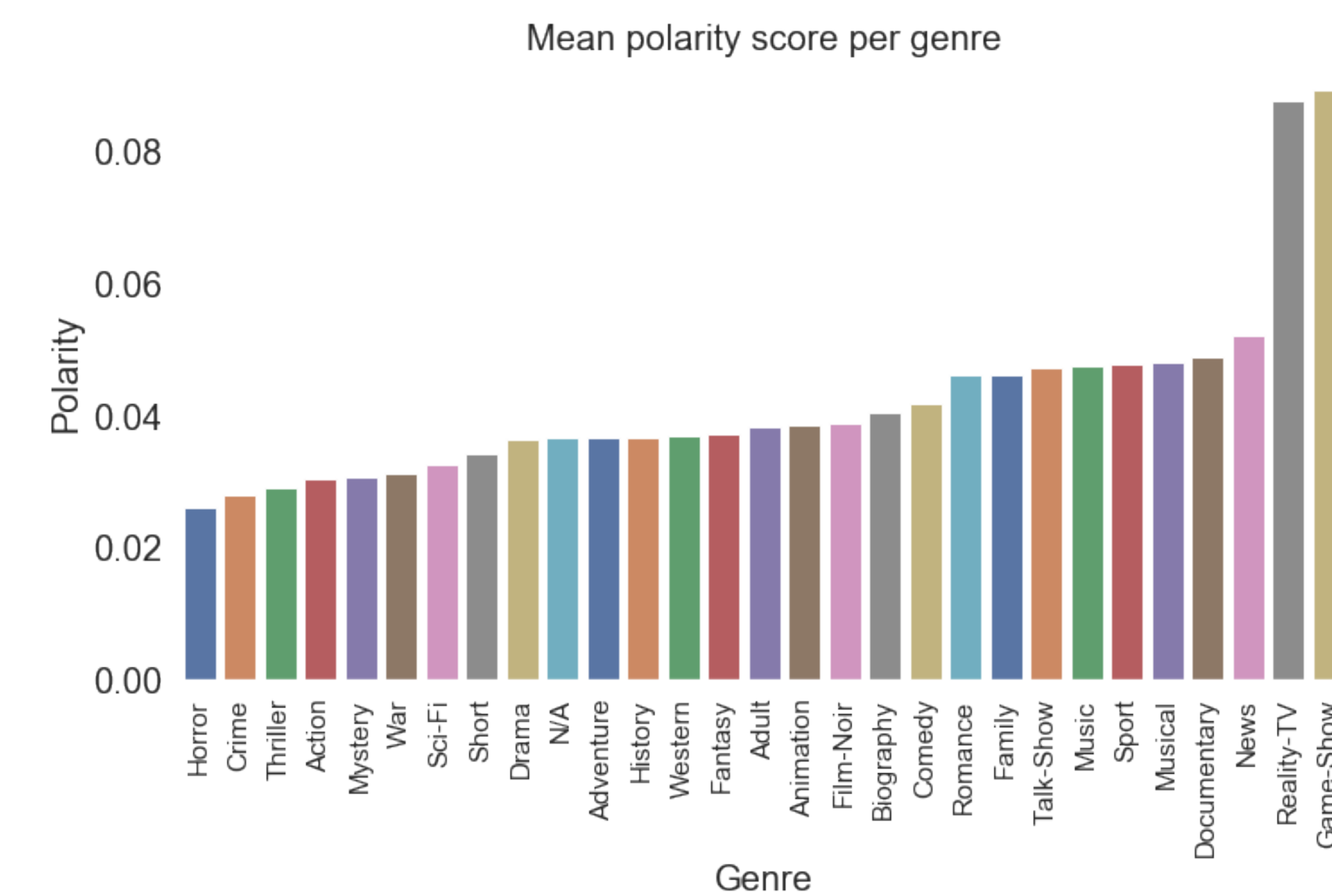
Unsurprisingly, game-shows and reality-TV are in general positive and horror, crime and thriller are negative. It is interesting to notice that documentaries and news-shows have one of the most positive sentiments - and they are mostly talking about the real world.

How to Make a Blockbuster

You now know that films are pretty readable, that swear words will soon become more common than non-swear words, and that horrors are not so happy.

But what is the worth of knowledge if you can't use it to make some money? We will look at the three film properties that we have already analyzed (complexity, proportion of offensive words, sentiment) and try to find correlation between them and movie ratings.

IMDB rating and readability relationship

Proportion of profanity and IMDB rating relationship

Polarity and IMDB rating relationship

There seems to be **no correlation** between IMDB ratings and any of the properties. Our data tells us that the quality of a movie is not simply explained by measuring complexity, sentiment or the number of profanities - it is something much more subtle, and at the end - subjective.