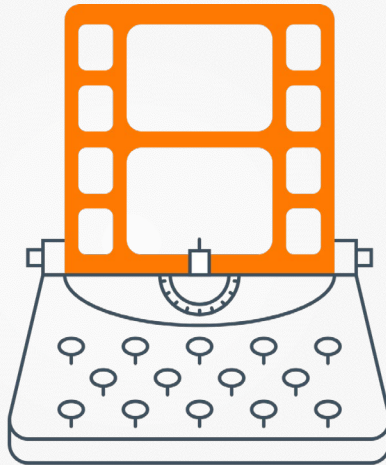


# JUDGING SCREENPLAYS BY THEIR COVERAGE

*An analysis of 12,000+ unproduced feature film screenplays and the scores they received,  
revealing what professional script readers think makes a good screenplay.*



BY  
**STEPHEN FOLLOWS AND JOSH COCKCROFT**  
WITH LIORA MICHLIN

IN PARTNERSHIP WITH  
**SCREENCRAFT**

# CONTENTS

<b>Foreword .....</b>	<b>3</b>
<b>TL;DR .....</b>	<b>5</b>
<b>Introduction .....</b>	<b>6</b>
<b>Section A: How To Impress Script Readers.....</b>	<b>8</b>
What Matters Most to Script Readers .....	9
Genre-specific Advice.....	10
Sentiment.....	17
Plot Arcs.....	19
Swearing.....	28
Number of Pages.....	29
Speech Descriptors.....	30
Use of Voiceover.....	31
<b>Section B: The Average Screenplay .....</b>	<b>32</b>
Pages.....	33
Scenes.....	35
Speaking Characters .....	36
Dialogue .....	37
Locations.....	38
Punctuation.....	40
Swearing.....	41
Age of Characters.....	44
Genre-skewed Words.....	48
<b>Section C: Screenwriters and The Act of Screenwriting ...</b>	<b>60</b>
Gender.....	61
The Working Habits of Screenwriters .....	63
Screenwriting Software .....	65

# FOREWORD

## A Note from the Authors

---

Screenwriting is a lot like ice skating. It's harder than it first appears, it takes a lot of practice to stop constantly getting hurt and success means spending your life on thin ice. Oh, and if you want to make a career of it, you need to impress a small number of judges.

In the case of screenwriting, the first line of judges are script readers. They read and rate scripts on behalf of producers, studios and competitions. Scoring well with readers will mean that your screenplay reaches the desks of the great and the good (who are hopefully also the rich and the powerful). Score poorly and all the countless hours you put into your screenplay will just have been "character building".

Given how important script readers are, it's vital that any aspiring screenwriter spends time understanding what readers think a good script looks like. This has been our mission in this project.

Using data on over 12,000 unproduced feature film screenplays, along with the scores they received from professional script readers, we can lift the lid on how to impress these vital conduits.

This project is not about measuring art or rating how good a story is; it's about decoding the industry's gatekeepers. Rather than suggesting "this is what a good script contains," instead we are saying "this is what readers think a good script contains".

In the real world, this distinction may not matter as readers are an integral part of the industry's vetting process. But it is important to remember that all the advice to screenwriters contained herein is in relation to the data and through the lens of what script readers have revealed in their scores.

The most talented writers can overcome most, if not all, of these correlations. They can make the impossible possible, spin an old tale a new way, induce real tears over imagined events and lead us to root for characters we know to be doomed.

We have thoroughly enjoyed conducting the research and we hope you find the final report useful.

One last thing – once you've finished reading the report, stop procrastinating and get back to that screenplay you're supposed to be writing...



*Stephen Follows*



*Josh Cockcroft*



*Liora Michlin*

## Acknowledgments

---

The authors would like to say a special thank you to John at ScreenCraft and Scot at Coverfly for making this project possible and for being there to support us at every stage. Their unwavering focus on helping screenwriters is palpable and helped us ensure this report was practical as well as data-driven.

We would also like to thank the wonderful, smart people who helped along the way, including Paul Thompson, Ben Aston, Edward Dark, Chris Jones, Lucy V Hay, Jasmin Peppiatt, Monique Charlesworth and Sophie Lifschutz.

We are grateful to Laurie Clouston and Greg Dawson at Celtx and Guy Goldstein at WriterDuet for helping us identify scripts created with their programs.

This report was funded by proceeds from our last publication, The Horror Report<sup>1</sup>. The Horror Report was published via a 'Pay What You Can' model, with all the income going to support future film data research. We are very grateful to everyone who purchased a copy and especially to the generous people who chose to give us way more than the minimum. The report you are reading now simply would not have been possible without such contributions.

## A Note from ScreenCraft

---

I simply want to say thank you to Stephen Follows and his talented team (Josh and Liora). ScreenCraft was so fortunate to find such a talented team to help us analyze this insurmountable amount of data from thousands of screenwriters and professional script readers.

They've helped us discover some fascinating correlations between elements of screenplays and how readers rate them.

And to all our screenwriters, script readers and event attendees out there – thank you all for entrusting us with your projects.

We hope this report is both entertaining and helpful for screenwriters and filmmakers everywhere.



*John Rhodes, Co-Founder, ScreenCraft*



---

<sup>1</sup> <https://stephenfollows.com/horrorreport/>

# TL;DR

Look, we get it – you’re busy. You’ve got irons in fires, buns in ovens, plates spinning and other kitchen-based metaphors which illustrate the underlying panic we all feel about underachieving in an ever more frenetic world.

But... breathe. Be calm. We’ve got you, fam. Here is a one-page summary of the project.

## What We Did

We analysed 12,309 feature film screenplays which were submitted to screenplay competitions and for script reports. All of the scripts were read by professional script readers, who gave the scripts an overall score as well as scores for other factors, including plot, characterization and voice. We looked for connections and correlations to discover what professional script readers think a good screenplay looks like.

## What We Found

Later sections go into more detail and more topics, but below are nine tips screenwriters should take on board to help improve their chances of impressing script readers.

1. **Know thy genre.** Your priorities should rest on the particular nature of your chosen genre. For example, Family films place the highest premium on catharsis, while for Action films it’s plot.
2. **Some stories work better than others.** The vast majority of scripts can be summarized using just six basic emotional plot arcs – and some perform better than others.
3. **If you’re happy and you know it, redraft your script.** Film is about conflict and drama and for almost all genres, the happier the scripts were, the worse they performed. The one notable exception was comedy, where the reverse is true.
4. **Swearing is big and it is clever.** There is a positive correlation between the level of swearing in a script and how well it scored, for all but the sweariest screenplays.
5. **It’s not about length, it’s what you do with it.** The exact length doesn’t matter too much, so long as your script is between 90 and 130 pages. Outside of those approximate boundaries scores drop precipitously.
6. **Don’t rush your script for a competition.** The closer to the deadline a script was finished, the worse it performed.
7. **Use flashbacks responsibly.** Scripts with more than fifteen flashbacks perform worse than those with few to no flashbacks.
8. **VO is A-OK.** Some in the industry believe that frequent use of voiceover is an indicator of a bad movie, however we found no such correlation. We suggest that any complaints on the topic should be sent to editors, rather than writers.
9. **Don’t worry if you’re underrepresented within your genre – it’s your superpower.** Female writers outperform male writers in male-dominated genres (such as Action) and the reverse is true in female-dominated genres (such as Family).

# INTRODUCTION

## Our Dataset and Research

---

This project is a collaboration with ScreenCraft, a screenwriting consultancy and competition platform. They provided anonymised scoring data on 12,309 feature film screenplays which had been submitted either to one of ScreenCraft's own script competitions or to their script review service.

The vast majority of these scripts will not have been produced into movies yet and a large number of the screenwriters will still be at entry level, rather than professional writers. That said, within the dataset are scripts which have won awards, been optioned by established producers and been written by writing professionals and Hollywood stars.

No matter their source or the background of the writer, all scripts were independently assessed and rated by professional script readers. This produces a collection of scores for each script, ranging from the all-important Review Score down to specific aspects of each script, such as plot, characterization, concept and voice.

We ran the screenplays through a series of bespoke algorithms, pulling out all manner of data from each script. This gave us a vast array of data to add to the score given by the script readers, ranging from basic metadata (such as length, number of scenes, and the software the script was written with) through to more complex analysis (such as the sentiment of each line of dialogue, genre-skewed language, and how characters interact with each other).

Using this dataset, we sought out connections and correlations to discover what professional script readers think a good screenplay looks like and how screenwriters can improve their chances of receiving a high Review Score.

## An Introduction to Script Coverage

---

Many producers and production companies are not able to personally read each and every screenplay which is submitted to them. They rely on script readers to filter their script mountain down to a select few projects which deserve additional attention.

A typical script coverage will cost the producer \$80 to \$150 and, in return, they will receive a short report containing a precis of the script and an opinion as to its quality.

The script reader's verdict is typically summarised by awarding the script one of three possible verdicts: Pass, Consider or Recommend. In addition, the report will contain scores for various aspects of the script and a breakdown of the script's strengths and shortcomings.

## An Introduction to Script Competitions

---

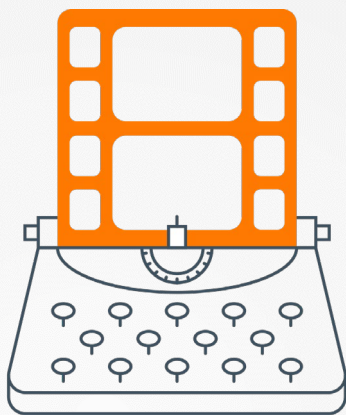
The vast majority of screenplay competitions are open to anyone who has written a qualifying screenplay and who pays the entry fee, typically around \$30 to \$80. Submitted scripts will be evaluated by script readers and in some cases a script report is generated and sent to the writer.

The highest scoring scripts progress to later rounds, between which screenwriters may be permitted to submit updated drafts of their shortlisted project. Eventually, a winner is announced and prizes distributed.

Given the vast number of aspiring screenwriters and the industry's reliance on 'who you know', script competitions offer industry outsiders a chance to have their work recognised, celebrated and perhaps produced.

---

# SECTION A: HOW TO IMPRESS SCRIPT READERS



*It's not possible to give you a secret formula which can guarantee success in screenwriting.  
However, it is possible to note significant correlations between the scores readers  
give screenplays and certain aspects of those scripts.  
This section contains a series of recommendations for screenwriters, based on the data,  
assuming one's goal is to increase the score awarded by script readers.*

What Matters Most to Script Readers .....	9
Genre-specific Advice.....	10
Sentiment.....	17
Plot Arcs.....	19
Swearing.....	28
Number of Pages.....	29
Speech Descriptors.....	30
Use of Voiceover.....	31

# What Matters Most to Script Readers

As well as the overall 'Review Score', ScreenCraft's script readers are asked to provide scores for a variety of specific factors such as plot, tone and concept.

We tracked how important each of these factors are in the success of scripts through this process.

There is a longer explanation in the footnotes<sup>2</sup> but suffice it to say that the higher the number, the greater the level of correlation between that factor and the script's overall Review Score.

The greatest correlations are within the subcategories of characterization, plot and style. Among the least important factors are formatting, originality and the script's hook.



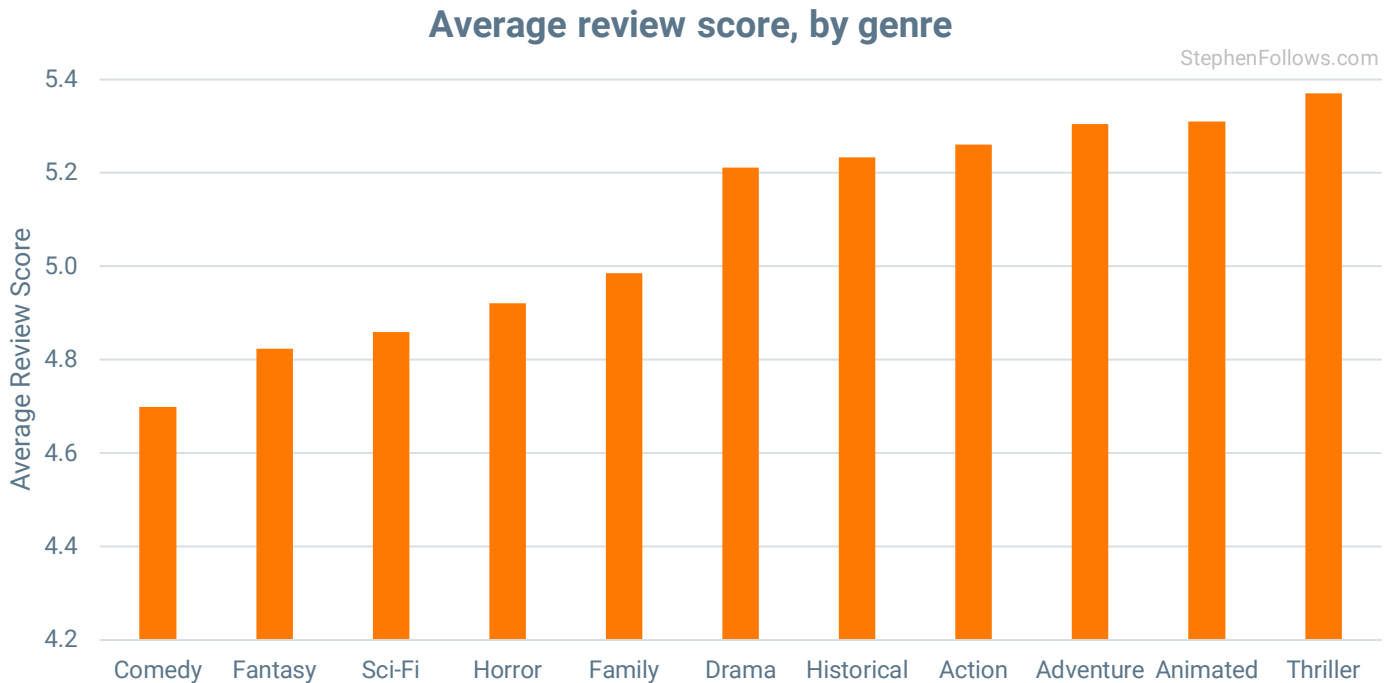
It should be noted that although some factors are less important than others, all are positively correlated within the overall Review Score and therefore no factor should be neglected.

<sup>2</sup> We used the Pearson correlation coefficient, which provides a value between 1 and -1 to indicate how linked two sets of numbers are. A value of 1 would reveal that the two factors are perfectly positively correlated (i.e. when one of the figures rises, so too does the other). A value of -1 would show perfect negative correlation (i.e. when one rises the other falls). Values of between 0.2 and -0.2 are not regarded as statistically significant.

# Genre-specific Advice

When submitting their screenplays, writers were asked to select a principal genre for their script<sup>3</sup>. This allows us to split the results by these self-reported genres.

Across our dataset, Thrillers receive the highest average scores from readers, with an average of 5.4 out of 10 across our dataset<sup>4</sup>. Comedies score the lowest, with an average of 4.7.



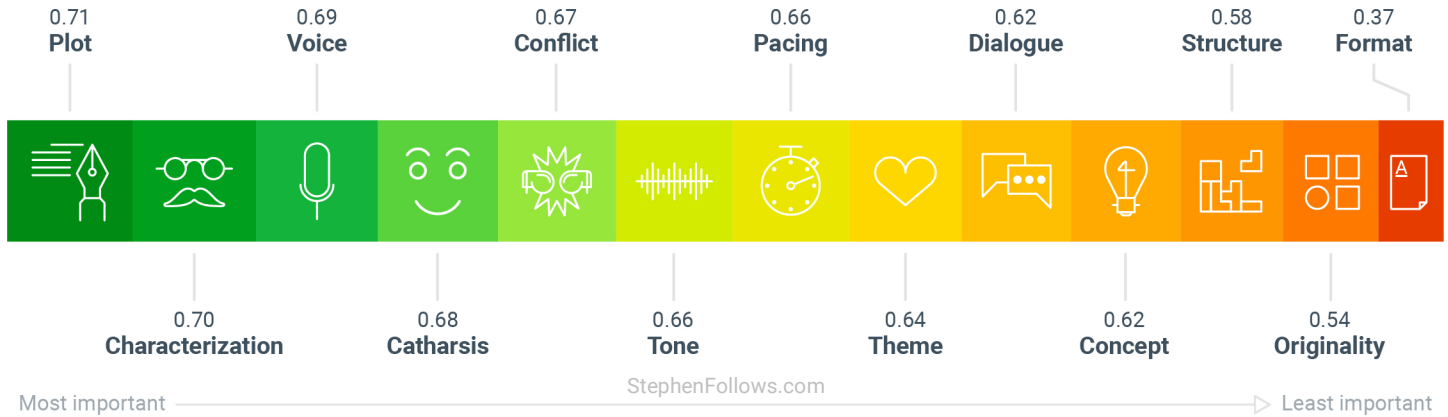
Most of the elements which are present in high-quality screenplays are universal no matter which type of film is being written. However, the priority order of these elements can differ between genres.

<sup>3</sup> Scripts with the genre classifications of "Other" and "Unknown" have been excluded from genre-breakdowns, as are the genres whereby there were too few scripts to make the results meaningful, such as "Faith", "Musical" and "Western". These scripts are, however, included in any analysis not specifically relating to genre.

<sup>4</sup> As we have no independent yardstick to measure of the quality separately from the script readers' scores, we cannot say if this is due to Thriller being an objectively better-written genre or if this is due to a prejudice on the behalf of script readers. As discussed previously, in the real world this distinction is irrelevant as the views of script readers are taken as fact by those commissioning reports and running competitions.

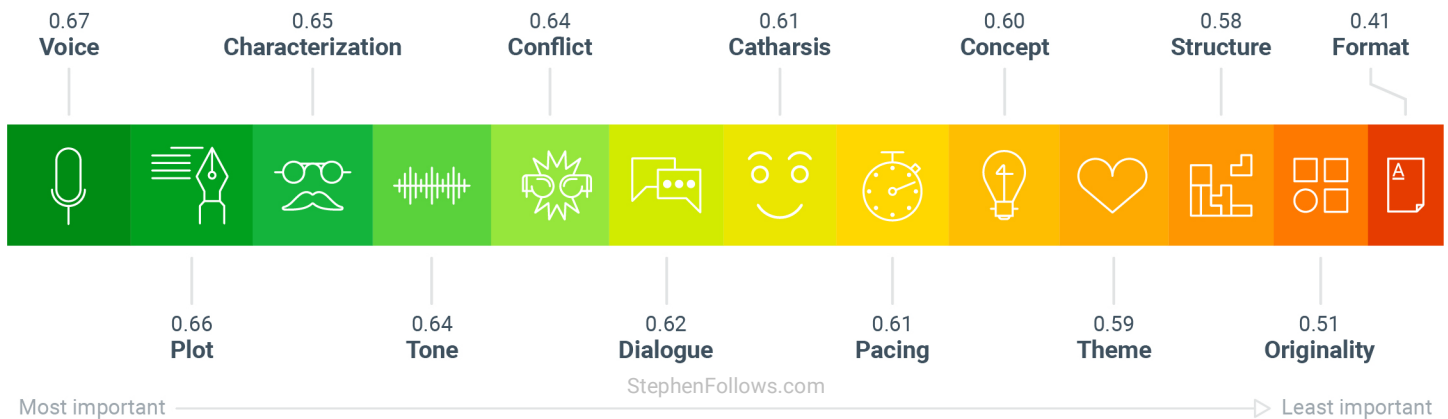
## Action

The most important indicator of the final score of an Action script is what the Reader thinks of the plot. The strength of the dialogue is not a key factor, and neither is originality or the script's concept.



## Adventure

Voice and tone have a much bigger role to play in Adventure scripts, when compared to Action scripts. Conversely, catharsis is less important than in almost all other genres.



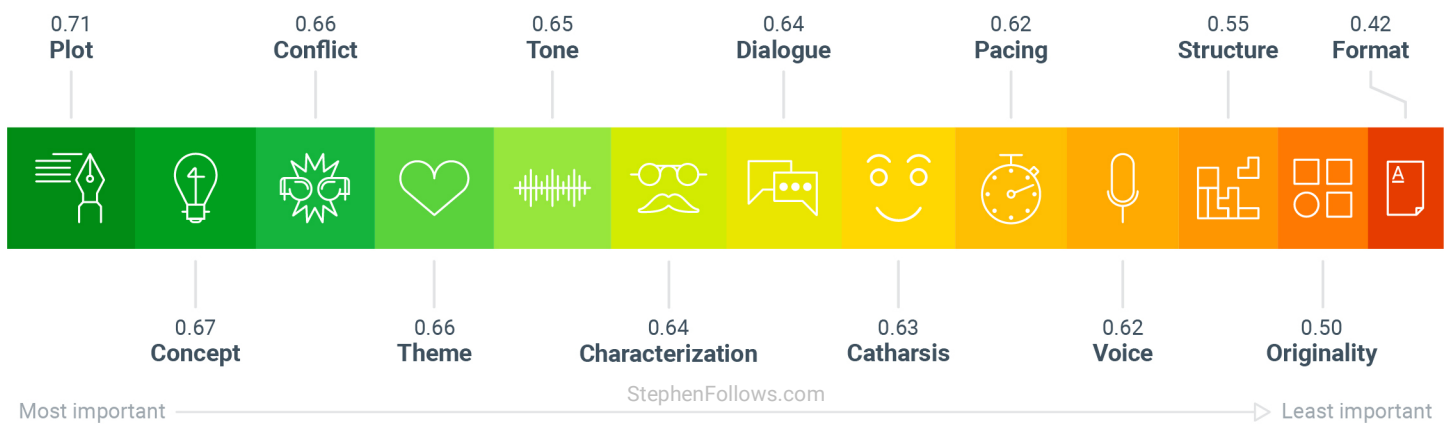
## Animated

The strength of the plot and concept are key to the success of Animated scripts.

When compared with scripts of other genres, Animated scripts appear to rely more heavily on broad ideas (such as concept and theme) rather than specific features of the writing (such as voice and structure).

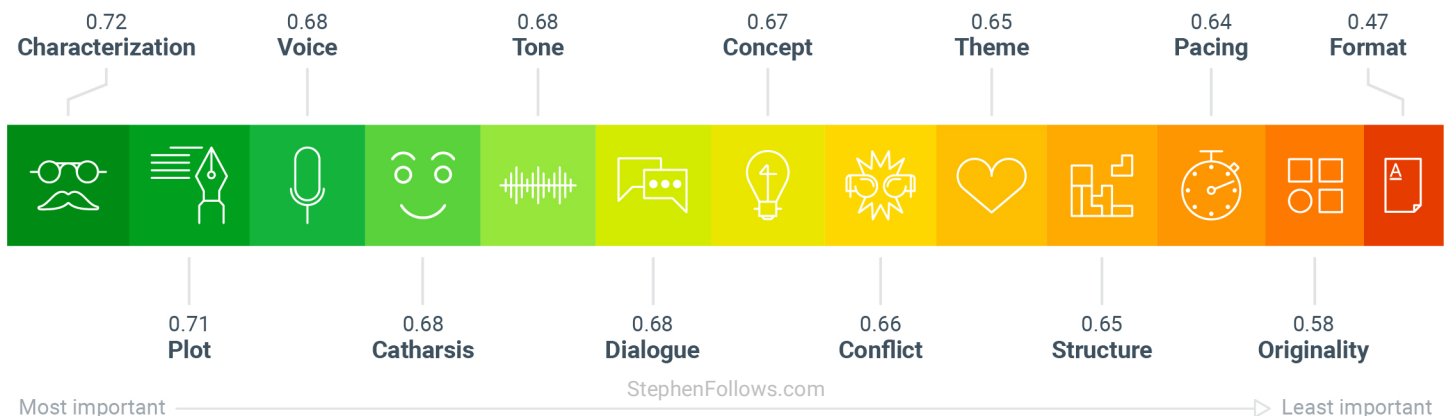
The “voice” of the screenwriter is far less important in an Animated screenplay, compared to other genres. Two-thirds of genres have voice as one of their top three factors, whereas, for Animated scripts, it’s towards the bottom of the list. A similar observation can be made in relation to catharsis, with Animated film the least dependent on it for overall success.

### Animated



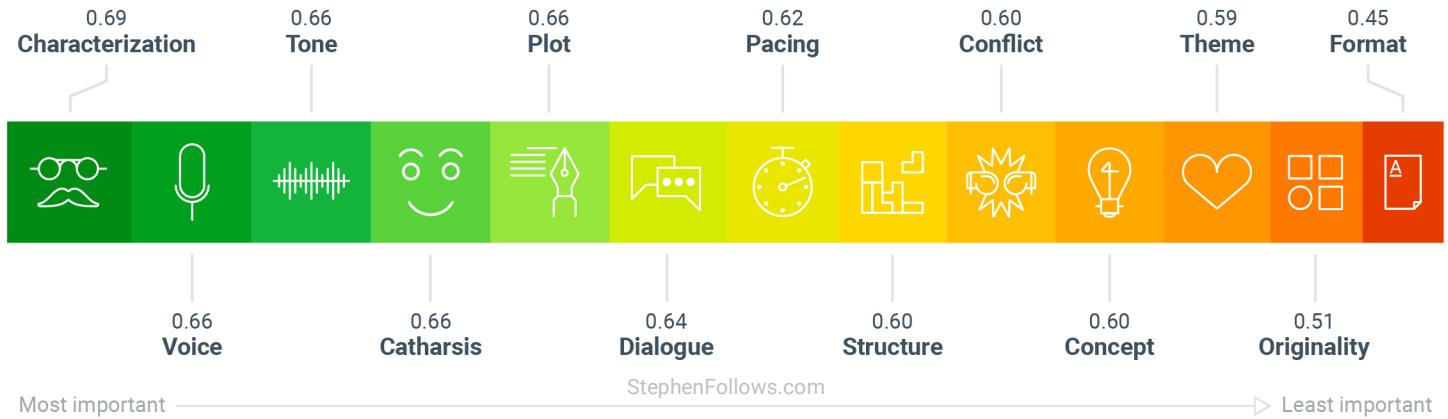
## Comedy

The most important factors for Comedies are characterization and plot. Interestingly, the pace of a comedy script has a far weaker connection to its overall score than that of any other genre. In this context, pacing refers to the speed of plot points moving forward, rather than fast-talking characters.



## Drama

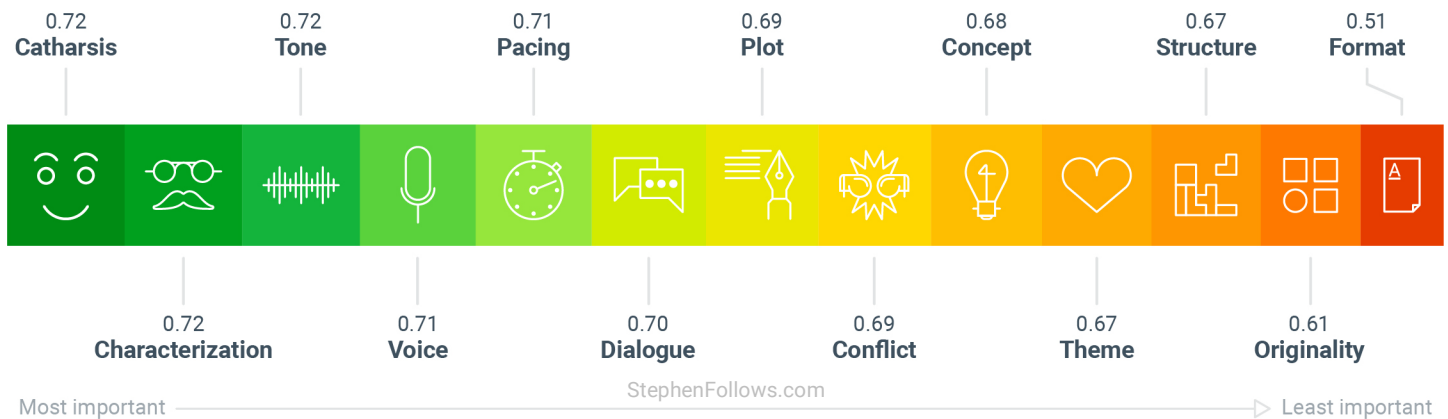
Drama scripts owe a large amount of their final score to the strength of their characterization. Interestingly, plot is far less important than with most other genres.



## Family

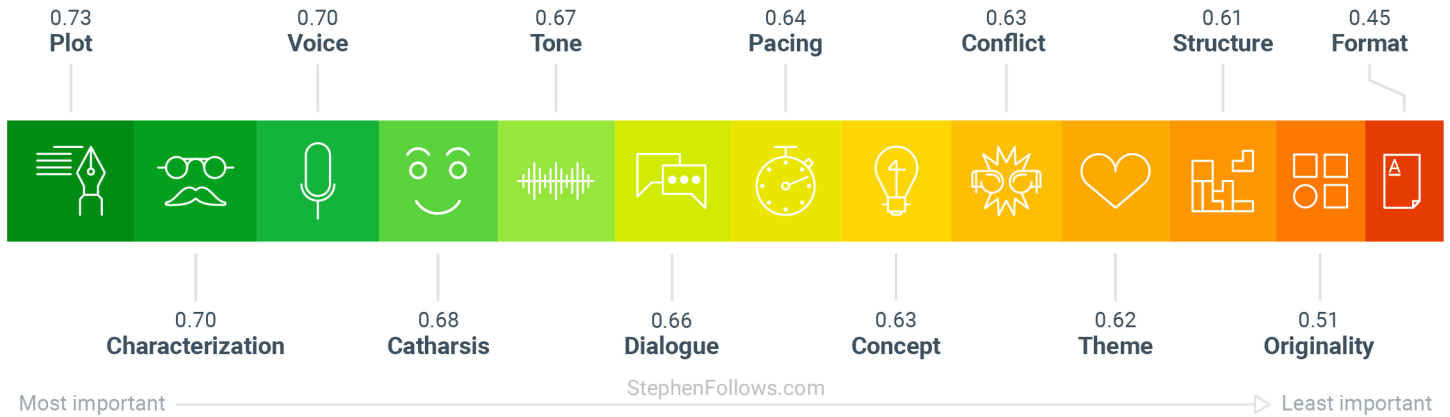
Family scripts rely most on catharsis and are the least dependent on the strength of their plot of all the major genres.

Family films are also rewarded for their pacing, keeping the story moving (even though it doesn't seem to matter what that story is!)



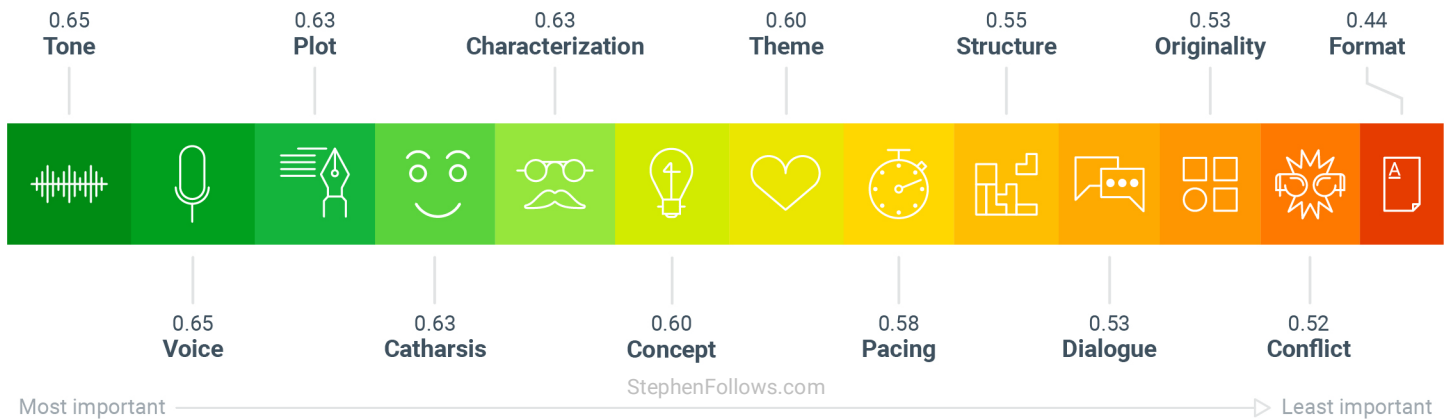
## Fantasy

Fantasy scripts are driven by the quality of their plot, characterization and voice. Fantasy scripts and Action scripts have many parallels, the principal difference being the diminished role of conflict in Fantasy scripts.



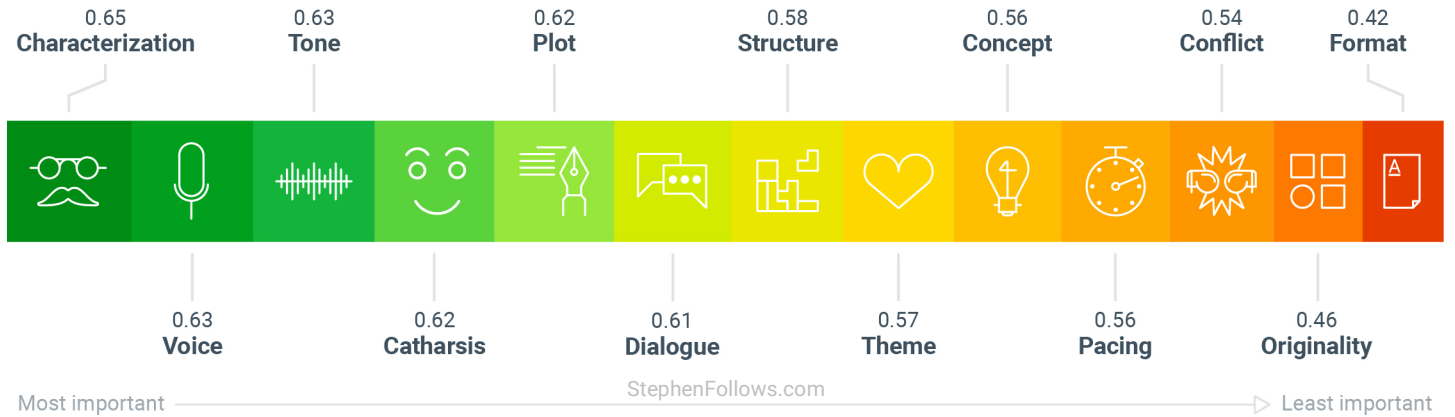
## Historical

Uniquely, historical scripts place the highest premium on tone, making it the most important factor in success with script readers.



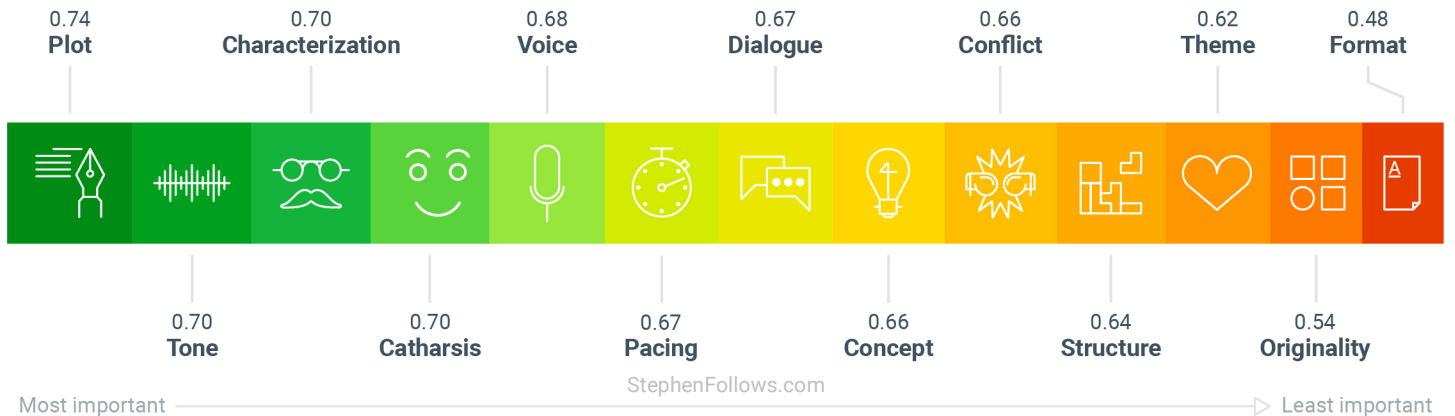
## Horror

Plot is not a comparatively key factor in the success of Horror scripts. The structure of a Horror script is more important than with any other genres.



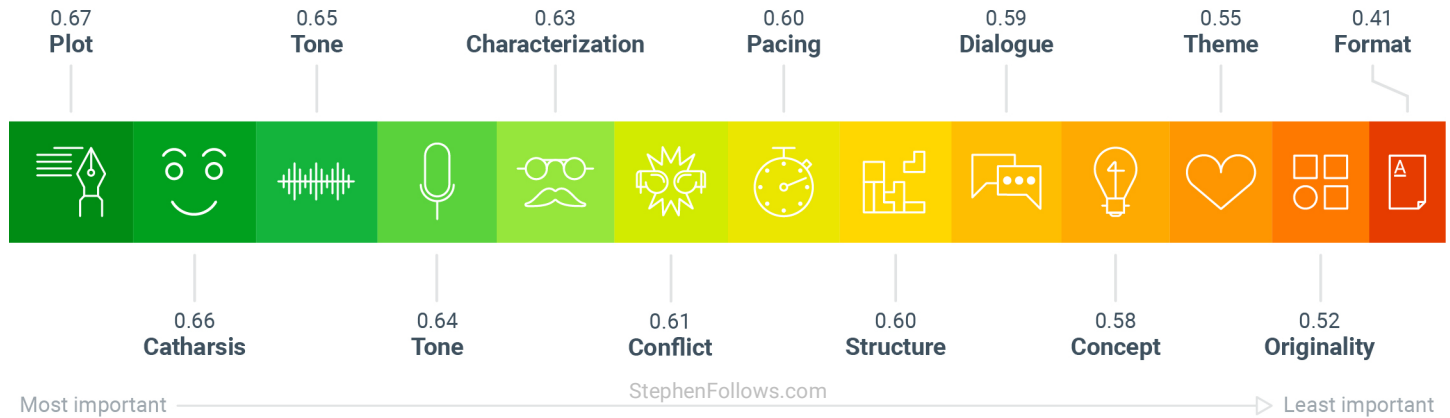
## Sci-Fi

Plot tops the Sci-Fi list with tone and characterization coming in a close second.



## Thriller

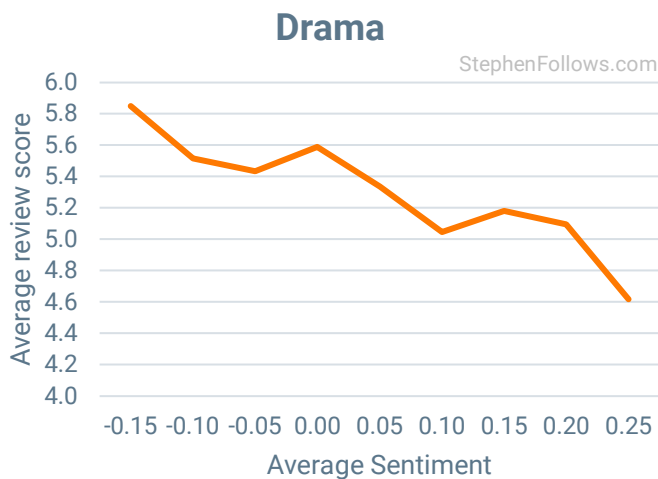
Thrillers are led by plot. However, when compared to other genres, they rate catharsis and tone higher than characterization and voice.



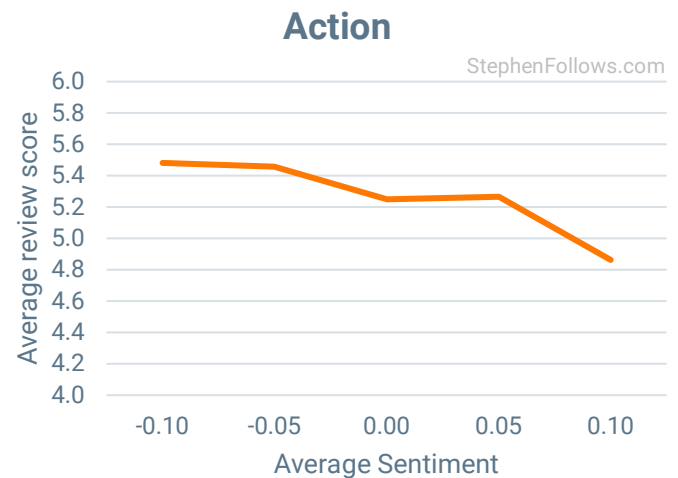
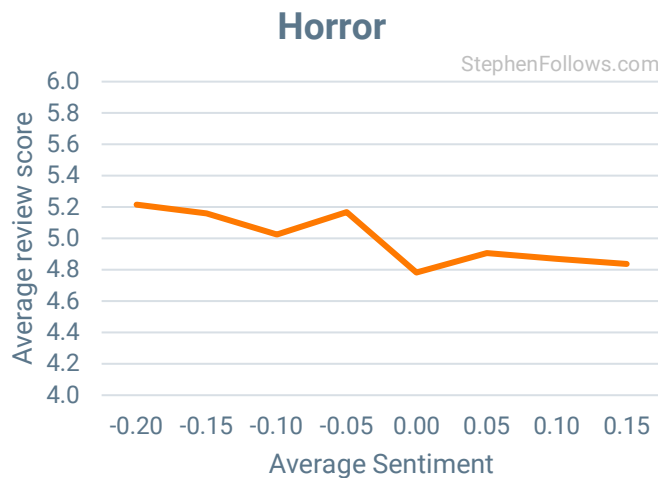
# Sentiment

Our system measures the average sentiment of each script<sup>5</sup> and provides an average value of between minus one (i.e. entirely negative) and one (i.e. entirely positive). A value of zero would indicate that the scripts contain an equal number of positive and negative elements.

Drama and Thriller scripts have the strongest negative connection between their average sentiment value and Review Score<sup>6</sup>. Dramas with a sentiment value of between 0.20 and 0.25 receive an average score of 4.68 out of 10, whereas much more negative films (i.e. those with a sentiment value between -0.20 and -0.15) receive an average score of 5.85.



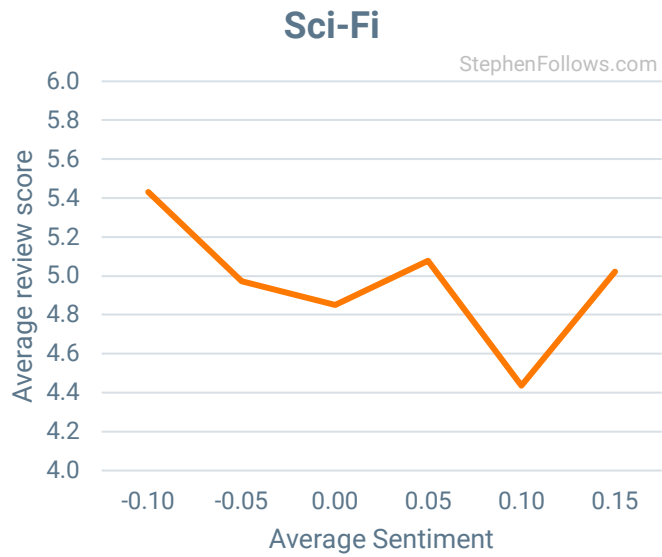
Horror, Action and Sci-Fi scripts show the same pattern, although it is less pronounced with Dramas and Thrillers.



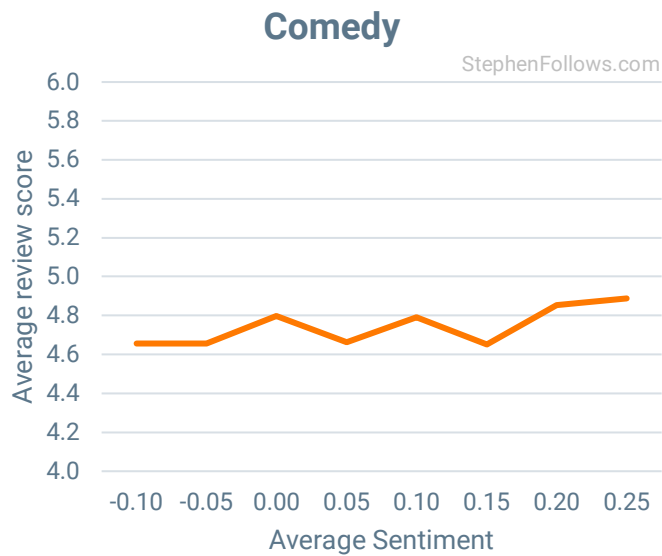
<sup>5</sup> Sentiment was calculated using the VADER (Valence Aware Dictionary and sEntiment Reasoner) lexicon and rules. This is a dictionary which assigns a 'sentiment' value to each word to indicate how positive/negative it is, validated by thousands of human checkers. The sentiment of each sentence is calculated by summing and normalising the values of the words in it, as well as applying a number of contextual rules; e.g. adjusting the value of words depending on qualifiers such as 'very' or punctuation such as '!' and examining sentence structure to take into account 'but' and other negators.

<sup>6</sup> The sentiment charts only show values for subsets in which there were at least 25 scripts (i.e. Drama scripts with an average sentiment value of between -0.20 and -0.15).

## JUDGING SCREENPLAYS BY THEIR COVERAGE



The only genre with the opposite relationship (i.e. where positive films receive slightly higher scores than negative films) is Comedy.



The other genres did not show a clear relationship between sentiment and readers' scores.

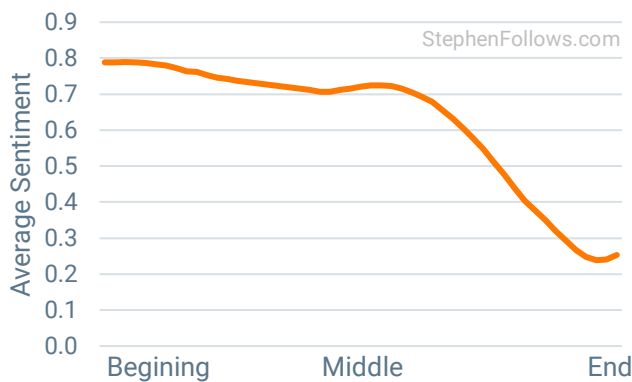
# Plot Arcs

By tracking the sentiment of the script over the length of a script, we are able to measure the emotional journey of the film. When things get dicey for the characters, the language becomes increasingly negative, and when things are going well, the sentiment values will rise accordingly.

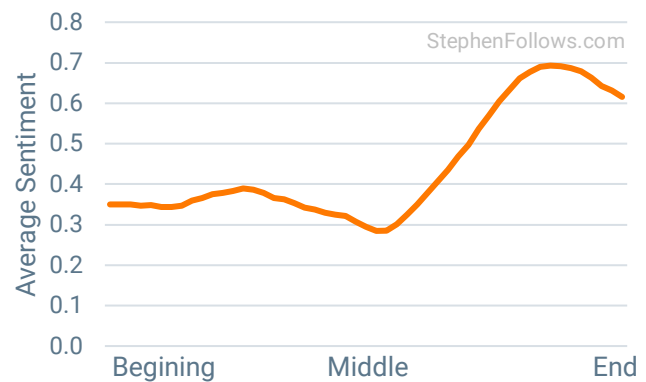
Past researchers have found that there are six commonly found plot arcs in books<sup>7</sup> and movies<sup>8</sup>. We found our scripts could be sorted into the same six arcs:

- **"Riches to rags"** – a continuing emotional fall. E.g. Blue Jasmine
- **"Rags to riches"** – a continuing emotional rise. E.g. Shawshank Redemption
- **"Oedipus"** – fall-rise-fall. E.g. Inglourious Basterds
- **"Cinderella"** – rise-fall-rise. E.g. Rushmore
- **"Man in a hole"** – fall-rise. E.g. Die Hard
- **"Icarus"** – rise-fall E.g. Se7en

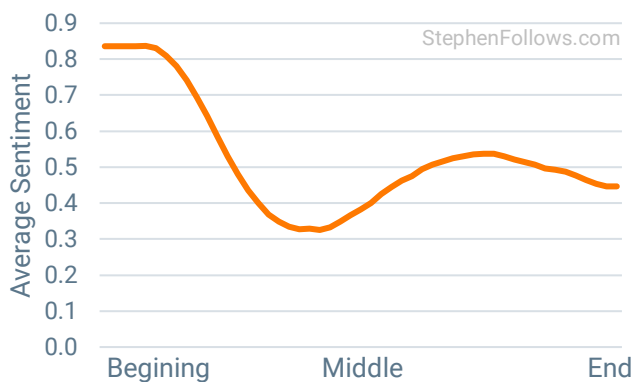
### Riches to Rags (fall)



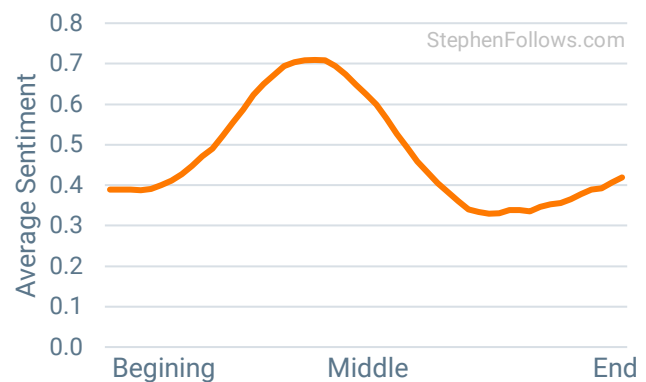
### Rags to Riches (rise)



### Oedipus (fall-rise-fall)



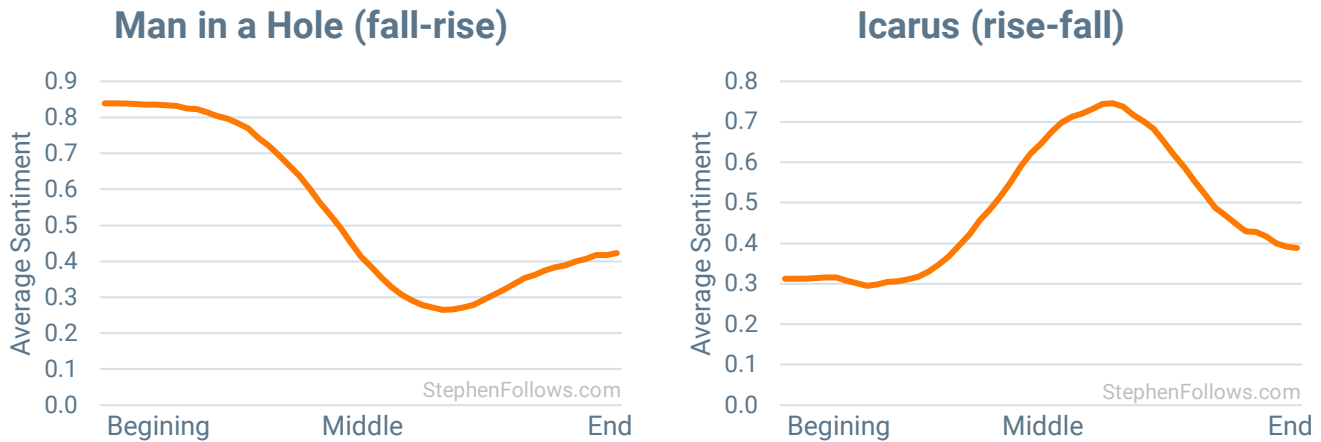
### Cinderella (rise-fall-rise)



<sup>7</sup> "Data Mining Reveals the Six Basic Emotional Arcs of Storytelling" <https://www.technologyreview.com/s/601848/data-mining-reveals-the-six-basic-emotional-arcs-of-storytelling/>

<sup>8</sup> "The Data Science Of Hollywood: Using Emotional Arcs Of Movies To Drive Business Model Innovation In Entertainment Industries" <https://arxiv.org/vc/arxiv/papers/1807/1807.02221v1.pdf>

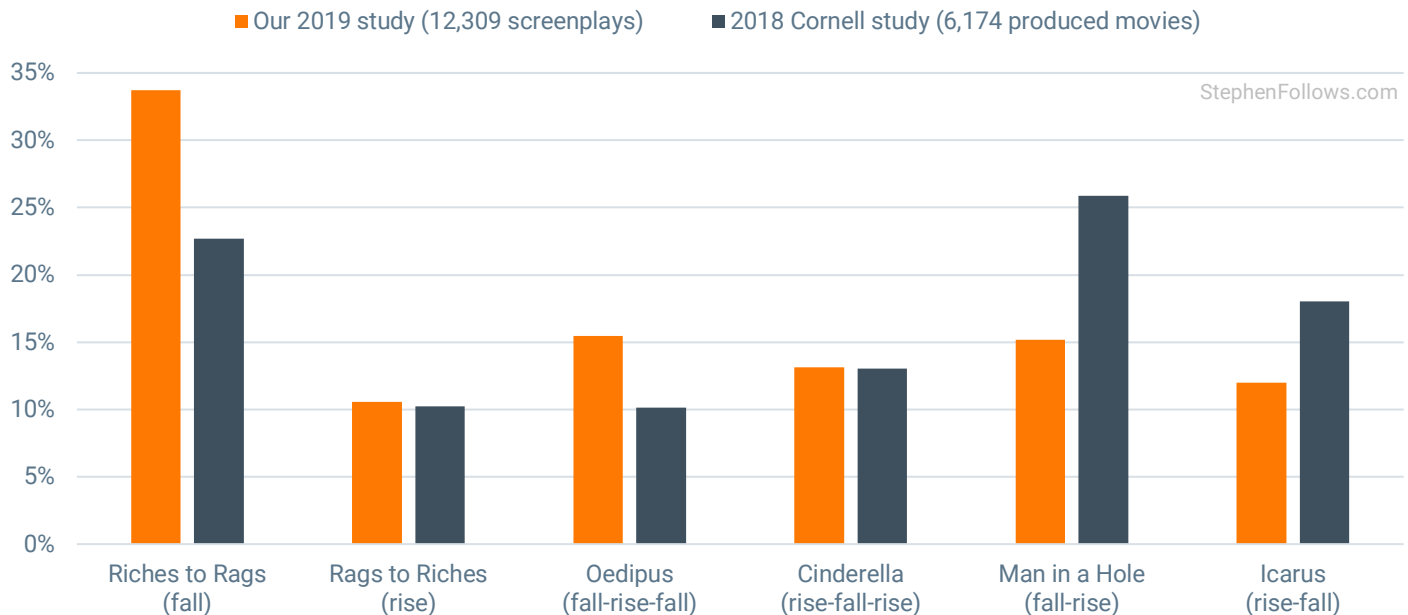
## JUDGING SCREENPLAYS BY THEIR COVERAGE



Naturally, individual scripts/movies have multiple emotional ups and downs. There are two reasons why we can generate smooth graphs as shown above. Firstly, the more significant/emphatic the emotional highs and lows are, the greater their impact on the sentiment graph. Secondly, we are averaging over a very large number of scripts.

A recent study<sup>9</sup> measured the emotional arcs of over 6,000 produced feature films and tracked how common each of the six plot arcs are. The most common arc found in the study was Man in a Hole, followed by Riches to Rags. However, within our set of amateur scripts, we found Riches to Rags is the most common by a long way, with over 30% of our films fitting that category.

### Prevalence of the six common plot arcs



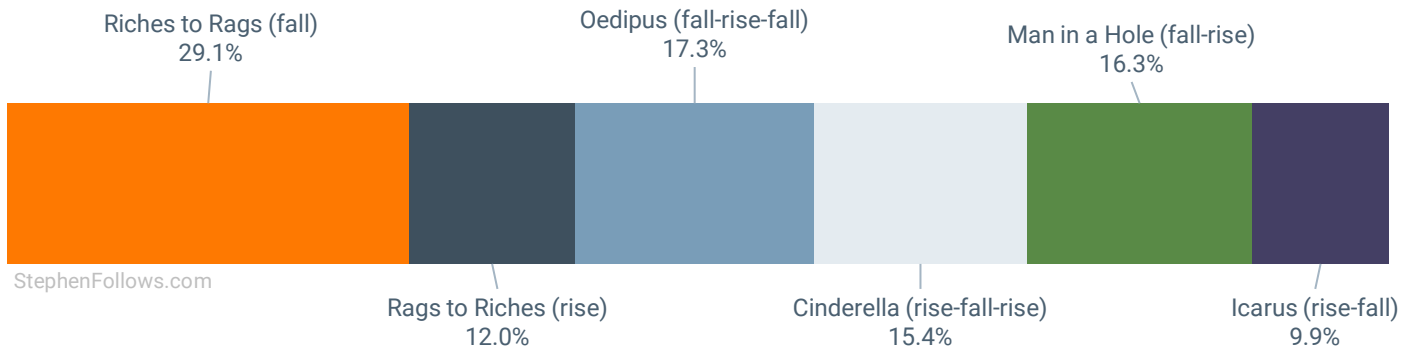
Naturally, different genres have different storytelling requirements and conventions. So it's more revealing to take a look at how prevalent these arcs are within different genres, and how they relate to reader scores.

<sup>9</sup> Vecchio et al. (2018). The Data Science of Hollywood: Using Emotional Arcs of Movies to Drive Business Model Innovation in the Entertainment Industry. <https://arxiv.org/abs/1807.02221>

## Action

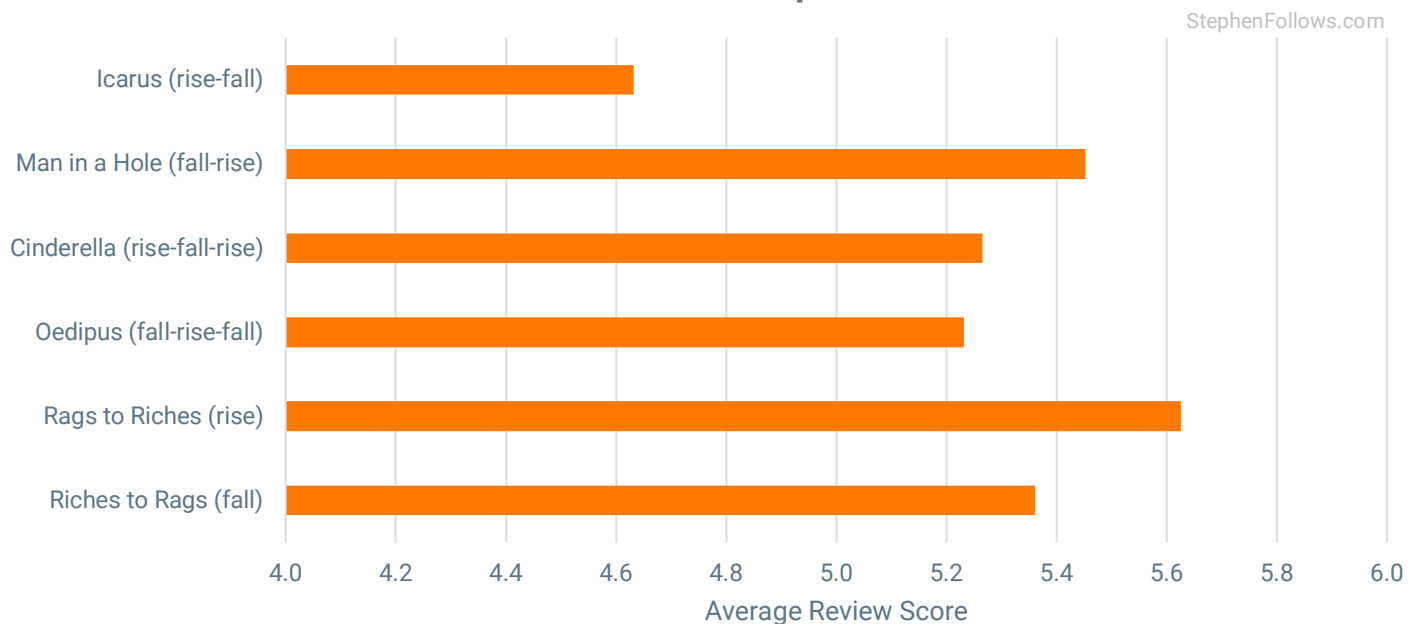
The majority of our Action scripts fell into the Riches to Rags arc (29.1%) and the least common was Icarus (9.9%).

### Prevalence of the six common plot arcs - Action



When we compare the mean scores of Action scripts within different arcs, we can see some significant differences<sup>10</sup>. Rags to Riches is associated with the greatest average review score (5.63), while the scripts within the Icarus arc had an average score of one point less (4.63). The difference between Man in a Hole (5.45) and Icarus was on the verge of statistical significance, so may also indicate a trend.

### Review scores of common plot arcs - Action

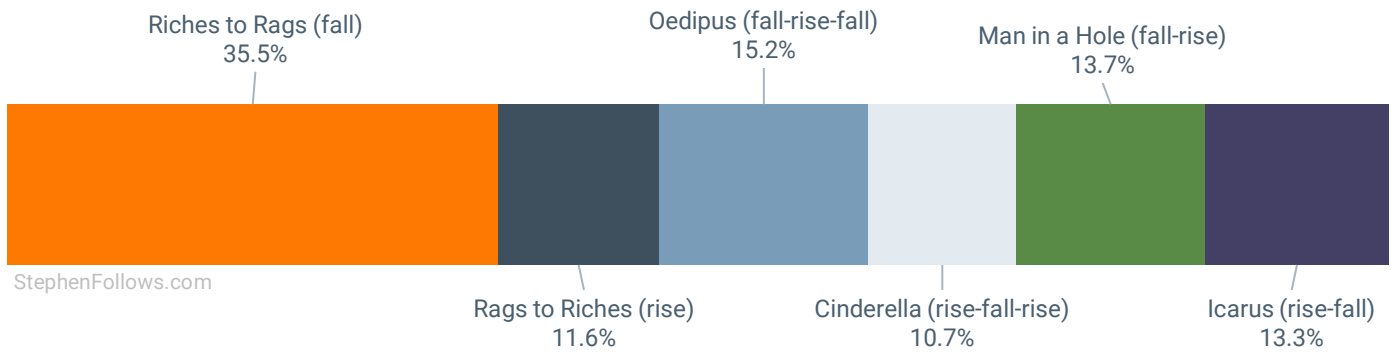


<sup>10</sup> Significance was established by conducting multiple comparison T-Tests. This test compares the mean and spread of the scores to determine the probability they represent distinct groups. We corrected for multiple comparisons by using the Bonferroni correction. Essentially, this adjusts the threshold you accept statistical significance at to be smaller, in order to account for the fact that by running more tests you increase the chance of a false significance.

## Comedy

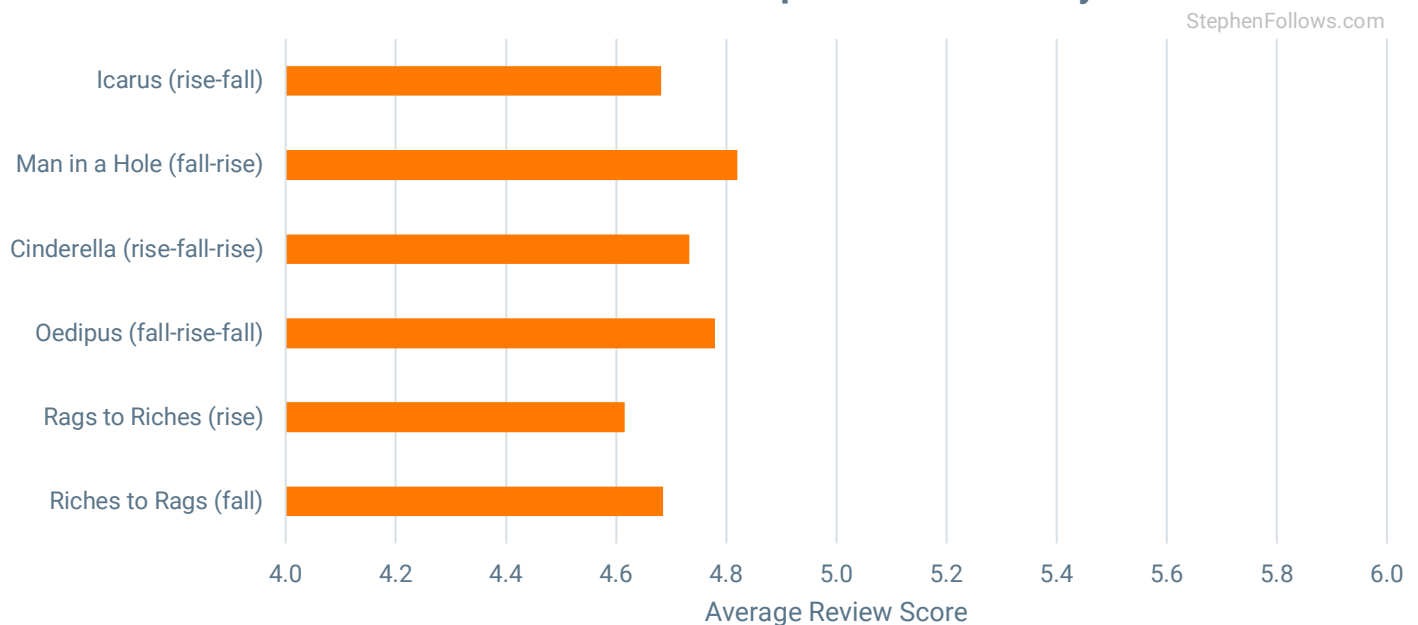
Within Comedy, we again found that Riches to Rags was the most common emotional arc (35.5%), while Cinderella was the least common (10.6%).

### Prevalence of the six common plot arcs - Comedy



There were some small differences in review scores between different emotional arcs, however tests showed that these were not statistically significant.

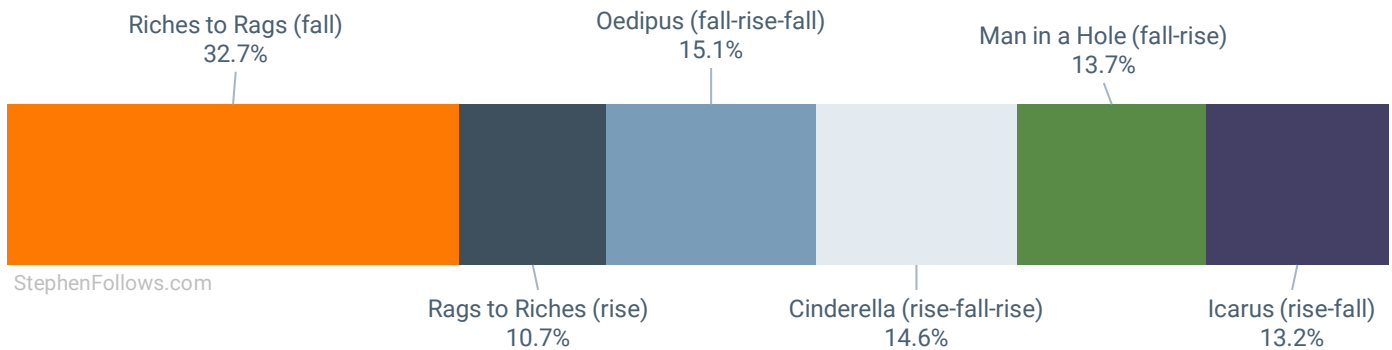
### Review scores of common plot arcs - Comedy



## Drama

Riches to Rags was once again the most popular arc within this genre (32.7%), while Rags to Riches was the least popular (10.7%).

### Prevalence of the six common plot arcs - Drama



We found some small differences in mean scores between the different arcs, and while none were statistically significant, we found a trend on the verge of significance: Cinderella arcs seemed to do better than Riches to Rags.

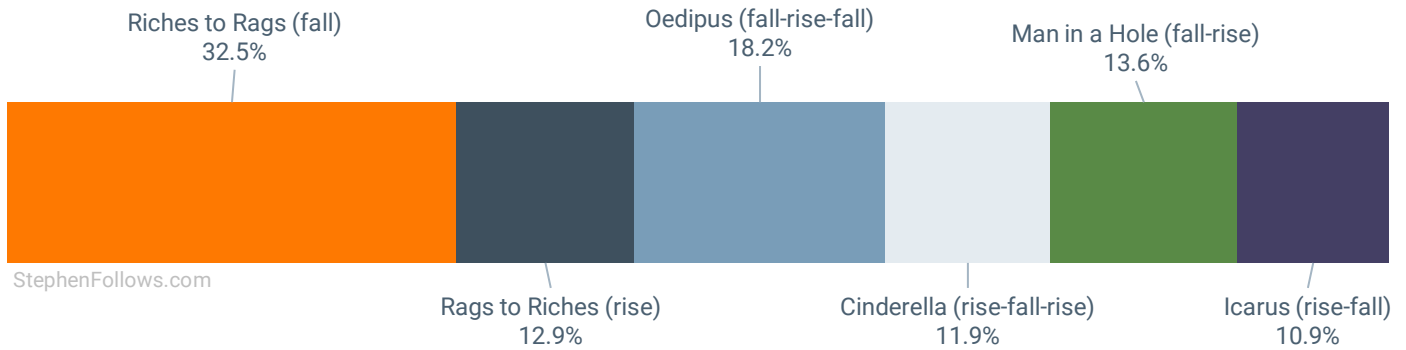
### Review scores of common plot arcs - Drama



## Fantasy

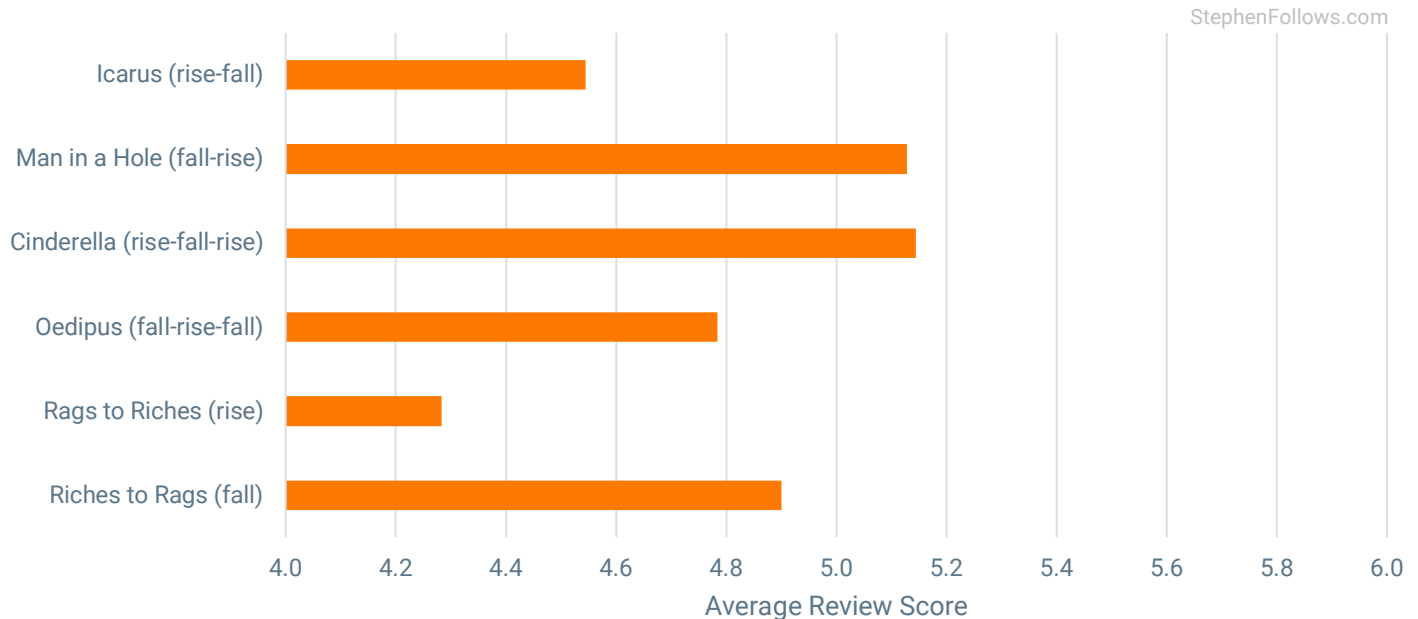
Riches to rags were once again the most common story arc (32.5%), while Icarus was the least common (10.9%).

### Prevalence of the six common plot arcs - Fantasy



We found significant differences in mean scores between different arcs. With Rags to Riches performing the worst (mean score of 4.28), and Cinderella (5.14), Man in Hole (5.13) and Riches to Rags (4.9) performing the best.

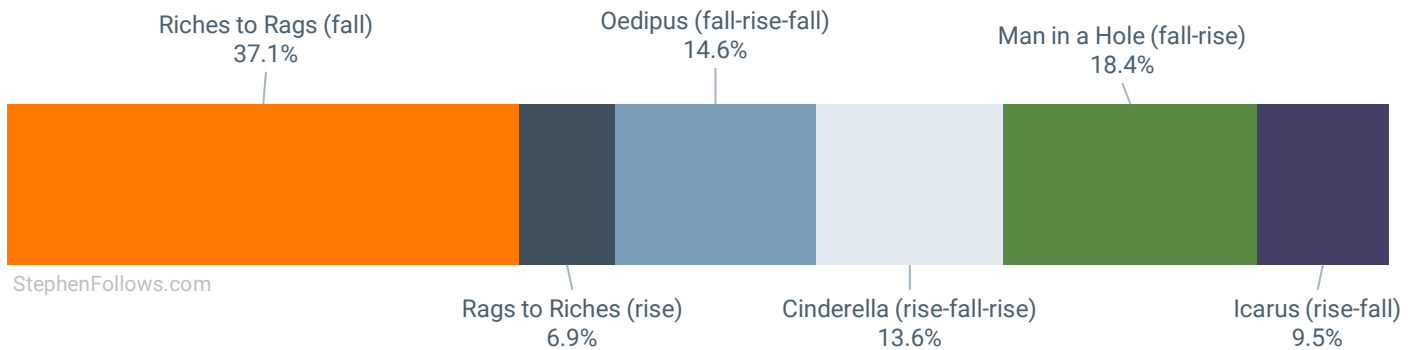
### Review scores of common plot arcs - Fantasy



## Horror

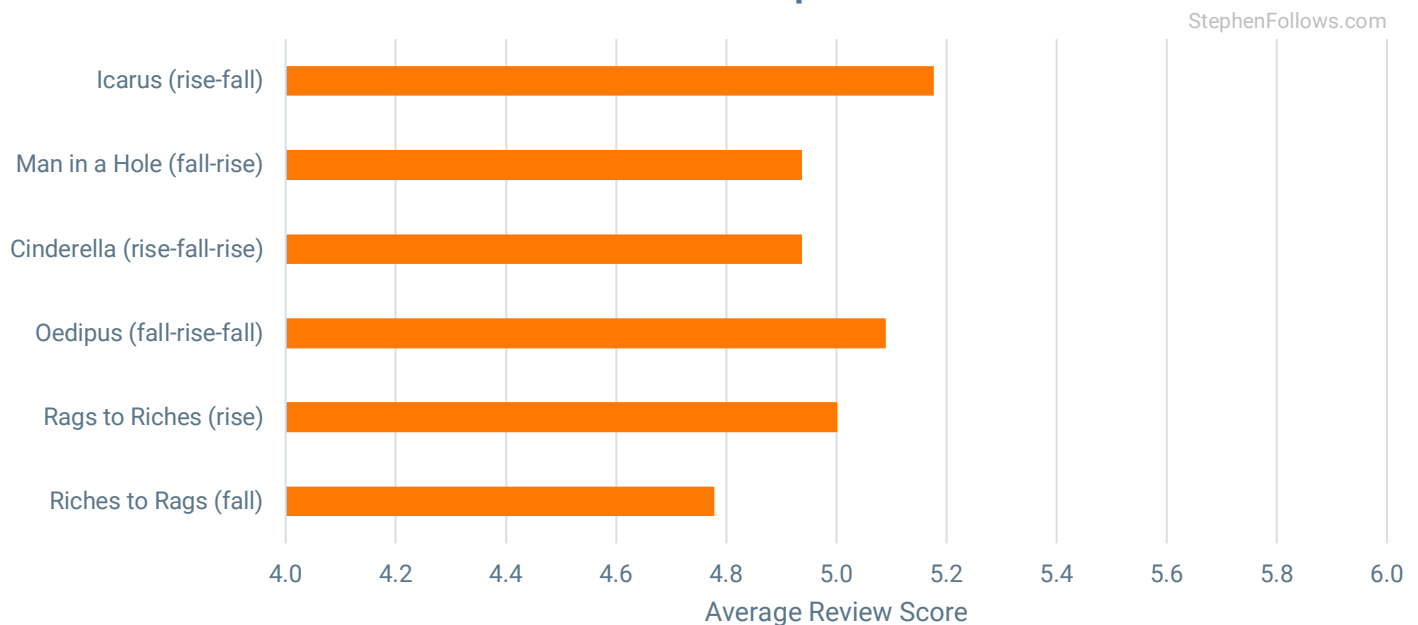
Horror had the highest proportion of stories with a continuing emotional fall (37.1%) and the lowest proportion with a continuing emotional rise (6.9%).

### Prevalence of the six common plot arcs - Horror



But despite being the most common, there was a significant difference in the scores of the Riches to Rags arc (4.78) compared to Icarus (5.18) and Oedipus (5.09) arcs.

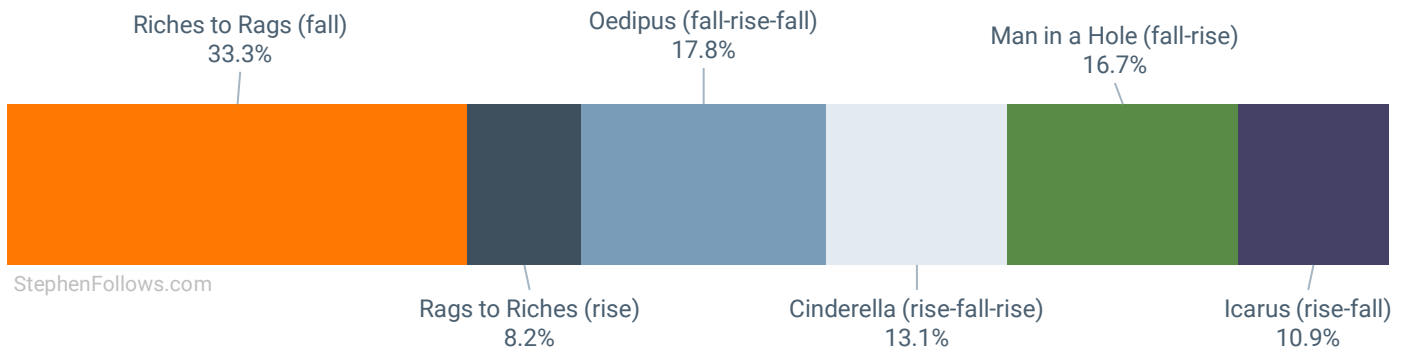
### Review scores of common plot arcs - Horror



## Sci-Fi

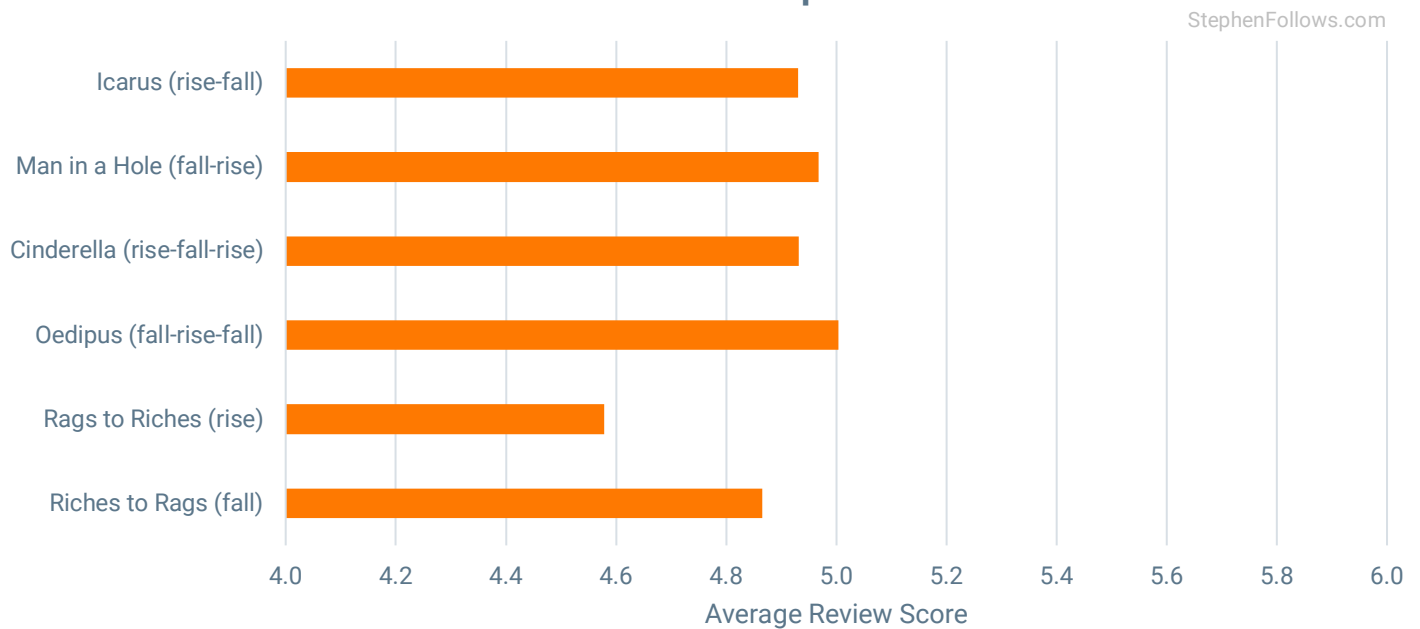
The most common arc in Sci-Fi is once-again the Riches to Rag arc (33.3%), and the least common is Rags to Riches (8.2%).

### Prevalence of the six common plot arcs - Sci-Fi



While not statistically significant, it's noticeable that Rags to Riches (4.58) performed worse than all other categories.

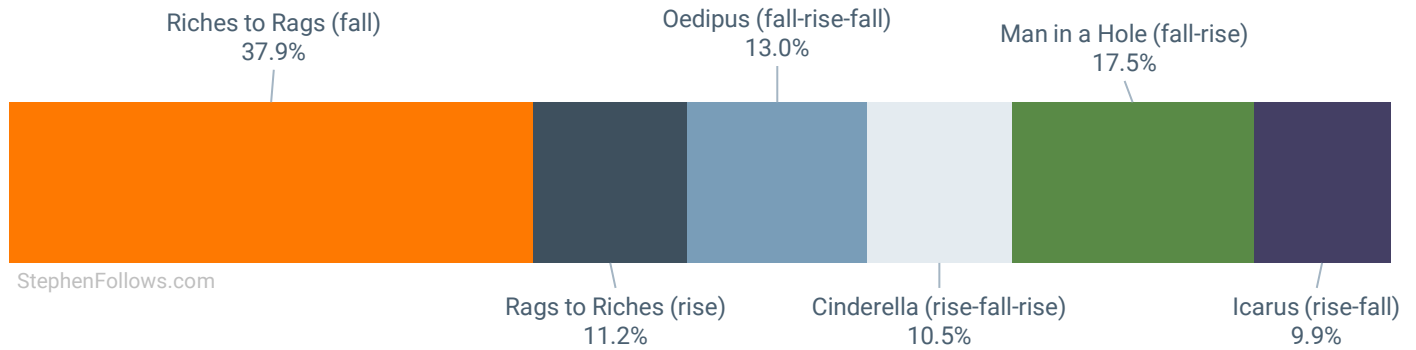
### Review scores of common plot arcs - Sci-Fi



## Thriller

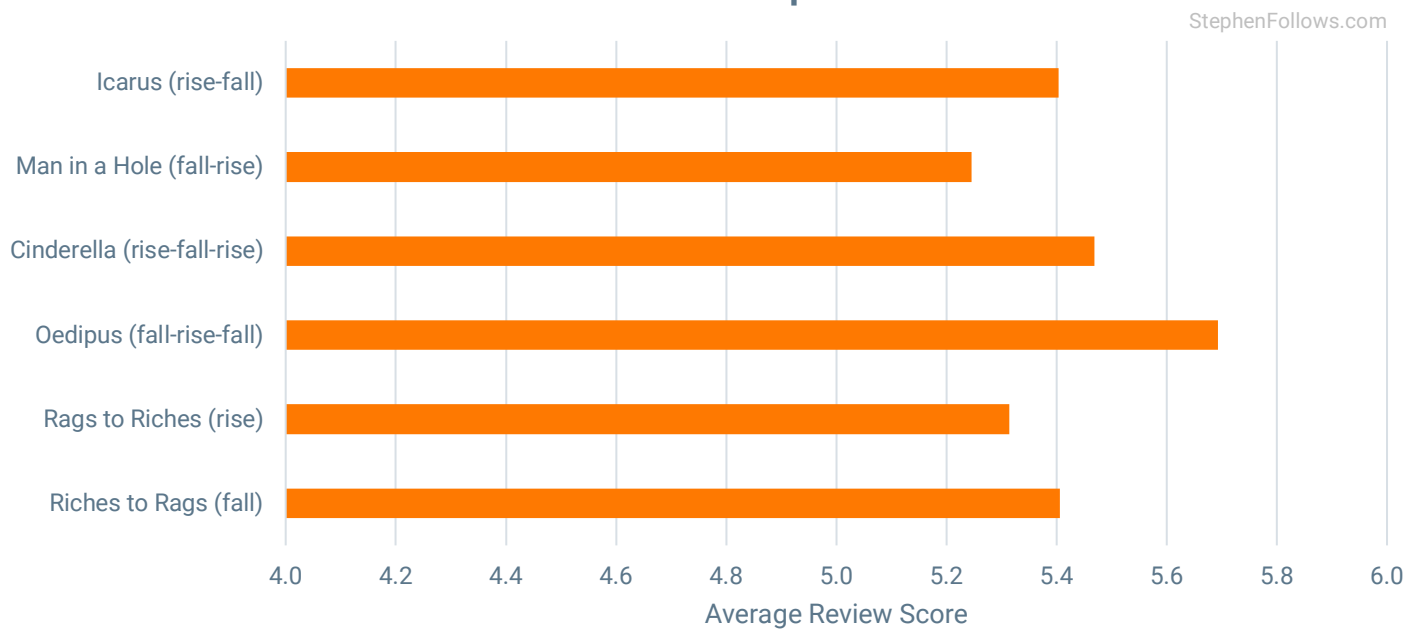
Riches to Rags was still the most popular arc (37.9%), while Icarus was the least popular (9.9%).

### Prevalence of the six common plot arcs - Thriller



While not statistically significant, we did find a trend that Man in a Hole was likely to perform worse than Oedipus in this genre.

### Review scores of common plot arcs - Thriller

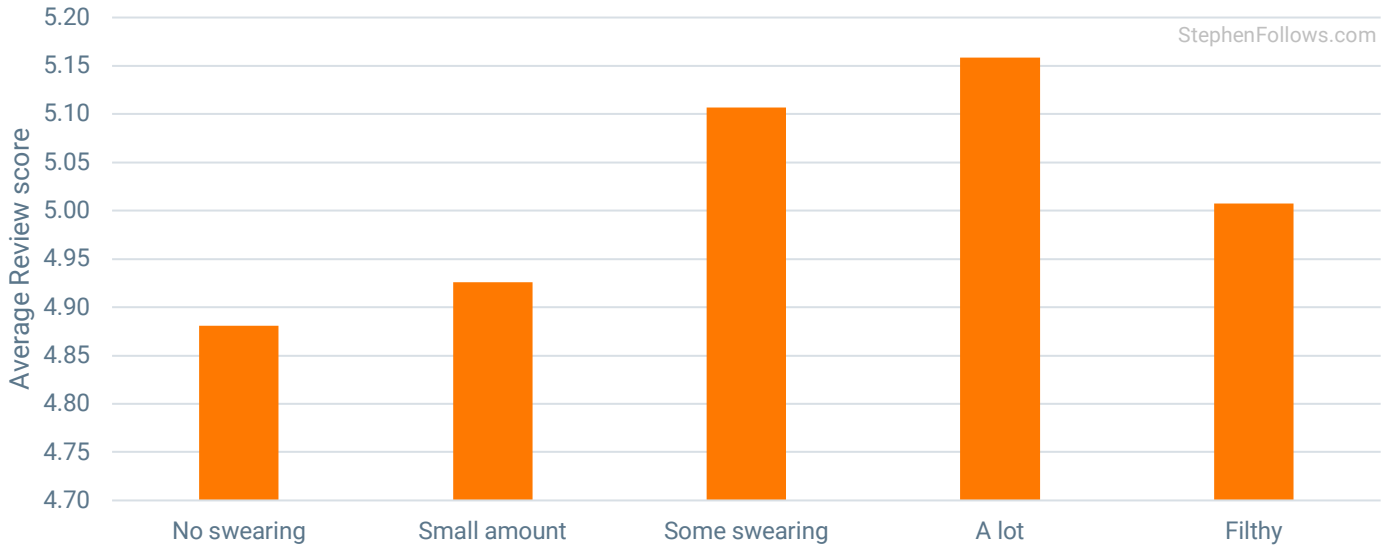


# Swearing

There is a mild correlation between the level of swearing and the overall score a script received.

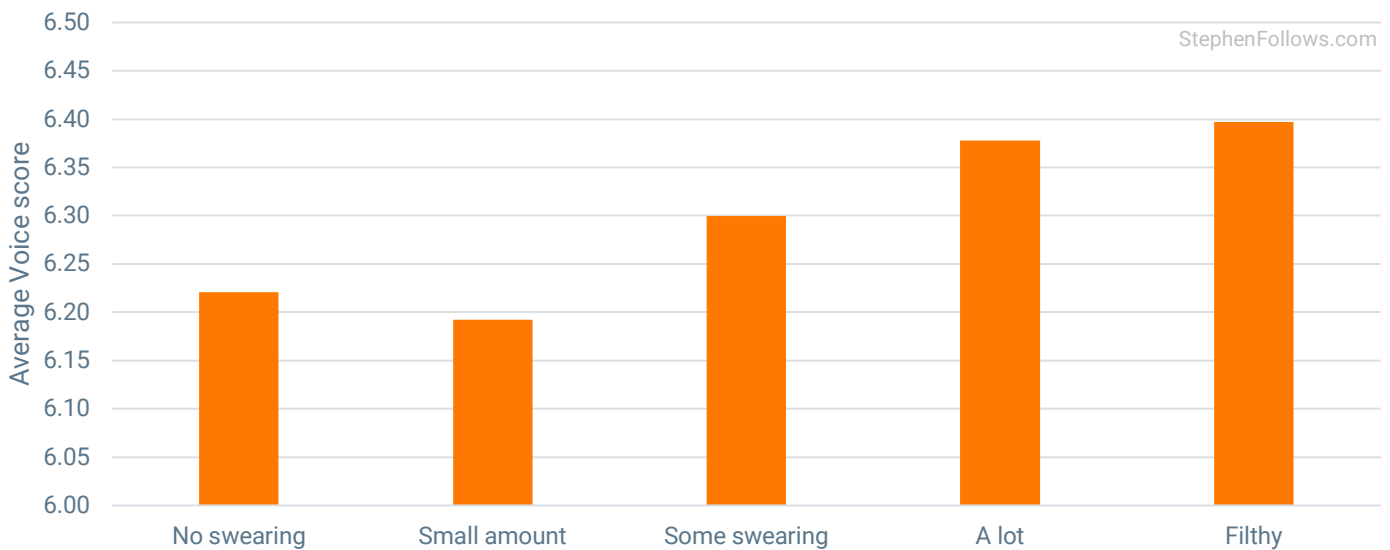
When the scripts are grouped into five levels of swearing<sup>11</sup>, the average score increases along with the level of profanity. The only exception to this rule are the filthiest scripts (i.e. the 20% sweariest scripts) which were mildly less successful.

**Average review score by amount of swearing**



Interestingly, one of the connections we discovered is a link between the level of swearing and scores the scripts received for "Voice".

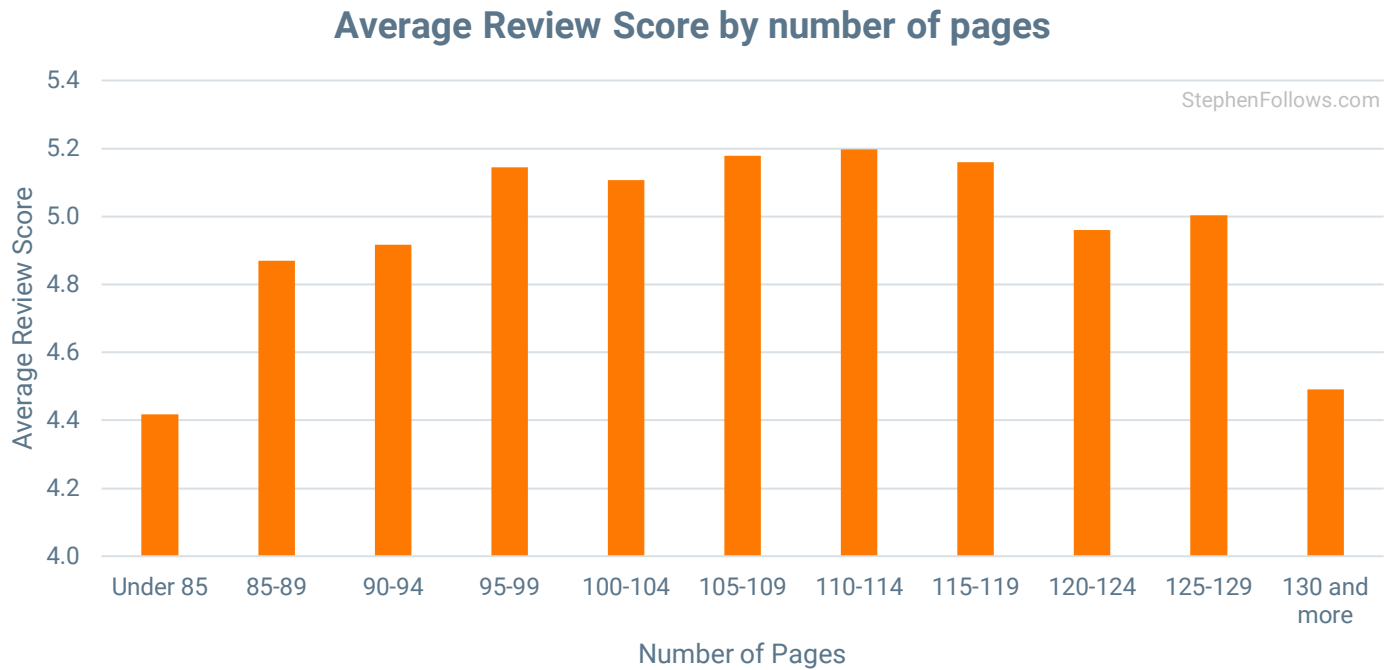
**Average 'Voice' score by amount of swearing**



<sup>11</sup> See a later section for details on our methodology for these classifications.

# Number of Pages

There is no statistically significant correlation between the number of pages and the overall score a script received. However, scripts at the margins (i.e. the very shortest and longest) do perform worse than the rest.



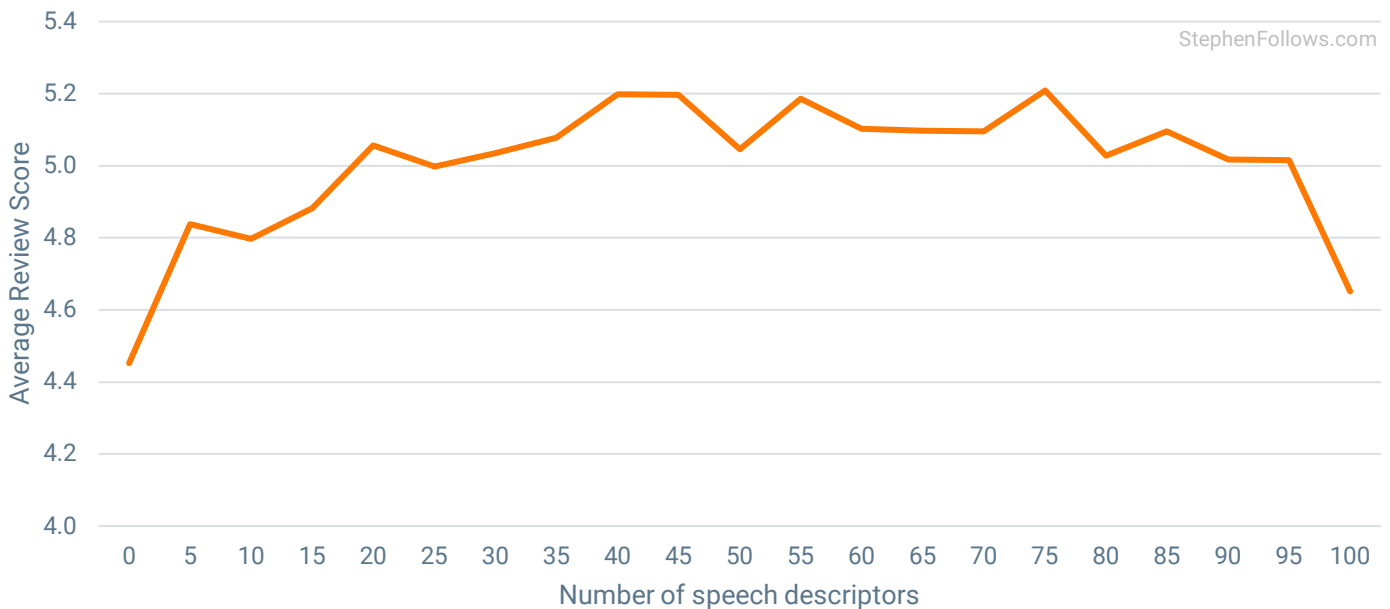
# Speech Descriptors

One complaint producers and directors level at screenwriters is that they overwrite the speech descriptions.

These appear below a character's name, ahead of any dialogue and are in parentheses, examples being “(loudly)”, “(angrily)”, etc. The complaint mostly comes down to writers overreaching and telling either the director or the actors how to do their jobs.

We found that speech descriptors are fairly common, averaging at 74.0 instances across our scripts. For the most part, they were not correlated with the scores readers gave the scripts, except at the margins. Scripts with an unusually low number of descriptors (below around 25) score poorly, as do the scripts with the most (above around 235).

**Average Review Score by usage of speech descriptors**



# Use of Voiceover

An old adage of the screenwriting craft is “show – don’t tell” and this is often cited when discussing the use of voiceover. Some contend that an over-reliance on voiceover is a sure sign of a poor script. Our script dataset is the perfect place to test this theory and the answer to the question “Is the amount of voiceover correlated to the quality of a script?”

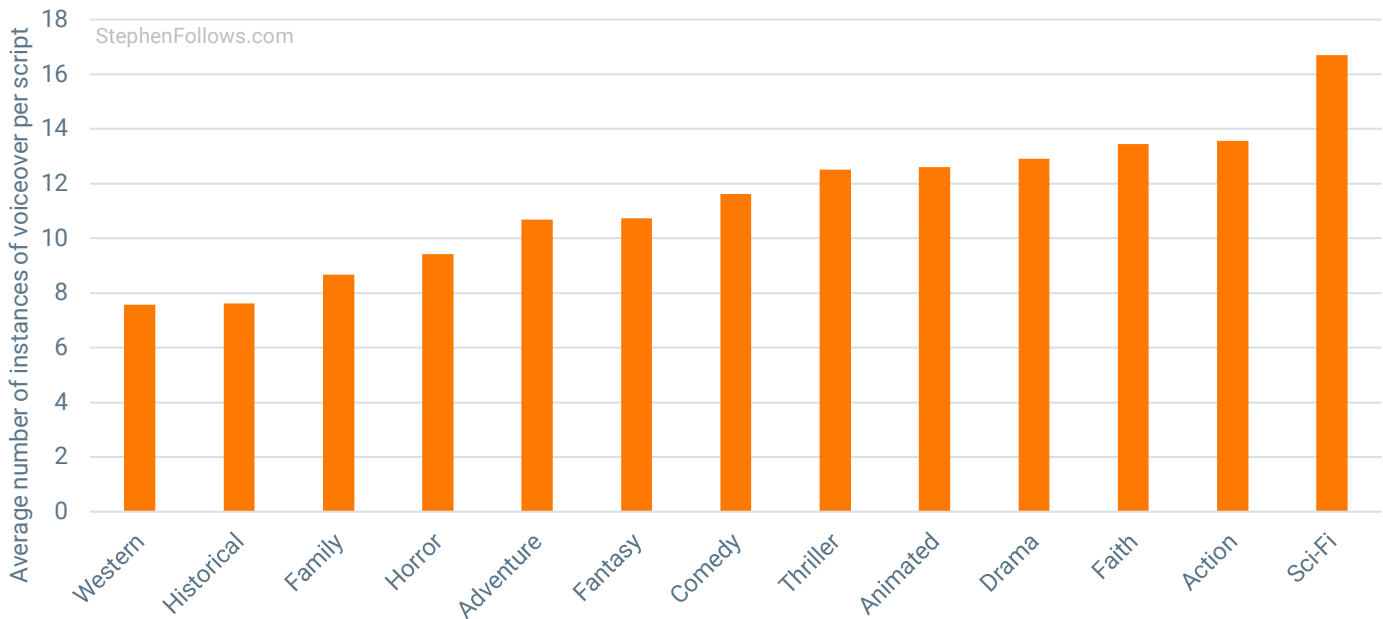
The answer is... no. There is almost no correlation between the scores scripts received and their reliance on voiceover<sup>12</sup>.

While this is true for our scripts, it may still be that finished movies which rely on voiceover are below par. This could be the result of frantic edits where editors, directors or even producers step in to “solve” issues in edits by adding extra lines of voiceover<sup>13</sup>.

Almost a third of all scripts didn’t feature any voiceover and across all scripts the average project featured just 12.1 instances of voiceover.

Sci-Fi scripts are the most likely to use voiceover (an average of 16.7 lines per script) with Westerns the least (7.6).

## Average number of instances of voiceover, by genre

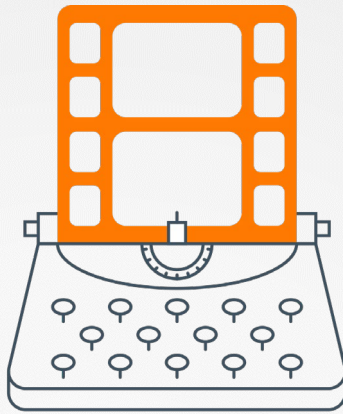


<sup>12</sup> The only exceptions to the finding that voiceover is not correlated with quality are the scripts which relied significantly on voiceover (i.e. over 70 instances), which do perform worse than the average script.

<sup>13</sup> An ‘instance’ of voiceover relates to a block of dialogue. This could contain one line or many.

---

# SECTION B: THE AVERAGE SCREENPLAY



*Analysing such a large dataset of feature film scripts also affords us  
a unique opportunity to measure what screenwriters are writing.*

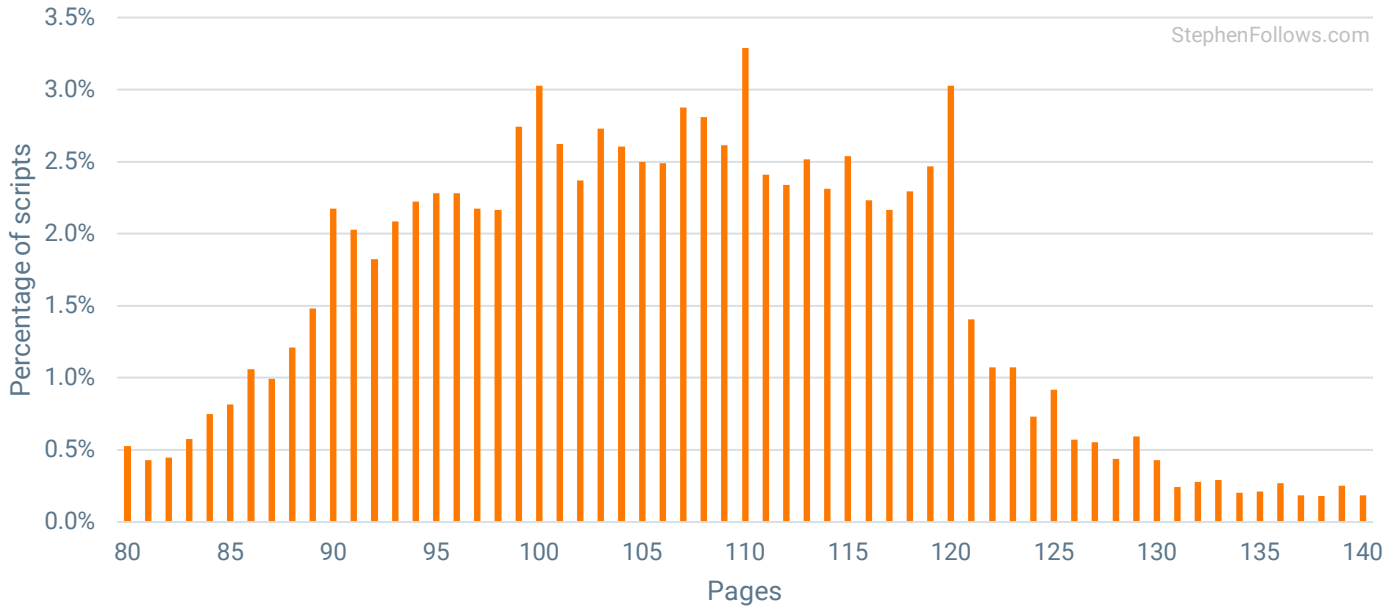
Pages.....	33
Scenes.....	35
Speaking Characters .....	36
Dialogue .....	37
Locations.....	38
Punctuation.....	40
Swearing.....	41
Age of Characters.....	44
Genre-skewed Words.....	48

# Pages

The median length across all of our scripts was 106 pages. However, there was a broad spectrum of lengths, with 68.5% of screenplays running between 90 and 120 pages long.

As the chart below shows, there are spikes on round numbers; namely pages 90, 100, 110 and 120.

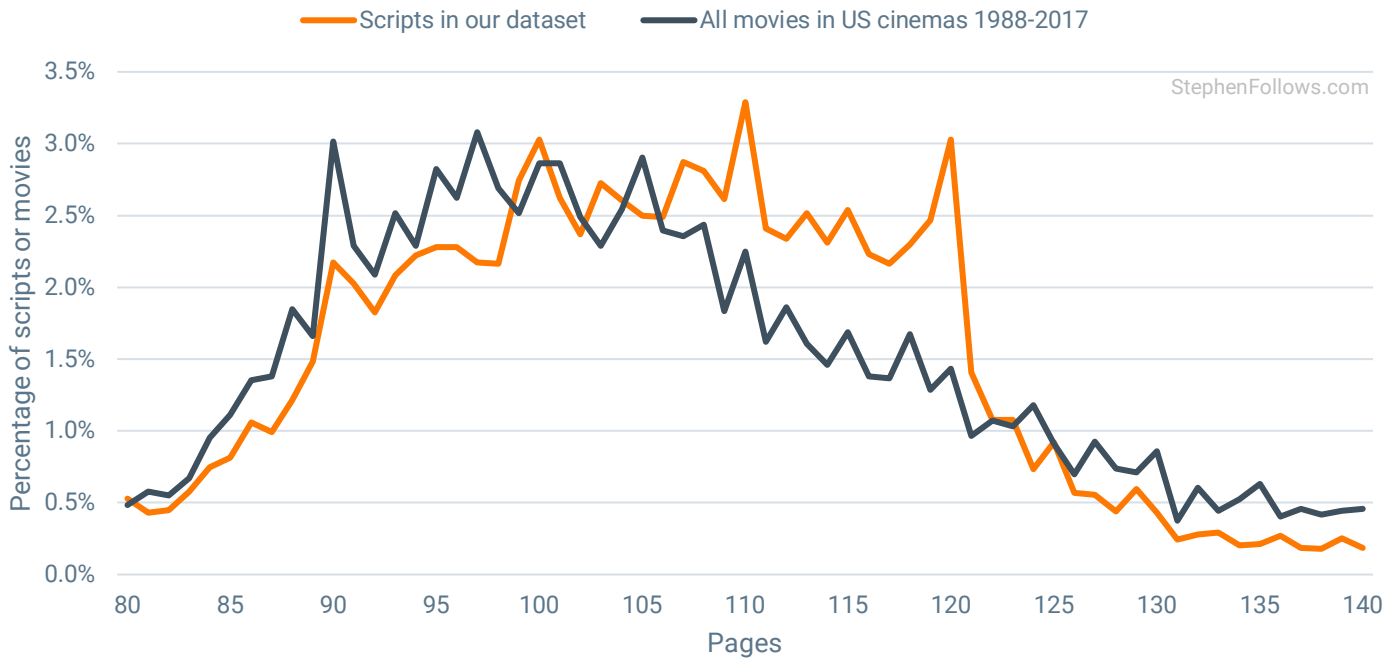
**Length of scripts in our dataset**



One page of correctly-formatted film script averages out at one minute of eventual screen time, meaning that a 120-page screenplay is likely to translate into a two-hour movie. This means we can compare our scripts to the running time of produced movies.

Movies in cinemas are generally slightly shorter than scripts in our dataset and are far less likely to run between 110 and 120 minutes.

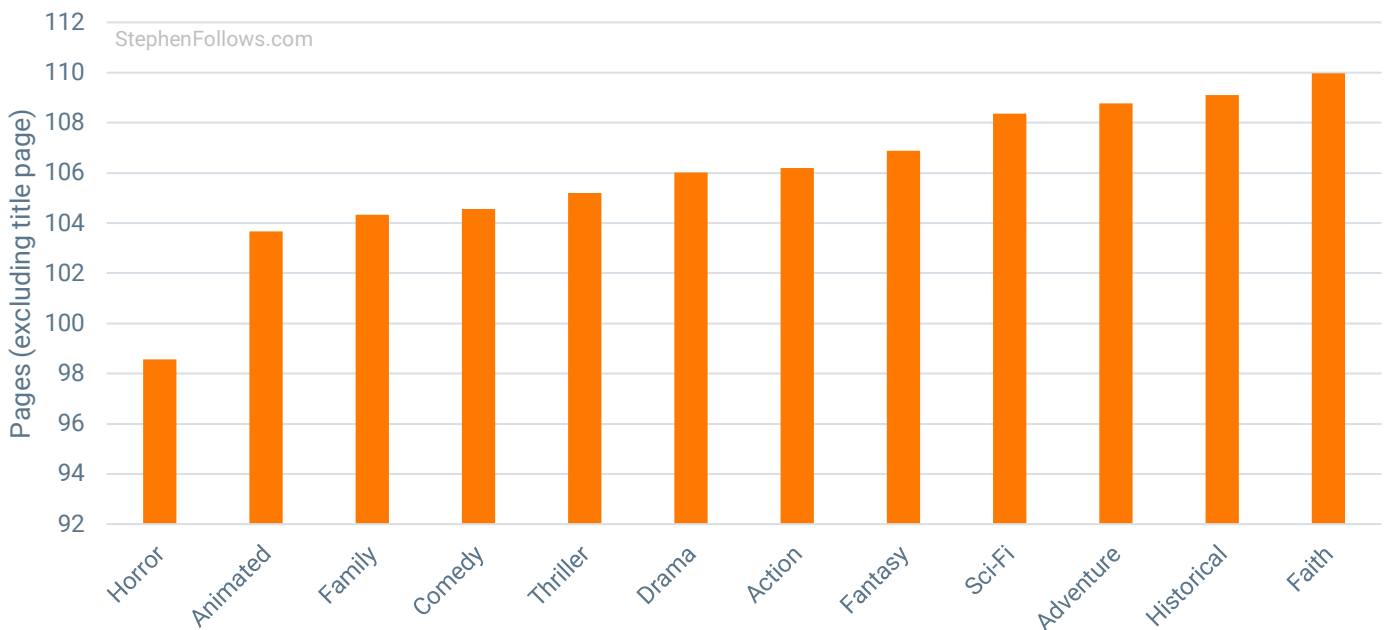
## Running time of produced movies and scripts in our dataset



Interestingly, at the longest end of the spectrum, the trend reverses. 5.8% of the movies released in US cinemas between 1988 and 2017 were over 140 minutes long, compared to just 1.4% of scripts in our dataset. Many such movies come from established directors and producers, such as James Cameron, Peter Jackson, Spike Lee, Oliver Stone and Steven Spielberg.

Horror scripts are the shortest, with an average page count of 98.6 while the longest were Faith scripts at 110.0 pages.

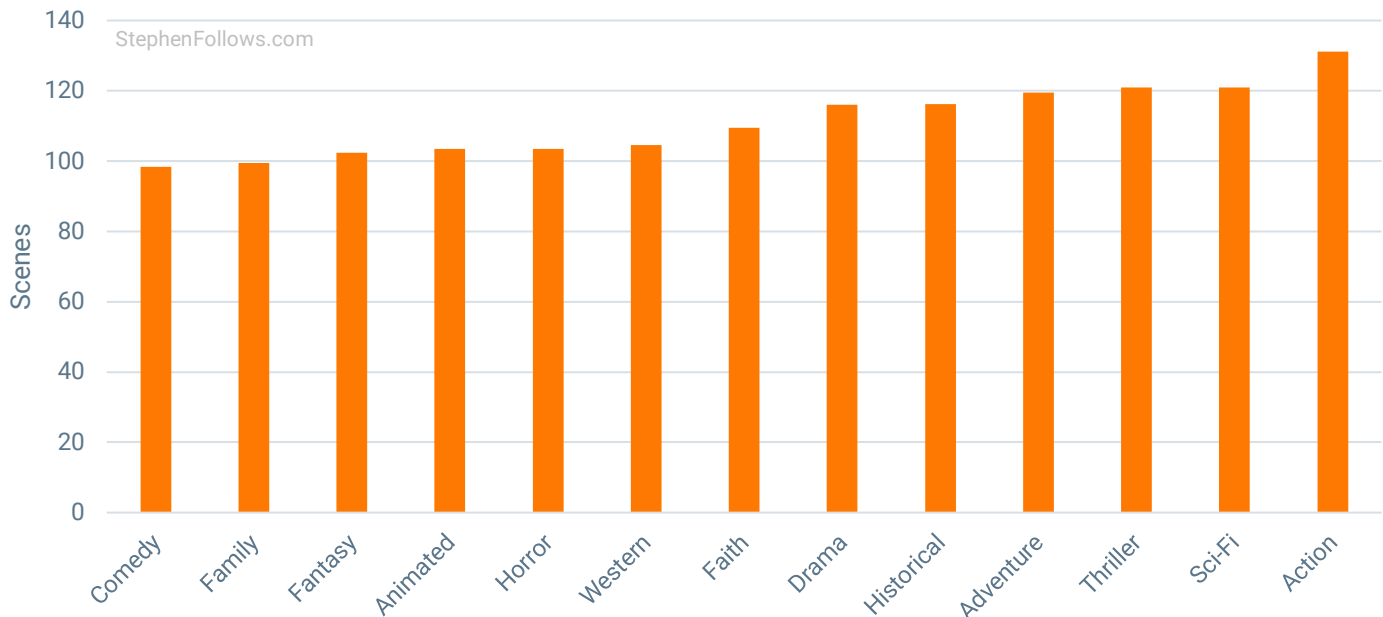
## Average page count, by genre



## Scenes

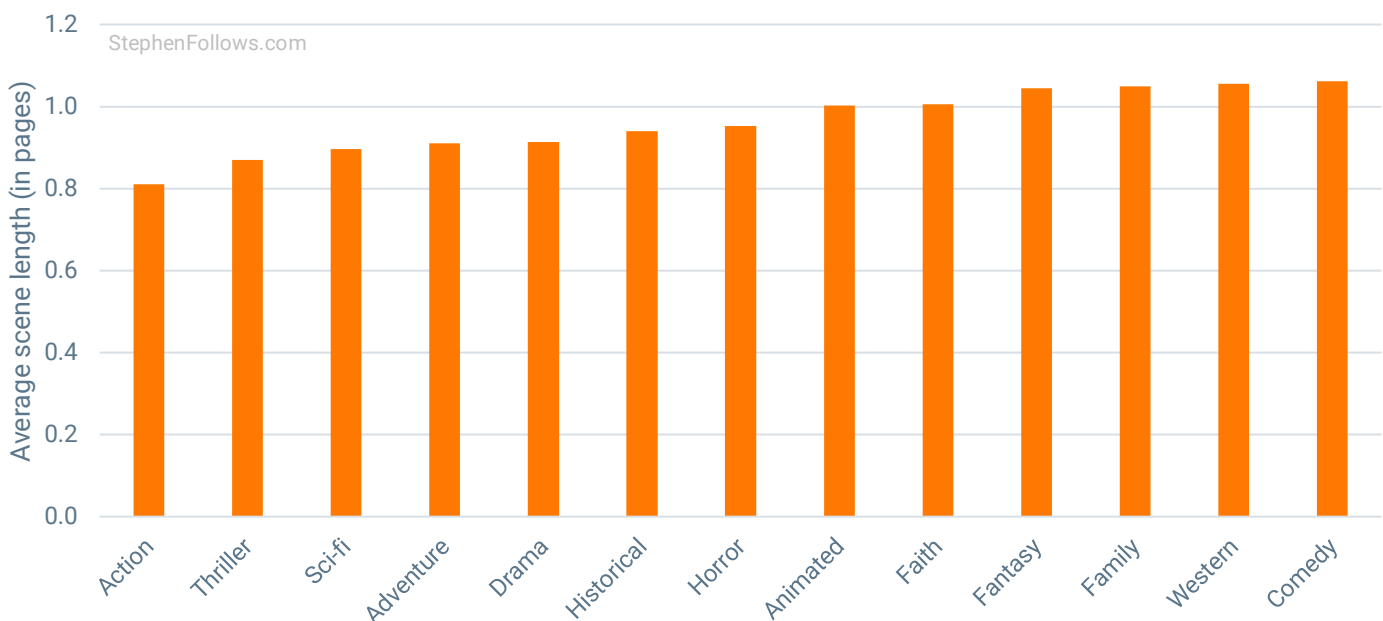
The average script has 110 scenes – just over one scene per page. Action scripts have the greatest number of scenes (an average of 131.2 scenes) with Comedies having the fewest (just 98.5).

**Average number of scenes, by genre**



If we bring together the data on the number of pages and scenes, we can calculate the average length of a scene by genre. Action scripts have the shortest scenes (an average of 0.87 pages) while Comedies have the longest (average of 1.06 pages per scene).

**Average number of pages per scene, by genre**

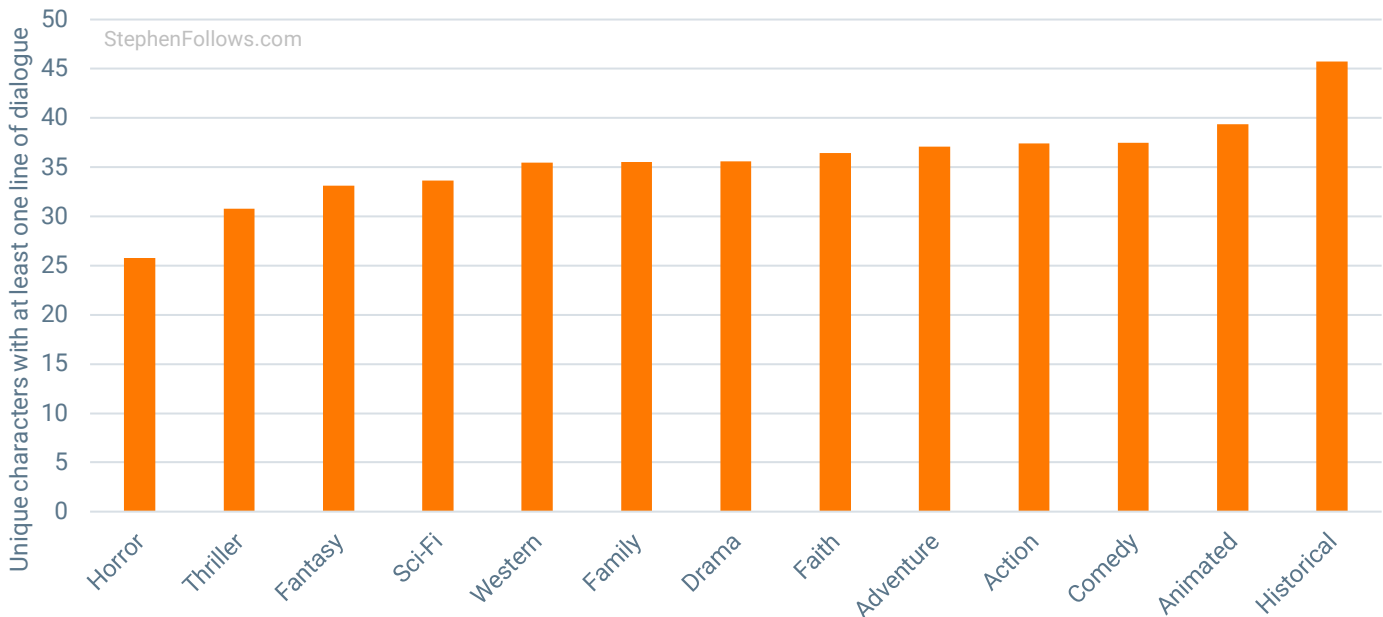


# Speaking Characters

Our analysis allows us to look at the number of unique characters who speak in each script, from our principal hero/heroine right through to background characters with single perfunctory lines.

Historical scripts have the greatest number of speaking characters (an average of 45.7) and Horror scripts have the fewest (25.8). Sadly, we were unable to track how many of those characters were still alive by the final page.

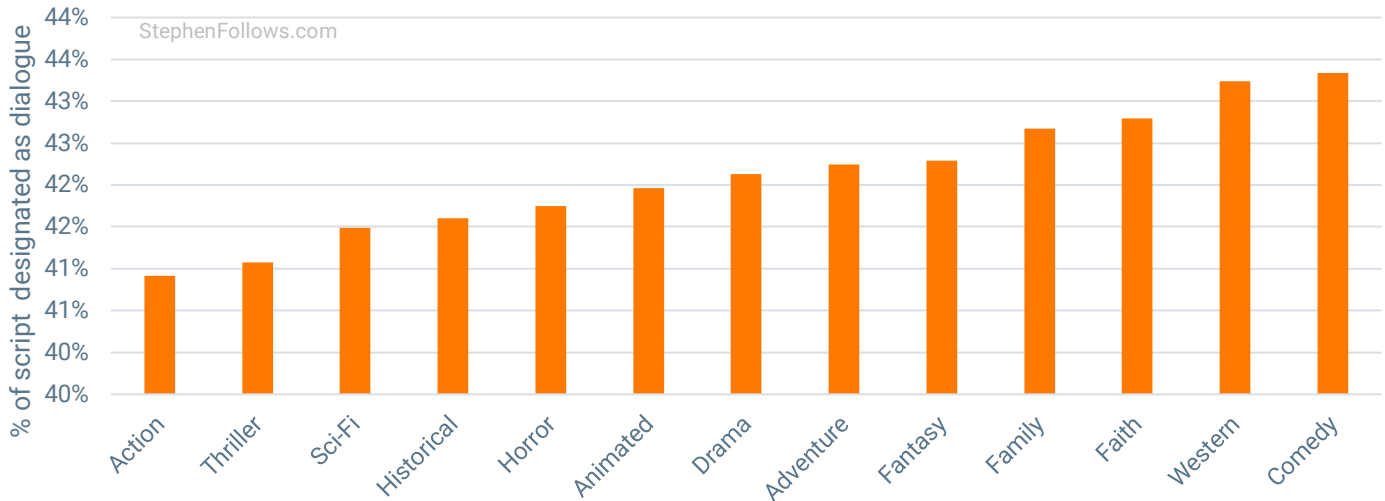
**Average number of speaking characters per scene, by genre**



# Dialogue

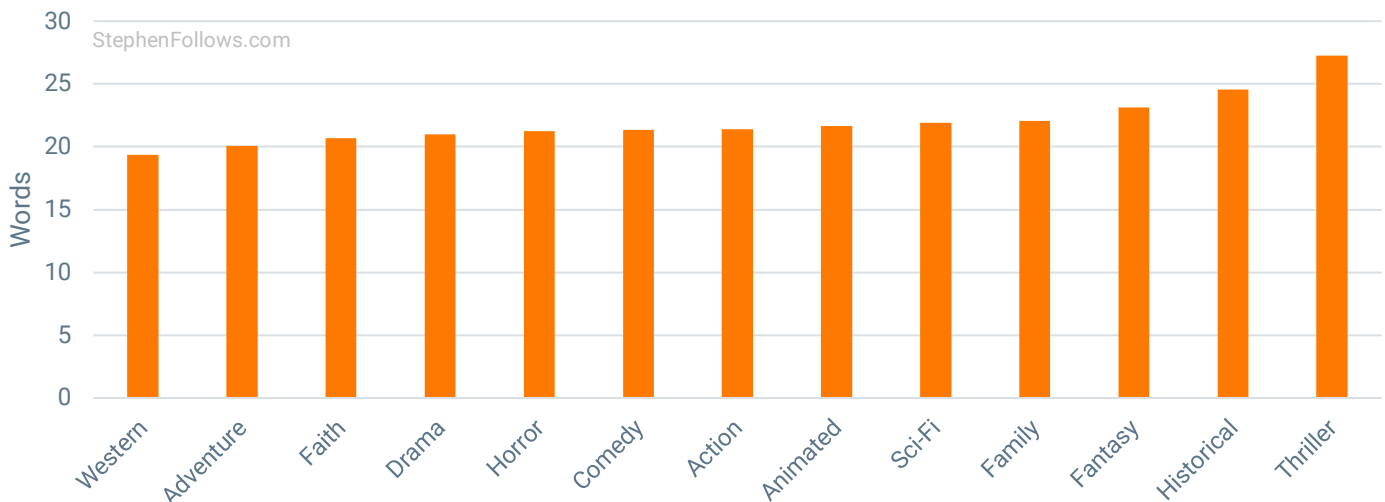
Speaking of speaking characters, let's look at the speeches they are speaking. By breaking each script down to its core components, we are able to track what percentage is made up of dialogue<sup>14</sup>.

**Dialogue as a percentage of the whole script, by genre**



The average line of dialogue is 21.7 words long, with Thrillers being the most verbose (an average of 27.2 words in each line of dialogue) and Westerns being the briefest (19.4 words).

**Average length of dialogue, by genre**



Because this genre is among the least dialogue heavy but also has the longest average length of dialogue, Thrillers feature a great deal more monologuing than other genres.

<sup>14</sup> To calculate the dialogue percentages, we broke each script down into its component 'blocks'. A 'block' is a group of text of the same designation (such as scene headings, action, character names or dialogue) irrespective of the number of sentences within each block.

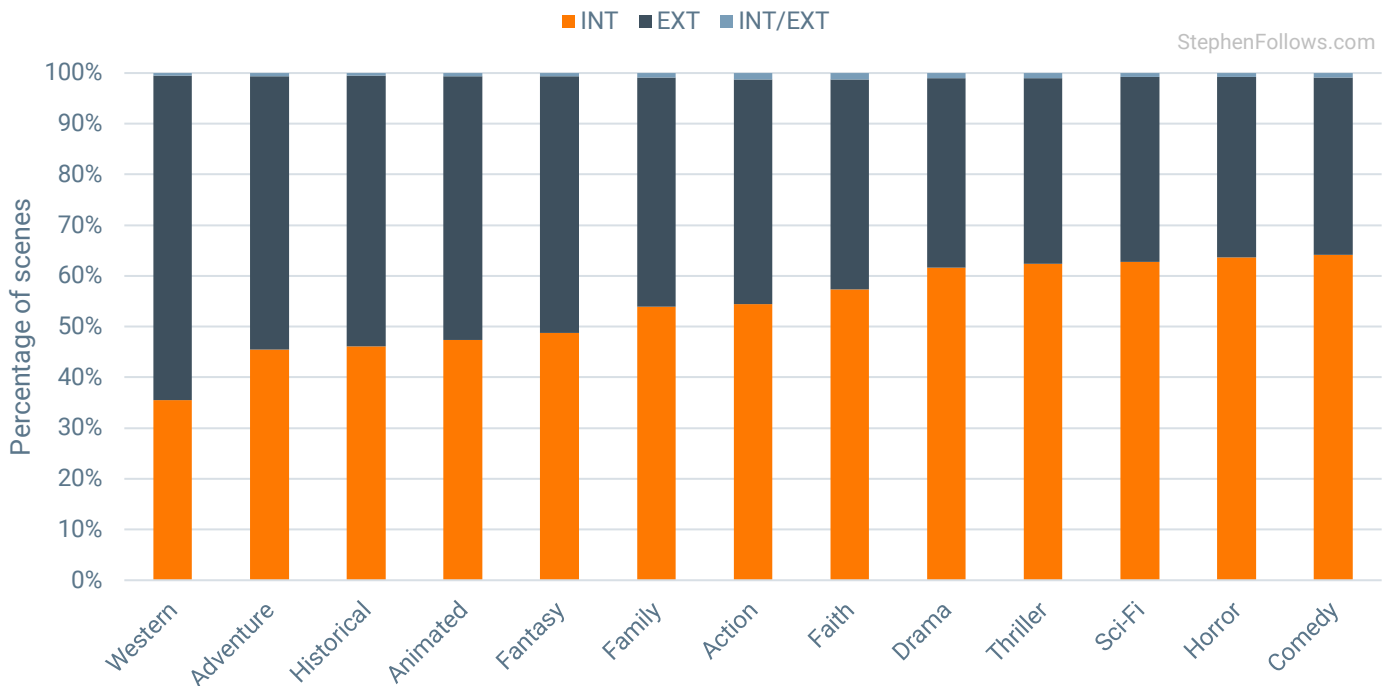
# Locations

Each scene heading starts with an indication as to whether the scene takes place inside (aka “INT” for interior), outside (“EXT” for exterior) or a hybrid (“INT/EXT”).

Across all scripts, 60.2% of scenes are interiors, 38.9% are exteriors and 0.9% are hybrid locations.

Westerns are mostly set outside, with 64.4% of their scenes taking place in exterior locations. At the opposite end of the scale we see 65.2% of Comedy scenes taking place indoors.

## Scene settings, by genre

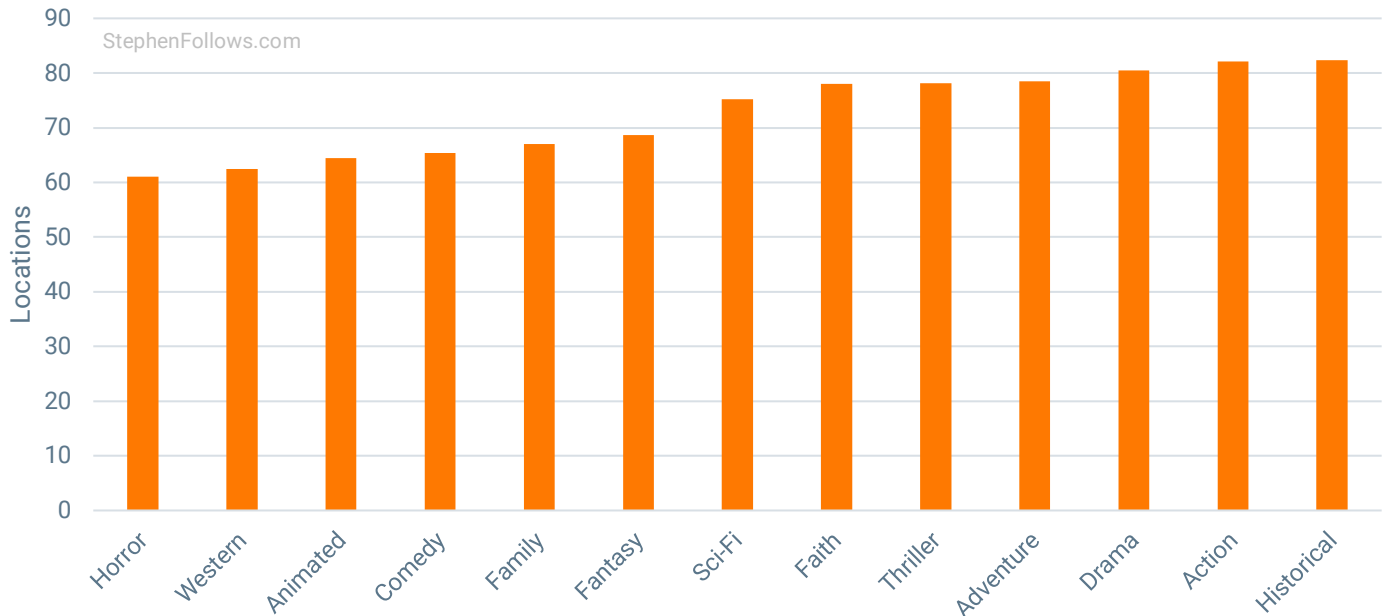


A fact that must make most producers wince is that the average location only appears in 1.5 scenes.

## JUDGING SCREENPLAYS BY THEIR COVERAGE

The average script takes place in 72.2 locations, with Action and Historical scripts covering the greatest number (an average of 82.3 and 82.1). Horror scripts have the fewest (61.1 per script).

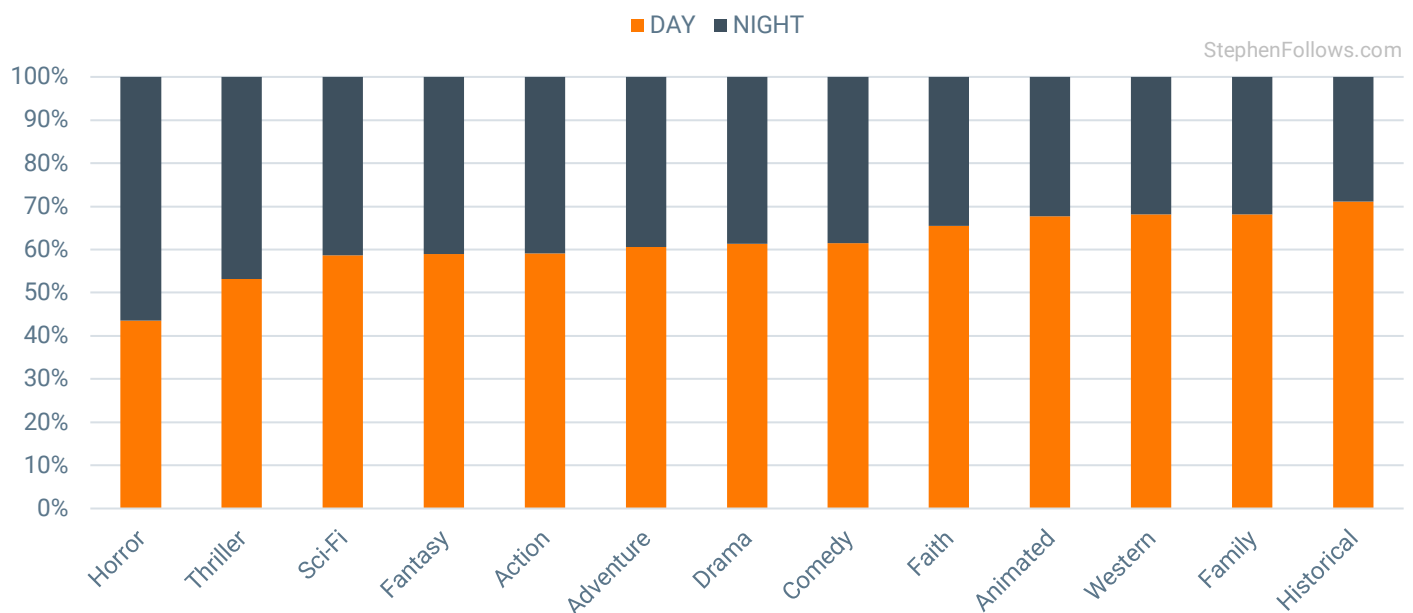
### Average number of locations, by genre



58.3% of scenes take place during the day and 41.7% take place at night.

Perhaps unsurprisingly, Horror scripts are much more likely to be set at night (56.5% of scenes) whereas Historical scripts are the most nyctophobic, with only 28.9% taking place at night.

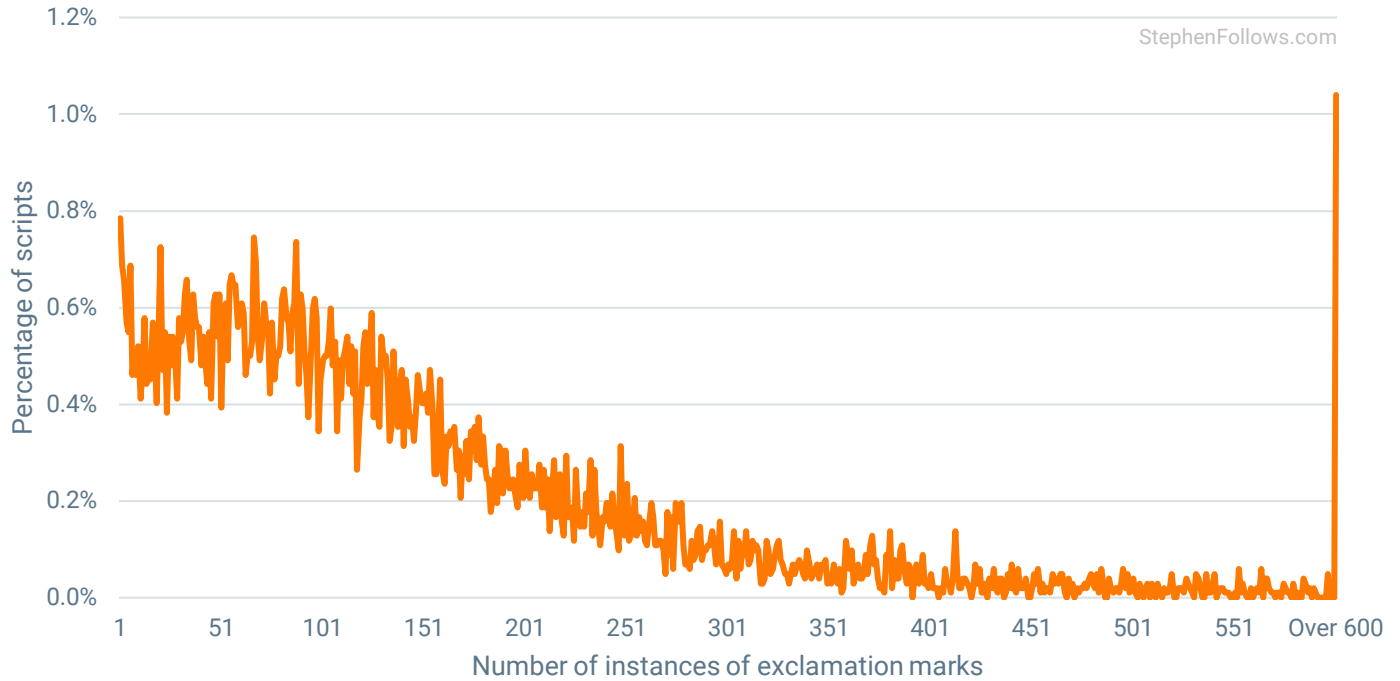
### Days scenes vs Night scenes, by genre



# Punctuation

The average script has 142.1 exclamation points, or an average of 1.3 per page. Five scripts feature over 1,000 exclamation points, with the highest containing 1,576!! That averages to just over 12.4 exclamation points per page.

## Number of exclamation marks used in scripts in our dataset



Question marks are used more sparingly. The average script contains just 9.6 question marks, or one every eleven pages. The most egregious use of question marks was far milder than we saw with exclamation marks, with the worst offender featuring only 156 question marks – just over one per page.

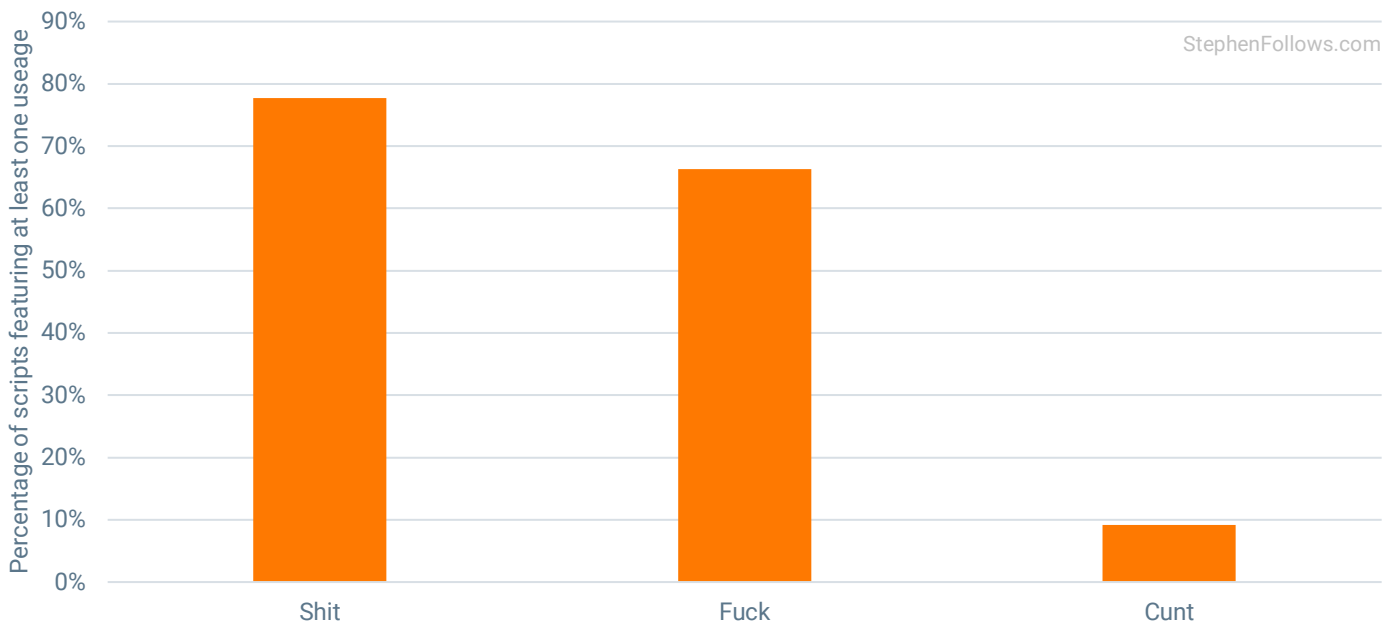
# Swearing

## Warning:

This section features repeated uncensored uses of s\*\*, f\*\*k and c\*\*. If you would rather not read such words, we suggest you skip to the next subsection.

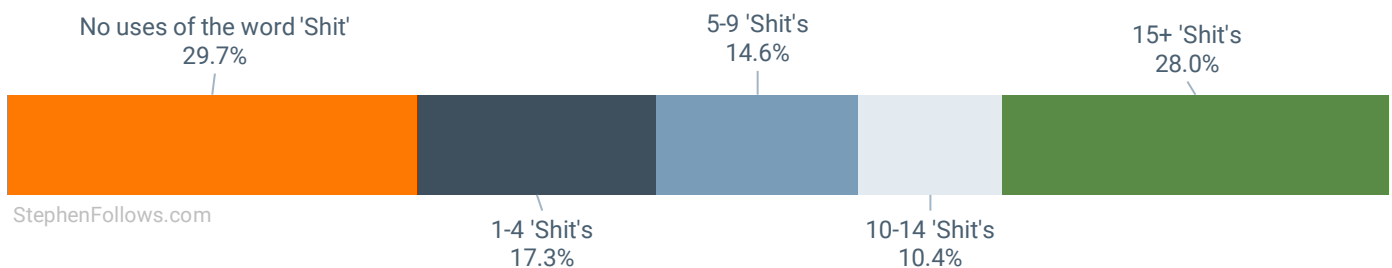
We tracked the usage of three key swear words – 'shit', 'fuck' and 'cunt'.

## Percentage of scripts which feature certain swear words

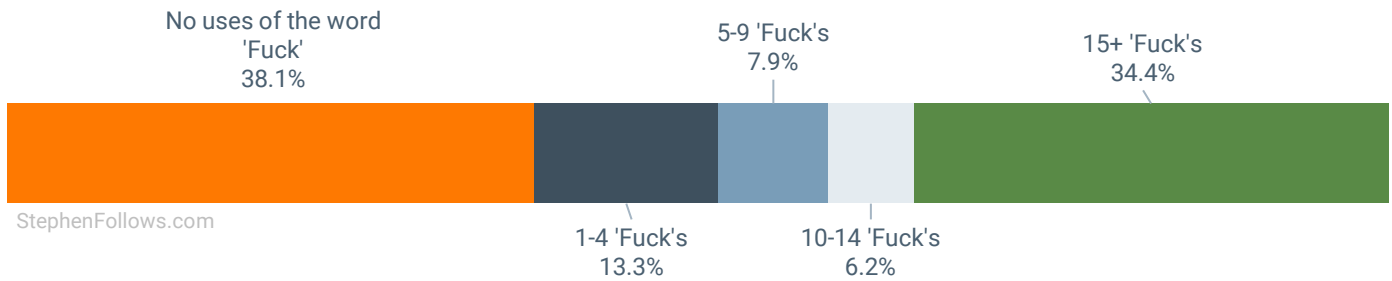


Although more scripts feature one 'shit' than those with one 'fuck', when a 'fuck' does appear it tends to be used more frequently than 'shit'. Across all our scripts, 'shit' is used an average of 13.2 times, 'fuck' 23.9 times and 'cunt' 2.1 times.

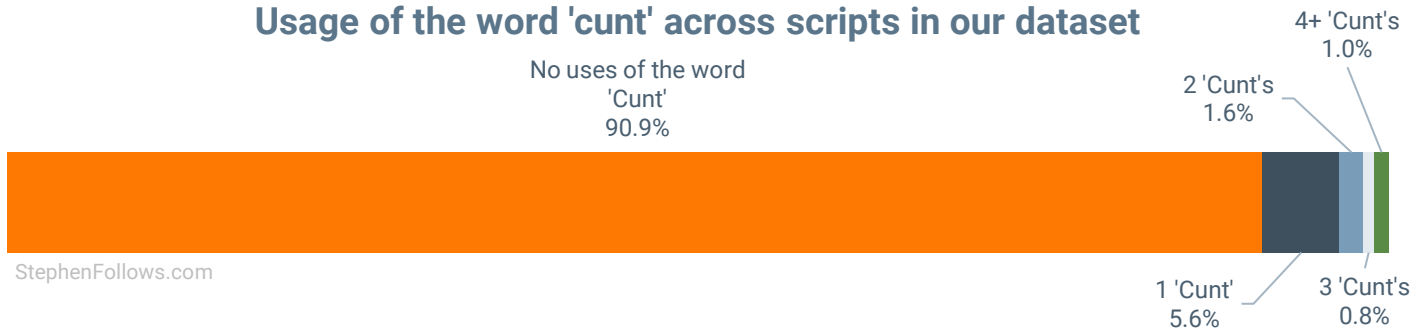
## Usage of the word 'shit' across scripts in our dataset



## Usage of the word 'fuck' across scripts in our dataset

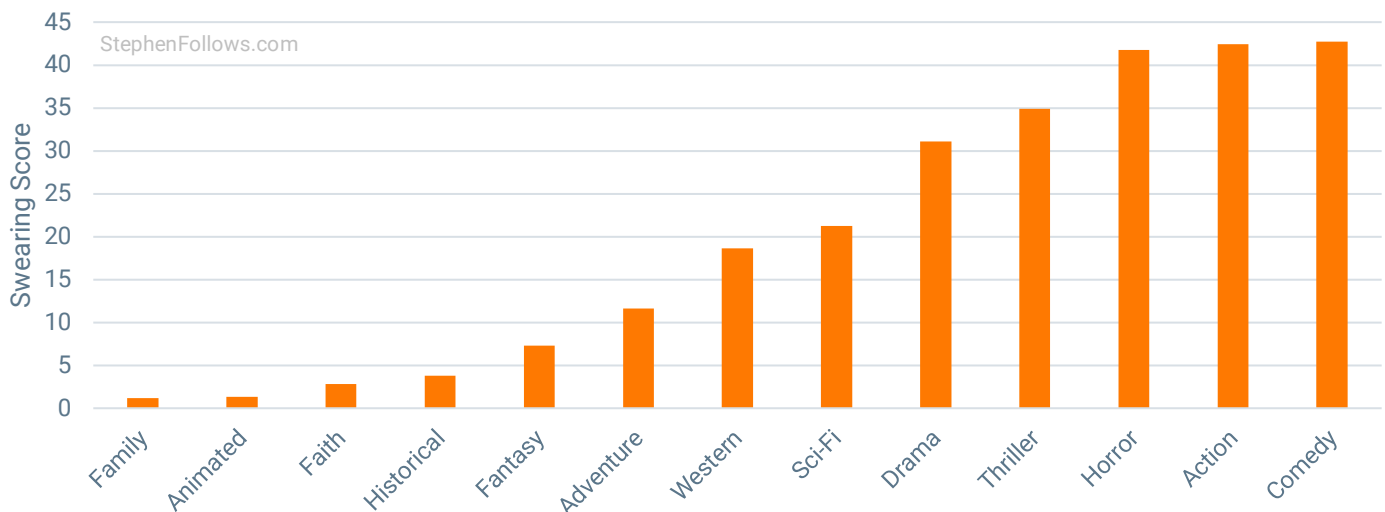


## Usage of the word 'cunt' across scripts in our dataset



Unsurprisingly, the swear words were not spread equally across all scripts. We developed a swearing score<sup>15</sup>, based on the frequency of the three swear words we tracked. This gave 1.00 for each use of 'shit', 1.17 for 'fuck' and 8.51 for 'cunt'. Comedies are the sweariest, beating Action and Horror scripts by a tiny margin (Comedy scores 42.8, Action scores 42.5 and Horror scores 41.8). The genres featuring the lowest levels of swearing are Family (1.2), Animated (1.3) and Faith-based scripts (2.8).

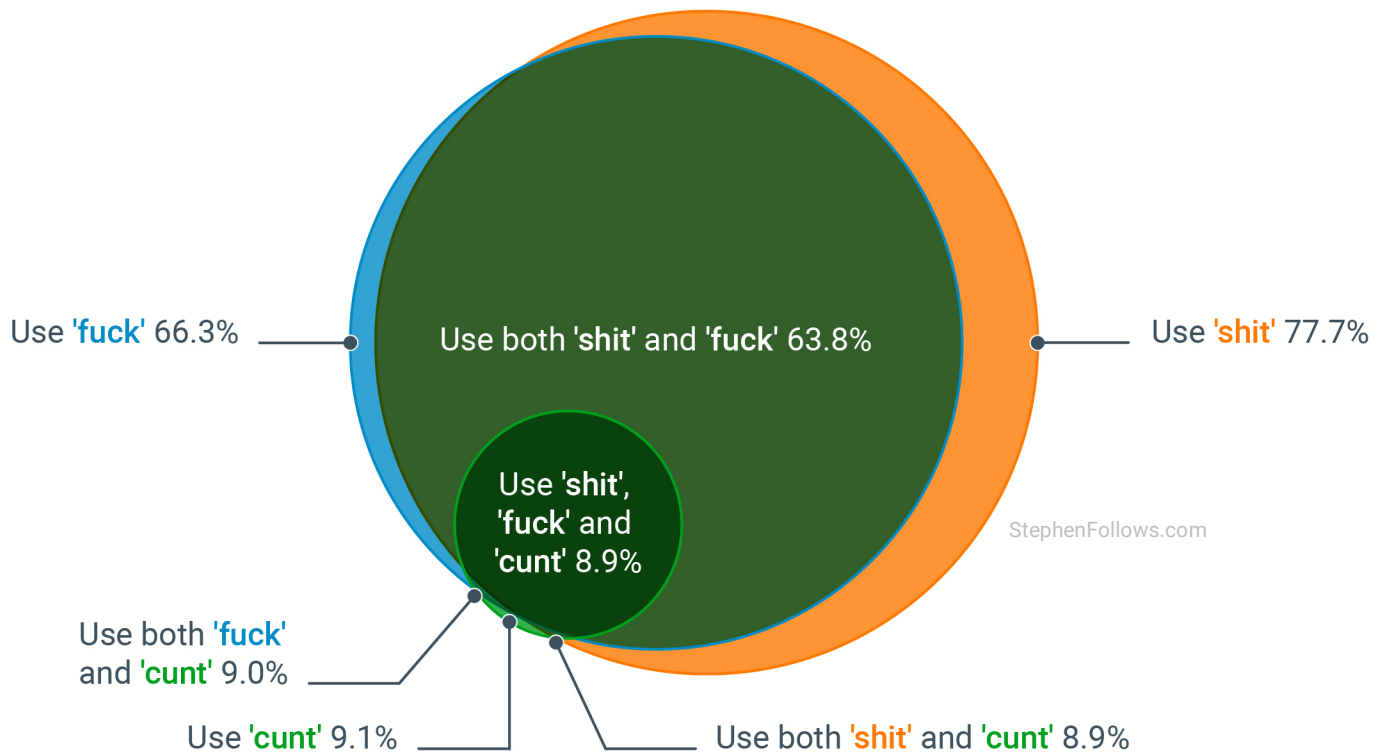
## Levels of swearing in scripts in our dataset, by genre



<sup>15</sup> Our weighting system calculates a single 'Swearing Score' for each script. We tracked the number of scripts which feature at least one mention of the three key swear words (shit, fuck and cunt). 71.8% of scripts feature at least one 'shit', 61.3% feature at least one 'fuck' and 8.4% have at least one 'cunt'. This gave us a weighting of 1.00 for 'shit', 1.17 for 'fuck' and 8.51 for 'cunt'.

## JUDGING SCREENPLAYS BY THEIR COVERAGE

Only sixteen scripts used 'cunt' without also using either 'shit' or 'fuck' at least once.



# Age of Characters

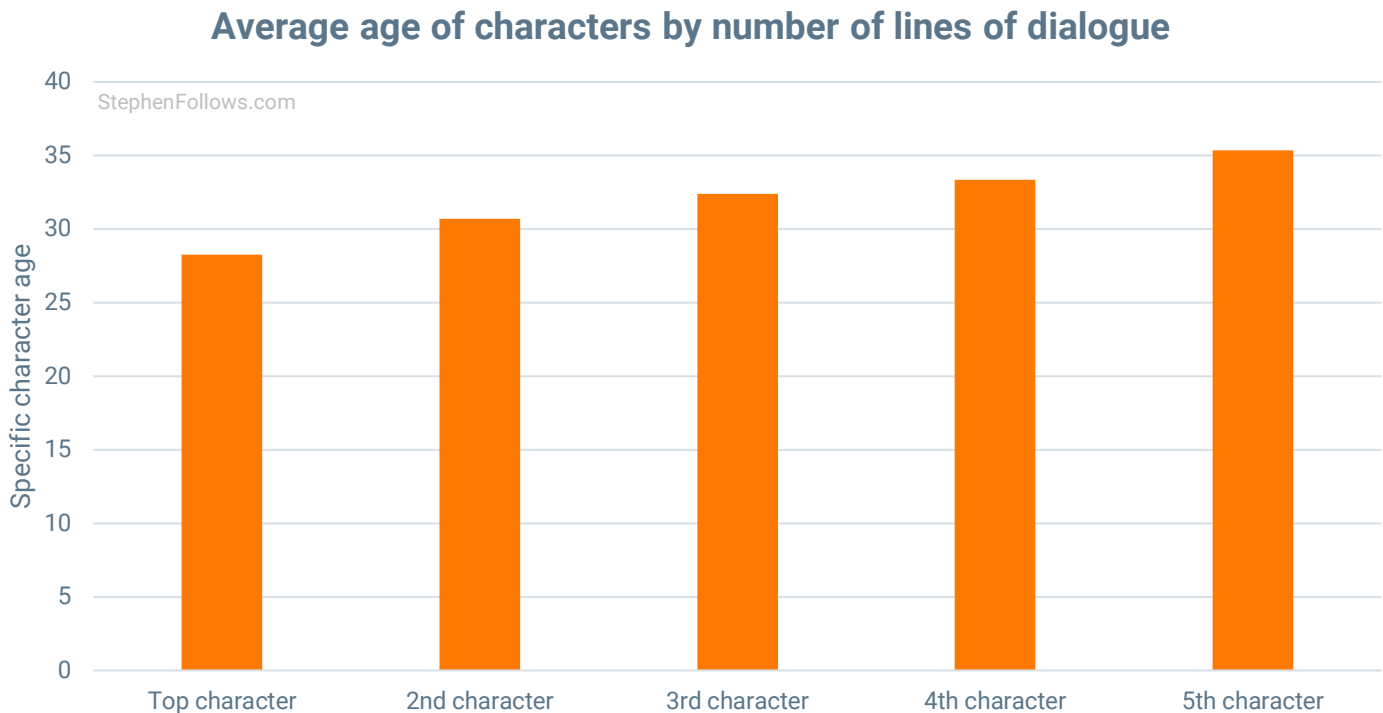
There are two ways a writer can express a character's age:

- **Specific**, such as "28 years old";
- **Non-specific**, such as "20s", "mid-20", "20-odd", "20 something", etc.

## Specific Ages

The average specific age of the top five characters across all our scripts is 31.8 years old<sup>16</sup>.

The character who speaks most often is typically a little younger (average age: 28.3) and as we move down to characters who speak less frequently the age increases slightly. The average age of the fifth most frequently-speaking character is 35.4.

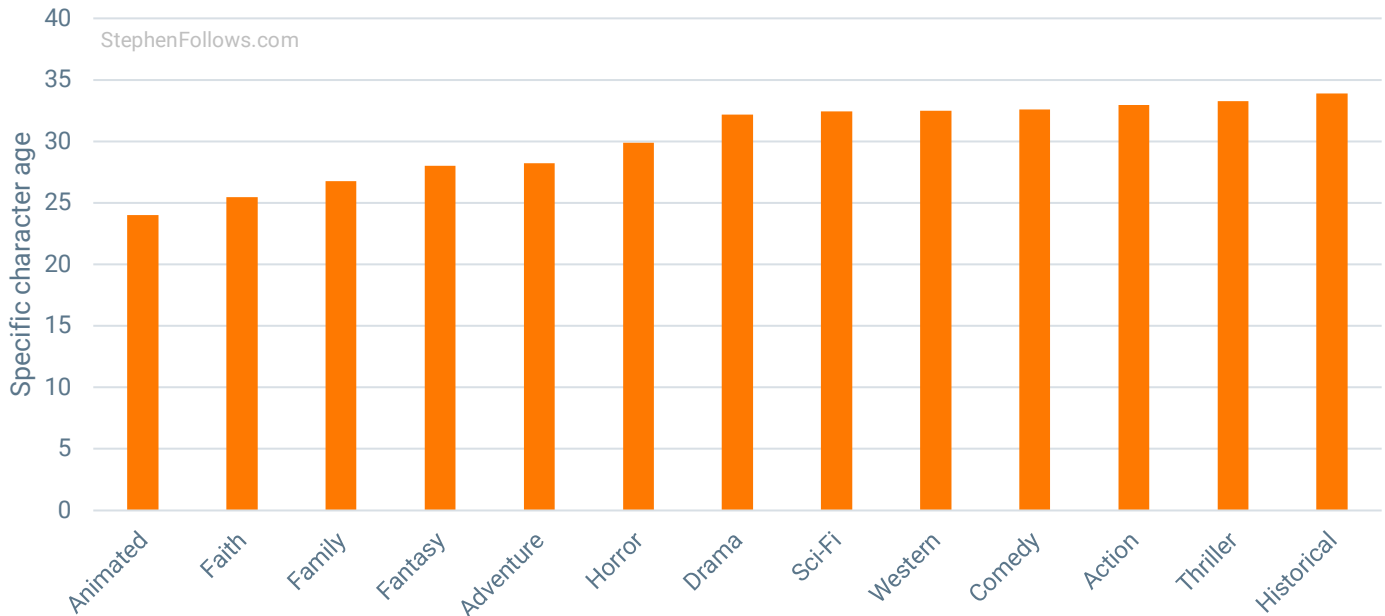


<sup>16</sup> This research on ages relates to the five characters who speak most frequently in scripts where the character ages are provided. If there was an age range we took the lowest of the ages. We excluded any characters with ages over 110 as these were all magical, mythical or fantastical characters. Ages were also excluded when they related to non-humans. Quite why the specific age of a dog is needed is unclear but this did not stop a number of writers defining their age.

## JUDGING SCREENPLAYS BY THEIR COVERAGE

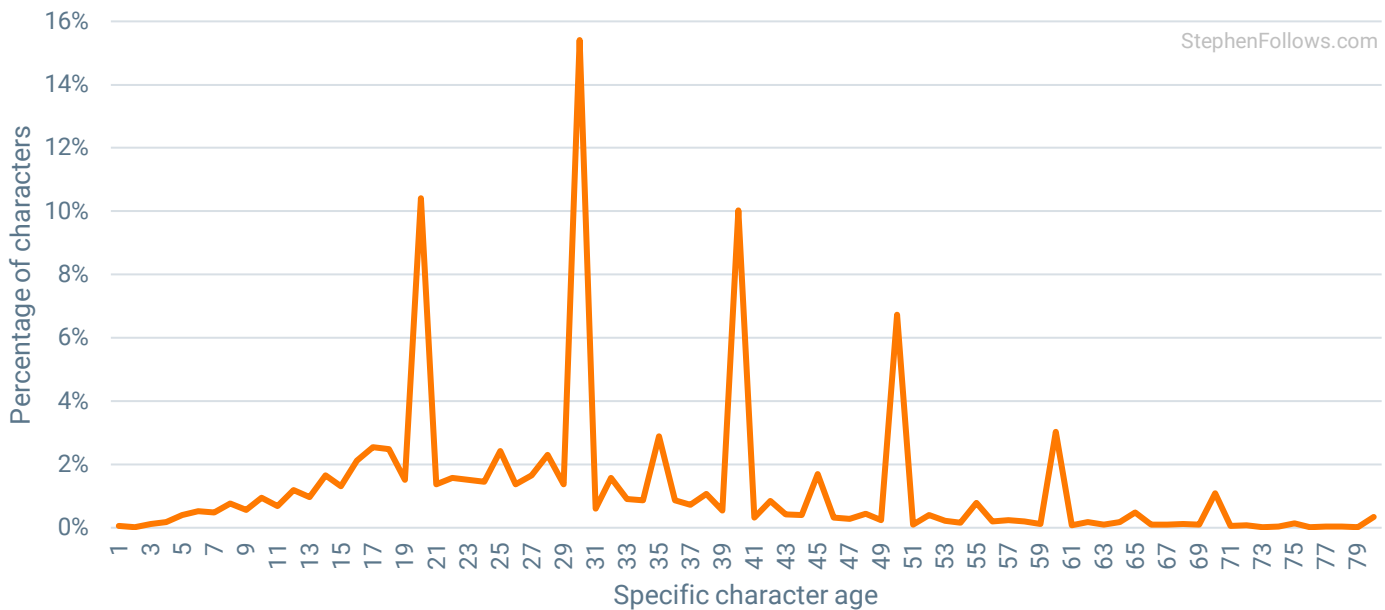
Historical and Thriller scripts have the oldest main characters, at an average of 33.9 and 33.3 years old, respectively. Characters in Animations and Faith-based scripts have the youngest average age (24.0 and 25.5 respectively)

### Average age of characters, by genre



The median age is 30 years old, with 15.4% of all characters being listed as exactly 30.

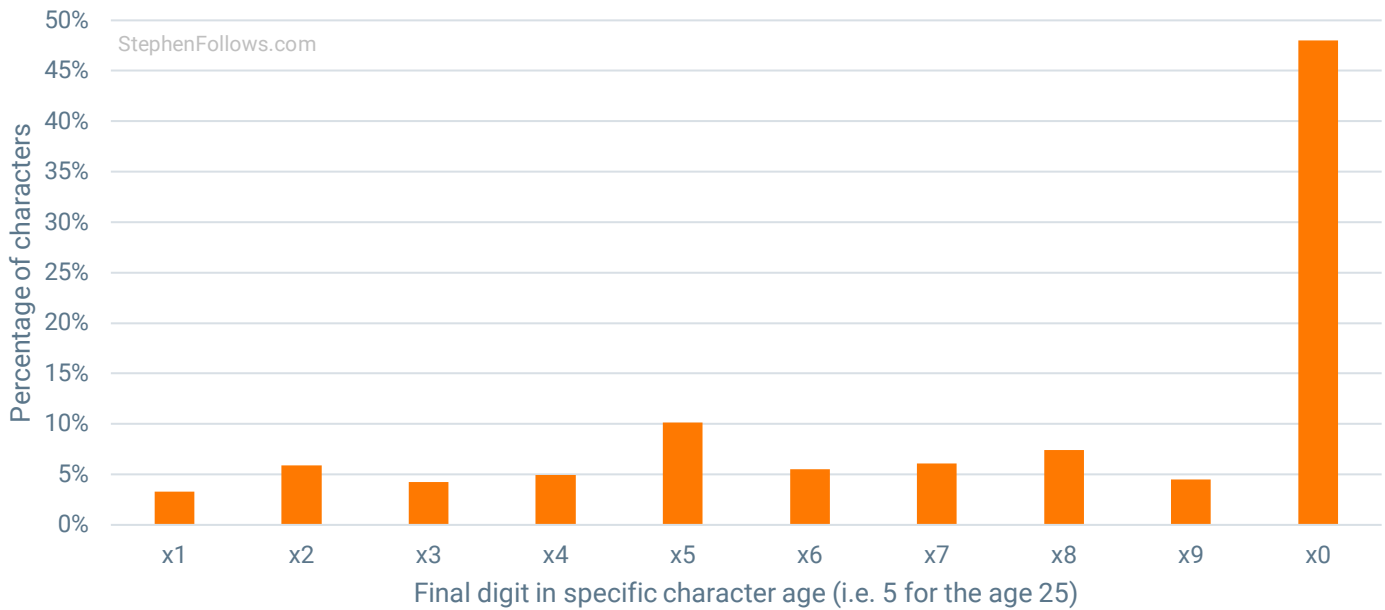
### Age of characters, when a specific age is provided



## JUDGING SCREENPLAYS BY THEIR COVERAGE

Almost half of all specific ages end with a zero (such as 20 years old). The second most frequent are those ending in a five (10.1% of ages) followed by those ending in an eight (7.4%).

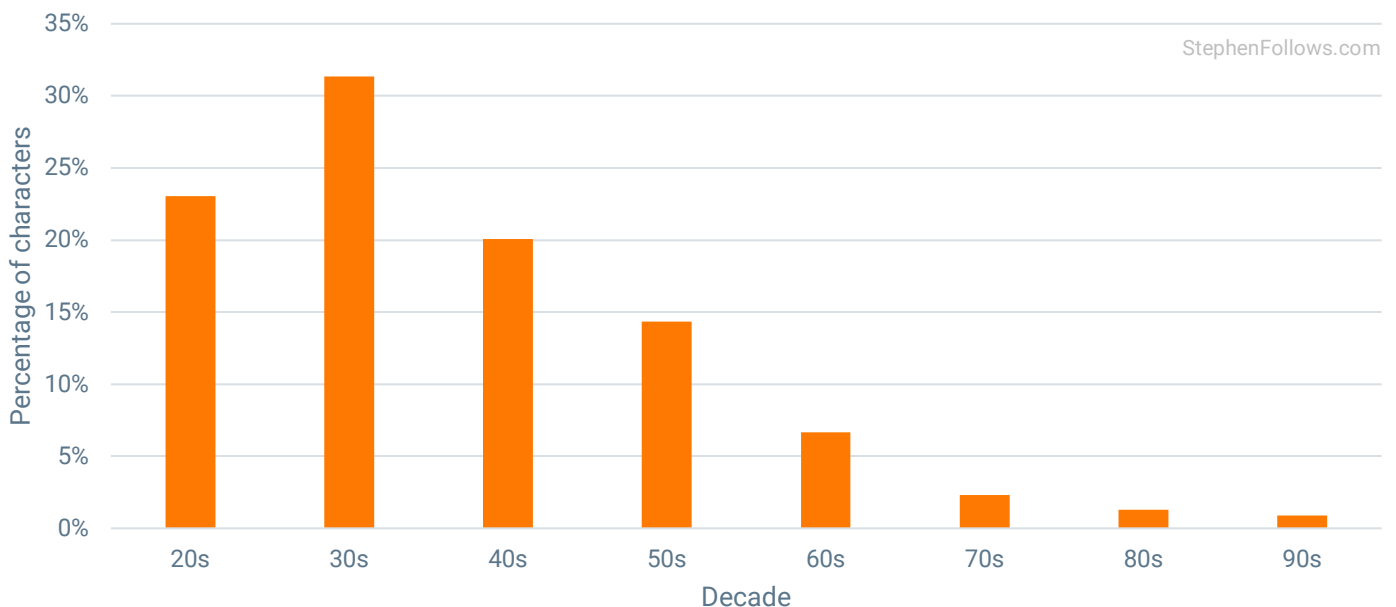
### Final digit in the age of characters with specific ages



## Non-specific Ages

Characters with non-specific ages follow a similar pattern to those with a defined age. 31.4% are in their 30s, 23.0% are in their 20s and 45.6% are either younger than 20 or older than 39.

### Decade of the age of characters with non-specific ages

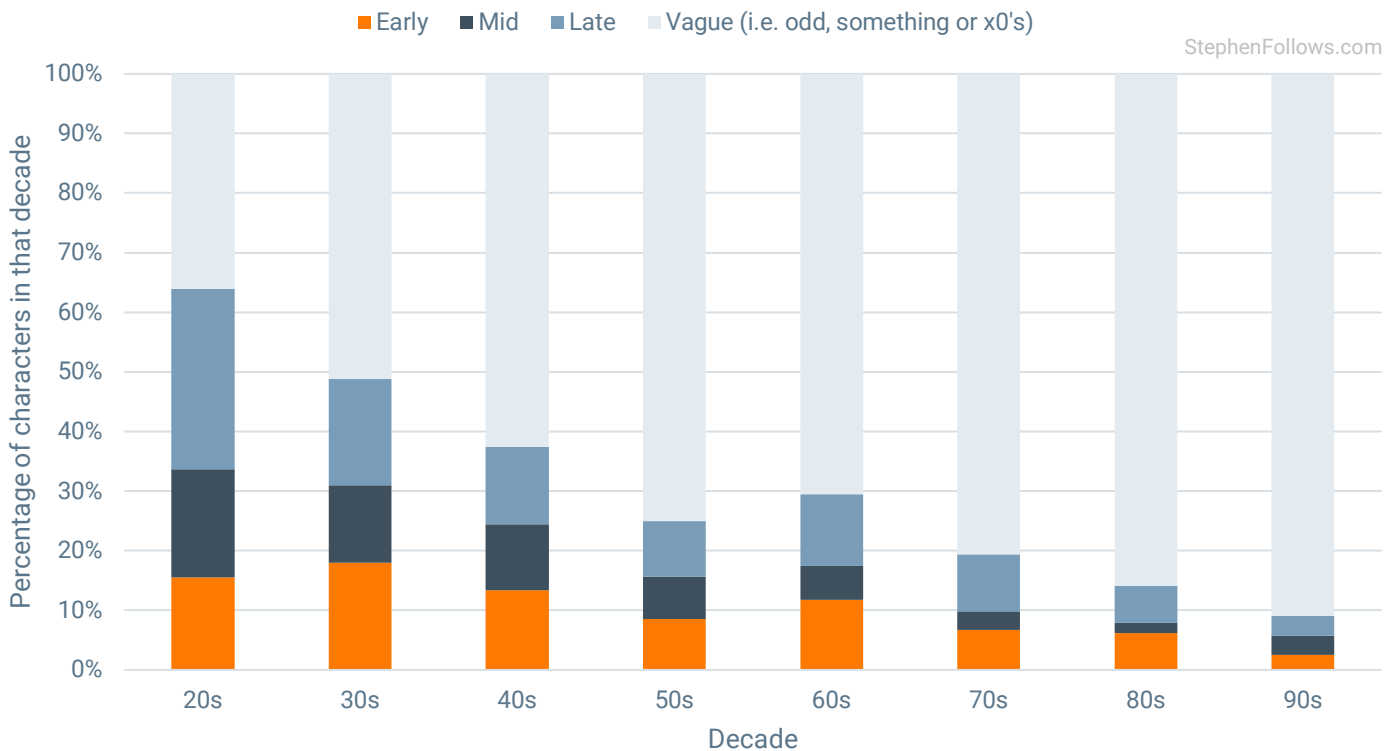


## JUDGING SCREENPLAYS BY THEIR COVERAGE

Not only do writers favour characters in their twenties and thirties, they also provide more detail as to roughly where in the decade they are when the script takes place.

When characters are given a non-specific age in their 20s, 15.6% are said to be in their “early 20s”, 18.1% in their “mid 20s”, 30.3% in their “late 20s” and 36.0% are given a vague age such as “20s”, “20-odd” or “20-something”. Compare this to characters in their 50s, where vague descriptions account for 70.5% of all non-specific ages.

### Age description of characters when a finite number is not given

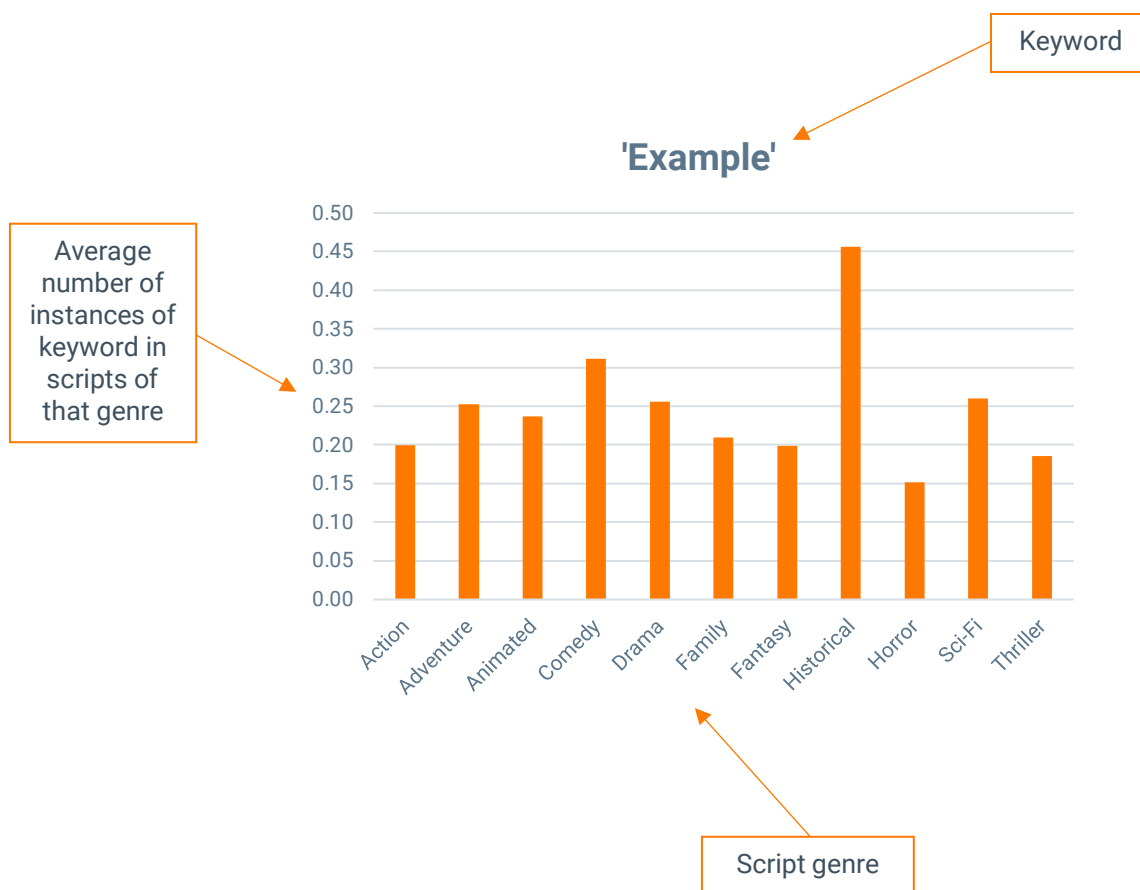


# Genre-skewed Words

The dataset offers us a unique opportunity to look at how the choice of words differs between genres. We analysed the scripts to pull out key words<sup>17</sup> for each major genre.

We are presenting this data in two different ways:

- **Word Clouds showing the most heavily-skewed words for each genre<sup>18</sup>**, with the size of the word reflecting their frequency of use within that genre (i.e. the bigger the word, the more it's used in the genre);
- **Bar charts showing the average number of usages of a certain word within each genre.**



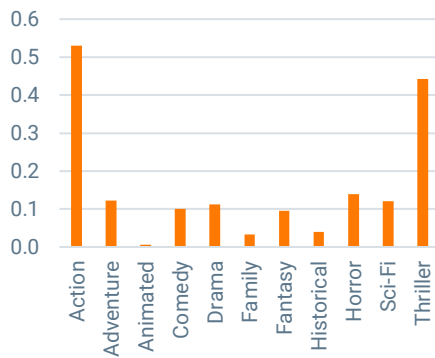
<sup>17</sup> Key words were determined using Term Frequency-Inverse Document Frequency. TF-IDF divides how frequent a word is in a document by how frequently the word is used across the text to highlight relevant words.

<sup>18</sup> The words in the Word Clouds were selected using the following criteria:

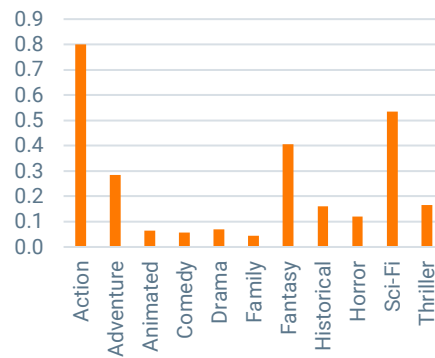
- They appear in at least 10% of scripts within the relevant genre (removing script-specific words such as character names);
- They don't appear in over 5,000 scripts across the dataset (thereby excluding the most common words);
- They are the most heavily-skewed towards the relevant genre. Skew is measured by the percentage of uses in a particular genre divided by all usages. For example, if a word has been used 100 times across all scripts and 50 of those were in Action scripts, then this has a 50% skew towards Action.

[illegible]

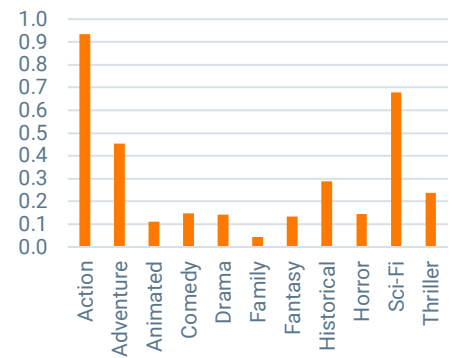
## 'Unmarked'



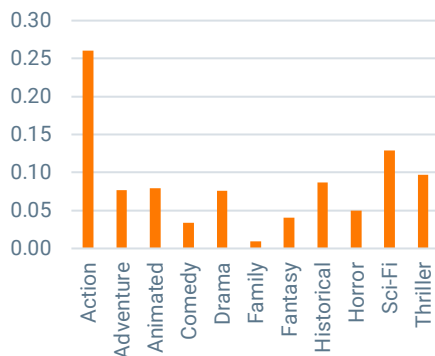
## 'Armored'



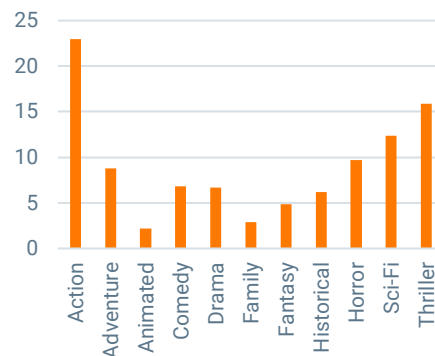
## 'Explosive'



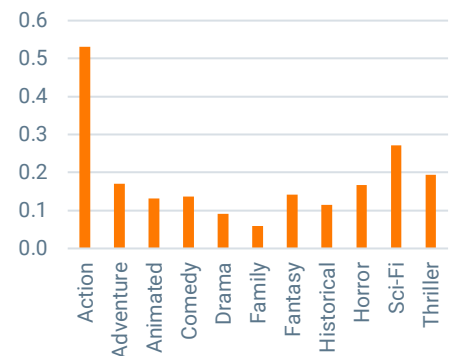
## 'Forces'



## 'Gun'

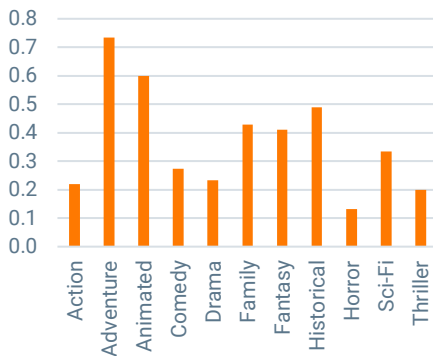


## 'Shoots'

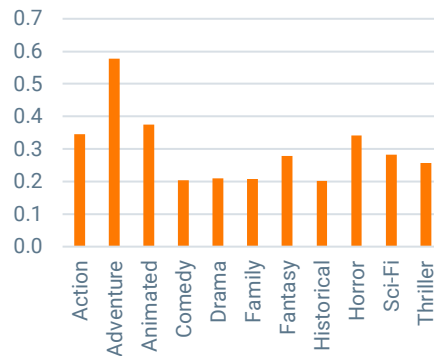


[illegible]

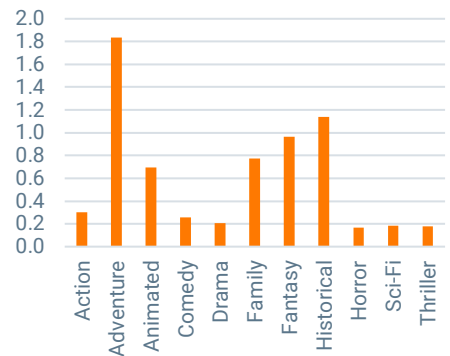
## 'Amazed'



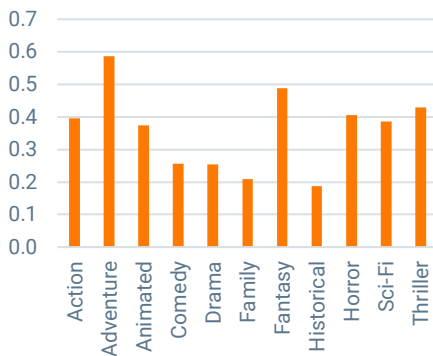
## 'Dazed'



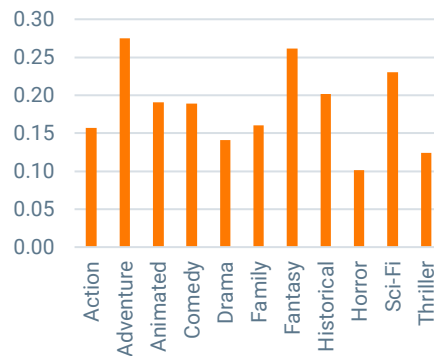
## 'Treasure'



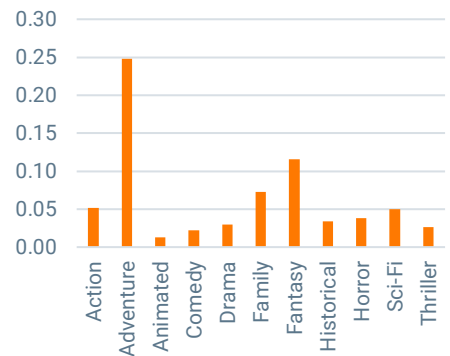
## 'Mysterious'



## 'Exotic'

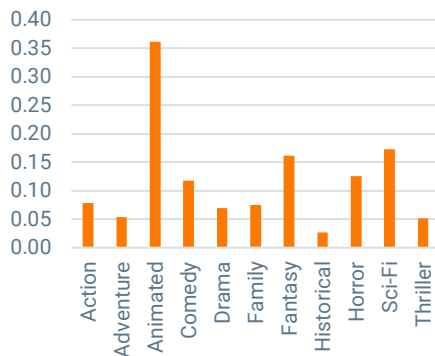


## 'Mystic'

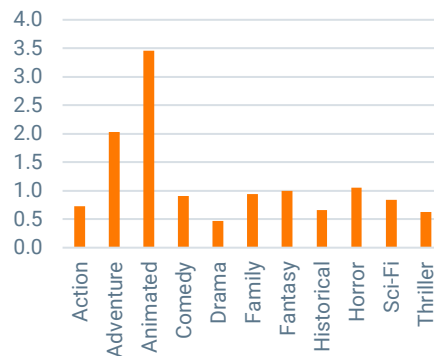


[illegible]

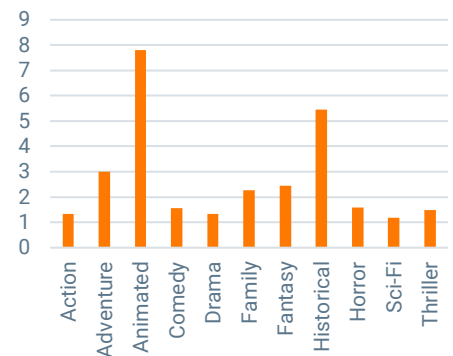
## 'Beetle'



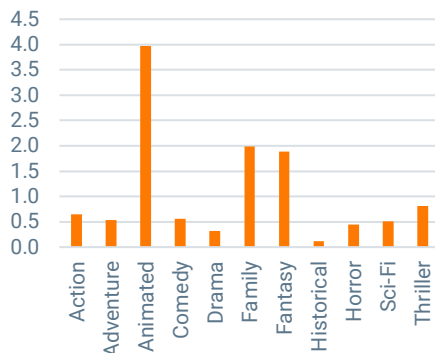
## 'Rat'



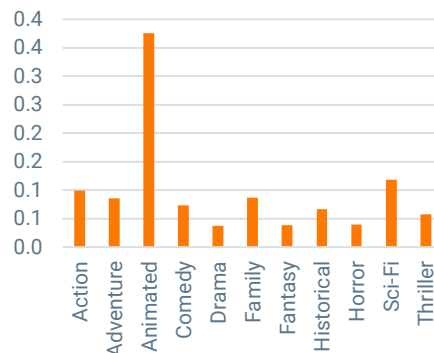
## 'Bird'



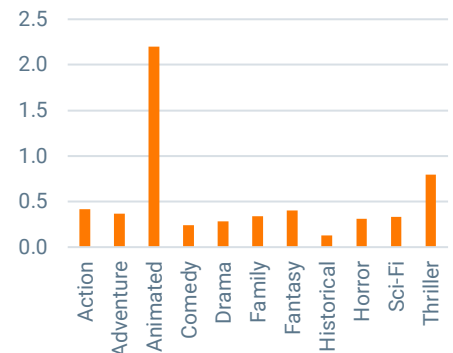
## 'Mouse'



## 'Leopard'



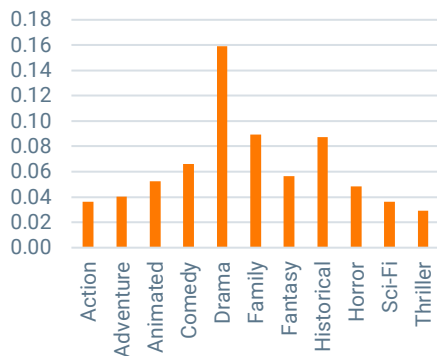
## 'Shark'



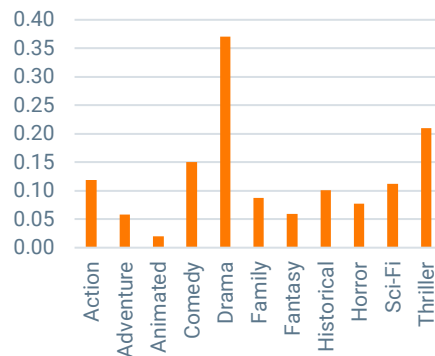


[illegible]

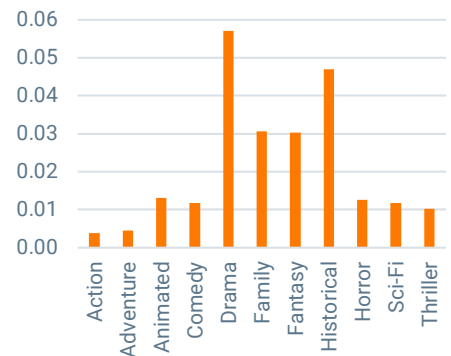
## 'Hugs'



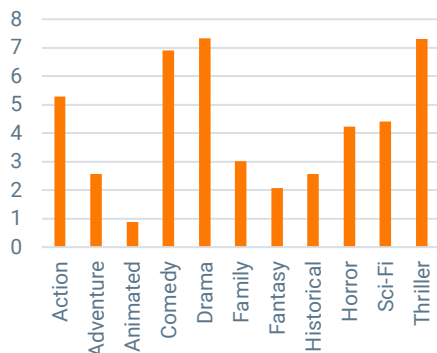
## 'Anonymous'



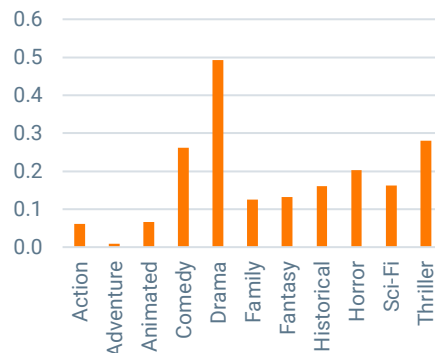
## 'Expressive'



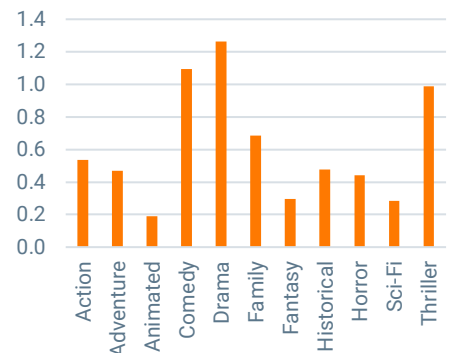
## 'Apartment'



## 'Therapist'



## 'Lawyer'



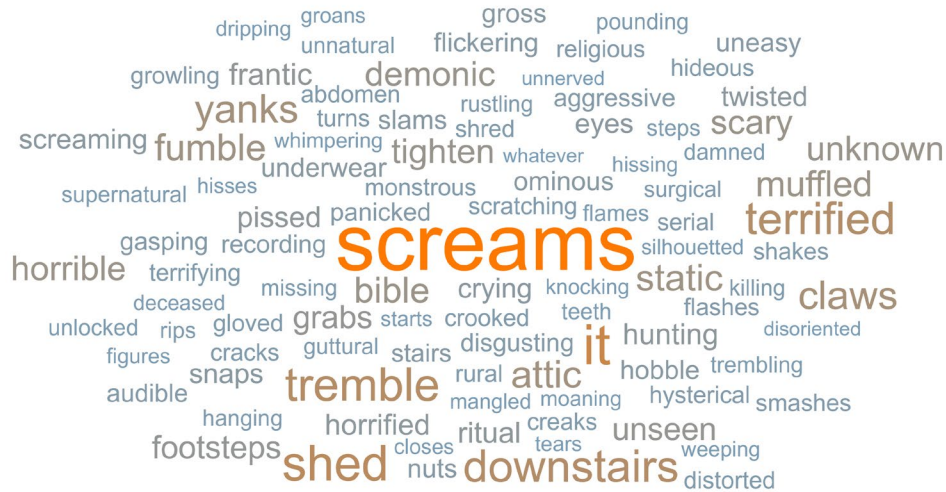






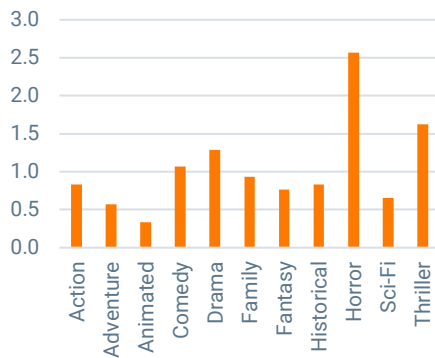
## Horror

Unsurprisingly, Horror scripts are dominated by words describing suffering, including screams, terrified, frantic, grabs and horrible.

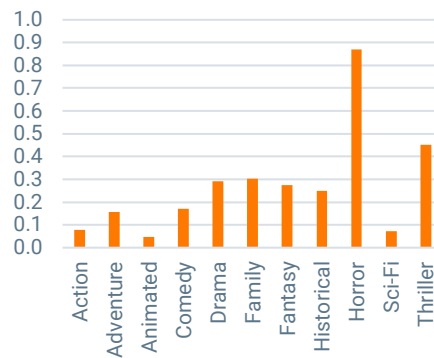


Amongst the most Horror-filled places are the upstairs of a house, an attic and the woods.

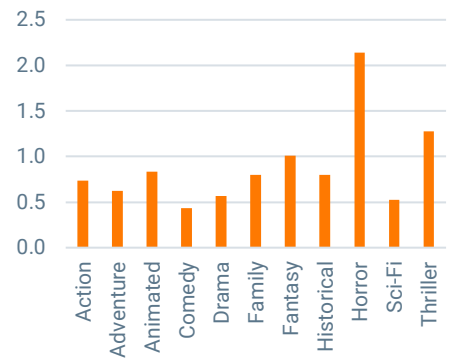
'Upstairs'



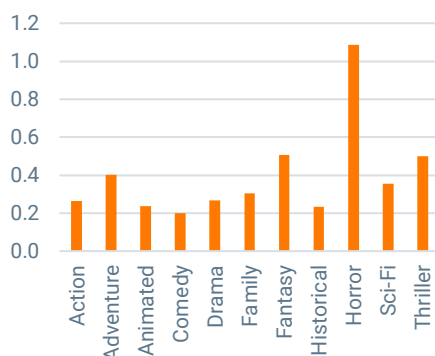
'Attic'



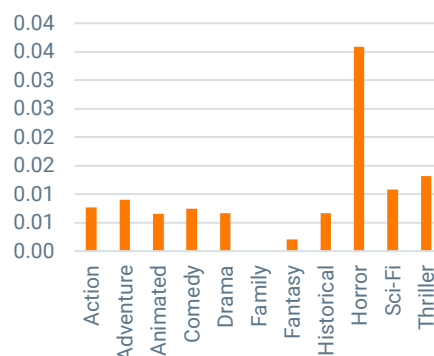
'Woods'



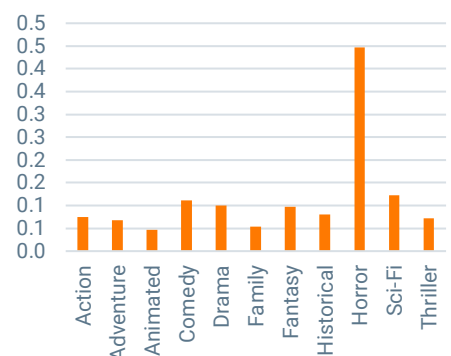
'Creak'



'Ritualistic'

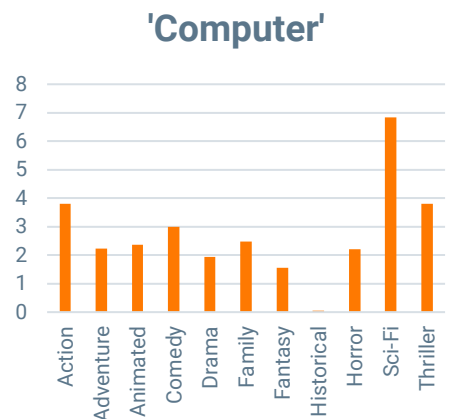
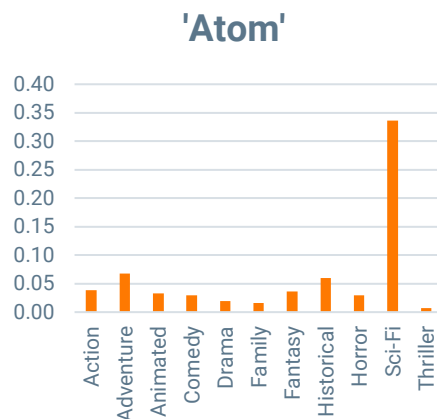
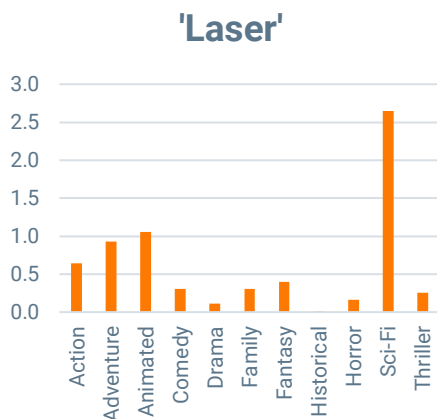
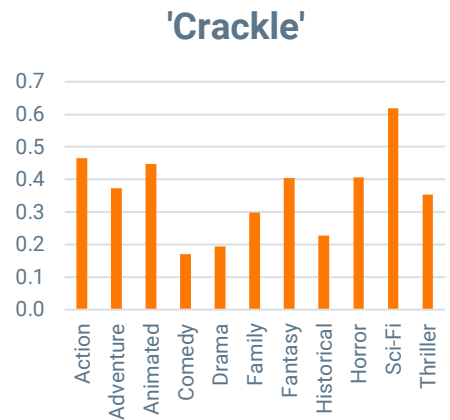
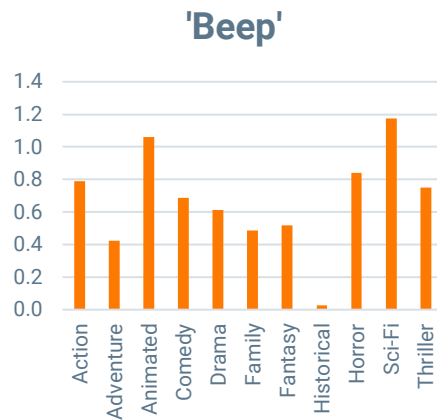


'Lunatic'



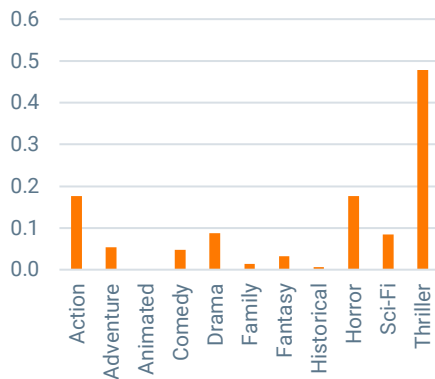
A word cloud of terms related to the movie 'The Matrix'. The words are arranged in a circular pattern, with 'The Matrix' at the top. Other prominent words include 'military', 'holographic', 'central', 'complex', 'warning', 'glowing', 'being', 'technician', 'landing', 'static', 'disable', 'magnetic', 'recording', 'visual', 'external', 'global', 'advanced', 'similar', 'effective', 'affirmative', 'organic', 'critical', 'incoming', 'programming', 'personnel', 'loading', 'operational', 'identical', 'cavernous', 'negative', 'incredible', 'technical', 'surgical', 'elevated', 'sophisticated', 'shimmering', 'tactical', 'firing', 'capable', 'intact', 'flashing', 'graphic', 'biological', 'potential', 'shaped', 'detailed', 'armored', 'unstable', 'shattered', 'docking', 'scanning', 'scientific', 'testing', 'robust', 'electric', 'cylindrical', 'genetic', 'tracking', 'lens', 'humming', 'industrial', 'physics', 'manifest', 'barren', 'active', 'electrical', 'internal', 'physical', 'blasts', 'diagnostic', 'artificial', 'connected', 'international', 'alarms', 'individual', 'beeps', 'functional', 'running', 'electronic', 'nuclear', 'virtual', 'setting', 'deafening', 'tremendous', 'futuristic', 'quarters', 'corporate', 'significant', 'damaged', 'vertical', 'lighting', 'explosive', 'mechanical'. The words are in various colors (blue, green, yellow, orange, red, purple, pink, brown, grey) and sizes, creating a vibrant and dynamic visual effect.

Genre	Percentage
Action	0.10
Adventure	0.11
Animated	0.11
Comedy	0.10
Drama	0.06
Family	0.12
Fantasy	0.08
Historical	0.03
Horror	0.12
Sci-Fi	0.27
Thriller	0.15

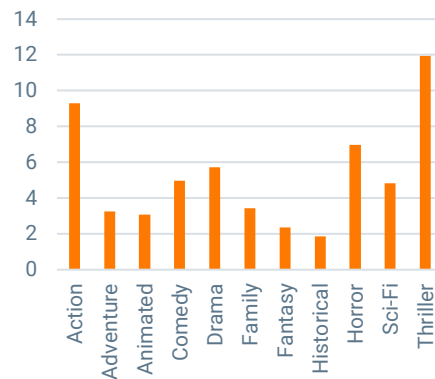


[illegible]

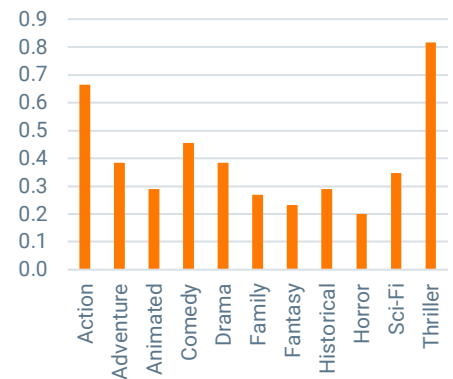
## 'Forensic'



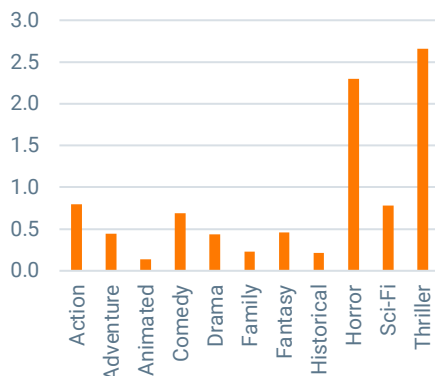
## 'Police'



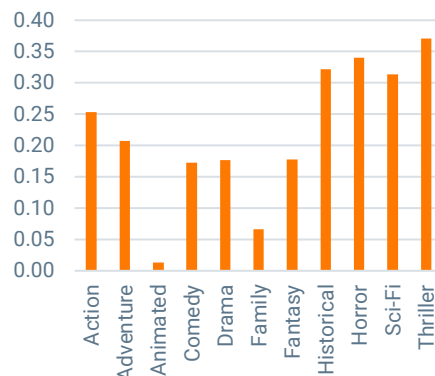
## 'Criminal'



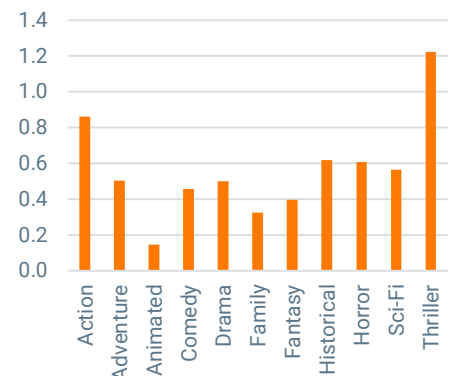
## 'Killer'



## 'Murderer'

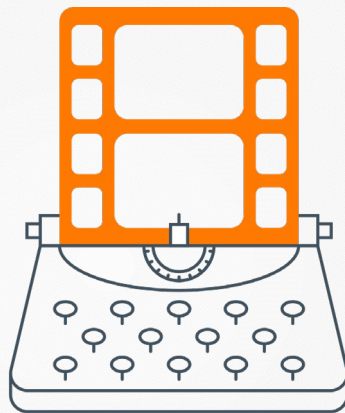


## 'Suspect'



---

# SECTION C: SCREENWRITERS AND THE ACT OF SCREENWRITING



*The dataset also allows us to understand some aspects of being a screenwriter.*

Gender .....	61
The Working Habits of Screenwriters .....	63
Screenwriting Software .....	65

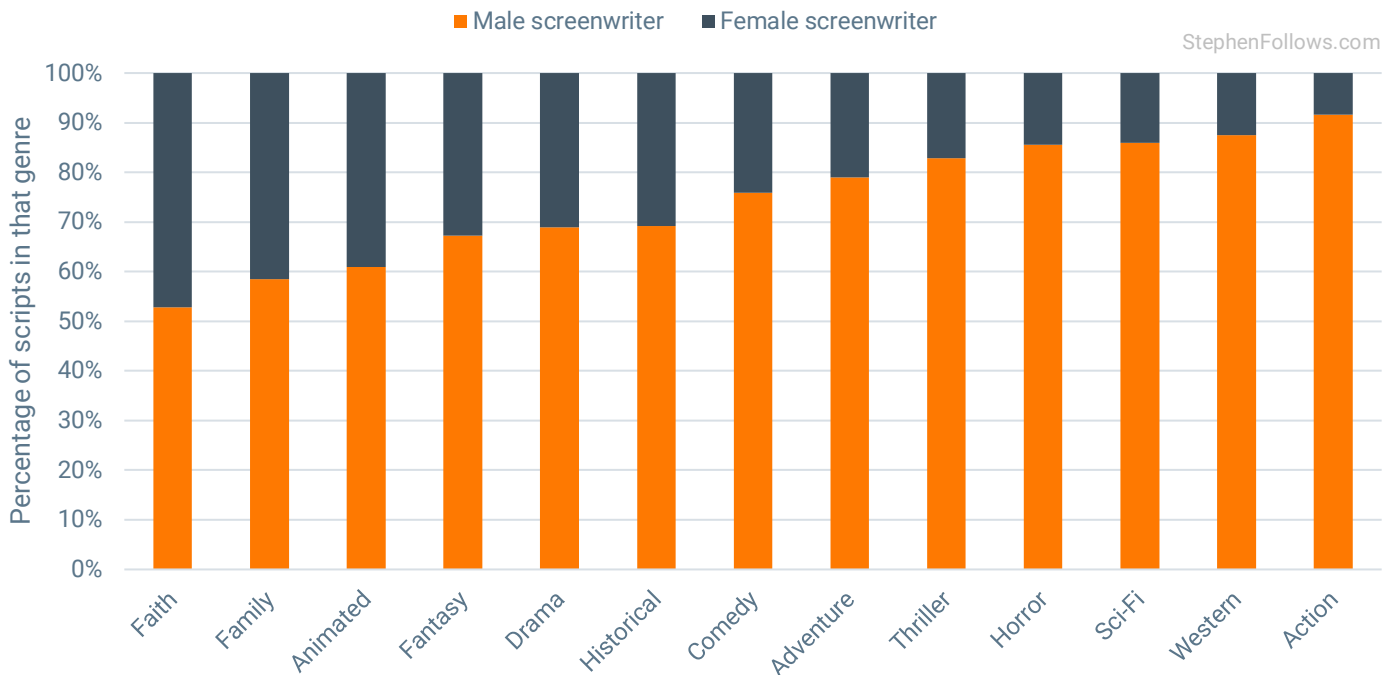
# Gender

Gender is a complicated (and sometimes contentious) issue within screenwriting. Years of under-representation of women in the film industry is looking ever more at odds with the changing face of society and gender “norms”.

Across our dataset, 23.7% of writers are women<sup>19</sup>.

The most male-dominated genres are Action (in which 8.4% of writers were women), Sci-Fi (14.1%) and Horror (14.5%). Women were best represented within Faith (47.2% female), Family scripts (41.5% female) and Animated (39.1%).

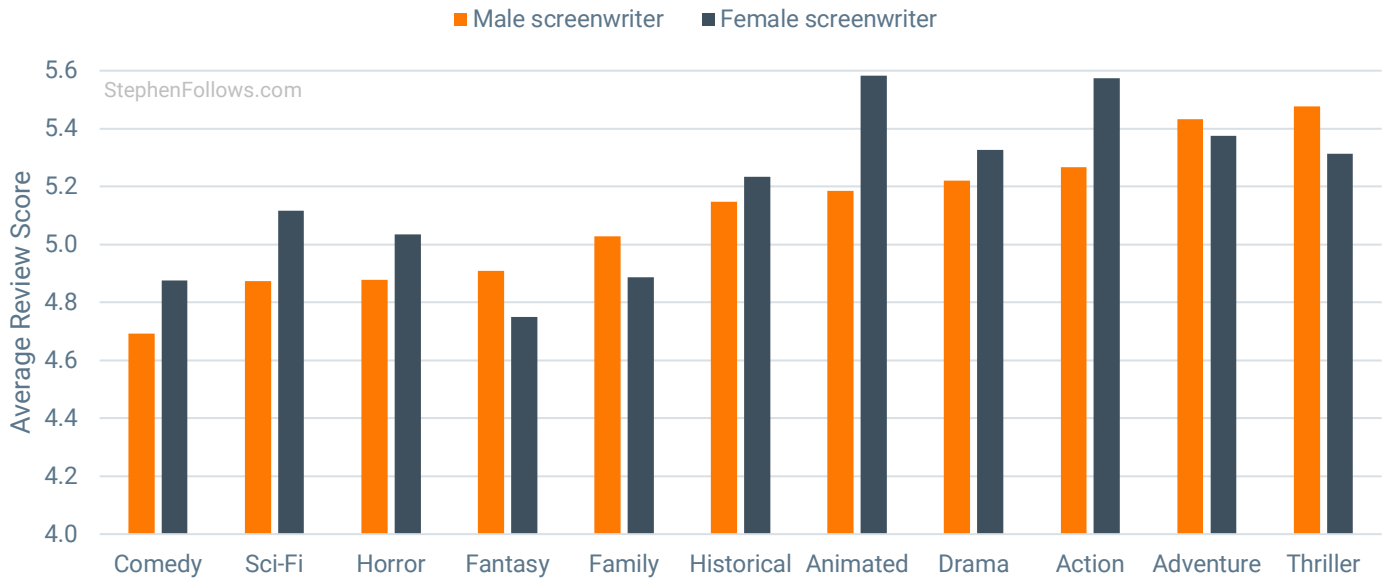
## Gender of screenwriters, by genre



Interestingly, when we look at the score given by readers, we see an advantage to writing in a genre dominated by another gender. Action is male-dominated but is also a genre in which female writers outperform their male counterparts by the second-largest margin. Likewise, Family films from men received higher ratings than those from women.

<sup>19</sup> We appreciate that gender is not always binary and that classifying gender as such is slightly reductive. Sadly, we do not have a way to account for gender fluidity in this dataset and feel that despite this limitation, the advantages of discussing gender outweigh the disadvantages of pursuing what appears at first glance to be such a binary approach. Gender of screenwriters was self-reported when the scripts were first submitted for coverage and/or to the competition.

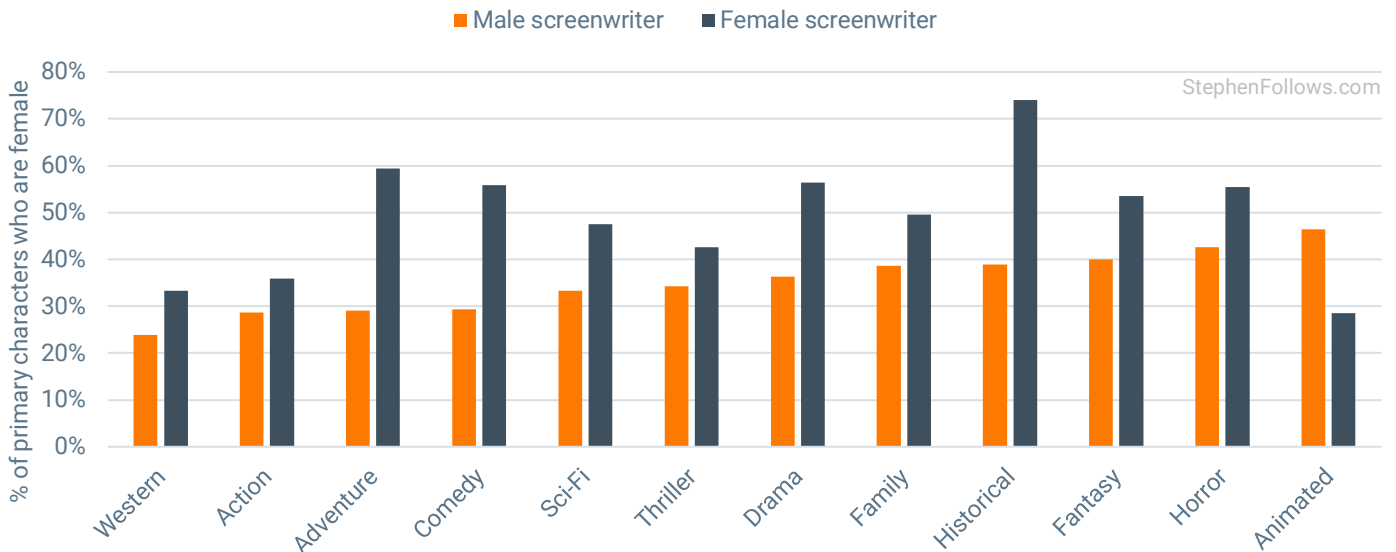
## Average Review Score by gender of the screenwriter and script genre



The data does not reveal why this is happening. It is plausible that the writers who have to overcome bigger barriers are more likely to be the hardest working and most tenacious – qualities which are correlated with success in screenwriting.

Women are much more likely to write scripts featuring female characters<sup>20</sup> in primary roles<sup>21</sup>. The starkest differences were with Historical, Adventure and Comedy scripts. Female characters accounted for 38.9 of primary characters in Historical scripts penned by men, whereas they made up 74.0% of character of such scripts written by women.

## Percentage of primary characters who are female by gender of the screenwriter and script genre

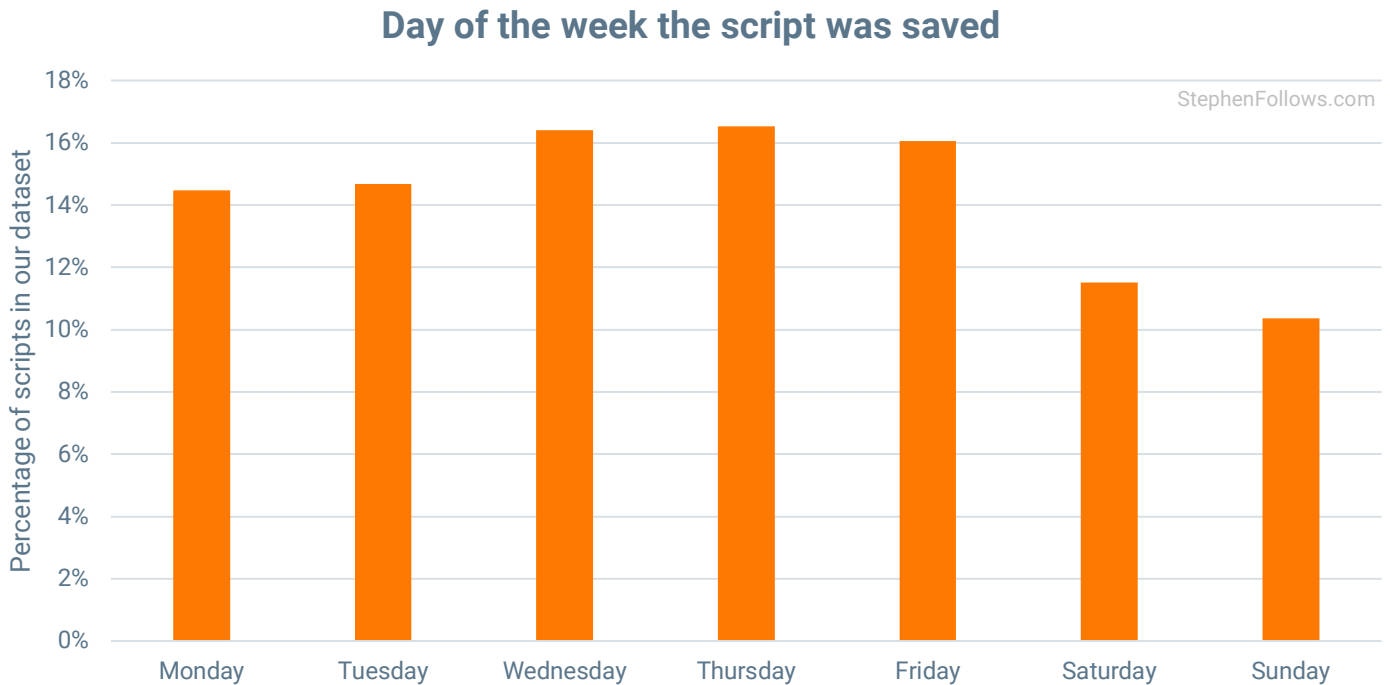


<sup>20</sup> The gender of characters was determined by comparing the first names against a database of 102,240 names and their typical gender. Unisex and unknown names were removed, leaving just those which are known to be typically male or female.

<sup>21</sup> Primary characters were determined via a character importance score, calculated thus: direction mentions/2 + dialogue mentions/2 + number of lines. Characters scoring over 0.08 were classed as primary, meaning that the average script has 2.2 primary characters.

# The Working Habits of Screenwriters

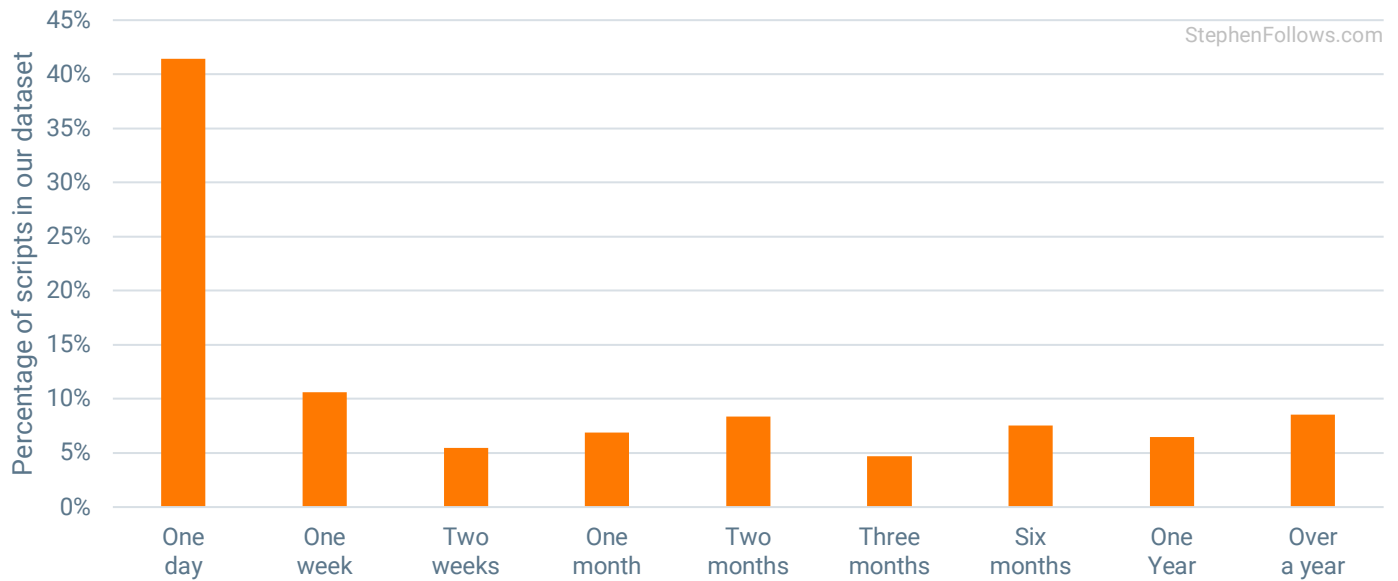
We were able to determine when the PDF was created by looking at the metadata of the scripts. Wednesdays and Thursdays appear to be the busiest days for screenwriters to export their work (and on the seventh day many of them rested).



There was no discernible difference in the quality of the scripts based on which day they were exported.

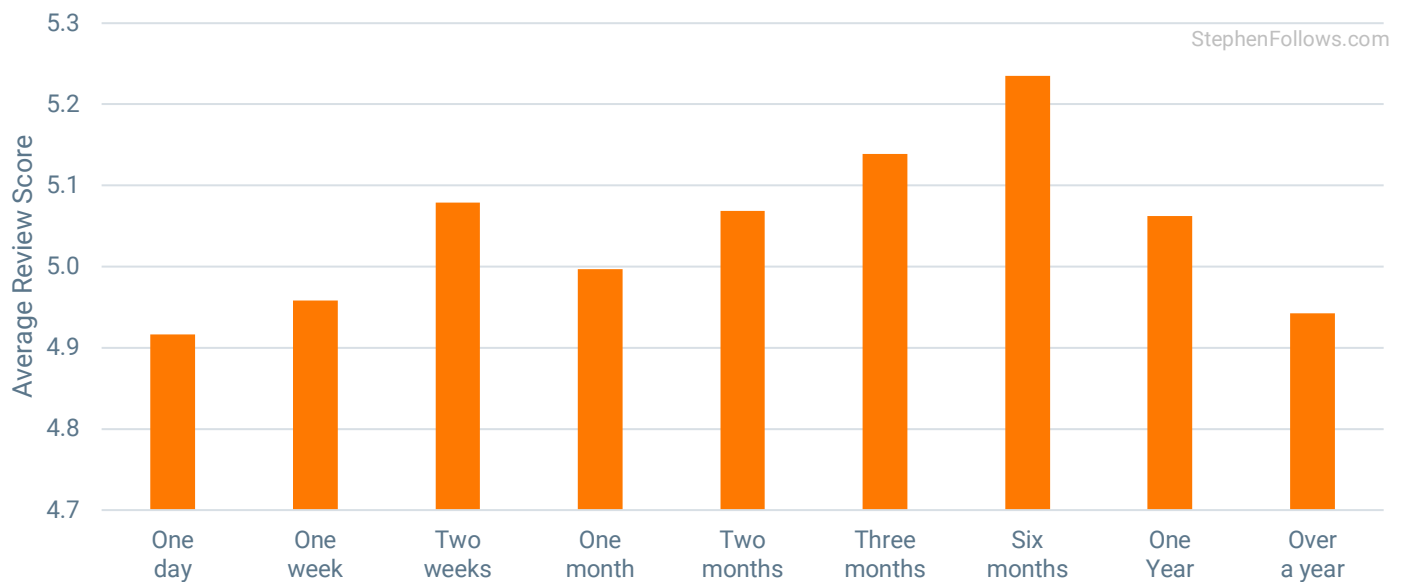
41.4% of the scripts submitted to a script competition were exported as a PDF within 24 hours of the competition submission. At the other end of the spectrum, 22.6% of scripts were created at least six months before submission.

### Time between when PDF of script was created and submission to a script competition



Scripts exported just before a competition deadline are less likely to impress script readers than those saved six months prior.

### Review Score split by time between when the PDF was created and submission to a script competition



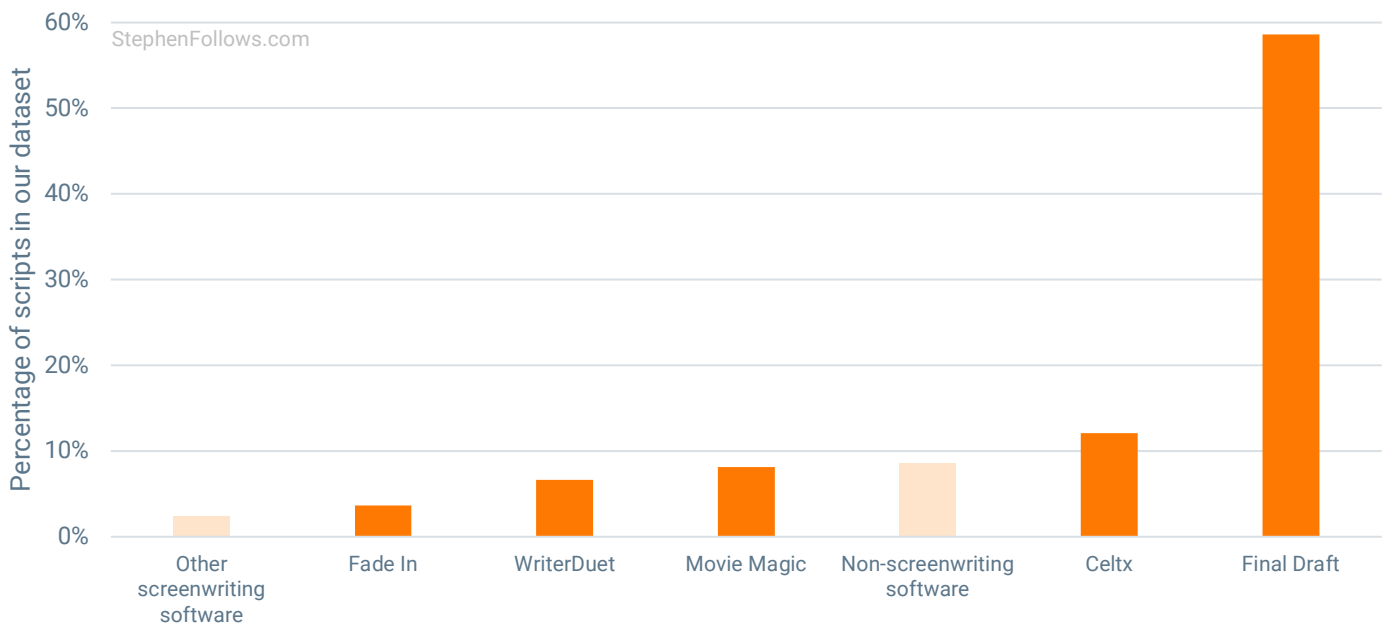
# Screenwriting Software

By analysing the metadata in each PDF file, we were able to determine which program was used to write the vast majority of scripts.

Final Draft dominates the competition and was used to write 58.6% of the scripts for which we could determine the software used<sup>22</sup>. The next most commonly used programs were Celtx (12.1%), Movie Magic Screenwriter (8.1%), WriterDuet (6.6%) and Fade In (3.6%). Other professional screenwriting programs made up a combined 3.3% of scripts.

8.6% of scripts were written in a non-screenwriting specific program, such as Microsoft Word, TextEdit and Notepad.

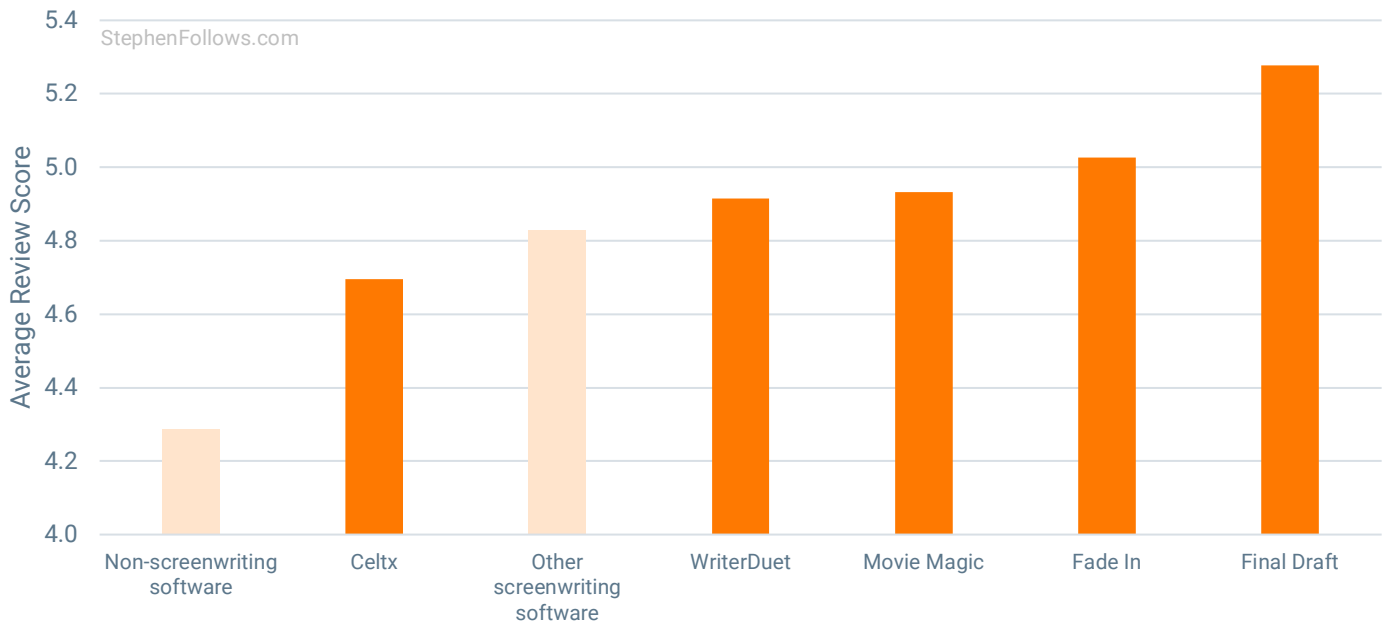
**Screenwriting software used to write scripts in our dataset**



Interestingly, there is a minor correlation between the screenwriting program used and the quality of the script. Scripts written using Final Draft scored the highest; the poorest-performing scripts were those written on non-screenwriting programs.

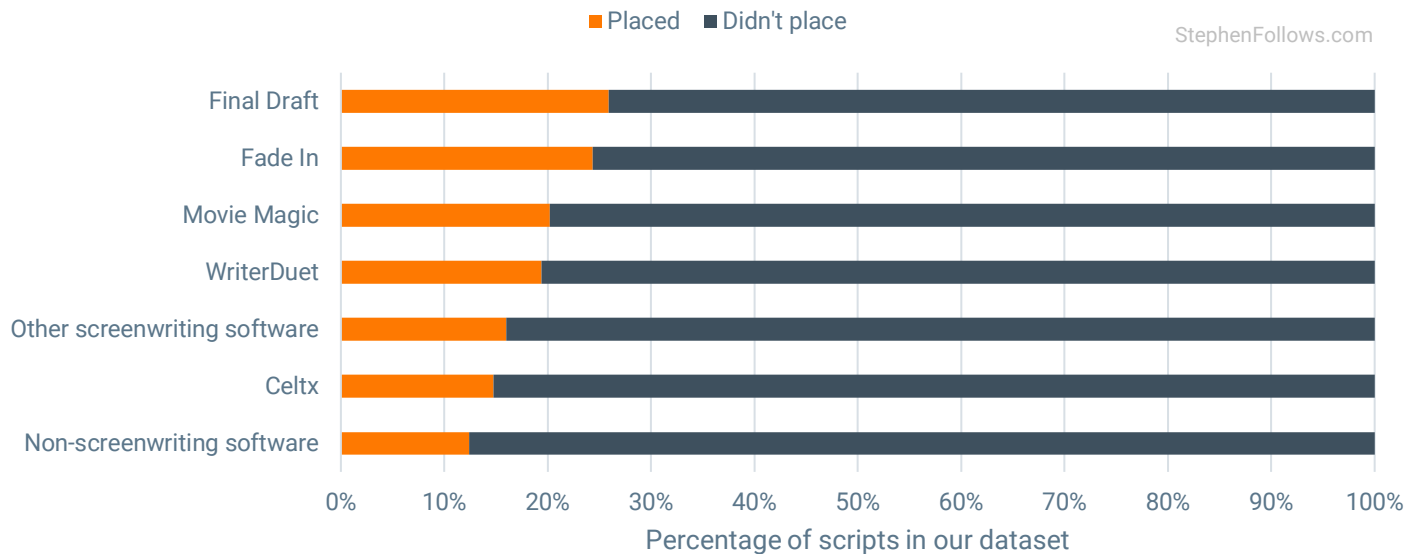
<sup>22</sup> We were able to determine the software used on over three-quarters of the scripts we studied. The charts showing market share relate only to the scripts for which a program could be determined. It accordingly presents a larger margin for error than our other findings, extrapolated to the wider screenwriting marketplace.

## Review Score by screenwriting software used



Scripts written in Final Draft and Fade In perform best in script competitions.

## Outcome of scripts entered into script competitions, by screenwriting software used



Screenwriters (and script software marketers) are reminded that correlation is not causation. Programs do offer different features and it's possible that some such features can make writing easier, more enjoyable and help identify possible problems with your work. However, just changing your screenwriting software choice is extremely unlikely to transform your writing.

***"Here I am paying big money to you writers and what for?  
All you do is change the words"***

*Samuel Goldwyn*

This report is dedicated to the people who spend their lives changing the words,  
whether they're paid the big bucks or not.

*Published by Stephen Follows, Erroneous Wit, Somerset House, London, UK*

*February 2019*

*© Stephen Follows. All rights reserved.*

*If you wish to use any part, please contact [mail@stephenfollows.com](mailto:mail@stephenfollows.com)*

*Whilst we have done our best to ensure that the report is accurate, the information it contains is not advice, and should not be treated as such. You must not rely on it as an alternative to advice from an appropriately qualified professional.*