Are films shorter when the scriptwriter directs?

Gordon Blackadder

July 23, 2015

1 Overview

In this short project I want to address an old question. When I was a student, I once had a debate with friends over how filmmakers convey ideas. I thought that films that were written and directed by the same person would be more likely to express ideas visually because, while writing, the filmmaker would be thinking along visual lines. In contrast, somebody who is only a scriptwriter may be more likely to express ideas through dialogue. This would have the effect of making films that the writer also directed shorter.

Of course there are many films that defy this simple logic, so the debate devolved into an exchange of examples ("The Unbelievable Trueth" written and directed by Hal Hartley with a runtime of 90 minutes on the one hand, "The Godfather" written and directed by Francis Ford Coppola at 175 minutes on the other) and we had to agree that the question could not be answered.

Now, many years later, I would like to do a statistical analysis of film lengths to find an answer. Using a database that I have compiled of film runtimes, categorized by country, language and genre, I would like to perform a statistical analysis to identify the trend in film lengths both when the writer and director are the same person and when they are different people.

This summary of the results (so far) is not intended to be a detailed review of statistical theory (for those interested I recommend Gelman et al, "Bayesian Data Analysis", *3rd ed.*, 2014, Taylor and Francis, Boca Raton, Florida, USA). However I will give a quick summary of my approach before discussing the results.

I expected film runtimes to follow a normal distribution, by which I mean that there will be a few films that are very short, a few films that are very long, with a majority of films having a runtime centered somewhere in the middle, probably around 90 minutes to two hours. What I would like to do is find out if this average is lower or higher when the director and scriptwriter are are the same person.

However the average runtime has clearly changed over the years. For example, films made before the widespread introduction of television are typically much shorter. Furthermore the analysis ought to be take into account that the average will likely change

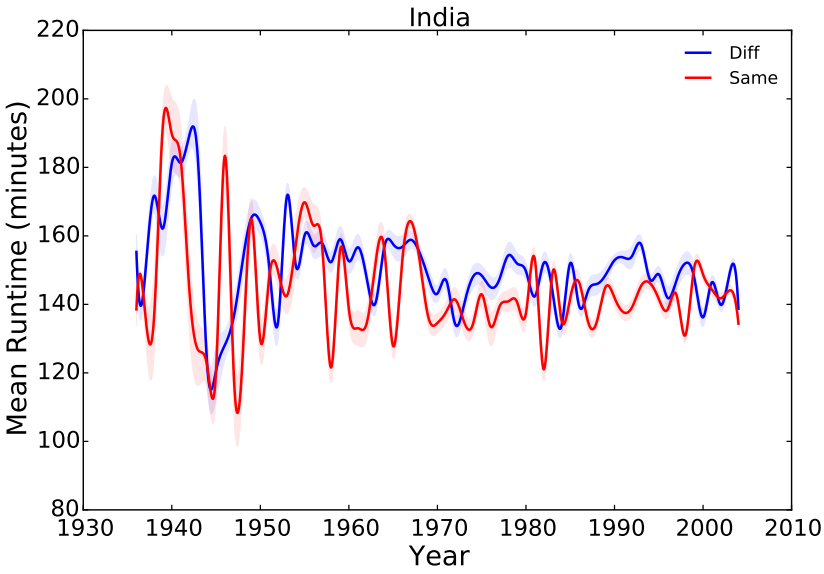

Figure 1: Average film runtime in India. Red denotes the average when one of the writers was also one of the directors, blue denotes the average of films for which this was not the case. The lighter shaded areas denote the 95% confidence interval of the average

by country (or language or genre) due to different cultural norms, for example Bollywood famously produces films much longer than those made in the America or Europe.

Therefore I modeled the possible averages as themselves following a normal distribution. This is called a hierarchical Bayesian model. I take this distribution to be centered around two hours but with a two hour standard deviation to ensure sufficient generality. (As I continue to work on this project, I may come back to generalize this.)

So for each country or language that I considered, I ran a Markov Chain Monte Carlo simulation to find the means along with the their 95% confidence intervals. These have then been plotted to clearly show how film lengths have evolved one time.

2 Results

India and the United States are the two largest film producing countries in the world. Figure 1 shows a plot of average film runtimes in India while figure 2 shows the results for films from the USA.

The plot of the Indian average film runtimes shows that there is no difference in the two groups. Both the red and blue lines substantially follow each other and overlap frequently.

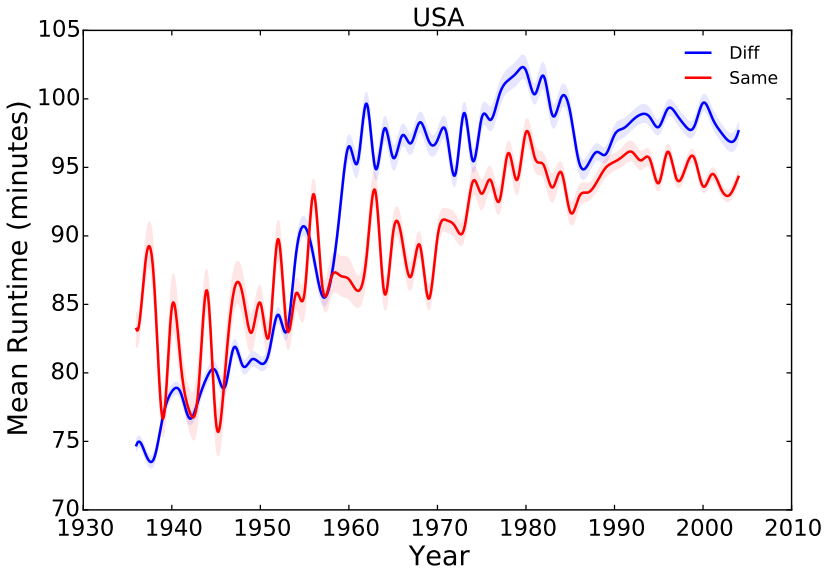

Figure 2: Average film runtimes in the USA. Colors as in figure 1.

On the other hand, the plot of American average film runtimes clearly shows a disparity. Over the last several decades films where the scriptwriter also directed have been significantly shorter.

Of course, the reason for this difference cannot be deduced by this analysis. While I originally conjectured that films where the writer also served as the director would be shorter because of the filmmaker expressing ideas more succinctly through visuals there could be other reasons. Such films may, for example, have smaller budgets.

Now lets consider films by language rather than by country. There are many reasons to do this. Spanish language movies are not just made in Spain but in many other countries in South America as well as a few productions taking place in the USA. It would be interesting to see how these film runtimes have evolved collectively. Alternatively we may wish to consider a films language to allow us to track the evolution of a particular film making culture even when borders rise and fall. By tracking Russian language films we can look at films made during the era of the Soviet Union and in modern Russia.

Figure 3 shows the evolution of runtime averages for Spanish movies while figure 4 shows the evolution for Russian language movies.

It is very clear from figure 3 that there is no difference between our two groupings in the runtimes of Spanish movies. Russian language films on the other hand appear to show that films where the writer was not the director are actually shorter from the 1960's up to the early part of the 1990's but after that the difference ends. This is an unexpected result

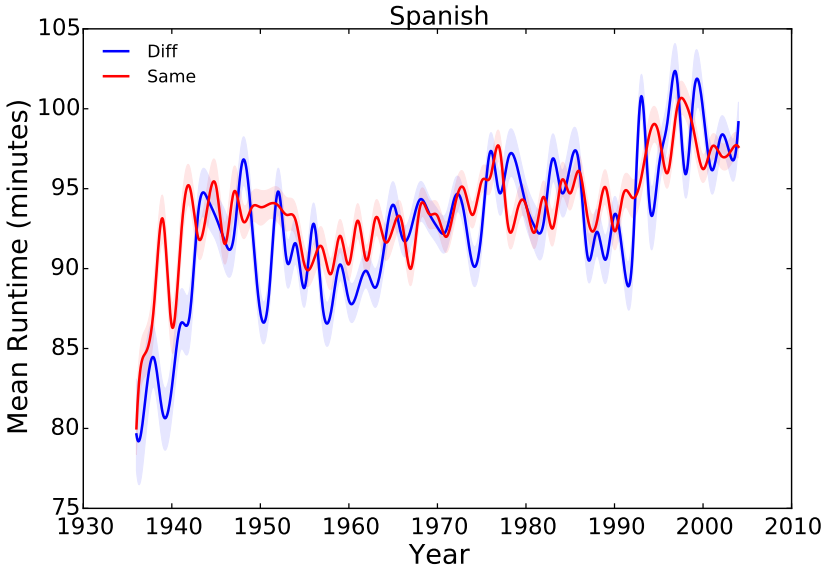

Figure 3: Average film runtimes for movies in Spanish. Colors as in figure 1.

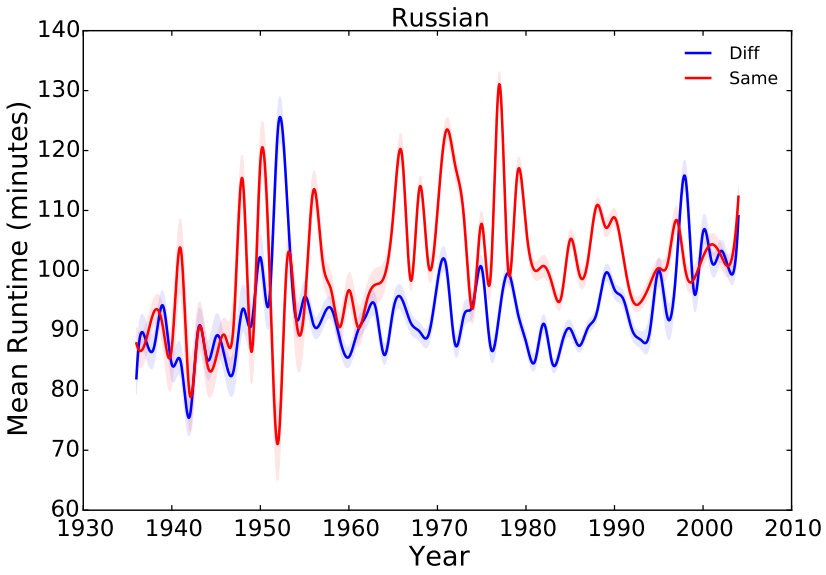

Figure 4: Average film runtimes for movies in Russian. Colors as in figure 1.

and it would be interesting to do more research to find the underlying causes.

3 Further Work

Obviously I could continue to run the analysis code for every country and language, however the problem with the model right now is that it considers every country, or every language, as being separate. This is clearly not the case as some films are made across several countries and, although less common, in several languages.

On a technical note, it would be better if the distribution of the averages were drawn from the data rather than being specified by the model.

And finally it would be nice to apply a linear regression to the plots in this paper to quantify differences (or similarities) in the runtimes. I will make this the priority for when I have more time to work on this project.