

What makes a Hollywood film successful?

Joanna Andari, Karim Awad, Jiye Ren, Nirbhay Sharma

8 October 2017

Introduction

The film industry is expected to generate global box office revenues of almost \$49.3bn by 2020, from a forecast of US\$38.3bn in 2016¹, representing a CAGR of over 32%. These substantial growth levels are likely to result in vast resources being poured into the industry, in search of the next “big hit”. Countries, recognising the value of developing a film industry to support a variety of local and national initiatives, have supported these activities, providing tax breaks to investors, to mitigate the material upfront costs involved.

Outside of India, the US continues to lead the way in producing a substantial volume of films, accounting for an estimated US\$10.7bn in box office revenues in 2016². Yet in contrast to Indian films, which principally serve a domestic audience, Hollywood films benefit from a global distribution network. This results in films having to cater for different audiences in different countries, overcoming issues varying from political censorship through to cultural sensitivities, to maximise their appeal and potential commercial success.

```
library("ggplot2")
```

```
## Warning: package 'ggplot2' was built under R version 3.4.2
```

```
names <- c(2016, 2017, 2018, 2019, 2020)
```

```
globalboxoffice16_20 <- c(38.3, 41.2, 44, 46.7, 49.3)
```

```
USboxoffice16_20 <- c(10.7, 10.6, 10.9, 11.1, 11.3)
```

```
boxoff <- data.frame(names, globalboxoffice16_20, USboxoffice16_20)
```

```
#####produce bar graph
```

With global audiences in mind, are there factors that can help predict a Hollywood film’s popularity, and with that, guide future decisions on the types of films produced? In turn, can this better inform investors on potential risks involved with backing particular projects?

This report seeks to examine a random sample of movie data pulled from IMDB data between 2011-16, concentrating on the US given its role in influencing the wider film industry. This will seek to understand how we should measure popularity, focusing on 2011-15 data, before devising a regression model that can help predict popularity. We will then test its capacity to predict 2016 trends, before noting the limitations of our work, and areas for further development.

As this report shall conclude, [...]

¹Statista (2016)

²Statista (2016)

Appendix 1 - Boxplots and Density graphs

Bibliography and references

We have made reference to various forums on coding best-practices, notably including Stackoverflow.com, and other relevant sources of R studio information available online. These have been cited during the main document, where they have been specific to our study, or represent the application of coding practices not within the Maths and Statistics Foundations for Analytics module. We have not referenced general R studio packages applied within this module (e.g. dplyr; stargazer; etc).

We obtained the original dataset, in conjunction with information on Oscar academy awards, from the following websites: <https://www.kaggle.com/tmdb/tmdb-movie-metadata/data> <https://www.kaggle.com/theacademy/academy-awards>

In terms of documents we have referenced:

[1-2] “Film industry in the US”, Statista, Dossier (2016) <https://www.statista.com/study/11472/film-industry-in-the-united-states-statista-dossier/>

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

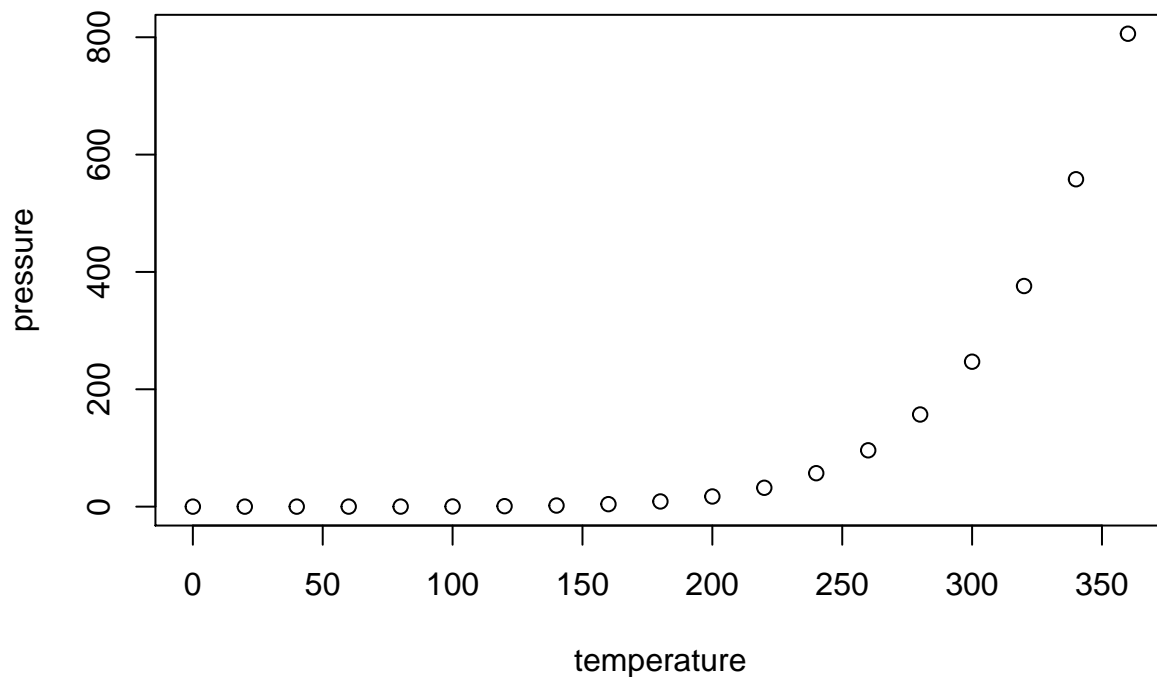
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   : 2.00
## 1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##   Mean  :15.4    Mean   : 42.98
## 3rd Qu.:19.0    3rd Qu.: 56.00
##   Max.  :25.0    Max.    :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.