Film Death Counts

N/A

Originally produced: 2023-05-04, latest update: 2023-05-15

Visualising the effects of on-screen deaths on movie data from a selection of films from 1949-2013

Research Questions

What Are The Top 15 Films With the Highest Number Of On-Screen Deaths?

Which genres depict the most on-screen deaths, on average?

How are age ratings distributed between genres that depict on-screen deaths?

Is there an association between number of on screen deaths and IMDB ratings within genres?

Data Origins

Dataset used in this project:

Olson, Randy (2013): On-screen movie kill counts for hundreds of films. *figshare. Dataset.* https://doi.org/10.6084/m9.figshare.889719.v1

The data contains on-screen death counts for 545 films within the date range of 1949-2013. Film title, release year, MPAA rating, genre(s), run time and director were also recorded.

```
filepath <- file.path("data", "filmdeathcounts.csv")
rawdata <- read.csv(filepath)
knitr::kable(head(rawdata), format = "markdown")
```

Film	Year	Body_Count	MPAA_Rating	Genre	Director	Length_Minutes	IMDB_Rating
24 Hour Party People	2002	7	R	Biography|Comedy|Drama|Music	Michael Winterbottom	117	7.3
28 Days Later	2002	53	R	Horror|Sci-Fi|Thriller	Danny Boyle	113	7.6
28 Weeks Later	2007	212	R	Horror|Sci-Fi|Thriller	Juan Carlos Fresnadillo	100	7.0

Film	Year	Body_Count	MPAA_Rating	Genre	Director	Length_Minutes	IMDB_Rating
30 Days of Night	2007	67	R	Horror|Thriller	David Slade	113	6.6
300	2007	600	R	Action|Fantasy|History|War	Zack Snyder	117	7.7
3:10 To Yuma	2007	45	R	Adventure|Crime|Drama|Western	James Mangold	122	7.8

Body count data was collected from www.MovieBodyCounts.com and tallied according to the website's guidelines.

All other data was scraped from each film's individual IMDB page and collated into this dataset by Randy Olson.

MovieBodyCounts Guidelines

The death count criteria were retrieved from the following forum page here.

*Please be aware that it is **not recommended** to follow any link provided on the forum as they may direct to unsafe websites.*

For ease, the detailed guidelines have also been restated verbatim below:

> 1. The "body counts" for this site are mostly "on screen kills/deaths" or fatal/critical/mortal shots/hits of human, humanoid, or creatures (ie monsters, aliens, zombies.) The rule of thumb is "do they bleed" which will leave the concept of cyborgs somewhat open and decided per film. The human and creature counts should be separate. These will be added together for a final tally.
>
> - Note on non-hominoid/animals - In films such as animations where animal/robots have the ability to speak their kills will sometimes be considered in the counts since they are "characters" to the story. This will be decided on a film by film basis.
>
> 2. Unspecified/displayed bodies (one's that do not actually have an "on screen" hit)
>
> - **Do not count** - unspecified piles or strewn "bodies" (ie-war films) unless they are clear and countable (head shots). These will be noted though in the misc. section.
> - **Do count** - any unspecified but clearly shown "displayed" bodies (IE crucifixions, hangings. . .)
>
> 3. Ships, buildings, vehicles ect can be counted but unless an "on screen kill/hit/death" is shown to a "body" it will not be added in the final "body count tally". A miscellaneous count will be added to the page for these. It is entirely optional to count any such.
>
> 4. Counts of kills by main characters will be added to the pages.
>
> 5. Animals: Depending on their type of death and their significance to the plot or main characters they will not be counted except under "Miscellaneous."
>
> 6. Implied Kills. . .
>
> - **Do not count** - if they are only talked about. Where nothing at all is shown in the film. Also photos of dead bodies do not count.
> - **Do count** - where everything is shown up until the final blow, fall, shot. . . These are called "cut aways." An example is the opening scene in Jurassic Park. If you look really closely at that sequence it actually shows nothing in the way of a "death" but it is clearly though "implied" that it is through editing.
>
> 7. Specify which cut of a film you are doing (theatrical, directors. . .), but please use an officially released extended/unrated/director's cut when available.
>
> 8. Specific guidelines pertaining to zombie deaths and other exceptional on-screen deaths can be resolved on a case-by-case basis under advisement from the site owner.

As implied by the guidelines above film body counts were counted by volunteers following this criteria and made available via this website.

Data Preparation

Preparing this data for analyses consisted of two steps.

1. Modernise outdated MPA ratings. This involved mutating any age ratings no longer used by the MPA into a modern equivalent.

This was only performed on films that did not have a modern rating attached, therefore, films where the rating may have changed over time from one modern rating to another were not unaffected. This may have introduced inconsistencies with the genres in this data set.

```
#---ORDINAL RATINGS---

    df_v2 <- df_v1 %>%
        mutate (rating_order = case_when(
            MPAA_Rating == "G" ~ "1",
            MPAA_Rating == "PG" ~ "2",
            MPAA_Rating == "PG-13" ~ "3",
            MPAA_Rating == "R" ~ "4",
            MPAA_Rating == "X" ~ "5",
            MPAA_Rating == "Unrated" ~ "0"
      )
    )
```

2. Genre tidying Genre data were then split into individual categories and the number of unique genres were calculated and printed.

```
genres <- unlist(strsplit(df_v2$Genre, split = "\\|"))
  num_unique_genres <- length(unique(genres))
  print(paste("There are",num_unique_genres, "unique genres in this dataset"))
```

```
## [1] "There are 22 unique genres in this dataset"
```

```
print(sort(unique(genres)))
```

```
##  [1] "Action"      "Adventure"   "Animation"   "Biography"   "Comedy"
##  [6] "Crime"       "Documentary" "Drama"       "Family"      "Fantasy"
## [11] "Film-Noir"   "History"     "Horror"      "Music"       "Musical"
## [16] "Mystery"     "Romance"     "Sci-Fi"      "Sport"       "Thriller"
## [21] "War"         "Western"
```

Six genres, chosen from the top 50% of the most frequently appearing genres, were subsequently dummy coded (indicating as to whether a film was or wasn't tagged as a specific genre) to allow for visualations to be made within each genre.

An example of the code used to dummy code for action movies is quoted below.

```
#---------DUMMY CODING BY INTERESTED GENRES---
#1. Identify key genre title within genres using grepl().

#2. Then generate new column tagging rows as being within that genre.

#3. Finally, create subset of only films in that genre
```

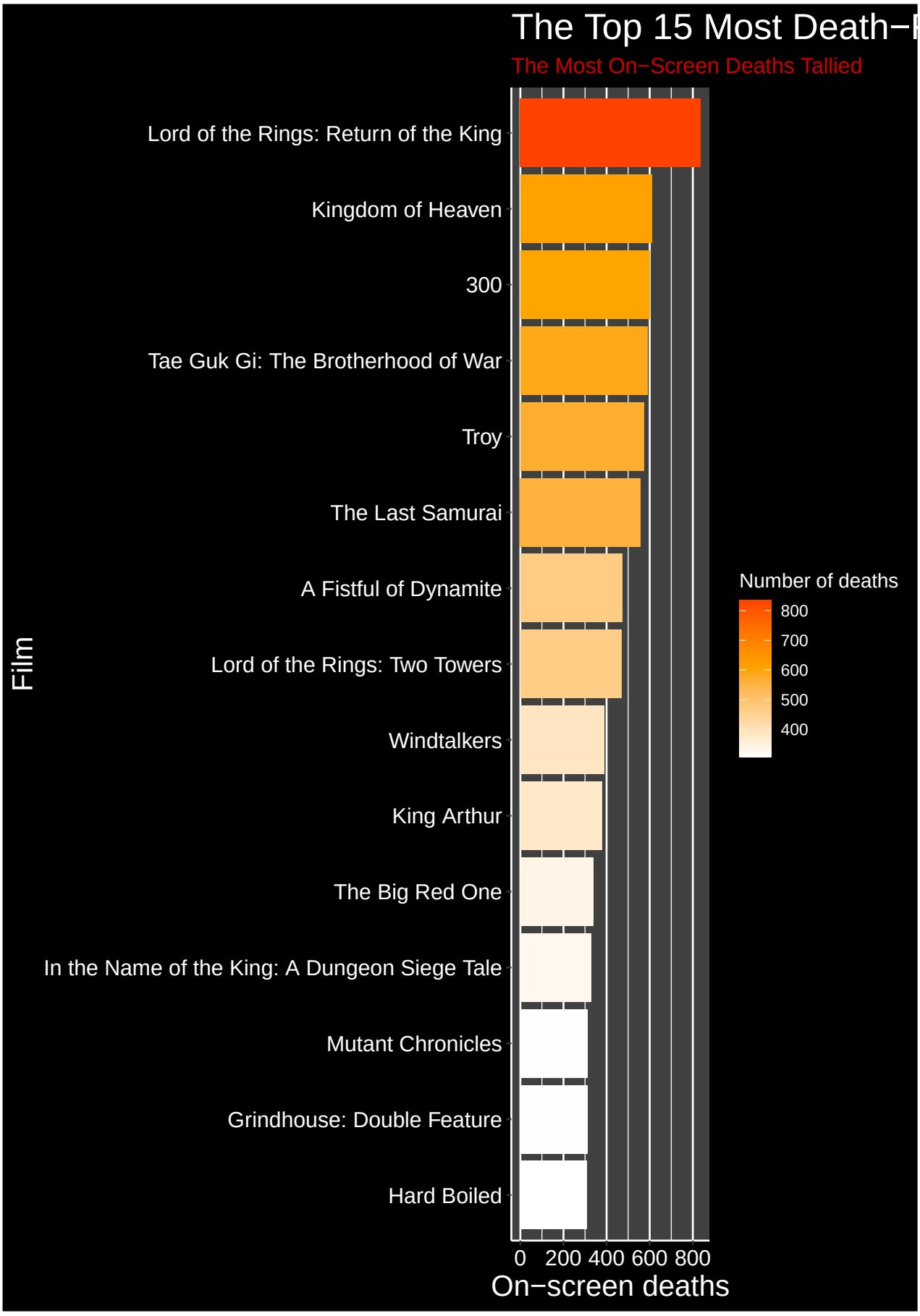
```r
#-Action-
df_v2$is_action <- grepl("Action", df_v2$Genre)
df_v2 <- df_v2 %>%
  mutate(is_action = as.numeric(is_action))
action_data <- subset(df_v2, is_action == "1")
```

The similar lines of code were used for the Fantasy, Horror, Thriller, Sci-Fi and War genres.

Visualisation 1: Top 15 Movies for On-Screen Deaths

```r
#--------------HIGHEST BODY COUNT MOVIES----------------

#Arrange in descending order and then select the top 10
df_HBCM <- arrange(df_v3, desc(Body_Count))
df_HBCM <- head(df_HBCM, 15)

# Make a Bar Chart from this data frame
      #Note reordered films by body count; displays graph in descending order
top15_HBCM <-
  ggplot(df_HBCM, aes(x = reorder(Film, Body_Count),
                      y = Body_Count,
                      fill = Body_Count)) +
  geom_bar(stat = "identity") +
          scale_fill_gradient2(low= "white",
                               mid = "orange",
                               high = "red",
                               midpoint = 600) +
  labs( title = "The Top 15 Most Death-Filled Movies (1949-2013)",
        subtitle = "The Most On-Screen Deaths Tallied",
        x = "Film",
        y = "On-screen deaths",
        fill = "Number of deaths"
       ) +
#Black Background
  theme(panel.background = element_rect(fill = "grey25"),
        plot.background = element_rect(fill = "black"),
        legend.background = element_rect(fill = "black"),
        panel.grid.major.y = element_blank(),
        panel.grid.minor.y = element_blank(),
#Text colour swap
          plot.title = element_text(colour = "white", size = "20"),
          plot.subtitle = element_text(colour = "red3", size = "12"),
          legend.text = element_text(colour = "white"),
          legend.title = element_text(colour = "white"),
#White Axes
           axis.line = element_line(colour = "white"),
           axis.text = element_text(colour = "white", size = "12"),
           axis.title = element_text(colour = "white", size = "16")) +
#Flip x and y axes for easier reading
  coord_flip()

#Print and Save
  print(top15_HBCM)
```

```r
ggsave("top15deathsmovies.png", top15_HBCM, width = 12, height = 6, dpi = 300)
```

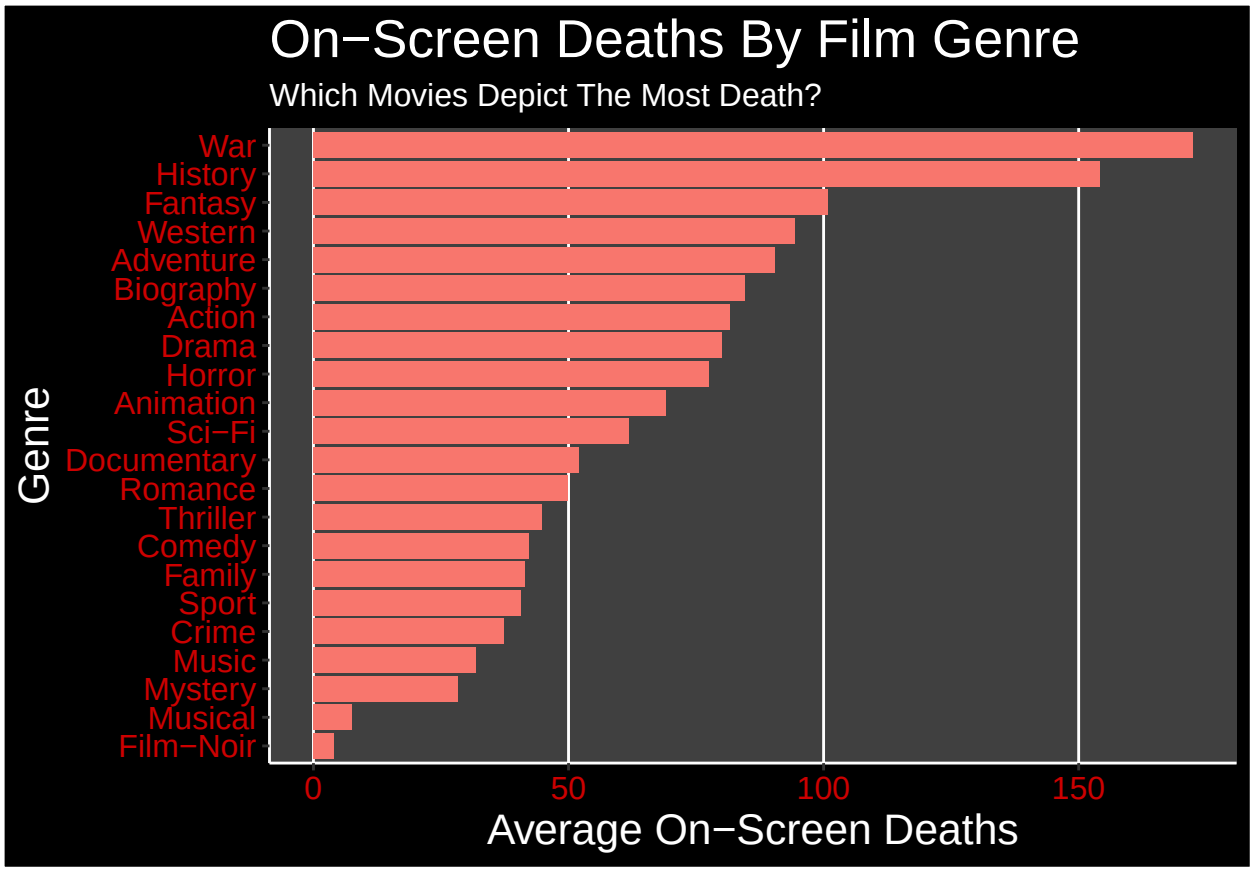

The Top 15 Most Death−
The Most On−Screen Deaths Tallied
Lord of the Rings: Return of the King
Kingdom of Heaven
300
Tae Guk Gi: The Brotherhood of War
Troy
The Last Samurai
A Fistful of Dynamite
Lord of the Rings: Two Towers
Windtalkers
King Arthur
The Big Red One
In the Name of the King: A Dungeon Siege Tale
Mutant Chronicles
Grindhouse: Double Feature
Hard Boiled
Film
Number of deaths
800
700
600
500
400
0
200
400
600
800
On−screen deaths

Data Preparation Part Two: Genre Analysis

```
#------------DATA PREP 2: HIGHEST BODYCOUNT GENRES-----------------------------
    #Split genres by row to duplicate films that have more than one genre
        df_split <- separate_rows(df_v3, Genre, sep = "\\|")
            df_HBCG <- arrange(df_split, desc(Genre))

#----------Average On-Screen Deaths Per Genre----------
avg_HBCG <- df_HBCG %>%
          group_by(Genre) %>%
          summarize(avg_body = mean(Body_Count))
```

Visualisation 2: Average Number of On-Screen Deaths by Genre

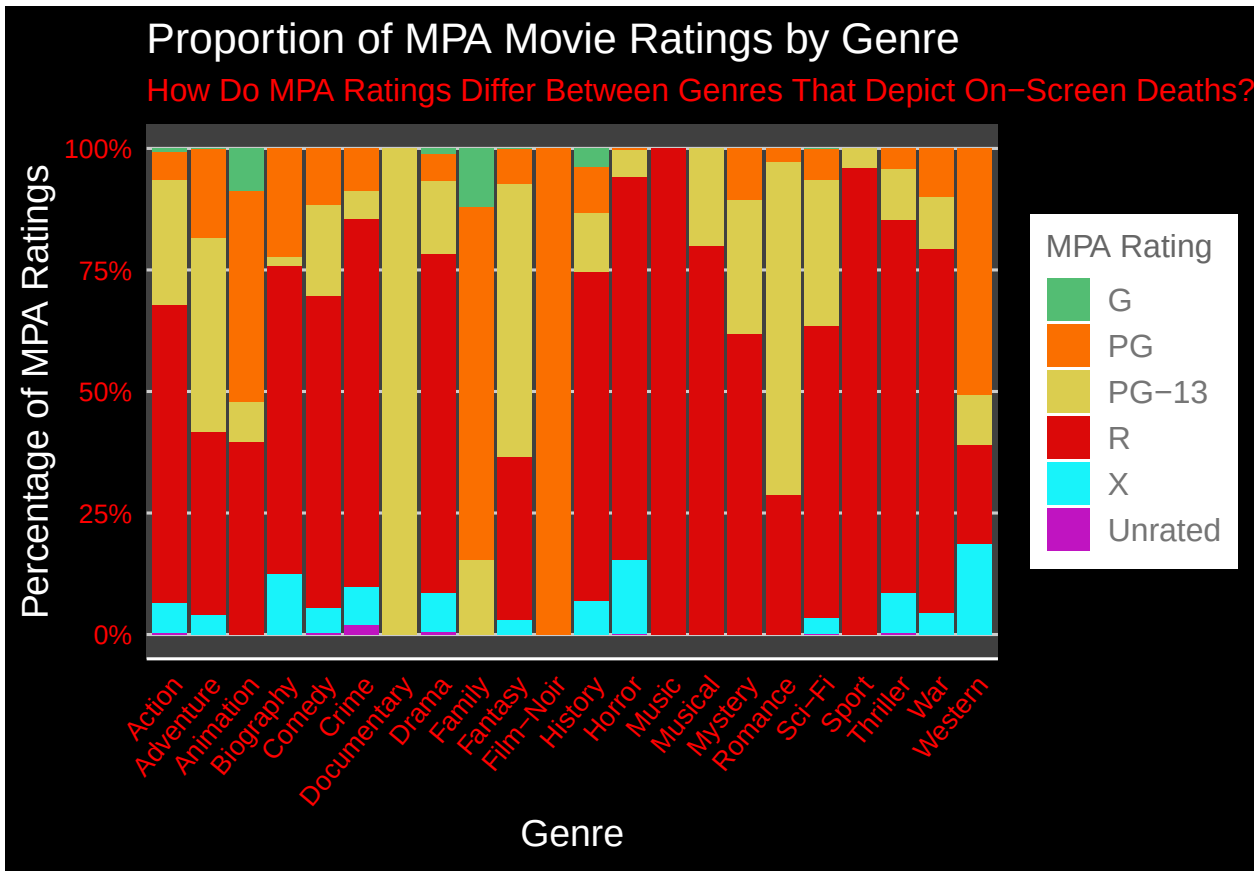
```
#1. Plot Average Body Count Per Genre

ABCG_plot <- ggplot(avg_HBCG, aes(x = reorder(Genre, avg_body),
                                  y = avg_body,
                                  fill = "white")) +
  geom_col(position = "stack") +
  labs(title = "On-Screen Deaths By Film Genre",
       subtitle = "Which Movies Depict The Most Death?",
       x = "Genre",
       y = "Average On-Screen Deaths"
  ) +
  theme(panel.background = element_rect(fill = "grey25"),
        plot.background = element_rect(fill = "black"),
        panel.grid.major.y = element_blank(),
        panel.grid.minor = element_blank(),
#Text colour swap
        plot.title = element_text(colour = "white", size = "20"),
        plot.subtitle = element_text(colour = "white", family = "", size = "12"),
#White Axes
        axis.line = element_line(colour = "white"),
        axis.text = element_text(colour = "red3", size = "12"),
        axis.title = element_text(colour = "white", size = "16"),
#Remove Legend
        legend.position = "none") +
#Flip x&y
        coord_flip()

#Print and Save plot
  print (ABCG_plot)
    ggsave("DeadliestGenres.png", ABCG_plot, width = 8, height = 5, dpi = 300)
```



Visualisation 3: Average Number of On-Screen Deaths by Genre

```
#2. Plot body count proportional to MPA rating by on screen deaths

HBCG_plot <-
  ggplot(df_HBCG, aes(x = Genre, y = Body_Count, fill = MPAA_Rating)) +
  geom_bar(stat = "identity", position = "fill") +
  theme_gdocs() +
  scale_y_continuous(labels = scales::percent) +
  labs(title = "Proportion of MPA Movie Ratings by Genre",
       subtitle = "How Do MPA Ratings Differ Between Genres That Depict On-Screen Deaths?",
       x = "Genre",
       y = "Percentage of MPA Ratings",
       fill = "MPA Rating"
  ) +
#Assign colours to MPA rating
   scale_fill_manual(values = c("#53bd73",
                                "#Fa6f00",
                                "#dbcd4f",
                                "#db0909",
                                "#18f3fa",
                                "#C014c1"),
                     labels = c("G",
                                "PG",
                                "PG-13",
                                "R",
                                "X",
```

```
                                "Unrated")) +

#Aesthetics similar to previous graphs
  theme(panel.background = element_rect(fill = "grey25"),
        plot.background = element_rect(fill = "black"),
        panel.grid.major.x = element_blank(),
        panel.grid.minor.x = element_blank(),
#Text colour swap
        plot.title = element_text(colour = "white", size = "16"),
        plot.subtitle = element_text(colour = "red", family = "", size = "12"),
#White Axes
        axis.line = element_line(colour = "white"),
        axis.text = element_text(colour = "red", size = "10"),
        axis.title = element_text(colour = "white", size = "14"),
#Bottom Text Align
        axis.text.x = element_text(angle = 50, hjust = 1))
#Print and Save Plot
    print (HBCG_plot)
      ggsave("MPAAdistribution.png", HBCG_plot, width = 8, height = 5, dpi = 300)
```



Visualisation 4: Average Number of On-Screen Deaths by Genre

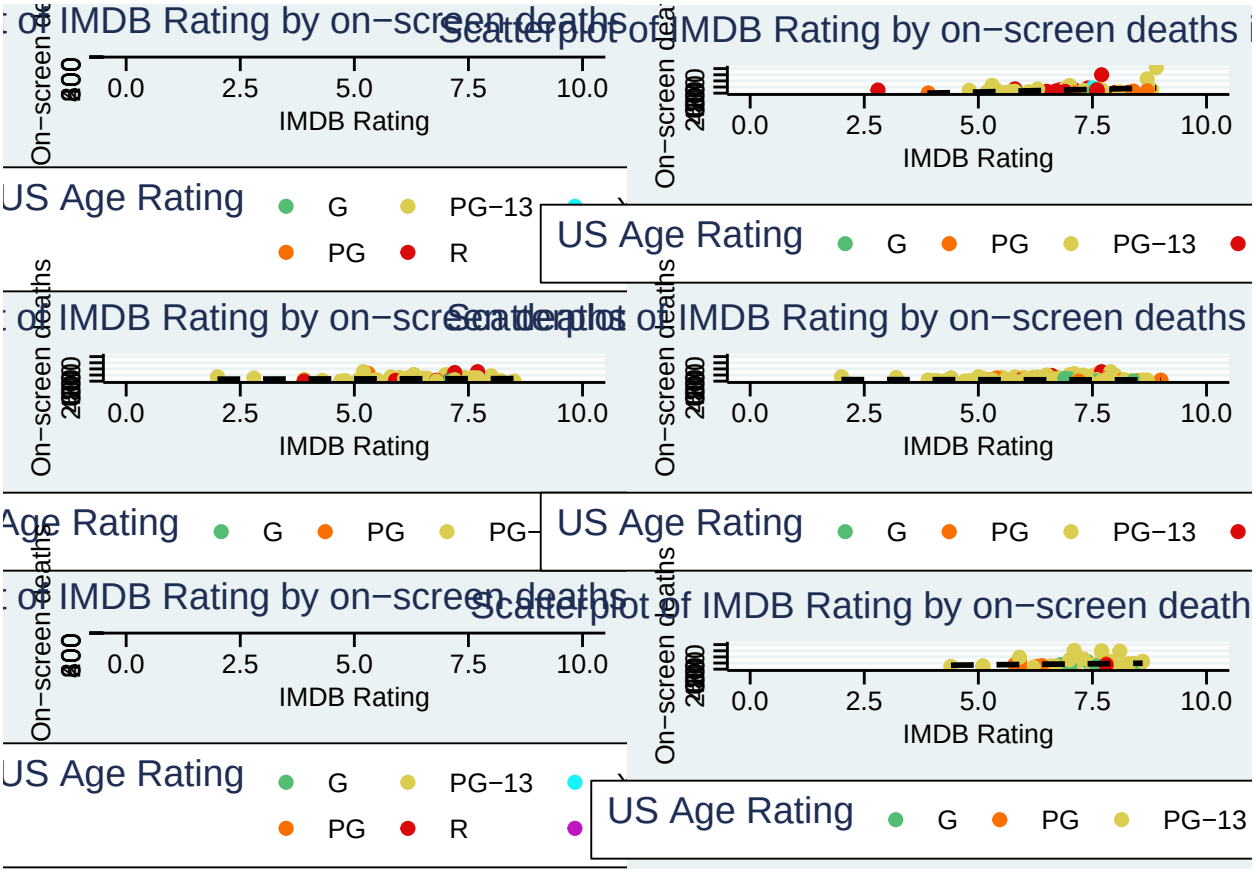
```
#----SCATTERPLOTS BY SINGLE GENRES-----

#ACTION ALONE
action_plot <- ggplot(action_data, aes(x = IMDB_Rating,
```

```
                        y = Body_Count,
                        colour = factor(MPAA_Rating))) +
  geom_point(size = 2) +
  labs(title = "Scatterplot of IMDB Rating by on-screen deaths in Action Films",
       x = "IMDB Rating",
       y = "On-screen deaths",
       colour = "US Age Rating") +
  xlim(0,10)+
  ylim(0, 850) +
  geom_smooth(method = "lm",
              color = "black",
              linetype = "dashed",
              se = FALSE) +
  theme_stata() +
  scale_colour_manual(values = c("#53bd73",
                                 "#Fa6f00",
                                 "#dbcd4f",
                                 "#db0909",
                                 "#18f3fa",
                                 "#C014c1"),
                      labels = c("G",
                                 "PG",
                                 "PG-13",
                                 "R",
                                 "X",
                                 "Unrated"))

#FANTASY ALONE
fantasy_plot <- ggplot(fantasy_data, aes(x = IMDB_Rating,
                          y = Body_Count,
                          colour = factor(MPAA_Rating))) +
    geom_point(size = 2) +
    labs(title = "Scatterplot of IMDB Rating by on-screen deaths in Fantasy Films",
         x = "IMDB Rating",
         y = "On-screen deaths",
         colour = "US Age Rating")+
  xlim(0,10)+
  ylim(0, 850) +
    geom_smooth(method = "lm",
                color = "black",
                linetype = "dashed",
                se = FALSE) +
    theme_stata() +
  scale_colour_manual(values = c("#53bd73",
                                 "#Fa6f00",
                                 "#dbcd4f",
                                 "#db0909",
                                 "#18f3fa",
                                 "#C014c1"),
                      labels = c("G",
                                 "PG",
                                 "PG-13",
                                 "R",
                                 "X",
                                 "Unrated"))
#HORROR ALONE
```

```r
horror_plot <-  ggplot(horror_data, aes(x = IMDB_Rating,
                        y = Body_Count,
                        colour = factor(MPAA_Rating))) +
  geom_point(size = 2) +
    labs(title = "Scatterplot of IMDB Rating by on-screen deaths in Horror Films",
         x = "IMDB Rating",
         y = "On-screen deaths",
         colour = "US Age Rating") +
  xlim(0,10)+
  ylim(0, 850) +
  geom_smooth(method = "lm",
              color = "black",
              linetype = "dashed",
              se = FALSE) +
  theme_stata() +
  scale_colour_manual(values = c("#53bd73",
                                 "#Fa6f00",
                                 "#dbcd4f",
                                 "#db0909",
                                 "#18f3fa",
                                 "#C014c1"),
                      labels = c("G",
                                 "PG",
                                 "PG-13",
                                 "R",
                                 "X",
                                 "Unrated"))

#THRILLER ALONE
thriller_plot <- ggplot(thriller_data, aes(x = IMDB_Rating,
                          y = Body_Count,
                          colour = factor(MPAA_Rating))) +
    geom_point(size = 2) +
    labs(title = "Scatterplot of IMDB Rating by on-screen deaths in Thriller Films",
         x = "IMDB Rating",
         y = "On-screen deaths",
         colour = "US Age Rating") +
  xlim(0,10)+
  ylim(0, 850) +
    geom_smooth(method = "lm",
                color = "black",
                linetype = "dashed",
                se = FALSE) +
    theme_stata() +
  scale_colour_manual(values = c("#53bd73",
                                 "#Fa6f00",
                                 "#dbcd4f",
                                 "#db0909",
                                 "#18f3fa",
                                 "#C014c1"),
                      labels = c("G",
                                 "PG",
                                 "PG-13",
                                 "R",
                                 "X",
                                 "Unrated"))
```

```
#SCI-FI ALONE
scifi_plot <- ggplot(scifi_data, aes(x = IMDB_Rating,
                          y = Body_Count,
                          colour = factor(MPAA_Rating))) +
    geom_point(size = 2) +
    labs(title = "Scatterplot of IMDB Rating by on-screen deaths in Sci-Fi Films",
         x = "IMDB Rating",
         y = "On-screen deaths",
         colour = "US Age Rating") +
  xlim(0,10)+
  ylim(0, 850) +
    geom_smooth(method = "lm",
                color = "black",
                linetype = "dashed",
                se = FALSE) +
    theme_stata() +
  scale_colour_manual(values = c("#53bd73",
                                 "#Fa6f00",
                                 "#dbcd4f",
                                 "#db0909",
                                 "#18f3fa",
                                 "#C014c1"),
                      labels = c("G",
                                 "PG",
                                 "PG-13",
                                 "R",
                                 "X",
                                 "Unrated"))

#WAR ALONE
war_plot <- ggplot(war_data, aes(x = IMDB_Rating,
                          y = Body_Count,
                          colour = factor(MPAA_Rating))) +
    geom_point(size = 2) +
    labs(title = "Scatterplot of IMDB Rating by on-screen deaths in War Films",
         x = "IMDB Rating",
         y = "On-screen deaths",
         colour = "US Age Rating") +
  xlim(0,10)+
  ylim(0, 850) +
    geom_smooth(method = "lm",
                color = "black",
                linetype = "dashed",
                se = FALSE) +
  theme_stata() +
  scale_colour_manual(values = c("#53bd73",
                                 "#Fa6f00",
                                 "#dbcd4f",
                                 "#db0909",
                                 "#18f3fa",
                                 "#C014c1"),
                      labels = c("G",
                                 "PG",
                                 "PG-13",
                                 "R",
```

```
                                  "X",
                                  "Unrated"))

# ------------------COMBINE WITHIN-GENRE VISUALISATIONS----------

combined_graph <- grid.arrange(action_plot,
                               fantasy_plot,
                               horror_plot,
                               thriller_plot,
                               scifi_plot,
                               war_plot,
                               ncol = 2)

#Print and save plot
  ggsave("CombinedGenres.png", combined_graph, width = 14, height = 15, dpi = 300)
```



Summary

Due to several visualisations clearly showing no association between genre and on-screen deaths, further visualisations could be made using this data (or a more up-to-date alternative). Specifically these could examine the associations between director and on-screen deaths, or even by country of production.

Discussion

Caveat #1: The purpose of this project was to demonstrate an ability to visualise a given dataset using RStudio effectively as part of a Data Analysis and Visualisation Course at the Universit y of Sheffield, rather than to provide any novel or groundbreaking data findings.

Caveat #2: The data set utilised in this study were collected several third-parties and made freely available via figshare. This dataset may be subject to several confounds such as inter-rater biases when determining body count, incorrect information retrieved from IMDB or simple human error during data collation. In particular, due to the collaborative nature of the MovieBodyCount forum, several data entries may have had inconsistent countings that were not revised when guidelines were altered or if guidelines were not strictly followed.

Caveat #3: MPAA ratings were simplified for ease of analysis. Due to rating systems in the United States having changed over the years several films may be inaccurately rated by today's standards or have been grouped into an incorrect category whilst tidying the data. Furthermore, it must be highlighted that two films with the same genre are not equivalent to one another in content. (Eg. One PG rated film may be significantly closer to a G rating than another) and therefore may limit the visualizations made.

Caveat #4: Despite average on-screen deaths per minute being calculated, it may be that the time of on-screen deaths into a film had an effect on IMDB scores. Would it be more entertaining to consistently show character deaths throughout a slasher film or for every character to die in the last few minutes?

Other interesting analyses based on this data

A visualisation of the top 25 deadliest movies by on-screen deaths

http://www.randalolson.com/2013/12/31/deadliest-films-of-all-time-by-on-screen-death-counts/

A visualisation of the top 25 deadliest movies by on-screen deaths per minute

http://www.randalolson.com/2013/12/31/most-violence-packed-films/

A visualisation showing the trend of on-screen deaths increasing by release year

http://www.randalolson.com/2013/12/31/violence-has-been-on-the-rise-in-the-film-industry/