The Rock in the Jungle with a Hart

William Lovejoy

7/8/2022

A common joke online states that if you want to make money off a movie, cast Dwayne "The Rock" Johnson and Kevin Hart, and then put them in a jungle. Variations of the joke claim that you can get the Rock to act in a film be telling him there's either a jungle or Kevin Hart. But how true is this?

```
library(rvest)
library(scales)
library(dplyr)
library(lubridate)
library(tidyverse)
library(ggrepel)
```

Start as always by loading in all the packages we'll need.

```
rock_wiki <- read_html("https://en.wikipedia.org/wiki/Dwayne_Johnson_filmography")
rock_films <- html_table(rock_wiki)[[2]] %>%
  as_tibble(.name_repair = "unique") %>%
  na.omit() %>%
  subset(select = -c(Notes, Ref.))

hart_wiki <- read_html("https://en.wikipedia.org/wiki/Kevin_Hart#Film")
hart_films <- html_table(hart_wiki)[[2]] %>%
  as_tibble(.name_repair = "unique") %>%
  na.omit() %>%
  subset(select = -c(Notes))

head(rock_films)
```

```
## # A tibble: 6 x 3
##    Year Title             Role
##   <int> <chr>             <chr>
## 1  1999 Beyond the Mat    "Himself"
## 2  2001 The Mummy Returns "Mathayus of Akkad / The Scorpion King"
## 3  2001 Longshot          "Mugger"
## 4  2002 The Scorpion King "Mathayus of Akkad / The Scorpion King"
## 5  2003 The Rundown       "Beck"
## 6  2004 Walking Tall      "Christopher \"Chris\" Vaughn, Jr."
```

For the filmography, the easiest way to get it is to scrape each actor's respective wiki page. So we load in those pages, and convert the table to a tibble. We'll also make sure to remove empty rows and unneccesary columns. We're doing this for both Dwayne "The

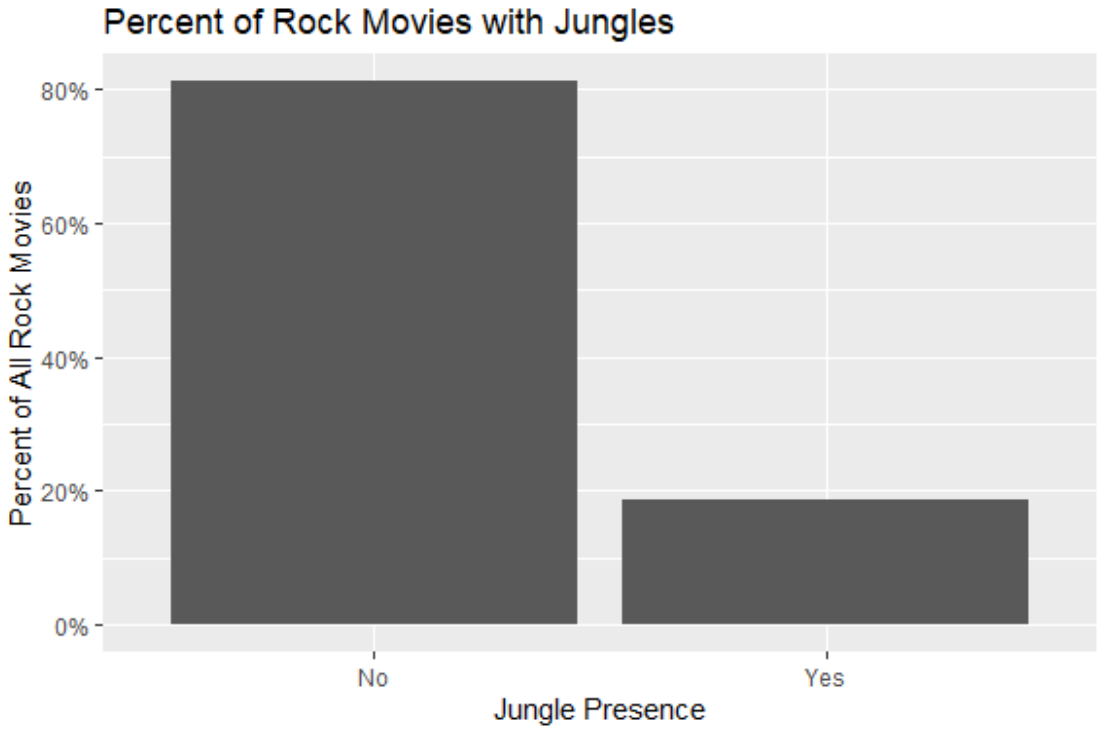
Rock" Johnson and Kevin Hart so we can see how often Hart is in the same movie as the Rock.

```
jungles <- c("Journey 2: The Mysterious Island", "Rampage", "Red Notice",
             "The Mummy Returns", "The Rundown", "Jumanji: The Next Level",
             "Moana", "Jungle Cruise", "Jumanji: Welcome to the Jungle")

movie_money <- read.csv("C:\\Users\\William Lovejoy\\Documents\\Codes\\R\\Dat
aScience\\movies.csv")
movie_money <- movie_money %>%
  subset(select = -c(...1, ReleaseDate))  %>%
  rename(Title = Movie)

rock_films <- rock_films %>%
  left_join(movie_money, by = "Title") %>%
  mutate(Hart = ifelse(Title %in% hart_films$Title, "Yes", "No"),
         Jungle = ifelse(Title %in% jungles, "Yes", "No"),
         Net = WorldwideGross-ProductionBudget)
rock_films$Year <- year(as.Date(as.character(rock_films$Year), format = "%Y")
)


head(rock_films)

## # A tibble: 6 x 9
##    Year Title  Role   ProductionBudget DomesticGross WorldwideGross Hart
Jungle
##   <dbl> <chr>  <chr>             <dbl>         <dbl>          <dbl> <chr>
<chr>
## 1  1999 Beyon~ "Hims~               NA            NA             NA No
No
## 2  2001 The M~ "Math~         98000000     202007640      435040395 No
Yes
## 3  2001 Longs~ "Mugg~               NA            NA             NA No
No
## 4  2002 The S~ "Math~         60000000      90580000      165890634 No
No
## 5  2003 The R~ "Beck"         85000000      47641743       80831893 No
Yes
## 6  2004 Walki~ "Chri~         56000000      46213824       47856383 No
No
## # ... with 1 more variable: Net <dbl>
```

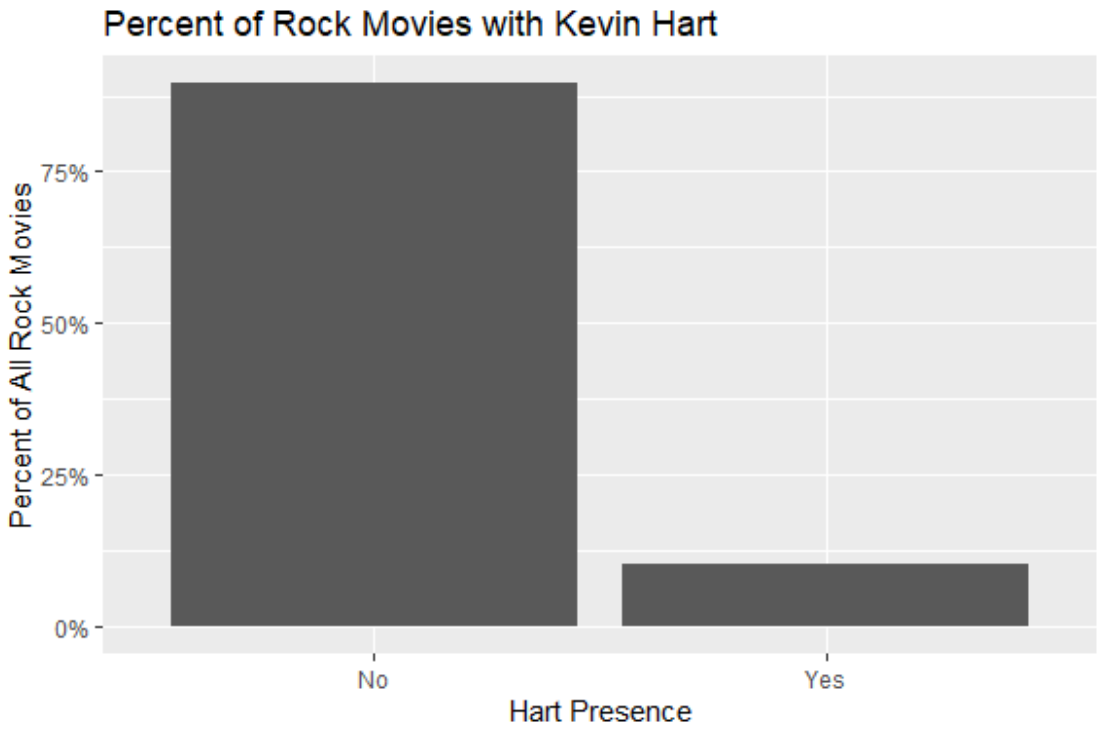
Next up we need to add more data to our list of Rock movies. To start, we'll create a vector with all of the rock movies that had a jungle or tropical forest in them. We'll also take the chance to load in a csv file from a previous project. Our "movie_money" dataset includes information about the release date, production budget, domestic gross income, and global gross income of over 5600 movies. We'll be merging this and our list of Kevin Hart movies with our Rock dataset. Now we know whether or not a given Dwayne Johnson

movie had a jungle and/or Kevin Hart in it. We also how much that movie made, and we can use that to find the rough Net income of a given movie.

```
ggplot(rock_films, aes(x = Jungle, y = (..count..)/sum(..count..))) +
  geom_bar() + labs(title = "Percent of Rock Movies with Jungles",
                    x = "Jungle Presence", y = "Percent of All Rock Movies")
+
  scale_y_continuous(labels = scales::percent)
```

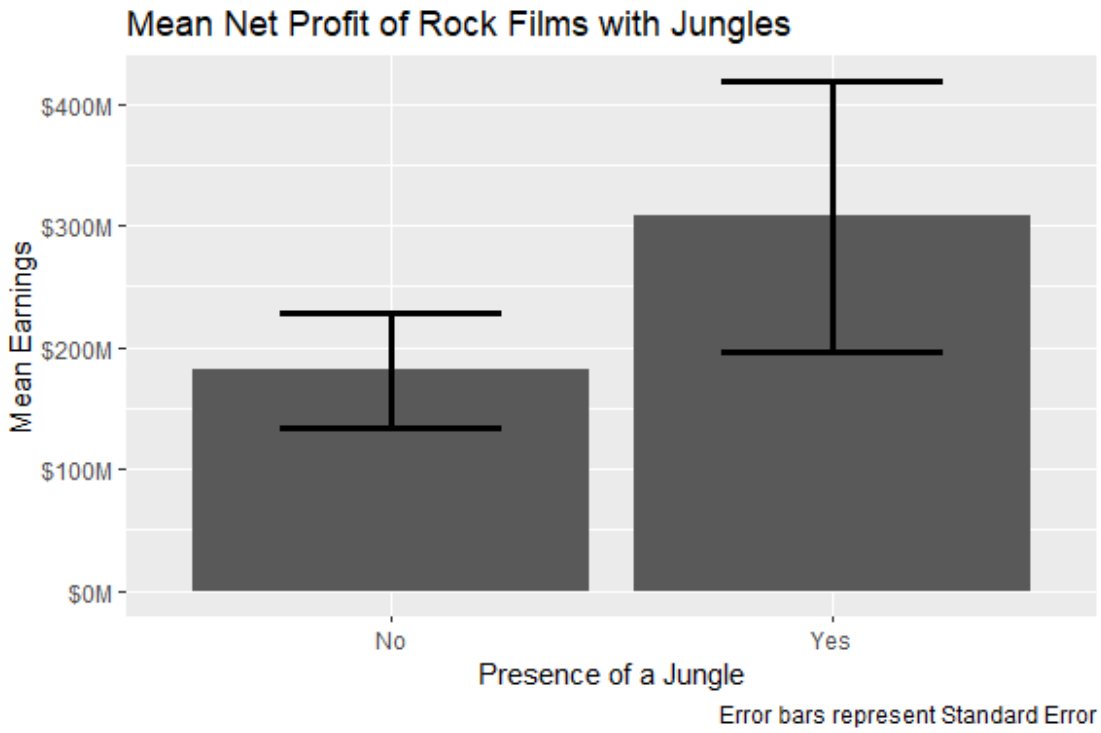



```
ggplot(rock_films, aes(x = Hart, y = (..count..)/sum(..count..))) +
  geom_bar() + labs(title = "Percent of Rock Movies with Kevin Hart",
                    x = "Hart Presence", y = "Percent of All Rock Movies") +
  scale_y_continuous(labels = scales::percent)
```

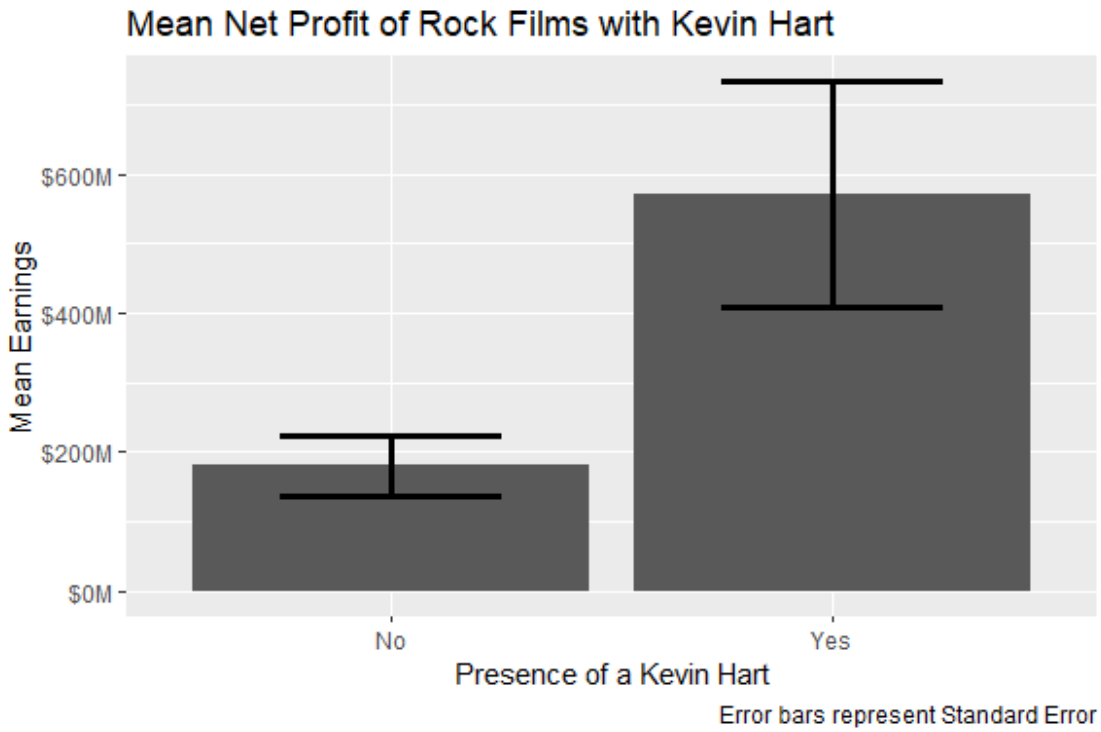


To start off, we can see that only about 20% of movies with the Rock have a jungle in them. And less than 15% of movies also have Kevin Hart. And while it's true that the Rock has been acting for longer than Hart, it's only by about 3 years. The earliest movie in the Rock's filmography was in 1999, while Hart's was in 2002. There were only 3 movies made with the Rock in them that came out before Kevin Hart's screen debut.

```
rock_films %>%
  group_by(Jungle) %>%
  summarize(mean = mean(Net, na.rm = TRUE), n = n(),
            sd = sd(Net, na.rm = TRUE), se = sd/sqrt(sum(n))) %>%
  ggplot(aes(x = Jungle, y = mean)) + geom_col() +
  geom_errorbar(aes(ymin = mean-se, ymax = mean+se), width = 0.5, size = 1.2)
+
  scale_y_continuous(labels = label_number(prefix = "$", suffix = "M",
                                           scale = 1 / 1e6)) +
  labs(title = "Mean Net Profit of Rock Films with Jungles",
       x = "Presence of a Jungle", y = "Mean Earnings",
       caption = "Error bars represent Standard Error")
```

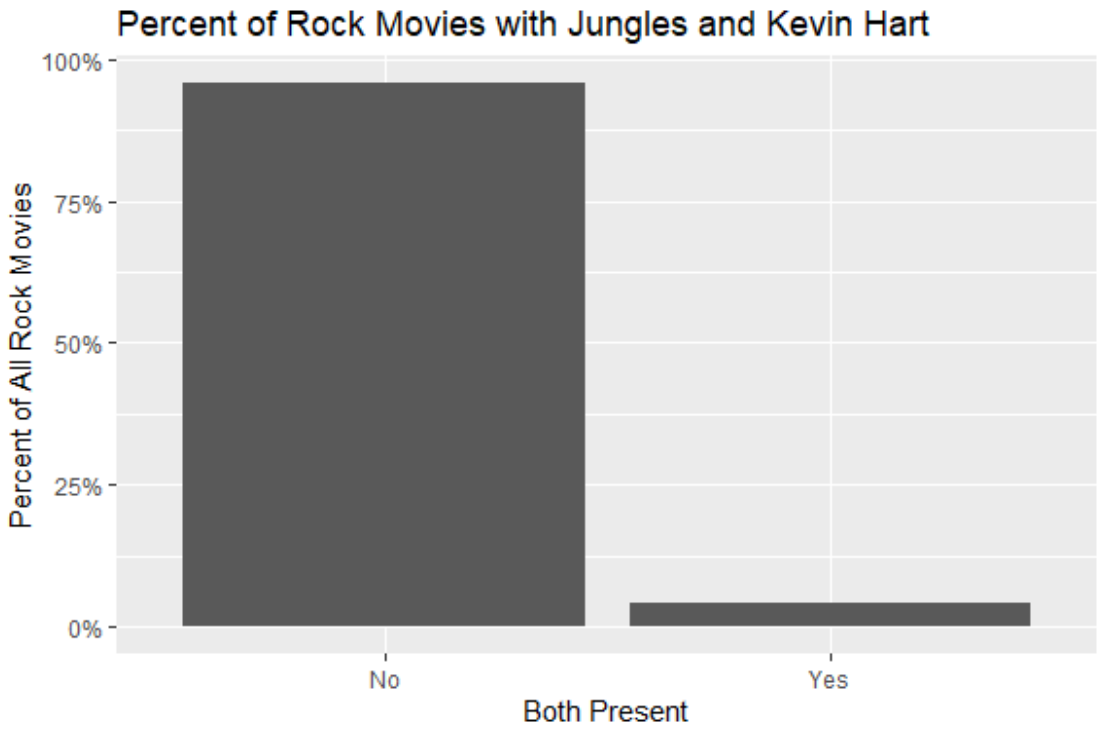



```
rock_films %>%
  group_by(Hart) %>%
  summarize(mean = mean(Net, na.rm = TRUE), n = n(),
            sd = sd(Net, na.rm = TRUE), se = sd/sqrt(sum(n))) %>%
  ggplot(aes(x = Hart, y = mean)) + geom_col() +
  geom_errorbar(aes(ymin = mean-se, ymax = mean+se), width = 0.5, size = 1.2)
+
  scale_y_continuous(labels = label_number(prefix = "$", suffix = "M",
                                           scale = 1 / 1e6)) +
  labs(title = "Mean Net Profit of Rock Films with Kevin Hart",
       x = "Presence of a Kevin Hart", y = "Mean Earnings",
       caption = "Error bars represent Standard Error")
```

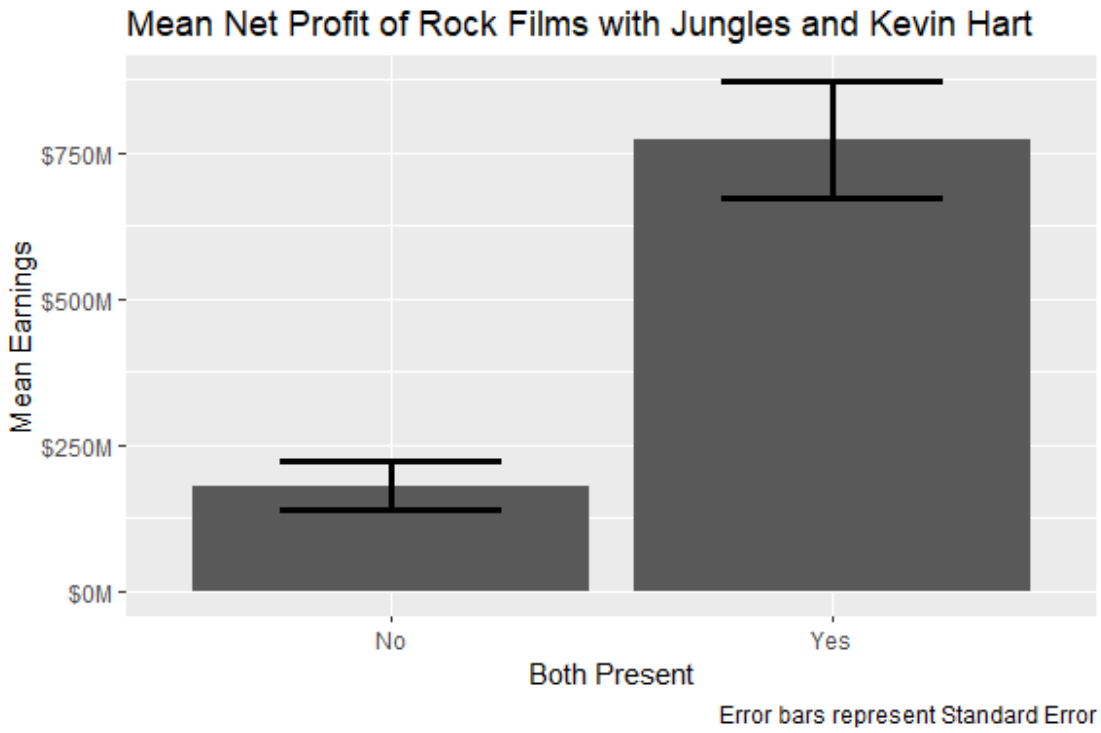


Looking into the net earnings of movies with the Rock, and the presence of Kevin Hart and/or a jungle, we can see that movies with a jungle don't earn that much more than movies without. However, adding Kevin Hart to a movie with the Rock seems like a surefire way to increase net earnings on a film. But this brings up a final question. How much more could a movie make with both a jungle and Kevin Hart?

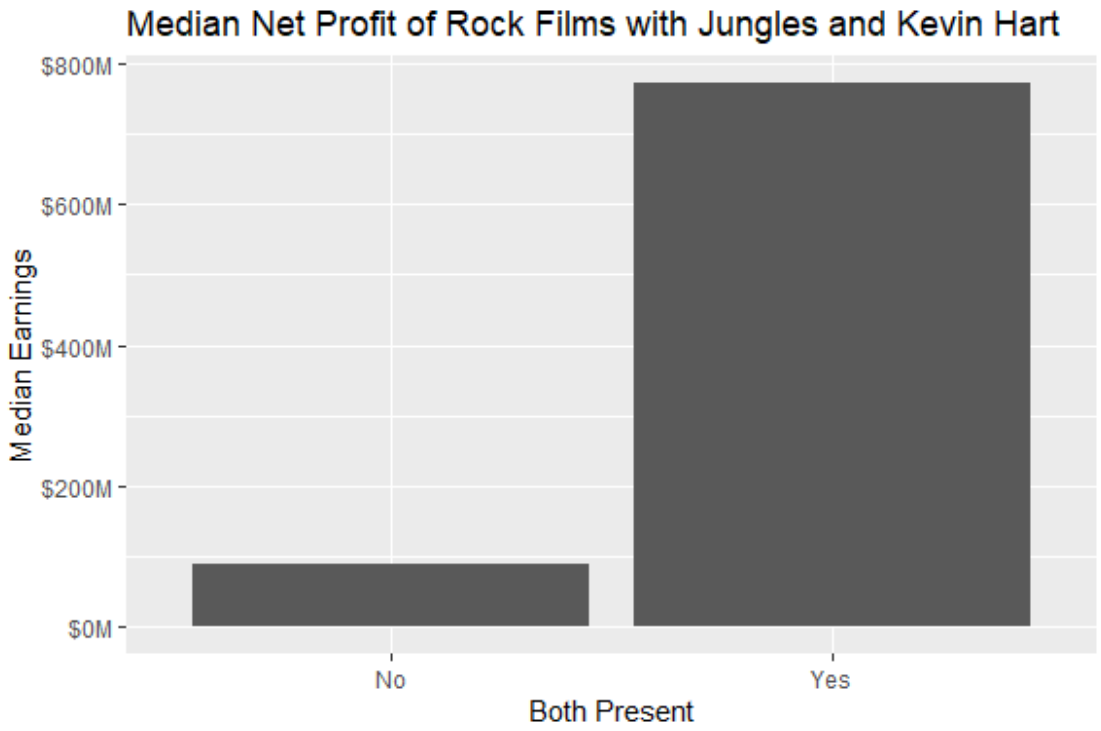
```
rock_films <- rock_films %>%
  mutate(Both = ifelse(Jungle == "Yes" & Hart == "Yes", "Yes", "No"))

rock_films %>%
  ggplot(aes(x = Both, y = (..count..)/sum(..count..))) +
  geom_bar() + labs(title ="Percent of Rock Movies with Jungles and Kevin Har
t",
                    x = "Both Present", y = "Percent of All Rock Movies") +
  scale_y_continuous(labels = scales::percent)
```



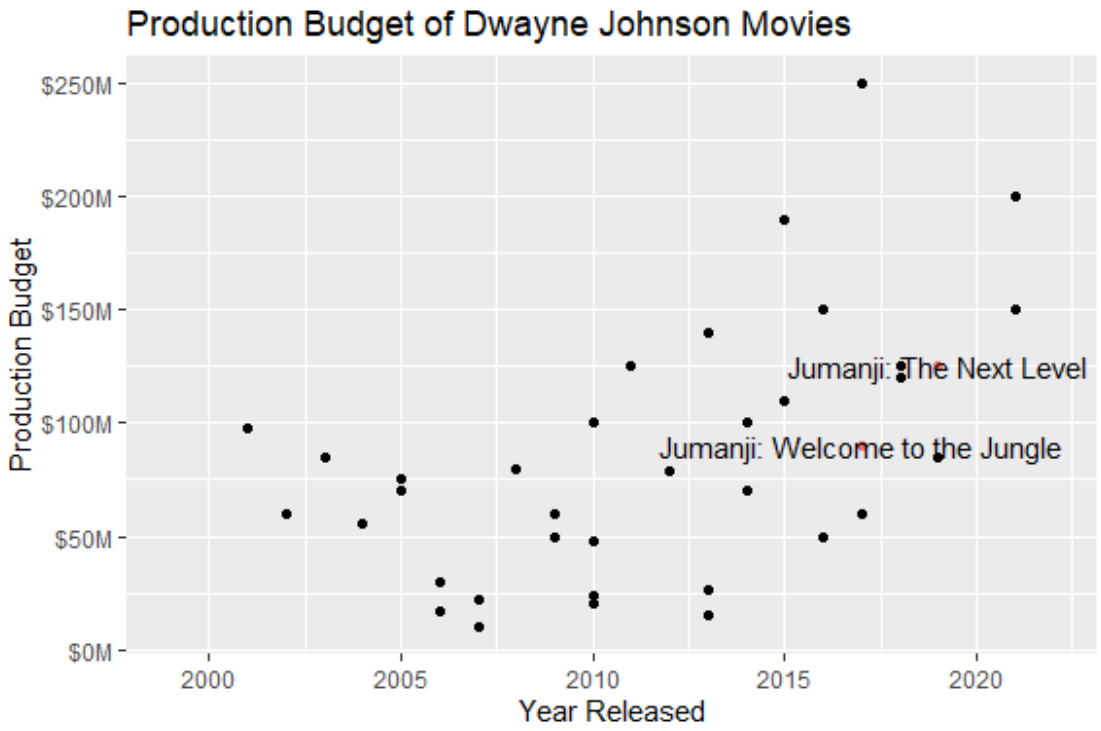

```
rock_films %>%
  group_by(Both) %>%
  summarize(mean = mean(Net, na.rm = TRUE), n = n(),
            sd = sd(Net, na.rm = TRUE),
            se = sd/sqrt(sum(n))) %>%
  ggplot(aes(x = Both, y = mean)) + geom_col() +
  geom_errorbar(aes(ymin = mean-se, ymax = mean+se), width = 0.5, size = 1.2)
+
  scale_y_continuous(labels = label_number(prefix = "$", suffix = "M",
                                           scale = 1 / 1e6)) +
  labs(title = "Mean Net Profit of Rock Films with Jungles and Kevin Hart",
       x = "Both Present", y = "Mean Earnings",
       caption = "Error bars represent Standard Error")
```



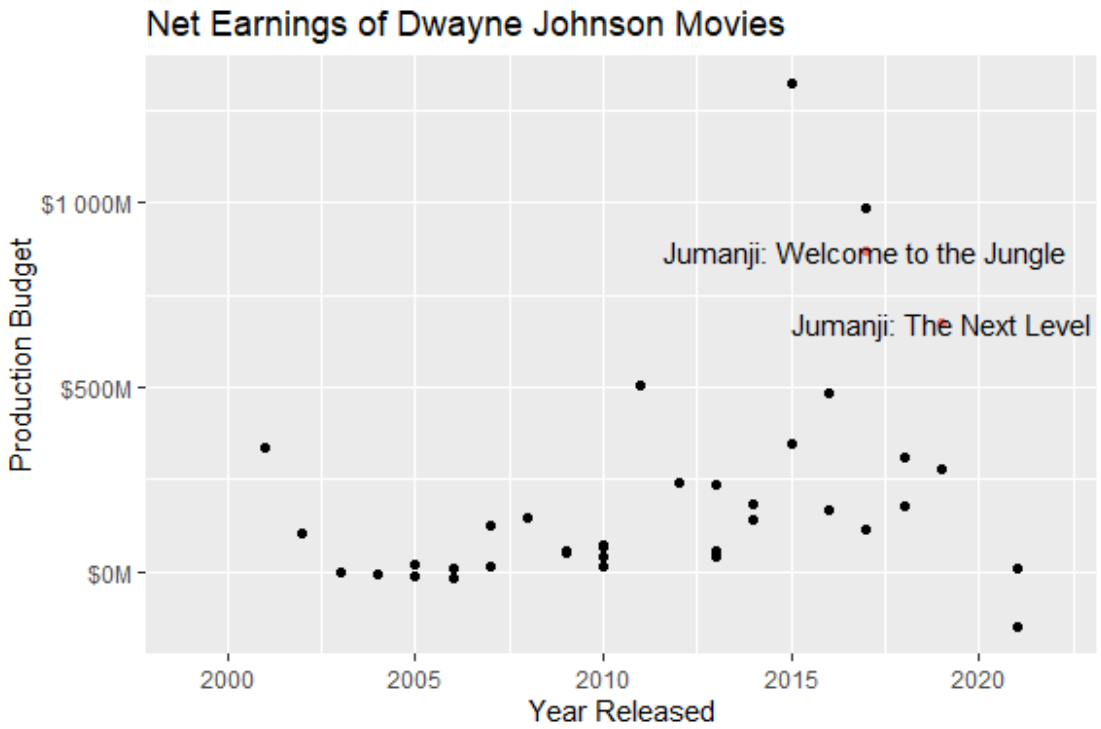

```
rock_films %>%
  group_by(Both) %>%
  summarize(mean = mean(Net, na.rm = TRUE), n = n(),
            median = median(Net, na.rm = TRUE),
            sd = sd(Net, na.rm = TRUE),
            se = sd/sqrt(sum(n))) %>%
  ggplot(aes(x = Both, y = median)) + geom_col() +
  scale_y_continuous(labels = label_number(prefix = "$", suffix = "M",
                                           scale = 1 / 1e6)) +
  labs(title = "Median Net Profit of Rock Films with Jungles and Kevin Hart",
       x = "Both Present", y = "Median Earnings")
```




```
both <- c("Jumanji: Welcome to the Jungle" , "Jumanji: The Next Level")

g <- rock_films %>%
  ggplot(aes(x = Year, y = ProductionBudget))
g + geom_point() +
  geom_point(data=subset(rock_films, Title %in% both),
             aes(Year, ProductionBudget, color = "firebrick")) +
  scale_y_continuous(labels = label_number(prefix = "$", suffix = "M",
                                           scale = 1 / 1e6)) +
  geom_text(data=subset(rock_films, Title %in% both),
            aes(Year, ProductionBudget, label = Title), check_overlap = T) +
  labs(title = "Production Budget of Dwayne Johnson Movies",
       x = "Year Released", y = "Production Budget") +
  theme(legend.position = "none")
```




```
h <- rock_films %>%
  ggplot(aes(x = Year, y = Net))
h + geom_point() +
  geom_point(data=subset(rock_films, Title %in% both),
             aes(Year, Net, color = "firebrick")) +
  scale_y_continuous(labels = label_number(prefix = "$", suffix = "M",
                                           scale = 1 / 1e6)) +
  geom_text(data=subset(rock_films, Title %in% both),
            aes(Year, Net, label = Title), check_overlap = T) +
  labs(title = "Net Earnings of Dwayne Johnson Movies",
       x = "Year Released", y = "Production Budget") +
  theme(legend.position = "none")
```



There are only 2 movies in The Rocks entire filmography that contain both a jungle and actor Kevin Hart. That's about 4.17% of all movies with the Rock. However, these movies (Jumanji: Welcome to the Jungle, and Jumanji: The Next Level) netted an average of over $750 million. This is far above the rest of the films which averaged less than $250 million. The gap only gets larger if you look at median net earnings. A potential cause for this could be due to the production budget of each movie. However, the more expensive of these two movies to produce (Jumanji: The Next Level), is only the 10th most expensive Dwayne Johnson movie ever made. Jumanji: Welcome to the Jungle comes it at 17th. It's unlikely that the production budget is the reason for the marked increase in net earnings from these two movies when compared to others in the Dwayne Johnson filmography.

So, while it's incorrect to say that every movie with Dwayne "The Rock" Johnson has a Jungle and/or Kevin Hart, it isn't incorrect to say that movies with Kevin Hart or Kevin Hart and a jungle tend to earn more on average. And it seems to be fair to say that if a company is producing a movie with the Rock and they want to earn good money, then they should also cast Kevin Hart and figure out why they're in a jungle.