Is There a Relationship between Disney Stock Value and when Movies are Released?

When I looked at the stock of Disney the Monday after Endgame was released, I saw a huge spike in the stock value. It was pretty much known that Endgame was going to be a huge box office success and had an amazing opening weekend in sales. This is what led to my idea for my second project. For my project, I am going to use some of my knowledge I learned from quant mod to see if there is a correlation where Disney stock value goes up when they release a blockbuster and where the Disney value goes down when a movie is released that is not from Disney. Additionally, I want to see if there is a correlation between how much the Disney stock goes up or down depending on how much the movie makes on opening weekend. Also, I will see if info about the movie (such as what type of movie it is, if it is a sequel, and if it is a popular saga) relates to how the movie affects the Disney stock value. It is important to note the exact stock values of Disney I will compare will be a week before a movie is released and first weekday after a movie is released.

Retrieving and Rearranging our data.

For this project we are going to need to use the packages quantmod,lubridate, and chron

```r
library(quantmod)
library(lubridate)
library(chron)
```

The info we are using is over the top 26 grossing movies since 2009. It is a mix of two spreadsheets, that can be found on https://www.boxofficemojo.com/alltime/world/ and https://www.boxofficemojo.com/alltime/weekends/

```r
setwd("~/Documents")
movies<-read.csv("movie data.csv",header=TRUE)
head(movies)
```

```
##                            Movie sales.millions. OWS.Millions.   Released
## 1                          Avatar          2788.0         241.6 2009/12/18
## 2                Avengers: Endgame        2192.0        1223.6   2019/4/26
## 3 Star Wars: The Force Awakens      2068.2         529.0 2015/12/18
## 4          Avengers: Infinity War       2048.0         640.5  2018/4/27
## 5                  Jurassic World        1671.0         525.5  2015/6/12
## 6            Marvel's The Avengers       1518.0         392.5   2012/5/4
##   Disney.
## 1      no
## 2     yes
## 3     yes
## 4      no
## 5      no
## 6     yes
```

This data includes how much the movie made in total, how much the movie made opening weekend, if it is a Disney movie, and when the movie was released.

I decided to change column names of the data.frame to make it easier to understand the information from each column.

```r
names(movies)<-c("movie","sales (millions)",
"Opening Weekend Sales (millions)","Release", "Disney Movie?")
```

Now put the data in order from when the film was released

```r
movies<-movies[order(as.Date(as.character(movies$Release))),]
```

Put `movies$Release` as dates. To do this, we have to first have the information as a character with as.character and then do as.Date on the character for all of the info. `movies$Release`

```r
movies$Release<-as.Date(as.character(movies$Release))
```

Getting Disney Stock Information

Bar plot

Before I am able to start pulling stock informaton from this data, I need to make sure the dates I will be using when pulling stock information fall on a weekday and/or not on holidays. To get started, I am going to see what weekday the movies were released. This will be done with the weekdays function which returns what day of the week the date falls on. Also, to make things a little easier to read, I will use a sorted table.

```r
sort(table(weekdays(movies$Release)))
```

```
##
## Wednesday    Friday
##         4        22
```

We see all of the dates of movie releases fall on either a Wednesday or Friday. This means we should most likely be able to use the dates the week before the movies were released because they fall on either Wednesday or friday.

```r
#create column in movies for a week before the movie was released
movies$"week before Release"<-movies[,4]-7


#get dates of the first weekday after the movies release. Most of the movies are released
#on Friday. This means we need to go three days ahead to get to the #first weekday
#after the movie is released
movies$"first weekday after Release"<-movies$Release+3

#find which movies were released on a Wednesday
which(weekdays(movies$Release)=="Wednesday")
```

```
## [1] 2 5 8 9
```

We need to change the `movies$"first weekday after Release"` where the movies released on Wednesday. The reason for this is so the day we have saved for the first weekday after the movie is released on these days is Thursday. For more clarification, we originally made it so all the movies had it so the first weekday was three days after the movies was released. This moved movies released on Wednesday to Saturday. To change this data so that movies that were released on Wednesday have it so the first week day after the movie was released is now Thrusday, we change the weekday column by moving it two days back for films that were released on a Wednesday.

```r
movies[c(2,5,8,9),6]<-movies[c(2,5,8,9),6]-2
```

The last thing we need to do is make sure all of these dates are not on holidays to make sure the only reason we possbile (and most likely) have some missing stock values dates is only due to quantmod not having the information. This can easily be done with the `is.holiday` function from the package chron.

```r
for( i in c(4,6,7)){
  print(which(is.holiday(movies[,i])==TRUE))
}
```

```
## integer(0)
```

```
## integer(0)
## integer(0)
```

no dates are on a holiday, so we are good to start pulling stock information.

```r
#Disney st0ck
getSymbols("DIS",from="2009-01-01",to="2019-05-04",auto.asign=FALSE)
```

```
## 'getSymbols' currently uses auto.assign=TRUE by default, but will
## use auto.assign=FALSE in 0.5-0. You will still be able to use
## 'loadSymbols' to automatically load data. getOption("getSymbols.env")
## and getOption("getSymbols.auto.assign") will still be checked for
## alternate defaults.
##
## This message is shown once per session and may be disabled by setting
## options("getSymbols.warning4.0"=FALSE). See ?getSymbols for details.
```

```
## [1] "DIS"
```

```r
head(DIS)
```

```
##            DIS.Open DIS.High DIS.Low DIS.Close DIS.Volume DIS.Adjusted
## 2009-01-02    22.76    24.03   22.50     23.92    9796600     20.84922
## 2009-01-05    23.51    24.00   23.29     23.50   11675100     20.48314
## 2009-01-06    23.81    24.83   23.80     24.31   12070800     21.18916
## 2009-01-07    23.89    23.89   22.92     23.18   12300700     20.20422
## 2009-01-08    23.10    23.22   22.51     22.90   12600700     19.96017
## 2009-01-09    23.12    23.12   22.23     22.31   11489700     19.44591
```

```r
# pull from DIS, the DIS (Disney) stock a week before the movies were released
movies$"stock week before"<-DIS[unlist(sapply(1:26,function(x) which(index(DIS)==movies[x,6]))),4]
```

```
## Error in `$<-.data.frame`(`*tmp*`, "stock week before", value = structure(c(31.700001, : replacement
```

We see that quantmod is missing stock values of Disney for two days because it says there are only 24 replacements. To fix this, we will run a loop that moves the two dates `movies$"week before Release"` that are missing in the DIS stock to one day earlier. The reason we move them one day earlier is because we know all `movies$"week before Release"`fall on a Wednesday and Friday so we know a day ealier will also be a weekday (Tuesday and Thursday). Agan this is necessary because stocks are only bought and sold Monday-Friday. It is important to note there are warning messages, but these are avoided with a for loop.

```r
#first get the dates in movies$ "week before Release" we have in DIS from quantmod
dates<-index(DIS[unlist(sapply(1:26,function(x) which(index(DIS)==movies[x,6]))),])


for(i in 1:2){
  #change the date of the first date in movies
  #movies$"week before Release" not in DIS info to a day earlier.
  movies[(which(movies$"week before Release"!=dates)),6]<-
    movies[(which(movies$"week before Release"!=dates)),6]-1

#updates the dates we have in movies$"week before
  #Release"we have in quantmond
  dates<-index(DIS[unlist(sapply(1:27,function(x) which(index(DIS)==movies[x,6]))),])
}
```

```
## Warning in `!=.default`(movies$"week before Release", dates): longer object
## length is not a multiple of shorter object length
```

```
## Warning in `!=.default`(movies$"week before Release", dates): longer object
## length is not a multiple of shorter object length
```

```
movies$"week before Release"
```

```
##  [1] "2009-12-11" "2011-06-20" "2011-07-08" "2012-04-27" "2012-06-18"
##  [6] "2012-11-02" "2013-04-26" "2013-11-18" "2014-06-09" "2015-03-27"
## [11] "2015-04-24" "2015-06-05" "2015-07-02" "2015-12-10" "2016-04-28"
## [16] "2016-12-08" "2017-03-09" "2017-04-06" "2017-12-07" "2018-02-08"
## [21] "2018-04-19" "2018-06-07" "2018-06-14" "2018-12-13" "2019-02-28"
## [26] "2019-04-18"
```

```
dates
```

```
##  [1] "2009-12-11" "2011-06-20" "2011-07-08" "2012-04-27" "2012-06-18"
##  [6] "2012-11-02" "2013-04-26" "2013-11-18" "2014-06-09" "2015-03-27"
## [11] "2015-04-24" "2015-06-05" "2015-07-02" "2015-12-10" "2016-04-28"
## [16] "2016-12-08" "2017-03-09" "2017-04-06" "2017-12-07" "2018-02-08"
## [21] "2018-04-19" "2018-06-07" "2018-06-14" "2018-12-13" "2019-02-28"
## [26] "2019-04-18"
```

```
which(movies$"week before Release"!=dates)
```

```
## integer(0)
```

We see the dates are the same so we will add the stock closing value from the dates in movies$"week before Release" into the movies data frame.

```
movies$"stock week before Release"<-DIS[dates,4]
```

Now we will get the opening stock prices for the first week day after movies were released. Note: There is a reason we are doing opening prices for the first weekday after the movies were released. If there is a correlation where Disney stock value goes up when they release a blockbuster, the idea would be to buy a week before the movie was released and then sell immediatly the first weekday after the movie was released.

```
#get stock vales the first weekday after the movies were released
movies$"stock first week day after Release"<-
  DIS[unlist(sapply(1:26,function(x) which(index(DIS)==movies[x,7]))),1]
```

```
## Error in `$<-.data.frame`(`*tmp*`, "stock first week day after Release", : replacement has 21 rows, c
```

Similary to before, we see there are 5 missing dates from `movies$"stock first week day after Release"`. However, from looking at how the Disney stock behaves when movies are released, they tend to go back down immediatly after that first week day after the movie was Released. Because of this, I do not want to change the dates of `movies$"stock first week day after Release"`, but rather just remove the movies that do no stock values for `movies$stock first week day after Release`. This will be done with a for loop. Again, there will be some warning messages, but again they can be ignored due to the for loop.

```
#get the dates in movies$" irst week day after Release" we have in DIS from quantmod.
dates<-index(DIS[unlist(sapply(1:26,function(x) which(index(DIS)==movies[x,7]))),])

#use for loop to remove the movies we do not have stock in DIS
for( i in 1:5){
  movies<-movies[-min(which(movies[,7]!=dates)),]
}
```

```
## Warning in `!=.default`(movies[, 7], dates): longer object length is not a
## multiple of shorter object length
```

```
## Warning in `!=.default`(movies[, 7], dates): longer object length is not a
## multiple of shorter object length

## Warning in `!=.default`(movies[, 7], dates): longer object length is not a
## multiple of shorter object length

## Warning in `!=.default`(movies[, 7], dates): longer object length is not a
## multiple of shorter object length

## Warning in `!=.default`(movies[, 7], dates): longer object length is not a
## multiple of shorter object length
```

Now, we can get the opening stock prices for the first weekday after movies were released.

```
movies$"stock first week day after Release"<-
  DIS[unlist(sapply(1:22,function(x) which(index(DIS)==movies[x,7]))),1]
```

I am going to clean up the movie data frame a little by moving the order of the columns around.

```
movies<-movies[,c(1,2,3,5,6,4,7,8,9)]
```

After collecting the stock information for a week before and a week after, we can calculate the returns between the stock information. This is simple and in generic terms done by subtracting the later stock value by the earlier stock value then diving them by the earlier stock times 100. We will put this as its own column in movies

```
movies$'stock returns (percentage)'<-(as.vector(movies$'stock first week day after Release')
-as.vector(movies$'stock week before Release'))/as.vector(movies$'stock week before Release')*100
```

Now I want to just add a column in movies that says if the stock went up by just being yes if the stock return is greater than or equal to zero and no if the stock return is negative. This is achieved with appending an empty character vector in a for loop and then using the factor of the vector to be our column in movies that tells us if the stock went up.

```
yn<-as.character()
for(i in 1:21){
  if(movies[i,10]>=0) {
      yn<-append(yn,"yes")
        } else {
          yn<-append(yn,"no")
              }
    }
movies$"did stock go up?"<-factor(yn)
```

Analyzing our Data from movies

It is finally time to start analyzing the data! However, before I start graphing anything. I want to see the amount of times that the Disney stock value relates to a disney movie or movie from another company. i.e. (the stock goes up when a Disney blockbuster is released and the Disney stock goes down when a blockbuser from another company is released. This is just done by letting a varaible (same) equal zero and creating a for loop that has the variable go up by one if the Disney stock goes up by one when Disney Movie is released or it goes up by one if the Disney Stock goes down when a non-Disney movie is released. The varaible goes up by one if `movie$"Disney Movie?"=movie$"did stock go up"` for the column we are looking at. After we have the amount of times this is true, we divide our variable by our total number of movies.

```
same<-0

for(i in 1:nrow(movies)){
```

```
  if(movies[i,4]==movies[i,11]){
    same=same+1
  }
}

same/nrow(movies)
```

[1] 0.8095238

We see that the amount of times that a disney stock's value relates to if a disney movie or movie from another company is released is about 80% of the time.

We will graphically show the amount of times that a disney stock's value relates to if a disney movie or movie from another company is released.

```
par(mar=c(3,3,3,3))
barplot(table(movies[,4],movies[,11]),beside=TRUE,col=c("red","Blue")
,names.arg=c("Disney Stock Went Down","Disney Stock Went up "),ylab="Amount of Time Event Occurred")
legend("topleft",fill=c("red","blue"),legend=c("Non Disney Movie Released","Disney Movie Released"))
```
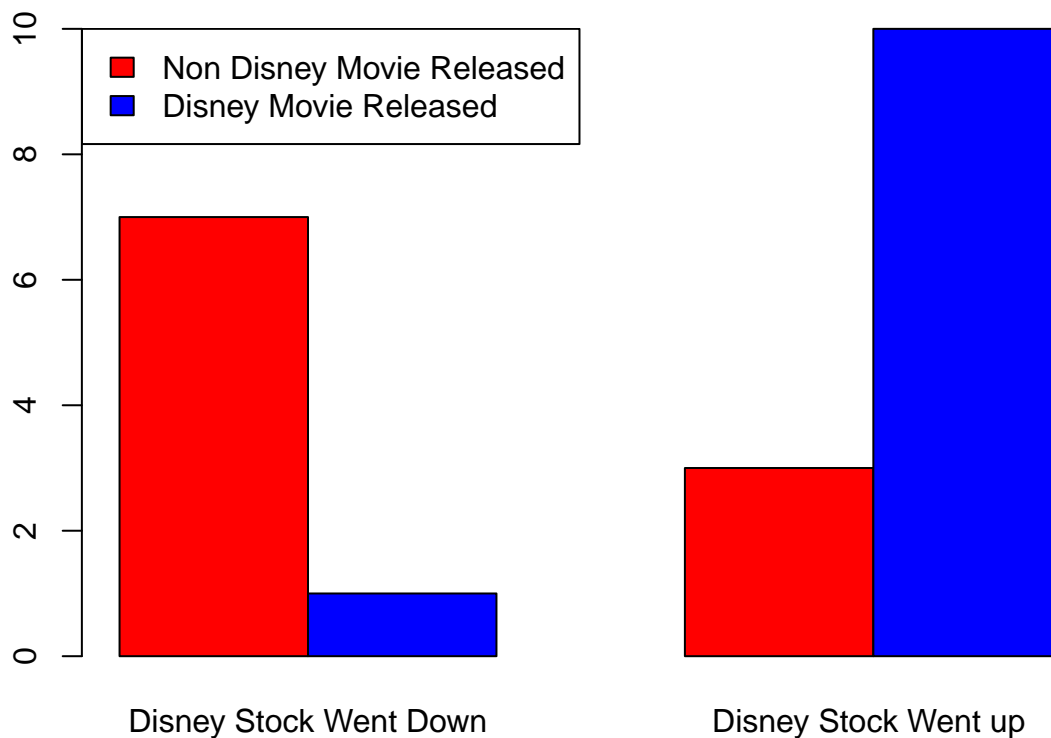

It is easy to see that the Disney stock value dropped only once when a Disney movie was released. This is easy to find by using which()

```
movies[which(movies$"Disney Movie?"=="yes" & movies$"did stock go up?"=="no"),1]
```

[1] Star Wars: The Force Awakens
26 Levels: Aquaman Avatar Avengers: Age of Ultron ... Transformers: Dark of the Moon

Now, we see a correlation of the stock price and type of movie(Disney and non Disney). It seems reasonable to see if there is a correlation between how much the Disney stock goes up or down depending on how much the movie makes on opening weekend?
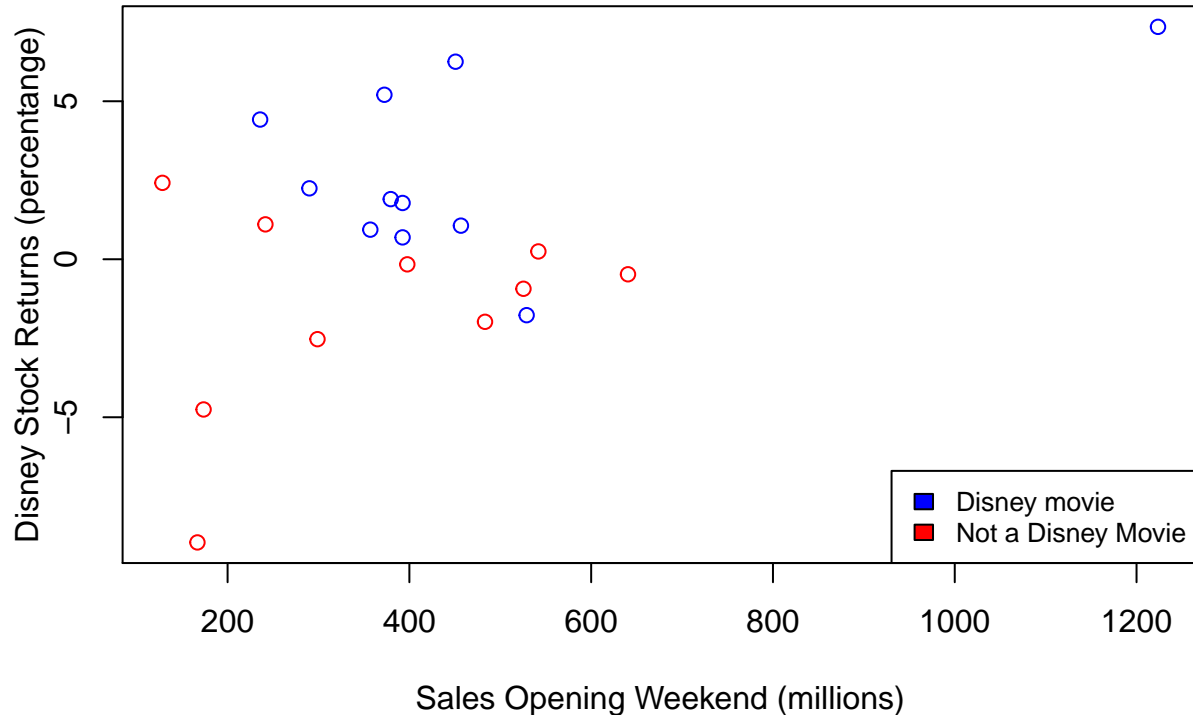
```
par(mfrow=c(1,1),mar=c(5,4,3,0.5))
cols<-c("red","blue")
```

```r
plot(movies[,3],movies[,10],col=cols[movies[,4]],xlab="",ylab="")

mtext(text="Sales Opening Weekend (millions)",side=1,line=3)
mtext(text="Disney Stock Returns (percentange)",side=2,line=2)
legend("bottomright",fill=c("blue","red"),legend=c("Disney movie","Not a Disney Movie"),cex=.8)
```


It is somehwat suprising to see there is no correlation between how much the Disney stock goes up or down depending on how much the movie makes on opening weekend. However, we can see groups for movies with returns greater than 4%, movies with returns lower than -2.5%, and then the movies in between. Although there is no correlation between how much the Disney stock goes up or down depending on how much the movie makes on opening weekend, it would be interesting to see if the information about the movies themselves relates to how the movie effects the Disney Stock returns. This can be done by creating a dendrogram that uses information about the movie.

Creating the Dendrogram

```r
setwd("~/Documents")
numMovies<-read.csv("pure movie info.csv",header=TRUE)

#put movies in order from when they were released
numMovies<-numMovies[order(as.Date(as.character(numMovies$Release))),]
head(numMovies)
```

```
##                                          Movie   Released Disney
## 1                                        Avatar 2009/12/18     no
## 8   Harry Potter and the Deathly Hallows Part 2  2011/7/15     no
## 6                           Marvel's The Avengers  2012/5/4    yes
## 18                                      Skyfall  2012/11/9     no
## 13                                     Iron Man 3  2013/5/3    yes
## 21                                      Furious 7  2015/3/27    no
##    action.movie cartoon.kids.movie MCU.movie Star.wars.movie Sequel
## 1          yes                 no        no              no     no
```

7

```
## 8              yes              no       no            no    yes
## 6              yes              no      yes            no    yes
## 18             yes              no       no            no    yes
## 13             yes              no      yes            no    yes
## 21             yes              no       no            no    yes
##    New.leading.actor.actors.
## 1                        yes
## 8                         no
## 6                         no
## 18                        no
## 13                        no
## 21                        no
```

numMovies contains information about the movies themselves, such as if it is in a saga, if it has a new lead actor/actors, and if it is a sequel. I decided to not include information about the time period when the movie was released. I wanted to only look at the information about the movie itself rather than what time of the year it was released.

First, we will change the rownames of numMovies to the actual movie names, and then we will just remove the columns `numMovies$Release` and `numMovies$movie` because they do not contain information that would be helpful when creating the dendrogram.

```r
#chagnge row names
rownames(numMovies)<-movies[,1]
#remove first and secodn column
numMovies<-numMovies[,-(1:2)]
```

Since a dendrogram is based on distance, all of the values in numMovies needs to be changed to numeric values.

```r
#change each individual column to numeric values
for( i in 1:ncol(numMovies)){
  numMovies[,i]<-as.numeric(numMovies[,i])
}
head(numMovies)
```

```
##                                        Disney action.movie
## Avatar                                      1          2
## Harry Potter and the Deathly Hallows Part 2 1          2
## Marvel's The Avengers                       2          2
## Skyfall                                     1          2
## Iron Man 3                                  2          2
## Furious 7                                   1          2
##                                        cartoon.kids.movie MCU.movie
## Avatar                                              1         1
## Harry Potter and the Deathly Hallows Part 2         1         1
## Marvel's The Avengers                               1         2
## Skyfall                                             1         1
## Iron Man 3                                          1         2
## Furious 7                                           1         1
##                                        Star.wars.movie Sequel
## Avatar                                              1      1
## Harry Potter and the Deathly Hallows Part 2         1      2
## Marvel's The Avengers                               1      2
## Skyfall                                             1      2
## Iron Man 3                                          1      2
## Furious 7                                           1      2
```

```
##                                                       New.leading.actor.actors.
## Avatar                                                                        2
## Harry Potter and the Deathly Hallows Part 2                                   1
## Marvel's The Avengers                                                         1
## Skyfall                                                                       1
## Iron Man 3                                                                    1
## Furious 7                                                                     1
```

The final thing we have to do before creating our dendrogram is find which movies had great postive returns, great negative returns, and the returns in between. We will seperate these based on what we saw in our xy plot with great returns>=4%, great negative returns<=-2.5%, and everything in the middle is together. This will use a for loop where the type of return gives us a 1,2,or 3 and then making that its own column in numMovies.

```r
#create empty vector
rets<-as.numeric()

for( i in 1:21){
  if(movies[i,10]>=4){ #great positive returns append rets with 1
    rets<-append(rets,1)
  } else{
    if(movies[i,10]<4 & movies[i,10]>=-2.5){
      rets<-append(rets,2) # in between returns append rets with 2
    }
    else {
      rets<-append(rets,3) #great negative returns append rets with a 3
    }
  }
}

rets
```

```
##  [1] 2 2 2 3 1 2 2 2 2 2 2 2 2 2 2 1 2 1 3 3 2 1
```

```r
numMovies$type.return<-rets
```

Finally we are able to plot our dendorgram

```r
par(mar=c(3,1,1,11),mfrow=c(1,1))
#note: we do not want to include any stock info for this dendrogram,
#so we are not including our type of return the films had for Disney
dend<-hclust(dist(scale(numMovies[,-ncol(numMovies)])))
plot(as.dendrogram(dend),leaflab="none",horiz=TRUE)
cols<-c("green","blue","red")
mtext(side=4,line=0.25,at=1:21,text=rownames(numMovies)
,las=1,cex=0.6,col=cols[numMovies[,ncol(numMovies)]]) #our color of the dendrogram is based on
#what type of return the movies had for Disney stock
legend("topleft",fill=c("green","blue","red"),
legend=c("great positive Returns","mild returns","negative Returns"))
```

Clearly, there is no correlation between how the movies themselves relate to how the Disney Stock returns are affected.

Concluding Thoughts

It should be noted there could be some flaws in my project, most specifically not having enough movies to use and enough information with my dendrogram. In the end, nothing else can be said except there is somewhat of a correlation where Disney stock value goes up when they release a blockbuster and where the Disney value goes down when a movie is released that is not from Disney. Specifically, when Disney releases a blockbuster film, the stock price most likely will go up the day after the movie is released from a week before the movie is released. It may be too early to say that a trader should buy Disney stock a week before they release a blockbuster film, but no one can deny the fact that it is impossible to say that there is absolutely zero reason to believe that Disney stock price changes when they release a motion picture.