

# **R STUDIO ASSIGNMENT:**

# RStudio

## Programming:

1) First, I imported  
my Dataset of Hollywood's  
Most Profitable into R  
Studio and re- named it to  
df so it's easier to write  
my lines of code

Import Dataset

Name: df

Encoding: Automatic

Heading: ☒ Yes ☐ No

Row names: Automatic

Separator: Comma

Decimal: Period

Quote: Double (")

Comment: None

na.strings: NA

☐ Strings as factors

Input File

Film,Genre,Lead Studio,Audience score %,Profitability,Rott

27 Dresses,Comedy,Fox,71,5.3436218,40,160.308654,2008

(500) Days of Summer,Comedy,Fox,81,8.096,87,60.72,2009

A Dangerous Method,Drama,Independent,89,0.44864475,79,8.972

A Serious Man,Drama,Universal,64,4.382857143,89,30.68,2009

Across the Universe,Romance,Independent,84,0.652603178,54,2

Beginners,Comedy,Independent,80,4.471875,84,14.31,2011

Dear John,Drama,Sony,66,4.5988,29,114.97,2010

Enchanted,Comedy,Disney,80,4.005737082,93,340.487652,2007

Fireproof,Drama,Independent,51,66.934,40,33.467,2008

Four Christmases,Comedy,Warner Bros.,52,2.022925,26,161.834

Ghosts of Girlfriends Past,Comedy,Warner Bros.,47,2.0444,27

Gnomeo and Juliet,Animation,Disney,52,5.387972222,56,193.96

Going the Distance,Comedy,Warner Bros.,56,1.3140625,53,42.0

Good Luck Chuck,Comedy,Lionsgate,61,2.36768512,3,59.192128

He's Just Not That Into You,Comedy,Warner Bros.,60,7.1536,4

High School Musical 3: Senior Year,Comedy,Disney,76,22.9131

I Love You Phillip Morris,Comedy,Independent,57,1.34,71,20

Data Frame

Film	Genre	Lead Studio
27 Dresses	Comedy	Fox
(500) Days of Summer	Comedy	Fox
A Dangerous Method	Drama	Independent
A Serious Man	Drama	Universal
Across the Universe	Romance	Independent
Beginners	Comedy	Independent
Dear John	Drama	Sony
Enchanted	Comedy	Disney
Fireproof	Drama	Independent
Four Christmases	Comedy	Warner Bros
Ghosts of Girlfriends Past	Comedy	Warner Bros
Gnomeo and Juliet	Animation	Disney
Going the Distance	Comedy	Warner Bros
Good Luck Chuck	Comedy	Lionsgate
He's Just Not That Into You	Comedy	Warner Bros
High School Musical 3: Senior Year	Comedy	Disney
I Love You Phillip Morris	Comedy	Independent

Import Cancel

# RStudio

## Programming:

2) I installed Tidyverse package which allowed me create visualisations using R studios

3) I used the programming language R to view, transform and clean up my dataset

```
#Take a look at the data:

View(df)

#Load library:

install.packages("tidyverse")

#Import library

library(tidyverse)

# Check data types:
str(df)

# Check for missing values:
colSums(is.na(df))

#Drop missing values

df <- na.omit(df)
library(dplyr)
df <- df %>% na.omit()

> str(df)
'data.frame': 74 obs. of 8 variables:
 $ Film      : chr  "27 Dresses" "(500) Days of Summer" "A Dangerous Metho
d" "A Serious Man" ...
 $ Genre      : chr  "Comedy" "Comedy" "Drama" "Drama" ...
 $ Lead.Studio : chr  "Fox" "Fox" "Independent" "Universal" ...
 $ Audience..score.. : int  71 81 89 64 84 80 66 80 51 52 ...
 $ Profitability : num  5.344 8.096 0.449 4.383 0.653 ...
 $ Rotten.Tomatoes.. : int  40 87 79 89 54 84 29 93 40 26 ...
 $ Worldwide.Gross : num  160.31 60.72 8.97 30.68 29.37 ...
 $ Year       : int   2008 2009 2011 2009 2007 2011 2010 2007 2008 2008 ...

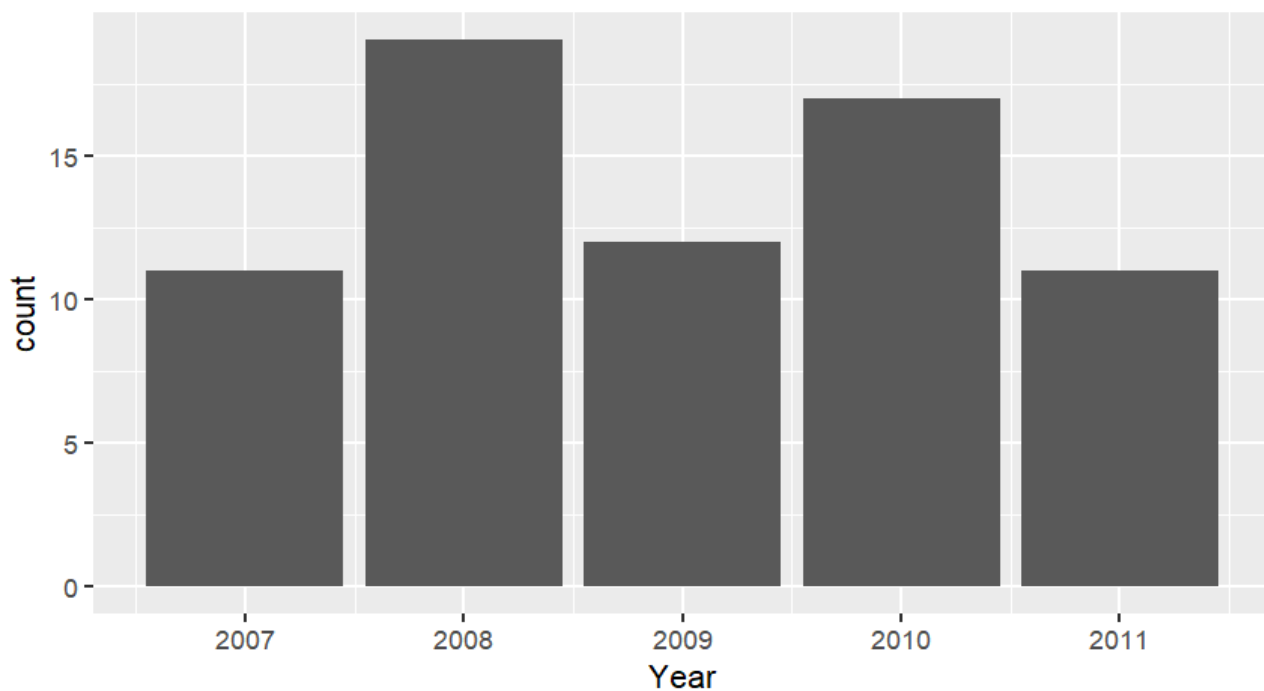
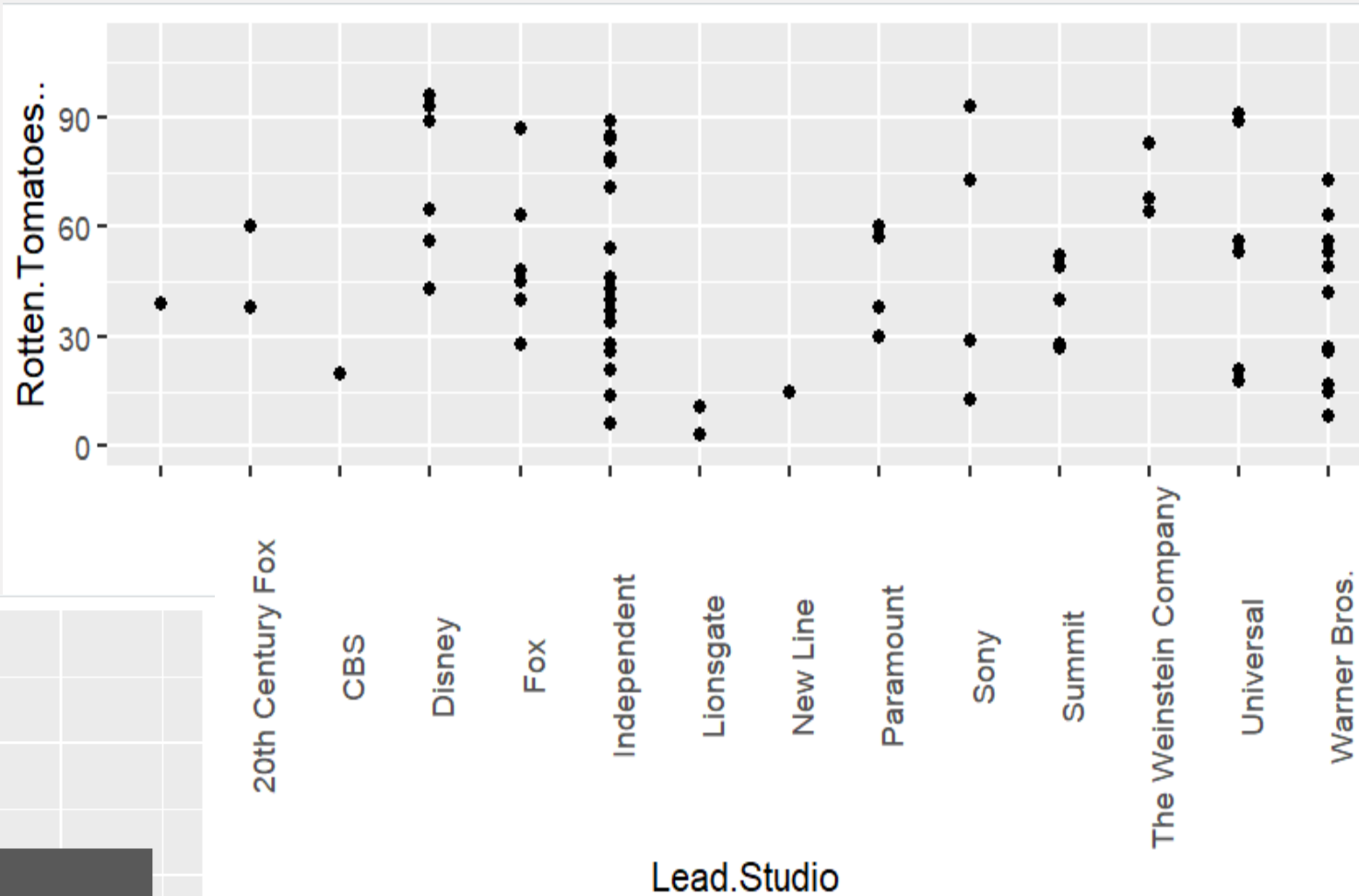
> install.packages("tidyverse")
WARNING: Rtools is required to build R packages but is not currently installed. P
lease download and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/UKoth/AppData/Local/R/win-library/4.3'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.3/tidyverse_2.0.0.zip'
Content type 'application/zip' length 430788 bytes (420 KB)
downloaded 420 KB

> colSums(is.na(df))
      Film      Genre      Lead.Studio Audience..score..
      0         0         0                 1
Profitability Rotten.Tomatoes.. Worldwide.Gross      Year
      3         1         0                 0

>
> #Drop missing values
>
> df <- na.omit(df)
> library(dplyr)
> df <- df %>% na.omit()
```

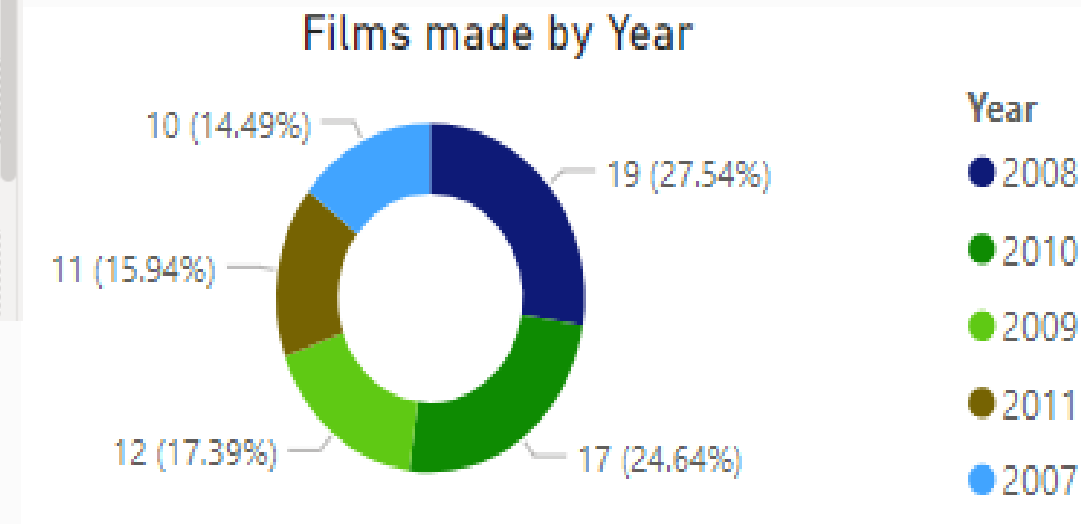
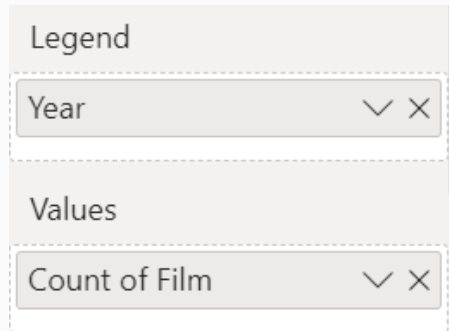
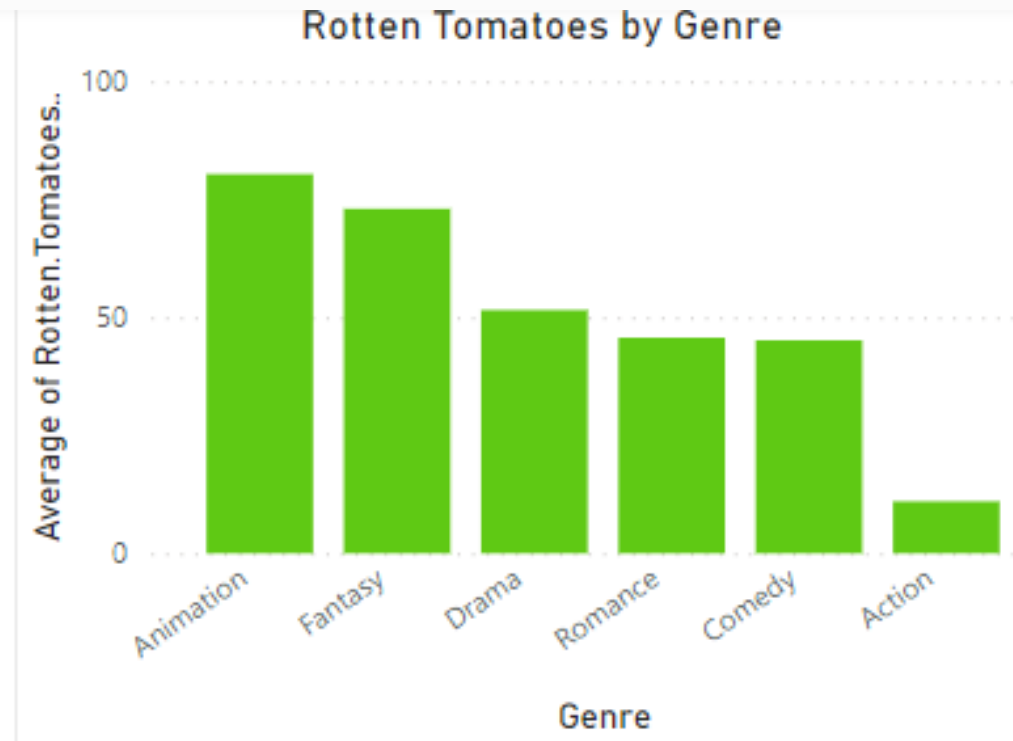
# RStudio Graphs:



# Power BI:

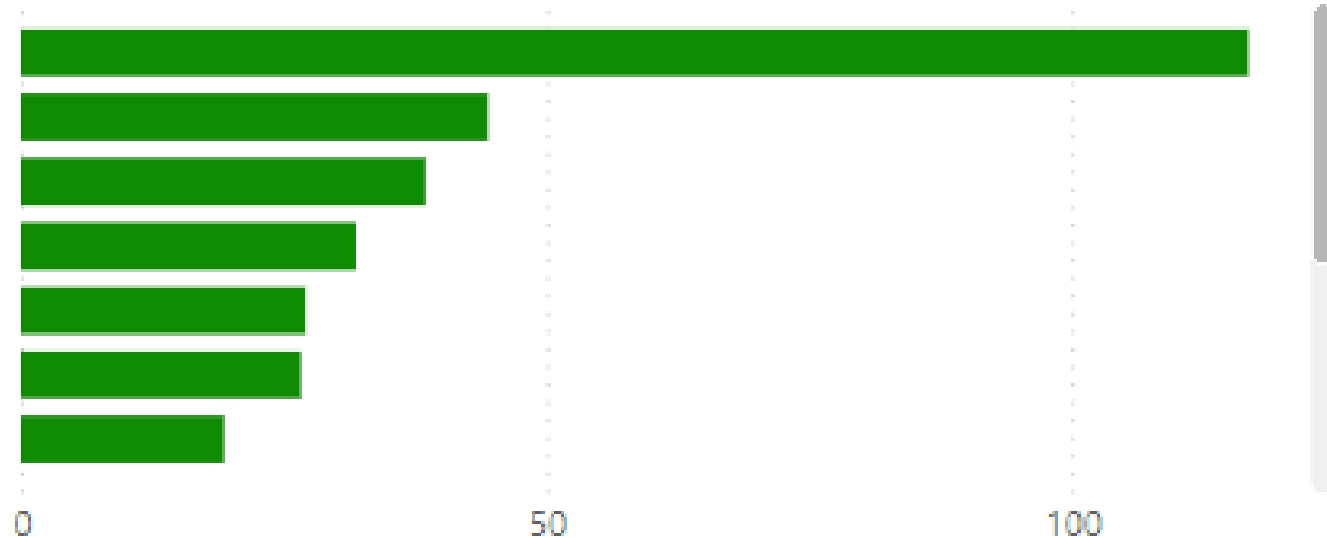
4) In Power BI I used the tools available to create different visualisations from the cleaned dataset to answer questions for the consumers.

5) I created the different visualisations in the client's colour scheme of blue, green and brown.



### Profitability by Studio

Independent  
Disney  
Warner Bros.  
Summit  
Fox  
Universal  
Sony



Sum of Profitability

Y-axis

Lead.Studio

X-axis

Sum of Profitability

Genre

Column y-axis

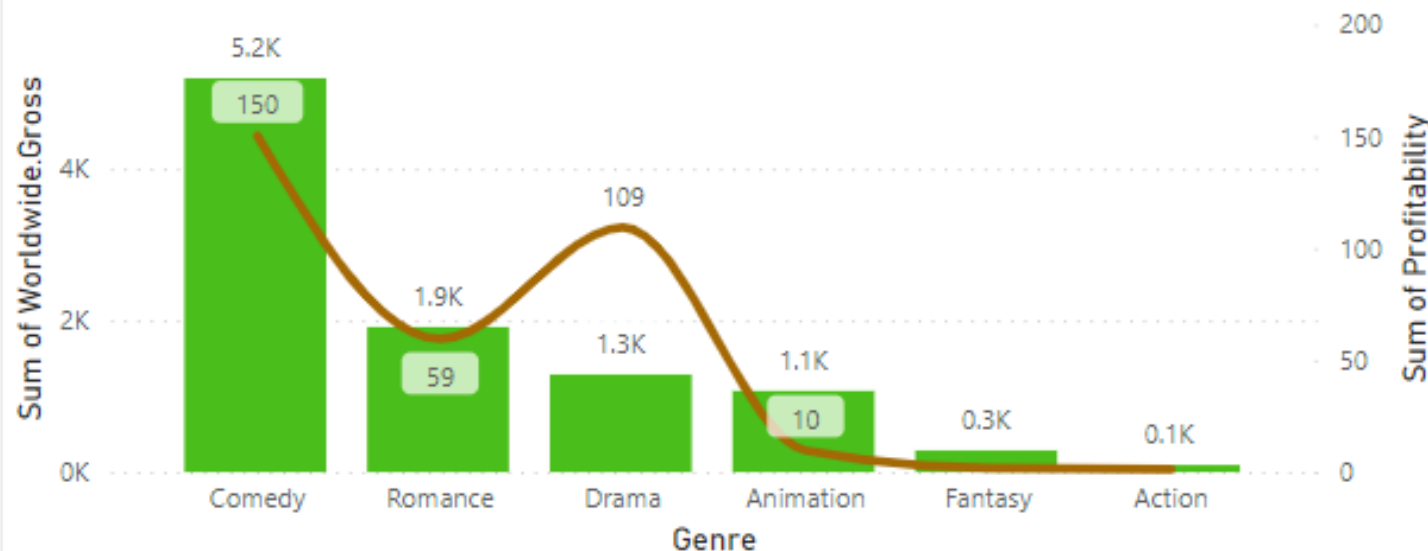
Sum of Worldwide.Gr...

Line y-axis

Sum of Profitability

### Worldwide Gross by Genre

Sum of Worldwide.Gross Sum of Profitability



# Power BI Dashboard:

