

# DATA SCIENCE PROJECT

Name - Aryaman Kalia

Roll no -

102003099

SUBGROUP -

Coe6

## R QUERIES

```
library(dplyr)
library(tidyverse)
amazon_prime<-read.csv('C:/Users/Aryaman
kalia/Desktop/finaltableau/amazon_prime_titles.csv')
head(amazon_prime,10)
netflix_title<-read.csv('C:/Users/Aryaman kalia/Desktop/finaltableau/netflix_titles.csv')
head(netflix_title,10)
#removing description and show_id
amazon_prime<-select(amazon_prime,-c(description))
netflix_title<-select(netflix_title,-c(description))
amazon_prime<-select(amazon_prime,-c(show_id))
netflix_title<-select(netflix_title,-c(show_id))
#-----
#adding ott name
amazon_prime["OTT"]<-"amazon"
netflix_title["OTT"]<-"netflix"
head(amazon_prime)
head(netflix_title)
colnames(amazon_prime)
colnames(netflix_title)
#checking for duplitcated values
sum(duplicated(amazon_prime))
sum(duplicated(netflix_title))
#_____
#finding unique values of type of shows
unique(amazon_prime[c("type")])
unique(netflix_title[c("type")])
#_____

tv_amazon_prime<- subset(amazon_prime,type=="TV Show")
head(tv_amazon_prime)
```

```
movie_amazon_prime<-subset(amazon_prime,type=="Movie")
```

```
#_____
```

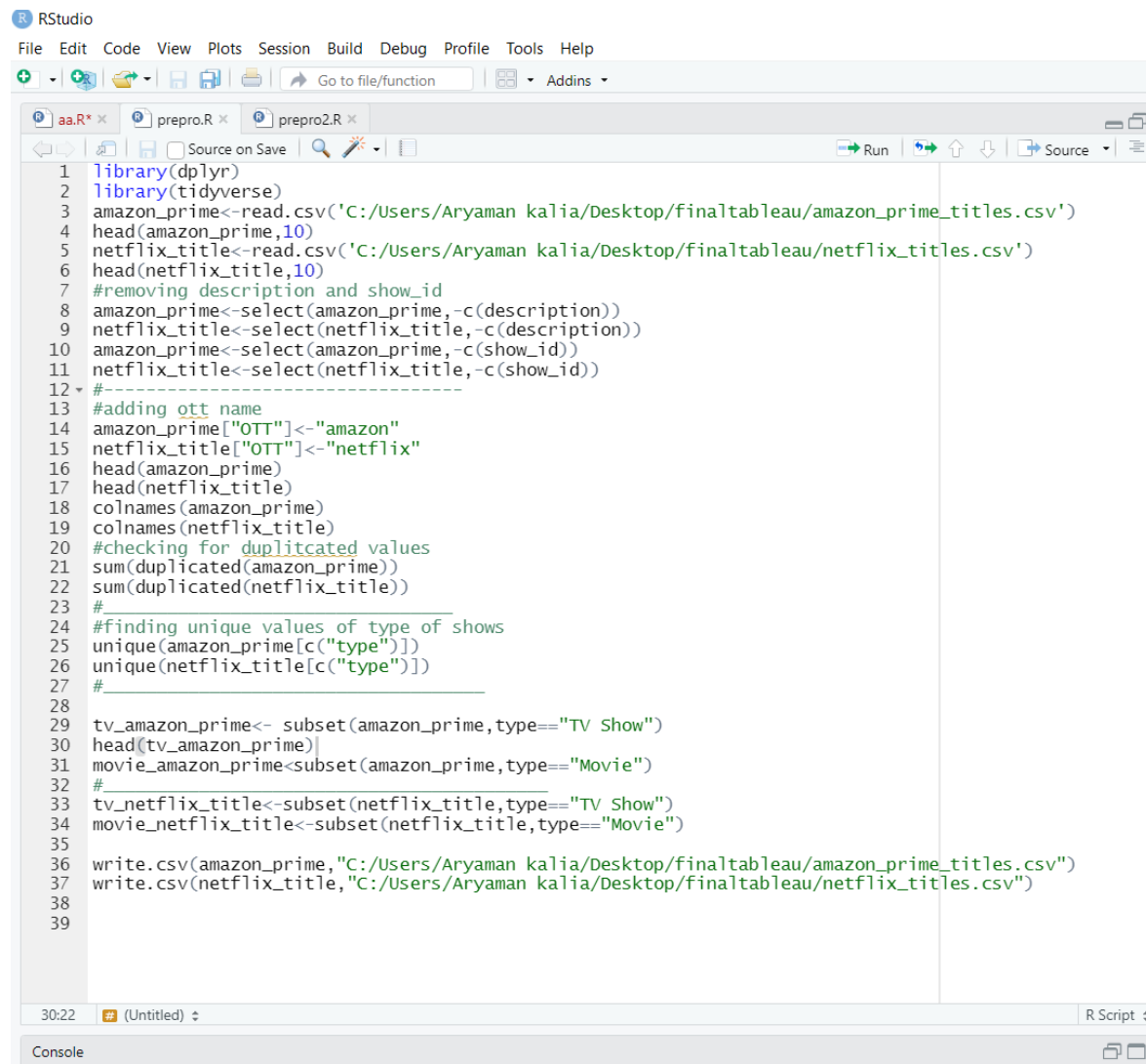
```
tv_netflix_title<-subset(netflix_title,type=="TV Show")
```

```
movie_netflix_title<-subset(netflix_title,type=="Movie")
```

```
write.csv(amazon_prime,"C:/Users/Aryaman
```

```
kalia/Desktop/finaltableau/amazon_prime_titles.csv")
```

```
write.csv(netflix_title,"C:/Users/Aryaman kalia/Desktop/finaltableau/netflix_titles.csv")
```

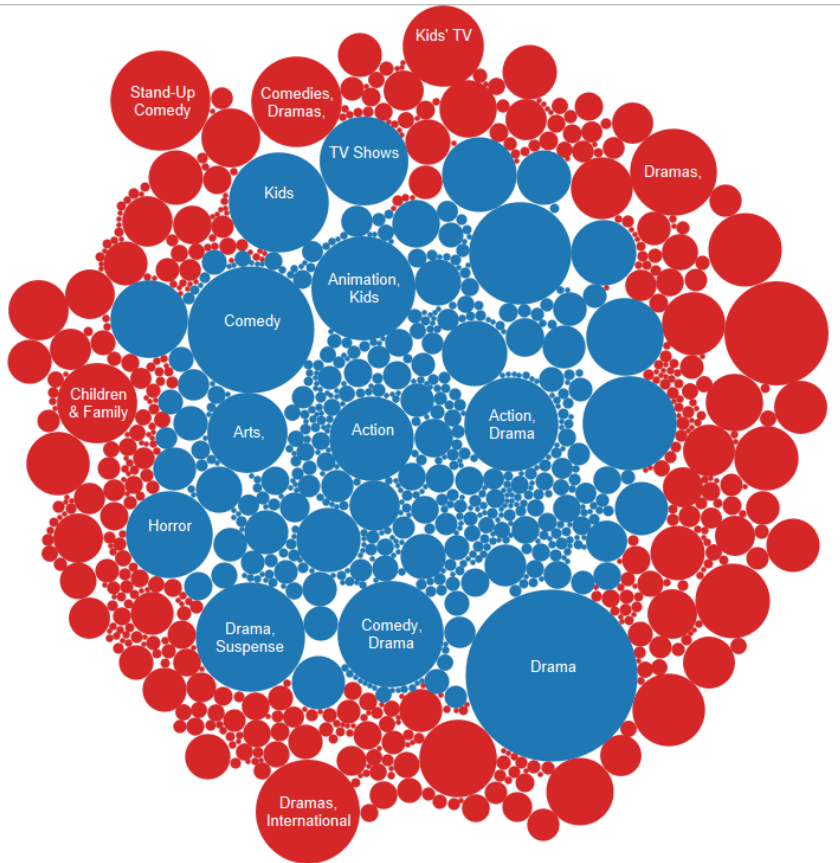


The screenshot shows the RStudio interface with a script editor containing R code. The code performs the following steps:

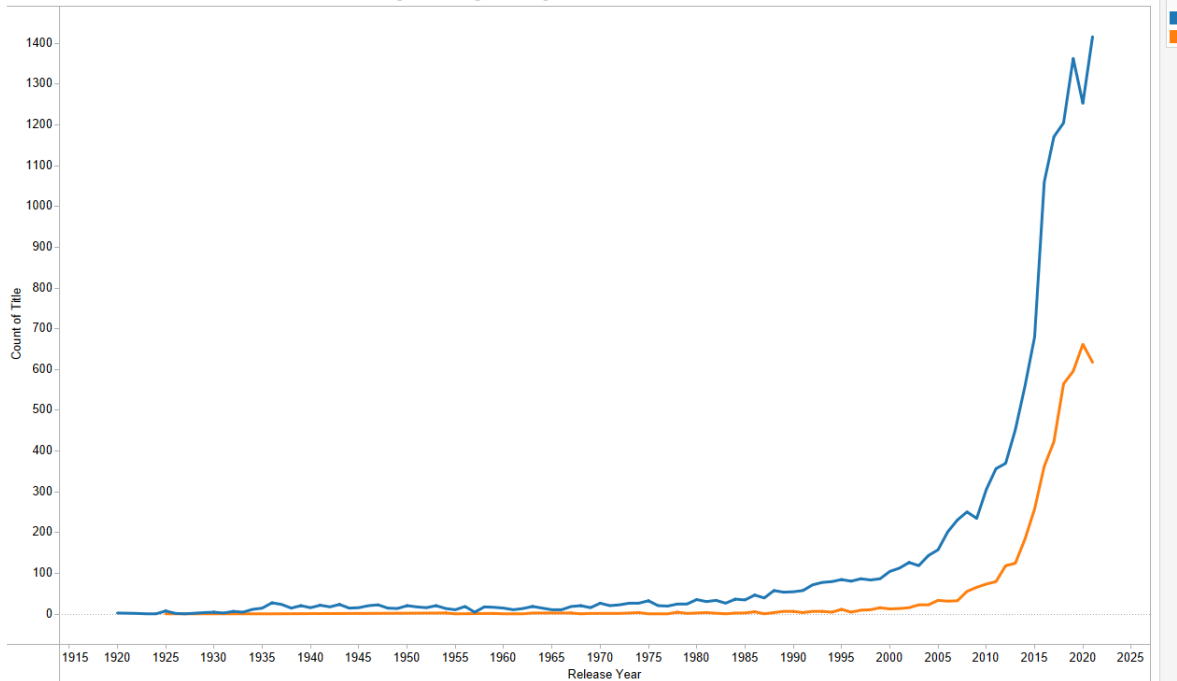
- 1. Loads the `dplyr` and `tidyverse` libraries.
- 2. Reads `amazon_prime` from a CSV file: `amazon_prime<-read.csv('C:/Users/Aryaman kalia/Desktop/finaltableau/amazon_prime_titles.csv')`
- 3. Prints the first 10 rows of `amazon_prime`: `head(amazon_prime,10)`
- 4. Reads `netflix_title` from a CSV file: `netflix_title<-read.csv('C:/Users/Aryaman kalia/Desktop/finaltableau/netflix_titles.csv')`
- 5. Prints the first 10 rows of `netflix_title`: `head(netflix_title,10)`
- 6. Removes the `description` and `show_id` columns from both datasets using `select`.
- 7. Prints the first 10 rows of both datasets.
- 8. Checks for duplicated values using `sum(duplicated())` on both datasets.
- 9. Finds unique values of the `type` column for both datasets using `unique()`.
- 10. Subsets the data by `type`: `tv_amazon_prime<- subset(amazon_prime,type=="TV Show")` and `movie_amazon_prime<subset(amazon_prime,type=="Movie")`.
- 11. Prints the first 10 rows of `tv_amazon_prime`.
- 12. Subsets the data by `type` for Netflix: `tv_netflix_title<-subset(netflix_title,type=="TV Show")` and `movie_netflix_title<-subset(netflix_title,type=="Movie")`.
- 13. Writes the `amazon_prime` dataset to a CSV file: `write.csv(amazon_prime,"C:/Users/Aryaman kalia/Desktop/finaltableau/amazon_prime_titles.csv")`
- 14. Writes the `netflix_title` dataset to a CSV file: `write.csv(netflix_title,"C:/Users/Aryaman kalia/Desktop/finaltableau/netflix_titles.csv")`

The console at the bottom shows the time 30:22 and the status "(Untitled)".

### Prevalent Genres in amazon and Netflix



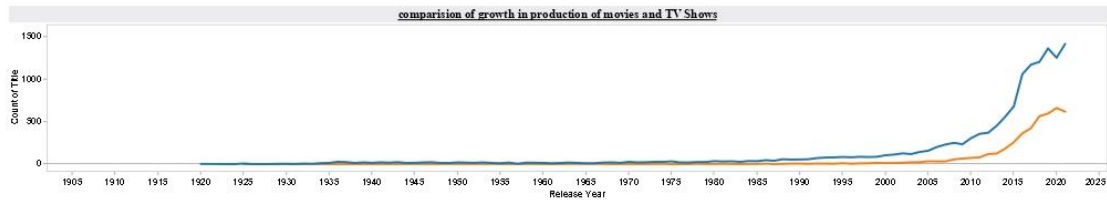
### comparison of growth in production of movies and TV Shows



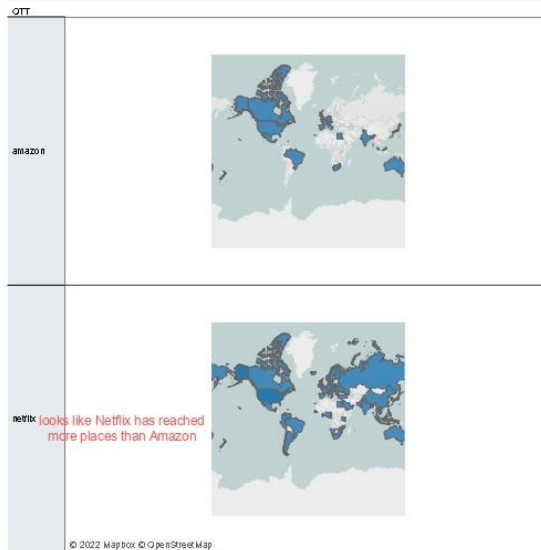
## NETFLIX VS AMAZON

Which do you prefer for your Entertainment .  
To get the Right answer let us use Statistics to  
answer your questions  
After using this Workbook , there will be no  
doubt in your mind

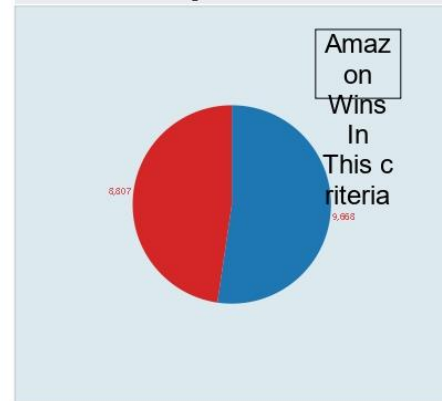
OTT  
amazon  
netflix  
Count of Type  
18,475  
Type  
Movie  
TV Show



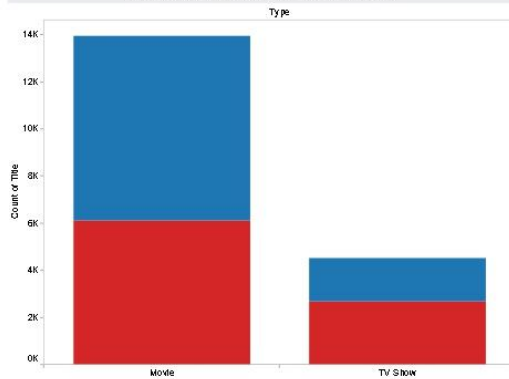
### countries where respective platforms are present



### comparison of number of titles present per ott platform



### overall and compartmentalized comparison of Type of content and its OTT Platform



### Prevalent Genres in amazon and Netflix

