Guadalupe Torres[1]

02/02/2025

–> The purpose of this document is to simulataneously analyze data on netflix titles. This data was given to us as a practice to become more familiar with datasets and give reports on our own.

```
##    show_id             type              title             director
##  Length:8807        Length:8807        Length:8807        Length:8807
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##       cast              country          date_added         release_year
##  Length:8807        Length:8807        Length:8807        Min.   :1925
##  Class :character   Class :character   Class :character   1st Qu.:2013
##  Mode  :character   Mode  :character   Mode  :character   Median :2017
##                                                           Mean   :2014
##                                                           3rd Qu.:2019
##                                                           Max.   :2021
##     rating            duration          listed_in          description
##  Length:8807        Length:8807        Length:8807        Length:8807
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
```

Shape and Structure of the Data

–> The data set contains statistics of Netflix movies and shows that were released between 1925-2021.

```
## spc_tbl_ [8,807 × 12] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ show_id     : chr [1:8807] "s1" "s2" "s3" "s4" ...
##  $ type        : chr [1:8807] "Movie" "TV Show" "TV Show" "TV Show" ...
##  $ title       : chr [1:8807] "Dick Johnson Is Dead" "Blood & Water" "Ganglands" "Jailbirds New Orleans" ...
##  $ director    : chr [1:8807] "Kirsten Johnson" NA "Julien Leclercq" NA ...
##  $ cast        : chr [1:8807] NA "Ama Qamata, Khosi Ngema, Gail Mabalane, Thabang Molaba, Dillon Windvogel, Na
tasha Thahane, Arno Greeff, Xolile "| __truncated__ "Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabiha Akkari, Sof
ia Lesaffre, Salim Kechiouche, Mourad$ne Farihi, G"| __truncated__ NA ...
##  $ country     : chr [1:8807] "United States" "South Africa" NA NA ...
##  $ date_added  : chr [1:8807] "September 25, 2021" "September 24, 2021" "September 24, 2021" "September 24, 20
21" ...
##  $ release_year: num [1:8807] 2020 2021 2021 2021 2021 ...
##  $ rating      : chr [1:8807] "PG-13" "TV-MA" "TV-MA" "TV-MA" ...
##  $ duration    : chr [1:8807] "90 min" "2 Seasons" "1 Season" "1 Season" ...
##  $ listed_in   : chr [1:8807] "Documentaries" "International TV Shows, TV Dramas, TV Mysteries" "Crime TV Show
s, International TV Shows, TV Action & Adventure" "Docuseries, Reality TV" ...
##  $ description : chr [1:8807] "As her father nears the end of his life, filmmaker Kirsten Johnson stages his d
eath in inventive and comical wa"| __truncated__ "After crossing paths at a party, a Cape Town teen sets out to p
rove whether a private-school swimming star is h"| __truncated__ "To protect his family from a powerful drug lor
d, skilled thief Mehdi and his expert team of robbers are pulled "| __truncated__ "Feuds, flirtations and toilet
talk go down among the incarcerated women at the Orleans Justice Center in New Or"| __truncated__ ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   show_id = col_character(),
##   ..   type = col_character(),
##   ..   title = col_character(),
##   ..   director = col_character(),
##   ..   cast = col_character(),
##   ..   country = col_character(),
##   ..   date_added = col_character(),
##   ..   release_year = col_double(),
##   ..   rating = col_character(),
##   ..   duration = col_character(),
##   ..   listed_in = col_character(),
##   ..   description = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
## NULL
```

–> The data set has 8,807 observations with 12 columns, the column categories consist of `show_id` `type` `title` `director` `cast` `country` `date_added` `release_year` `rating` `duration` `listed_in` and `description`. We can see that all our variables are character types except for `release_year` this is a numeric type. Because our variables are categorical types we can use models such as logistic regression, and decision trees.

Missing data

```
which(is.na(data))
```

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

```
sum(is.na(netflix_titles))
```

```
## [1] 4307
```

–> The data set contains 4,307 missing values. We can create a function to deal with NAs for the variable we assign into the function. Then we assign multiple variables to the function at the same time. The result would clean up the data set to replace the NAs to "UNK".

```
NAs <- function(x) {x = as.factor(ifelse(is.na(as.character(x)),'UNK',as.character(x)))}
```

```
netflix_titles <- netflix_titles %>%
  mutate_at(c('show_id','type','title','director','cast', 'country', 'date_added', 'release_year', 'rating', 'dur
ation', 'listed_in', 'description'), NAs)
```

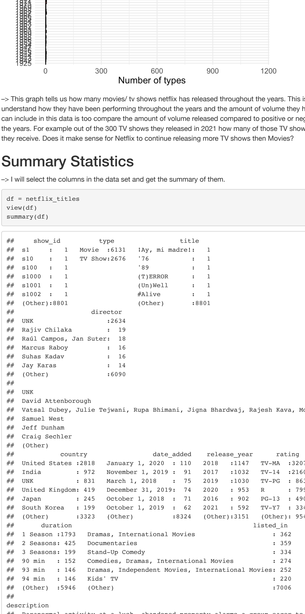
```
netflix_titles %>%
  keep(is.factor) %>%
  summary()
```

```
##     show_id          type                title
##  s1     :   1   Movie  :6131   ¡Ay, mi madre!:   1
##  s10    :   1   TV Show:2676   '76            :   1
##  s100   :   1                  '89            :   1
##  s1000  :   1                  (T)ERROR       :   1
##  s1001  :   1                  (Un)Well       :   1
##  s1002  :   1                  #Alive         :   1
##  (Other):8801                  (Other)        :8801
##                     director
##  UNK                  :2634
##  Rajiv Chilaka        :  19
##  Raúl Campos, Jan Suter:  18
##  Marcus Raboy         :  16
##  Suhas Kadav          :  16
##  Jay Karas            :  14
##  (Other)              :6090
##                                                                                                          cast
##  UNK                                                                                                      : 825
##  David Attenborough                                                                                       :  19
##  Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jigna Bhardvaj, Rajesh Kava, Mousam, Swapnil:  14
##  Samuel West                                                                                              :  10
##  Jeff Dunham                                                                                              :   7
##  Craig Gechler                                                                                            :   6
##  (Other)                                                                                                  :7926
##                country                     date_added    release_year       rating
##  United States :2818   January 1, 2020  : 110   2018   :1147   TV-MA  :3207
##  India         : 972   November 1, 2019 :  91   2017   :1032   TV-14  :2160
##  UNK           : 831   March 1, 2018    :  75   2019   :1030   TV-PG  : 863
##  United Kingdom: 419   December 31, 2019:  74   2020   : 953   R      : 799
##  Japan         : 245   October 1, 2018  :  73   2016   : 902   PG-13  : 490
##  South Korea   : 199   October 1, 2019  :  62   2021   : 592   TV-Y7  : 334
##  (Other)       :3323   (Other)          :8324   (Other):3151   (Other): 954
##          duration                     listed_in
##  1 Season :1793   Dramas, International Movies                        : 362
##  2 Seasons: 425   Documentaries                                       : 359
##  3 Seasons: 199   Stand-Up Comedy                                     : 334
##  90 min   : 152   Comedies, Dramas, International Movies              : 274
##  93 min   : 146   Dramas, Independent Movies, International Movies: 252
##  94 min   : 146   Kids' TV                                            : 220
##  (Other)  :5946   (Other)                                             :7006
##
description
##  Paranormal activity at a lush, abandoned property alarms a group eager to redevelop the site, but the eerie e
vents may not be as unearthly as they think.:   4
##  A surly septuagenarian gets another chance at her 20s after having her photo snapped at a studio that magical
ly takes 50 years off her life.              :   3
##  Challenged to compose 100 songs before he can marry the girl he loves, a tortured but passionate singer-songw
riter embarks on a poignant musical journey.:   3
##  Multiple women report their husbands as missing but when it appears they are looking for the same man, a poli
ce officer traces their cryptic connection. :   3
##  A budding politician has devious plans to rise in the ranks — until an unexpected new presence begins to inte
rfere with his every crooked move.           :   2
##  A scheming matriarch plots to cut off her disabled stepson and his wife from the family fortune, creating a d
ivision within the clan.                     :   2
##  (Other)
:8790
```

Visualization

–>

```
netflix_titles %>%
  ggplot() +
  geom_bar(mapping = aes(x = release_year, fill = type), color = "black") +
  labs(title = "Stacked Bar Chart of Type released by Year", x = "release_year", y = "Number of types") +
  coord_flip() +
  theme_minimal() +
  theme(text = element_text(size = 14))
```



–> This graph tells us how many movies/ tv shows netflix has released throughout the years. This is a good understanding for Netflix to understand how they have been performing throughout the years and the amount of volume they have been releasing. An improvement Netflix can include in this data is too compare the amount of volume released compared to positive or negative reviews they have received throughout the years. For example out of the 300 TV shows they released in 2021 how many of those TV shows performed well and what type of rating did they receive. Does it make sense for Netflix to continue releasing more TV shows then Movies?

Summary Statistics

–> I will select the columns in the data set and get the summary of them.

```
df = netflix_titles
view(df)
summary(df)
```

```
##     show_id          type                title
##  s1     :   1   Movie  :6131   ¡Ay, mi madre!:   1
##  s10    :   1   TV Show:2676   '76            :   1
##  s100   :   1                  '89            :   1
##  s1000  :   1                  (T)ERROR       :   1
##  s1001  :   1                  (Un)Well       :   1
##  s1002  :   1                  #Alive         :   1
##  (Other):8801                  (Other)        :8801
##                     director
##  UNK                  :2634
##  Rajiv Chilaka        :  19
##  Raúl Campos, Jan Suter:  18
##  Marcus Raboy         :  16
##  Suhas Kadav          :  16
##  Jay Karas            :  14
##  (Other)              :6090
##                                                                                                          cast
##  UNK                                                                                                      : 825
##  David Attenborough                                                                                       :  19
##  Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jigna Bhardvaj, Rajesh Kava, Mousam, Swapnil:  14
##  Samuel West                                                                                              :  10
##  Jeff Dunham                                                                                              :   7
##  Craig Gechler                                                                                            :   6
##  (Other)                                                                                                  :7926
##                country                     date_added    release_year       rating
##  United States :2818   January 1, 2020  : 110   2018   :1147   TV-MA  :3207
##  India         : 972   November 1, 2019 :  91   2017   :1032   TV-14  :2160
##  UNK           : 831   March 1, 2018    :  75   2019   :1030   TV-PG  : 863
##  United Kingdom: 419   December 31, 2019:  74   2020   : 953   R      : 799
##  Japan         : 245   October 1, 2018  :  73   2016   : 902   PG-13  : 490
##  South Korea   : 199   October 1, 2019  :  62   2021   : 592   TV-Y7  : 334
##  (Other)       :3323   (Other)          :8324   (Other):3151   (Other): 954
##          duration                     listed_in
##  1 Season :1793   Dramas, International Movies                        : 362
##  2 Seasons: 425   Documentaries                                       : 359
##  3 Seasons: 199   Stand-Up Comedy                                     : 334
##  90 min   : 152   Comedies, Dramas, International Movies              : 274
##  93 min   : 146   Dramas, Independent Movies, International Movies: 252
##  94 min   : 146   Kids' TV                                            : 220
##  (Other)  :5946   (Other)                                             :7006
##
description
##  Paranormal activity at a lush, abandoned property alarms a group eager to redevelop the site, but the eerie e
vents may not be as unearthly as they think.:   4
##  A surly septuagenarian gets another chance at her 20s after having her photo snapped at a studio that magical
ly takes 50 years off her life.              :   3
##  Challenged to compose 100 songs before he can marry the girl he loves, a tortured but passionate singer-songw
riter embarks on a poignant musical journey.:   3
##  Multiple women report their husbands as missing but when it appears they are looking for the same man, a poli
ce officer traces their cryptic connection. :   3
##  A budding politician has devious plans to rise in the ranks — until an unexpected new presence begins to inte
rfere with his every crooked move.           :   2
##  A scheming matriarch plots to cut off her disabled stepson and his wife from the family fortune, creating a d
ivision within the clan.                     :   2
##  (Other)
:8790
```

–> The summary tells us the number of values in each category and what those values consist of. This gives Netflix insight of what they are putting on their platform, for example there were 3,207 TV-MA ratings being put out throughout the years making it the most rating being put released. Netflix also put out more Movies then they did TV shows. There were 6,131 movies released and 2,676 TV shows released.

1. **Email** gtorressanchez01@hamline.edu, **Position** Student↩