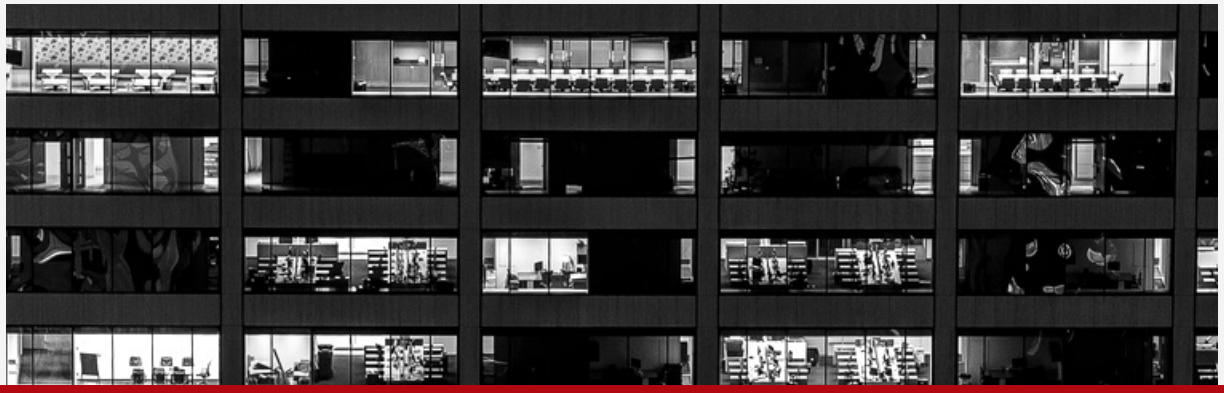
INTRODUCTION TO R



Netflix Movie Dashboard

APRIL 2023

STUDENT
Yu-Chieh, YU
DMDS

TEACHER
Guillaume COQUERE

EPISODES

01

PRE-VIEW TRAILOR OF THE DATASET

Introducing the chosen dataset, and providing the background knowledge of the source, context, size, variables, variable types, and step-by-step cleaning process. Also provide some simple statistic charts of the dataset.

02

BEHIND THE SCENE OF THE DATASET

The reason why I'm looking into this dataset & how it's going to help people.

03

EXPECTED OUTCOME OF THE DATASET

Provide future vision of the data.

04

ACTUAL OUTCOME & DEPLOYMENT ON SHINY APP

An overview of the final Shiny APP Netflix Movie Dashboard.

Pre-View Trailor of the Dataset

source, context, size, variables, and variable types

- Source: After a long search through the different datasets, the final decision fell on the Kaggle "Netflix Movies and TV Shows" dataset.
- Context: One of the biggest reasons I use this dataset is because of my passion for watching shows and movies, which further develops into the purpose of my whole project that I will later on explain.
- Size: 8808 lines, 12 columns

Name of Variables	Variable Types(before cleaning)	Variable Types(after cleaning)
show_id	Character	DELETE
type	Categorical Character	Categorical Character
title	Character	Character
director	Character	Character
cast	Character	Character
country	Categorical Character	Categorical Character
date_added	Character	Character
release_year	Categorical Character	Categorical Character
rating	Character	Categorical Character
duration	Character	Integer
listed_in	Character	Character
description	Character	Character

Pre-View Trailor of the Dataset

Step-by-step cleaning process

- Dataset cleaning steps:

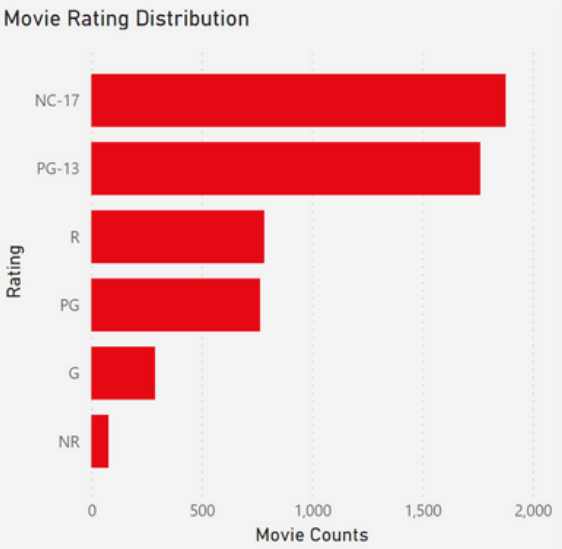
1. Load the dataset into the R environment
2. Explore: Generate a summary to have a clearer picture of the dataset, and see the variables that we have right now.
3. Deal with the missing values: Drop the NA data in our dataset, which turns the data from 8807 rows to 5332 rows. It is crucial to dropping the missing rows in the column director, cast, country, date_added, rating, and duration because we will need these columns to have value in order to generate graphs or charts later on.
4. Remove unnecessary columns & rows: Such as the show_id, which is only a sequence order of the dataset. Remove the "TV show" in the 'type' column to keep only the "Movie" type and that turns the data frame from 5332 to 5185 rows.
5. Correct the column into the right variable: change the "duration" column from character to integer by removing the 'min' text in the rows, and use function as.integer to change the variable.
6. Clean the rows: remove weird text '\80x' in the "description" column.
7. Double check the categorical data are clean: use the unique function to run through the "rating" column and found out there are two rating systems in the dataset, so I unified them into the "Motion Picture Association of America" rating system because this rating system focus on movie genre.
8. Separate the collapsed column "country" & "listed_in": in order to allow the further filter function in the shiny app, separate all the collapsed columns in "country" like: 'United States, Ghana, United Kingdom' into different rows.
9. Change the date format: change from 2021-Sep-24 to 2021-09-24, take out the month abbreviation and replace it with a number to allow deployment in the shiny app.
10. Save the final result into a .RData file!

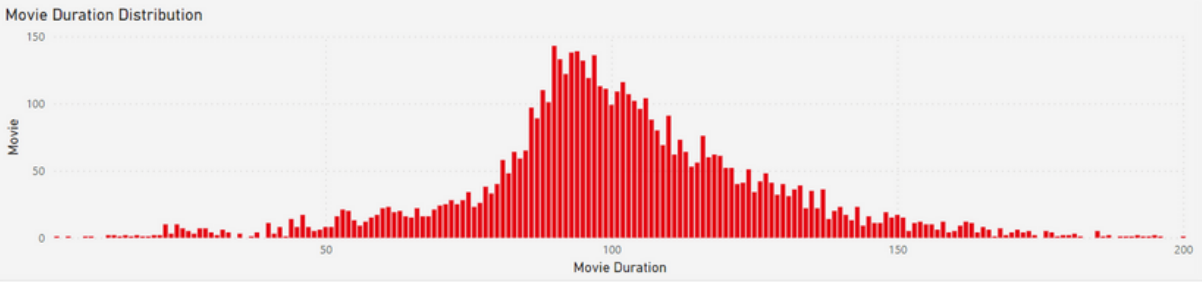
After this cleaning process, the dataset went from 12 columns to 11 columns and 8807 rows to 14872 rows with the final result shown below.

type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra Duah, Nick...	United States	2021-09-24	1993	NC-17	125	Dramas	On a photo shoot in Ghana, an American model slips b
Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra Duah, Nick...	United States	2021-09-24	1993	NC-17	125	Independent Movies	On a photo shoot in Ghana, an American model slips b
Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra Duah, Nick...	United States	2021-09-24	1993	NC-17	125	International Movies	On a photo shoot in Ghana, an American model slips b
Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra Duah, Nick...	Ghana	2021-09-24	1993	NC-17	125	Dramas	On a photo shoot in Ghana, an American model slips b
Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra Duah, Nick...	Ghana	2021-09-24	1993	NC-17	125	Independent Movies	On a photo shoot in Ghana, an American model slips b

Pre-View Trailor of the Dataset

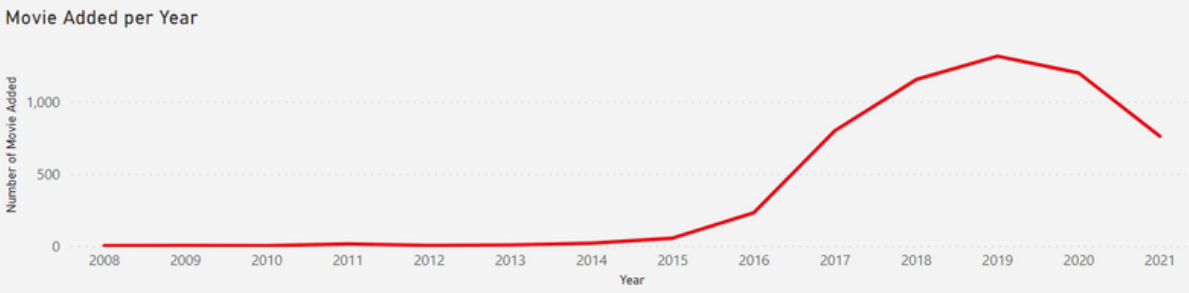
Statistic charts





From the various charts, we can see that among all the movies on Netflix, most of them are rated NC-17, in another word, adult only, but they also balance it with some teenager-age-appropriate movies that are rated PG-13.

Most of the movie on Netflix comes from the US, India, and the UK. If you are looking for a movie that lasts about 80-120 mins then you can find a lot of options on Netflix! Nevertheless, an interesting insight is that Netflix released most of its movies in 2019, and slowly dropped each year after that.



Behind the scene of the dataset

Why do I choose this dataset?

As a movie-lover, I grow up watching all kinds of different movies. I still remember when I was a kid, walking into the DVD stores with my family, we rent a movie and bought some delicious food for the perfect movie night.

Now that I grow up, move to another country, I still kept the old habit of watching movies. The only differnce is that I now watch it on Netflix with my boyfriend on another side of the world, but It can be a struggle to find the right movie for the perfect movie night sometimes.

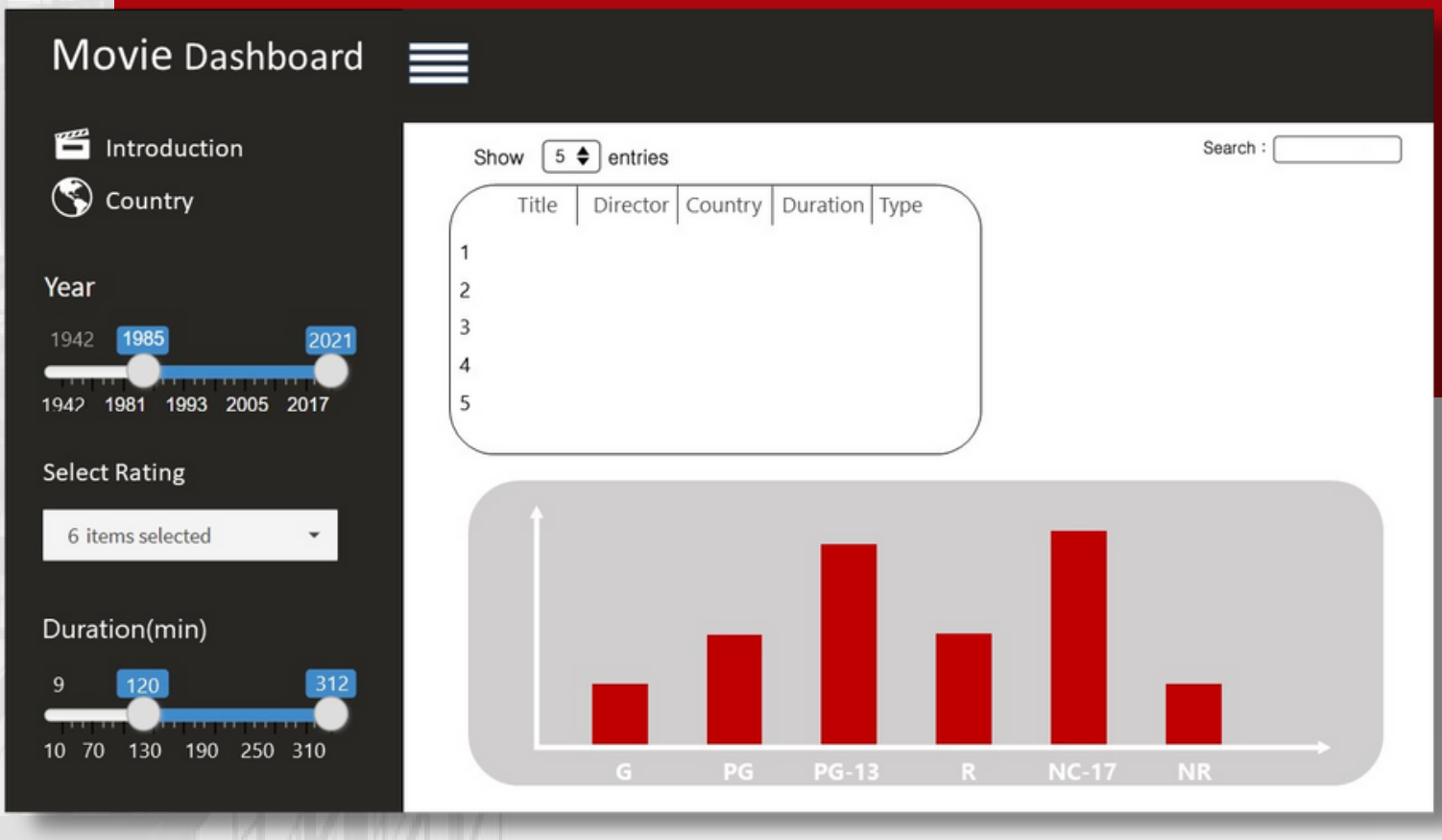
This is why I'm hoping to create a website for people to choose the right movie on Netflix that fits their needs. Whether you're picking the movie according to the length, country, year or director, you can find the best matching movie for your perfect movie night!

Expected Outcome of The Dataset

Ideally the first verion of the website will look like the following design.

The whole website will be seperate into left and right section. First if they click into introduction, I will tell them the story of why I create this dashboard and how it's going to help them to find the best matching movie. Next, user will be able to slide through the different years of movie that was released on Netflix. For the selcting section, they can pick the rating according to what they want. Lastly, the duration slide will help them find just the right length for the perfect movie night.

On the right hand side, it's the display of the information inculding graph and more details about the movies including the title, director, country, duration and type. Hopefully this website will be the best tool when it comes to picking the right movie:)



Actual Outcome & Deployment on Shiny App

Overview of the Shiny App Netflix Movie Dashboard

The objective for this dashboard is to create a website for people to choose the right movie to watch on Netflix. Whether you're picking the movie according to the duration, country, or type, you can find the best matching movie for your perfect movie night! The dataset includes movie releases from 1925 to 2021, giving you the opportunity to rediscover childhood favorites and explore the latest films.

On the left hand side you have the navigation bar, namely "Dashboard Info", "Movie List", "Movie Statistics" and three filters respectively "Movie Duration(mins)", "Country". and "Movie Type". The three filters can work across all different pages cater to the needs of user.

If you just want to find the right movie, you can use the filter on the left hand side and go straight to the "Movie List" tab to see all the movies that were listed out for you!

If you are interested about some fun fact and statistics of the movie on Netflix, you can go to the "Movie Statistics" tab for three different dashboard: histogram, bar chart and heatmap. Where you can find out what are the most common movie duration(mins), what are most of the movie rating on Netflix and when do Netflix upload most of their movie.

Have fun using the dashboard and hope you find your perfect match movie!!