# Report -  CS5044 Practical 2

200024763 | 200015498 |160015074

## DATA & QUESTION

This practice is based on the dataset -  'Netflix Movies and TV Shows'
(https://www.kaggle.com/shivamb/netflix-shows), which consists of 12 columns with different
attribute types:

| attributes | type | title | show_id | director | cast | country | date_added | release_year | rating | duration | listed_in | description |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| types | categorical | | ordinal | | | categorical | ordinal | ordinal | categorical | quantitative | categorical | |

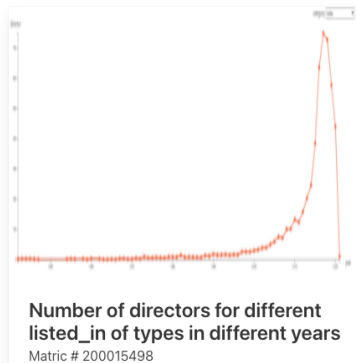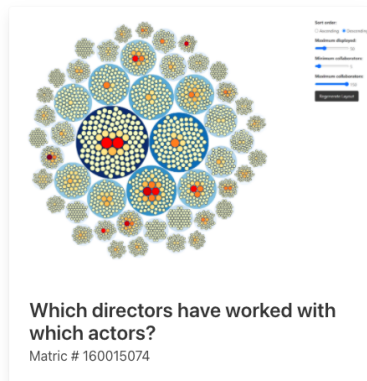Every group members chose one topic related to 'director'l:

1. the number of times an actor has worked with a director;
2. the number of directors for different types released in different years;
3. the number of directors in each country and the number of products of each director.

## DESCRIPTION OF VISUALISATION & IMPLEMENTATION

The visualisation results can be reached at - https://ikl.host.cs.st-andrews.ac.uk/index .
Three charts on the home page were presented to answer the three subjects of investigation
mentioned above separately - circle packing hierarchical visualisation, line chart and world
map. By clicking the miniatures or using the top bar, the description of this visualisation is
provided to help the viewers to read the visualisation and interact with it. Also, Tippy.js was
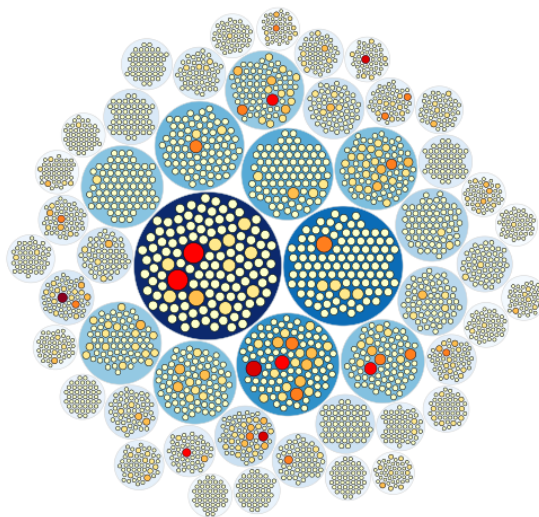used to provide tooltips. Bulma was used to provide readymade CSS stylings.

## Visualising Netflix Data



**Which directors have worked with which actors?**
Matric # 160015074

**Number of directors for different listed_in of types in different years**
Matric # 200015498

**Director & Country**
Matric # 200024763

## Circle Packing Hierarchical Visualisation



In this visualization, **the size of circles** indicates the number of times an actor has worked with a director. Thus, outer circles, representing the directors, are proportionately sized to the number of unique actors the director has worked with. Circles within the director circles, representing the actors that have worked with the specific director, are proportionately sized

to the number of times the actor has worked with the director.
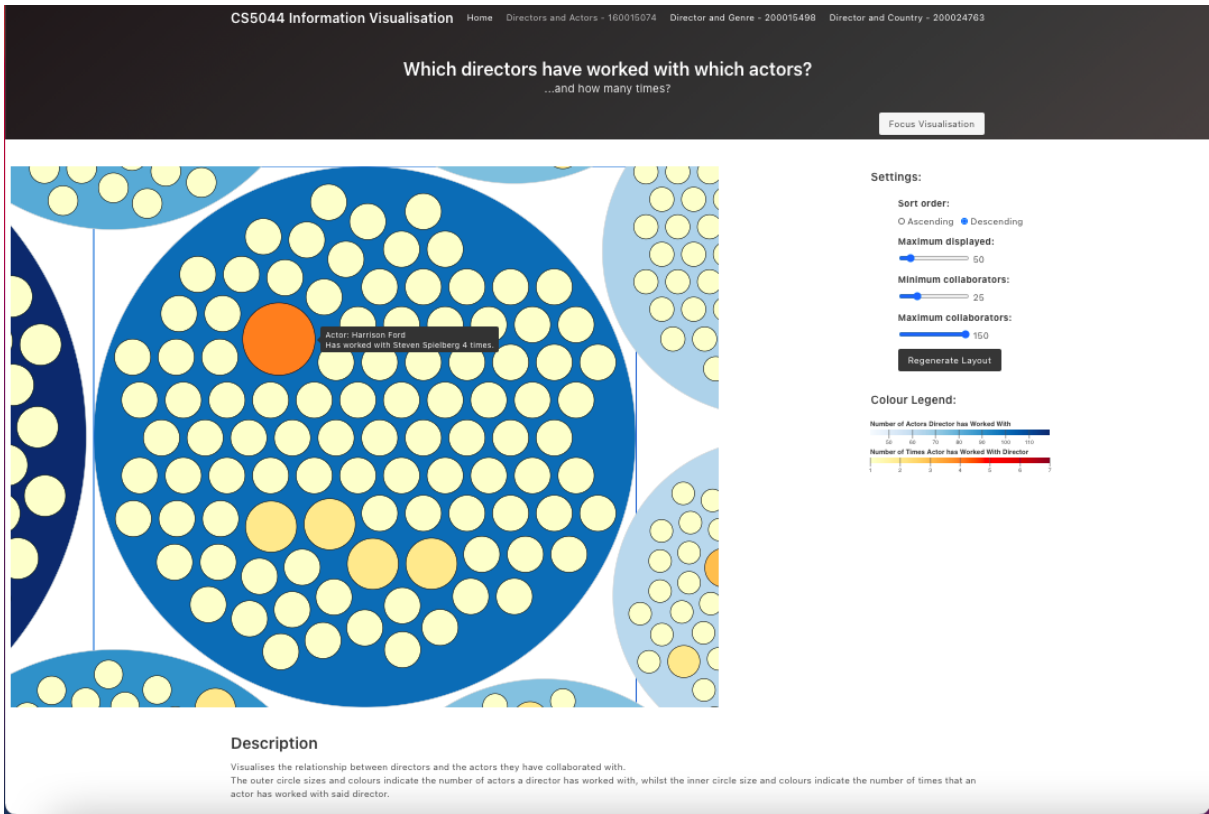


An additional visual encoding of these same values was implemented, whereby **two different colour scales** are used to colourise the circles and make it slightly easier to distinguish different circles. Colour encoding of the number of times actors have worked with directors also helps to make it more evident in the scenario such that an actor and a director have worked together many times but that director has not collaborated with many actors in total.

| visual variables | size | colour scales | |
|---|---|---|---|
| intinteractive features | a filter bar to control the size of data to present | hover to see more information related | click the big circle to zoom in to the small circle |
| visualisation techniques | D3's hierarchy pack | colour interpolation algorithms | |

**Line Chart**
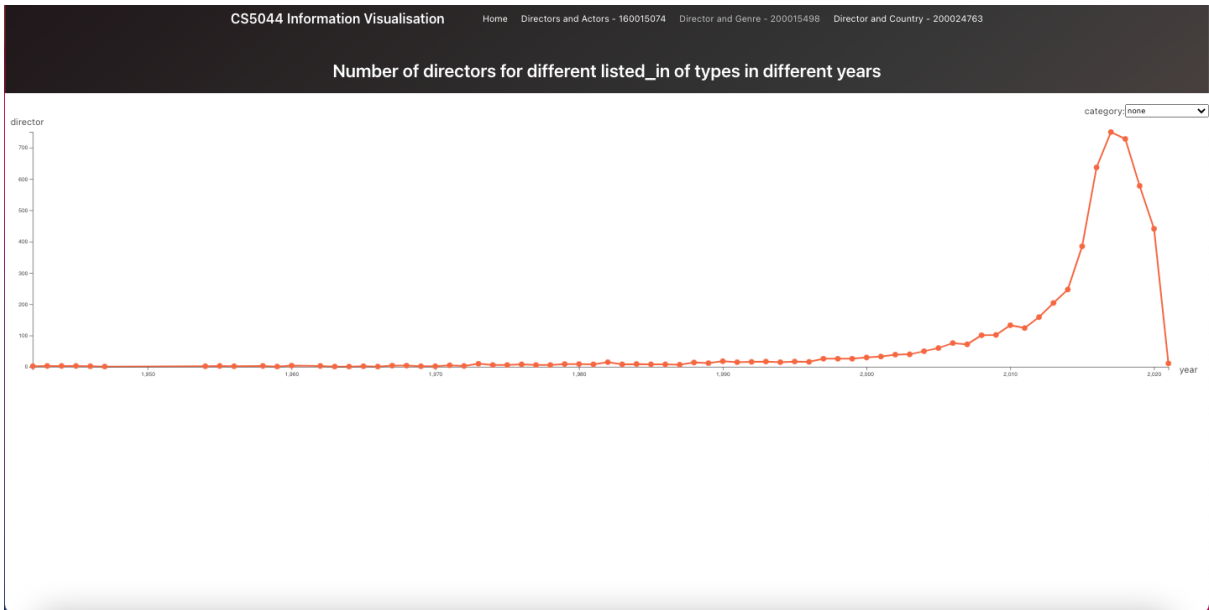


The line chart uses three different properties: director, release_year, and listed_in. The release_year was treated as the abscissa, the number of directors was considered as the ordinate, and the listed_in was used as a filter. All of these were to complete an interactive line chart. "None" in the filter indicates the sum of numbers for different types. By selecting the list of different movies in the filter, one can see the number of directors who released the selected movies in different years under the same list condition.
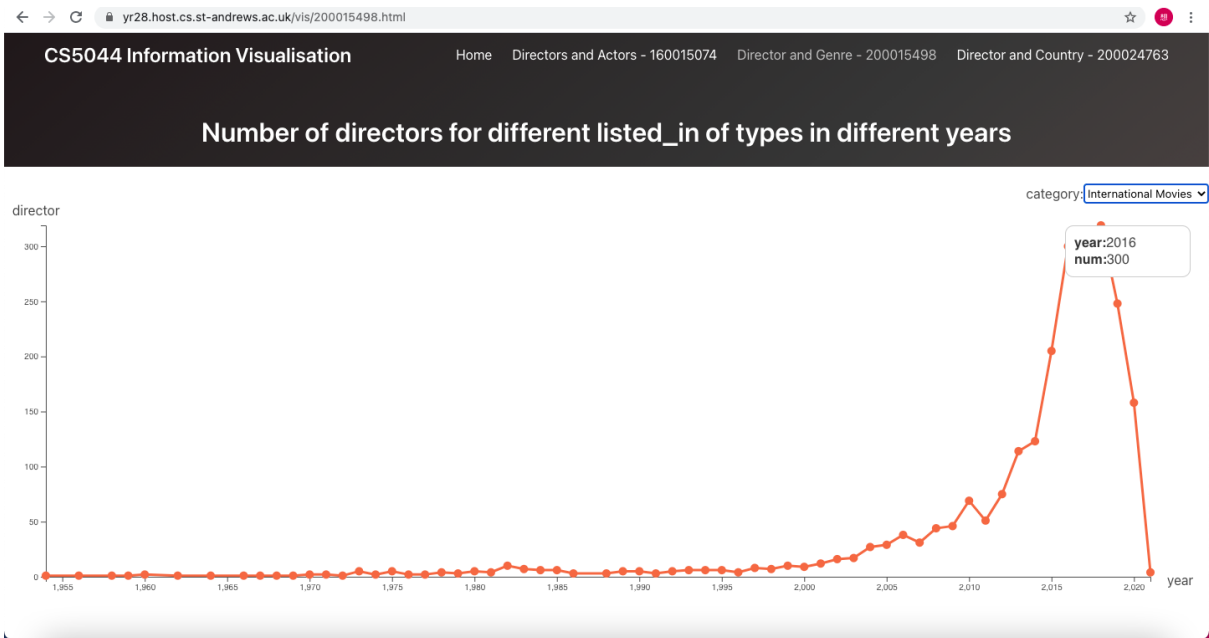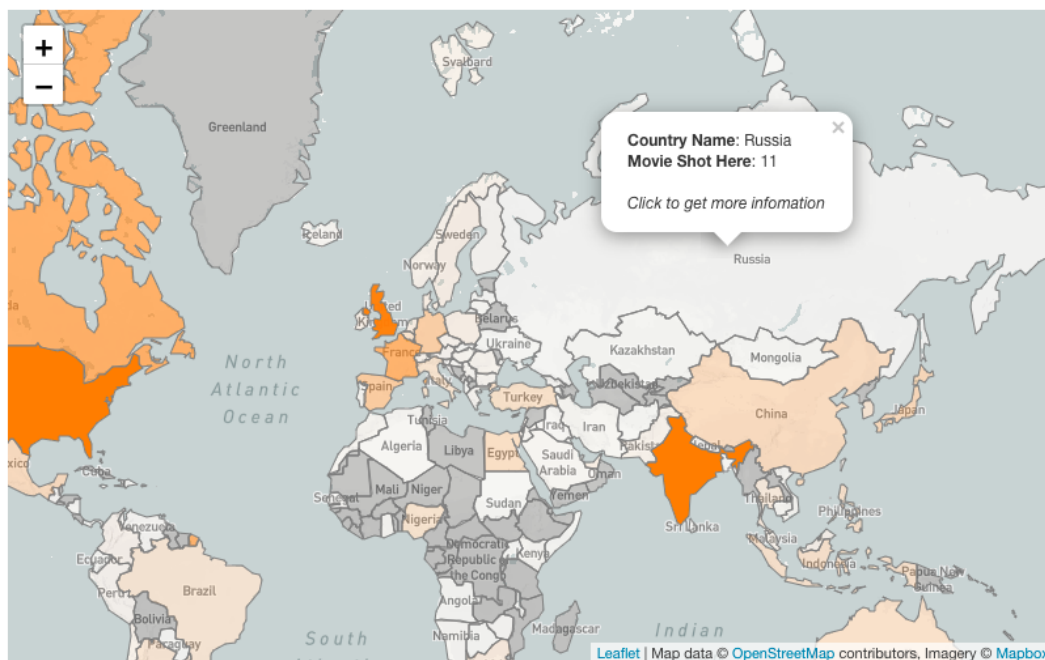


| visual variables | abscissa for release_year | ordinate for number of directors |
|---|---|---|
| | | |

| intinteractive features | a select list as filter | hover to see more information |
|---|---|---|
| visualisation techniques | D3's line chart technique | |

**World Map**



The world map provides the viewer with an overview of the number of directors in a specific country. The colour scale was used to present the number of directors in this country. Also, by hovering your mouse on a country, there will be a popup with detailed information. The viewer can drag the map and zoom in/out.
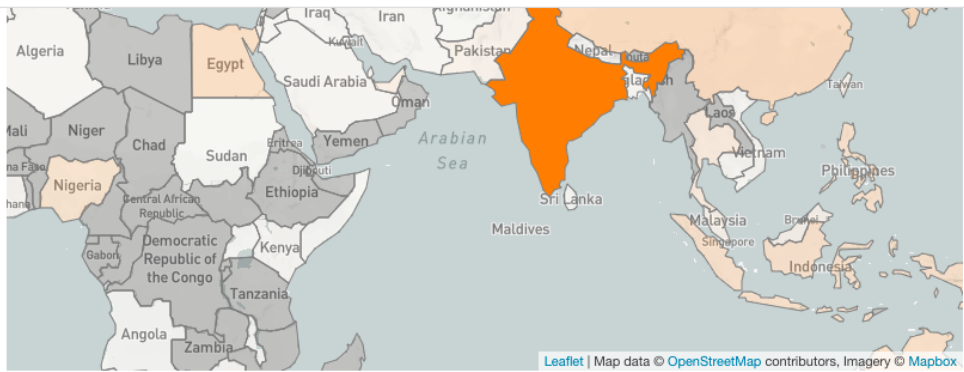
An additional interactive function is to click the country and there will be a list with names in different sizes for the directors in the country selected.

| visual variables | color scale for number of movies shot in different countries | size of name for number of movies shot by a director | |
|---|---|---|---|
| intinteractive features | click different countries on map to get a name list | hover to see more information | zoom in/out the map |
| visualisation techniques | map from leaflet and with world map GeoJson file | use word size to present quantitative attribute - idea from word cloud | color scale to present quantitative attribute |

## INSIGHT FROM THE VISUALISATION & CRITICAL DISCUSSION

Overall, the visualizations give viewers an insight of the information of Netflix directors from different aspects.

For the circle chart, acknowledging the dataset source being biased to due how and what Netflix can acquire, there are a few immediate conclusions. The top three directors by actors worked with are Martin Scorsese, Steven Spielberg and Cathy Garcia-Molina. Approximately 394 directors have worked with only one actor, and the vast majority of them only once. In general, actors do not work with the same actor more than once, though there are a significant minority of directors that shirk this pattern and have consistently worked with the majority of all actors they have worked with more than once. The latter is possible due to recurring characters appearing across franchises directed by a single director. There are also many cases of directors repeatedly "working with themselves" - in other words, starring in their films, and often.

As you can see in the line chart, there are very few directors for TV shows, because not all TV shows are provided with the name of the director. Therefore, in this case, the line chart only shows movies and TV shows with directors. With drama, there are directors in different years, and there are a lot of them. According to this line chart and filter type, you can see which listed director shot the most, and how many people released this type of director in different years.

In terms of the map and name list, we may tell that America, India and the UK are the biggest Netflix sources from the colour scale, while Canada and France follow behind. By hovering your mouse on it, America was on the top of the list with more than 2,000 works. Then Click it and we can see Jay Karas, Marcus, Raboy and Chapman are the directors with the highest occurrence rate in Netflix's works among all of them. For India, Anurag Kashyap, David Dhawan are the most active directors.