Assignment 1

Johnson ODEJIDE

2022-10-31

# 1. R and Descriptive Statistics

## a.) 12th Movie in the dataset

*Answer*: ***Ocean’s Thirteen***

Damon[12 , ]

## # A tibble: 1 x 3  
## Title Rating BoxOffice  
## <chr> <dbl> <dbl>  
## 1 Ocean's Thirteen 6.9 311

## b.) Average value of box office

*Answer*: ***175.3548***

mean(Damon$BoxOffice)

## [1] 175.3548

## c.) Lowest rating in the dataset

*Answer*: ***5.6***

min(Damon$Rating)

## [1] 5.6

## d.) Rows of the dataset containing movies with a rating of 8 or higher.

Damon[(Damon["Rating"] >= 8), ]

## Title Rating BoxOffice  
## <chr> <dbl> <dbl>  
## 1 The Departed 8.5 291  
## 2 The Bourne Ultimatum 8 444  
## 3 Inside Job 8.3 8  
## 4 The Martian 8 631

## c.) Resulting 3 centroids

*Centroids* - **5.9**, **6.9**, and **7.7**

Damon %>%   
 select(Rating, BoxOffice) %>%   
 scale() -> Damon.norm  
  
set.seed(12345)  
  
Damon.kmcluster <- kmeans(Damon.norm, 3, nstart = 10)  
  
unscale(Damon.kmcluster$centers, Damon.norm)

## Rating BoxOffice  
## 1 5.871429 104.42857  
## 2 6.913333 93.66667  
## 3 7.666667 366.66667

## c.) What the centroids tell us

By looking at the centroids for cluster 1, we can conclude that on average, all the observations have 5.9 ratings and made $104.4M at the box office worldwide. Similarly, for cluster 2, the average ratings for the observations within this cluster are 6.9 and they made $93.7M at the box office worldwide, and for cluster 3, there is an average of 7.7 ratings among the observations, and $366.7M was made at the box office worldwide.

## c.) Number of movies in each of the three clusters

**Cluster 1** has ***7 movies,*** **Cluster 2** has ***15 movies, and*** **Cluster 3** has ***9 movies***

Damon.kmcluster$size

## [1] 7 15 9