Top Actors

Noah Mott

12/15/2020

Run necessary libraries

library(tidyr)

## Warning: package 'tidyr' was built under R version 4.0.3

library(dplyr)

## Warning: package 'dplyr' was built under R version 4.0.3

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.0.3

import Data

df<-read.csv("C:/Users/noahm/OneDrive/Desktop/IMDb Movie Titles Cleaned.csv")

clean data

country<-separate\_rows(df, country, sep=",")  
df1<-separate\_rows(country, actors, sep=",")  
df1$country<-trimws(df1$country)  
df1$actors<-trimws(df1$actors)  
#filter for us actors  
df2<-df1 %>% filter(country=="USA")

#filter for number of votes  
df2<-df2 %>% filter(votes>=10000)  
#filter for time  
df2<-df2 %>% filter(year>=1970)

#create count of films for each actor  
df3<-count(df2, actors)  
#filter new df for at least 25 films  
df4<-subset(df3, df3$n>=25)  
#create an average of average score for each actor  
vec<-aggregate(df2$avg\_vote, by=list(df2$actors), FUN=mean, na.rm=TRUE)  
#merge actor averages and movie counts dataframes along actors  
together<-merge(df4, vec, by.x=c("actors"), by.y=c("Group.1"))  
#change columnnames  
colnames(together)<-c("Actors", "Count\_of\_Films", "AVG\_Rating")  
#sort df  
sorted<-together[order(-together$AVG\_Rating),]

Plot Data

plotstuff<-rbind(head(sorted, 10), tail(sorted, 10))  
plotstuff1<-plotstuff[order(-plotstuff$AVG\_Rating),]

plot1<-ggplot(data=plotstuff1, aes(size=Count\_of\_Films, y=AVG\_Rating, x=reorder(Actors,AVG\_Rating)))+geom\_hline(yintercept = mean(together$AVG\_Rating), linetype="dashed", color="red", size=1)  
plot1<-plot1+geom\_point(shape=21, color="black", fill="green", stroke=2)+coord\_flip()+annotate("text", x="Jean-Claude Van Damme", y=mean(together$AVG\_Rating), label="Mean Actor Rating", size=4, angle=90, vjust=-0.6, hjust=-.1)  
plot1<-plot1+labs(x="Actor", y="Average Movie Rating", size="Count of Films", title="Top & Bottom 10 US Actors by IMDb Average Film Rating Since 1970")  
plot1<-plot1+theme(plot.background = element\_rect(fill = "grey80", colour = "black"), panel.background=element\_rect(fill="grey60"), legend.backgroun=element\_rect(fill="grey80"), legend.key = element\_rect(fill="grey80"))+theme(plot.title=element\_text(hjust=.5))  
plot1

