R Notebook

This is an [R Markdown](http://rmarkdown.rstudio.com) Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Cmd+Shift+Enter*.

# Setup data.  
makes = read.csv("makematrix\_binary.csv", header = TRUE)  
str(makes)

## 'data.frame': 44 obs. of 45 variables:  
## $ MakeName : Factor w/ 44 levels "","Acura","Alfa Romeo",..: 30 28 32 10 35 2 31 15 16 39 ...  
## $ McLaren : int 1 0 1 0 1 1 1 1 0 1 ...  
## $ Maserati : int 0 1 0 0 0 0 0 0 1 0 ...  
## $ MINI : int 1 0 1 0 0 0 0 0 1 1 ...  
## $ Chevrolet : int 0 1 0 1 1 0 0 0 1 0 ...  
## $ Porsche : int 0 1 0 0 1 0 0 0 1 0 ...  
## $ Acura : int 0 1 0 1 1 1 1 1 1 0 ...  
## $ Mercedes.Benz: int 0 0 0 0 0 0 1 0 0 0 ...  
## $ Ford : int 0 0 0 0 0 0 0 1 0 0 ...  
## $ Genesis : int 0 0 0 0 0 0 0 0 1 0 ...  
## $ Smart : int 0 0 0 0 0 0 0 0 1 1 ...  
## $ Dodge : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Scion : int 0 0 0 0 0 0 0 0 1 0 ...  
## $ Rolls.Royce : int 0 0 0 0 0 0 0 1 1 0 ...  
## $ Cadillac : int 0 0 0 0 0 0 0 1 0 0 ...  
## $ Honda : int 1 1 1 0 1 1 1 1 0 1 ...  
## $ Hyundai : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Volkswagen : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Mazda : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Jeep : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Infiniti : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ LandRover : int 0 0 0 0 0 0 0 1 0 0 ...  
## $ Kia : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Mitsubishi : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ BMW : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ FIAT : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Lincoln : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Lamborghini : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Jaguar : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ GMC : int 0 0 0 0 1 0 0 1 1 0 ...  
## $ Toyota : int 1 0 1 1 0 0 1 0 0 0 ...  
## $ Nissan : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ AstonMartin : int 0 1 0 1 1 1 0 0 1 0 ...  
## $ Volvo : int 0 0 0 0 1 0 0 0 1 0 ...  
## $ Ferrari : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Chrysler : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Bentley : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Tesla : int 0 0 0 0 1 0 0 0 0 0 ...  
## $ Ram : int 0 0 0 0 0 0 0 0 0 1 ...  
## $ AlfaRomeo : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Subaru : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Buick : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Lexus : int 0 0 0 1 0 0 0 0 0 0 ...  
## $ Audi : int 0 0 0 0 1 0 1 0 0 0 ...  
## $ X : int 0 0 0 0 0 0 0 0 0 0 ...

colnames(makes) = c('MakeName', 'McLaren', 'Maserati', 'MINI', 'Chevrolet', 'Porsche', 'Acura', 'Mercedes-Benz', 'Ford', 'Genesis', 'Smart', 'Dodge', 'Scion', 'Rolls-Royce', 'Cadillac', 'Honda', 'Hyundai', 'Volkswagen', 'Mazda', 'Jeep', 'Infiniti', 'Land Rover', 'Kia', 'Mitsubishi', 'BMW', 'FIAT', 'Lincoln', 'Lamborghini', 'Jaguar', 'GMC', 'Toyota', 'Nissan', 'Aston Martin', 'Volvo', 'Ferrari', 'Chrysler', 'Bentley', 'Tesla', 'Ram', 'Alfa Romeo', 'Subaru', 'Buick', 'Lexus', 'Audi')  
str(makes)

## 'data.frame': 44 obs. of 45 variables:  
## $ MakeName : Factor w/ 44 levels "","Acura","Alfa Romeo",..: 30 28 32 10 35 2 31 15 16 39 ...  
## $ McLaren : int 1 0 1 0 1 1 1 1 0 1 ...  
## $ Maserati : int 0 1 0 0 0 0 0 0 1 0 ...  
## $ MINI : int 1 0 1 0 0 0 0 0 1 1 ...  
## $ Chevrolet : int 0 1 0 1 1 0 0 0 1 0 ...  
## $ Porsche : int 0 1 0 0 1 0 0 0 1 0 ...  
## $ Acura : int 0 1 0 1 1 1 1 1 1 0 ...  
## $ Mercedes-Benz: int 0 0 0 0 0 0 1 0 0 0 ...  
## $ Ford : int 0 0 0 0 0 0 0 1 0 0 ...  
## $ Genesis : int 0 0 0 0 0 0 0 0 1 0 ...  
## $ Smart : int 0 0 0 0 0 0 0 0 1 1 ...  
## $ Dodge : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Scion : int 0 0 0 0 0 0 0 0 1 0 ...  
## $ Rolls-Royce : int 0 0 0 0 0 0 0 1 1 0 ...  
## $ Cadillac : int 0 0 0 0 0 0 0 1 0 0 ...  
## $ Honda : int 1 1 1 0 1 1 1 1 0 1 ...  
## $ Hyundai : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Volkswagen : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Mazda : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Jeep : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Infiniti : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Land Rover : int 0 0 0 0 0 0 0 1 0 0 ...  
## $ Kia : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Mitsubishi : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ BMW : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ FIAT : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Lincoln : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Lamborghini : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Jaguar : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ GMC : int 0 0 0 0 1 0 0 1 1 0 ...  
## $ Toyota : int 1 0 1 1 0 0 1 0 0 0 ...  
## $ Nissan : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Aston Martin : int 0 1 0 1 1 1 0 0 1 0 ...  
## $ Volvo : int 0 0 0 0 1 0 0 0 1 0 ...  
## $ Ferrari : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Chrysler : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Bentley : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Tesla : int 0 0 0 0 1 0 0 0 0 0 ...  
## $ Ram : int 0 0 0 0 0 0 0 0 0 1 ...  
## $ Alfa Romeo : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Subaru : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Buick : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Lexus : int 0 0 0 1 0 0 0 0 0 0 ...  
## $ Audi : int 0 0 0 0 1 0 1 0 0 0 ...  
## $ NA : int 0 0 0 0 0 0 0 0 0 0 ...

# Remove columns  
# makes$MakeName = NULL  
# Remove duplicates  
makes = unique(makes)  
str(makes)

## 'data.frame': 44 obs. of 45 variables:  
## $ MakeName : Factor w/ 44 levels "","Acura","Alfa Romeo",..: 30 28 32 10 35 2 31 15 16 39 ...  
## $ McLaren : int 1 0 1 0 1 1 1 1 0 1 ...  
## $ Maserati : int 0 1 0 0 0 0 0 0 1 0 ...  
## $ MINI : int 1 0 1 0 0 0 0 0 1 1 ...  
## $ Chevrolet : int 0 1 0 1 1 0 0 0 1 0 ...  
## $ Porsche : int 0 1 0 0 1 0 0 0 1 0 ...  
## $ Acura : int 0 1 0 1 1 1 1 1 1 0 ...  
## $ Mercedes-Benz: int 0 0 0 0 0 0 1 0 0 0 ...  
## $ Ford : int 0 0 0 0 0 0 0 1 0 0 ...  
## $ Genesis : int 0 0 0 0 0 0 0 0 1 0 ...  
## $ Smart : int 0 0 0 0 0 0 0 0 1 1 ...  
## $ Dodge : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Scion : int 0 0 0 0 0 0 0 0 1 0 ...  
## $ Rolls-Royce : int 0 0 0 0 0 0 0 1 1 0 ...  
## $ Cadillac : int 0 0 0 0 0 0 0 1 0 0 ...  
## $ Honda : int 1 1 1 0 1 1 1 1 0 1 ...  
## $ Hyundai : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Volkswagen : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Mazda : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Jeep : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Infiniti : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Land Rover : int 0 0 0 0 0 0 0 1 0 0 ...  
## $ Kia : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Mitsubishi : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ BMW : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ FIAT : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Lincoln : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Lamborghini : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Jaguar : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ GMC : int 0 0 0 0 1 0 0 1 1 0 ...  
## $ Toyota : int 1 0 1 1 0 0 1 0 0 0 ...  
## $ Nissan : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Aston Martin : int 0 1 0 1 1 1 0 0 1 0 ...  
## $ Volvo : int 0 0 0 0 1 0 0 0 1 0 ...  
## $ Ferrari : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Chrysler : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Bentley : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Tesla : int 0 0 0 0 1 0 0 0 0 0 ...  
## $ Ram : int 0 0 0 0 0 0 0 0 0 1 ...  
## $ Alfa Romeo : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Subaru : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Buick : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Lexus : int 0 0 0 1 0 0 0 0 0 0 ...  
## $ Audi : int 0 0 0 0 1 0 1 0 0 0 ...  
## $ NA : int 0 0 0 0 0 0 0 0 0 0 ...

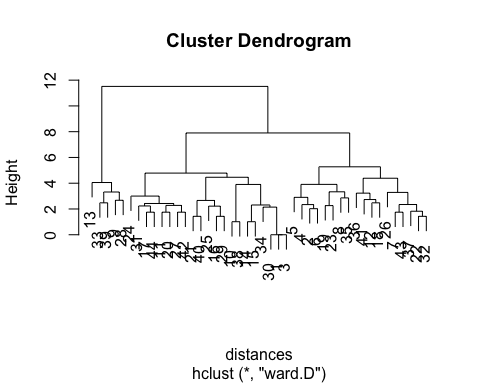
distances = dist(makes, method = "euclidean")

## Warning in dist(makes, method = "euclidean"): NAs introduced by coercion

# For dendrogram  
clusterMovies = hclust(distances, method = "ward")

## The "ward" method has been renamed to "ward.D"; note new "ward.D2"

plot(clusterMovies)



# 10 clusters, and average value for each cluster.  
clusterGroups = cutree(clusterMovies, k = 11)  
# Mean of each cluster, of which are "Action""  
tapply(makes$Honda, clusterGroups, mean)

## 1 2 3 4 5 6 7   
## 1.0000000 0.7500000 0.8333333 0.5000000 0.5000000 1.0000000 0.1250000   
## 8 9 10 11   
## 1.0000000 0.0000000 1.0000000 1.0000000

tapply(makes$Toyota, clusterGroups, mean)

## 1 2 3 4 5 6 7 8 9 10 11   
## 1.000 0.250 1.000 0.000 0.250 0.000 0.625 0.500 0.000 1.000 0.500

tapply(makes$BMW, clusterGroups, mean)

## 1 2 3 4 5 6 7 8 9 10 11   
## 0.000 0.000 0.000 0.000 0.000 0.000 0.125 0.000 0.000 0.000 0.000

# Find the cluster that a movie has fallen in.  
# This gives us the row number.  
subset(makes, MakeName == "Porsche")

## MakeName McLaren Maserati MINI Chevrolet Porsche Acura Mercedes-Benz  
## 5 Porsche 1 0 0 1 1 1 0  
## Ford Genesis Smart Dodge Scion Rolls-Royce Cadillac Honda Hyundai  
## 5 0 0 0 0 0 0 0 1 0  
## Volkswagen Mazda Jeep Infiniti Land Rover Kia Mitsubishi BMW FIAT  
## 5 0 0 0 0 0 0 0 0 0  
## Lincoln Lamborghini Jaguar GMC Toyota Nissan Aston Martin Volvo Ferrari  
## 5 0 0 0 1 0 0 1 1 0  
## Chrysler Bentley Tesla Ram Alfa Romeo Subaru Buick Lexus Audi NA  
## 5 0 0 1 0 0 0 0 0 1 0

# Which clusterGroup is movie n in?  
clusterGroups[5]

## 5   
## 2

cluster2 = subset(makes, clusterGroups == 2)  
head(cluster2)

## MakeName McLaren Maserati MINI Chevrolet Porsche Acura Mercedes-Benz  
## 2 Maserati 0 1 0 1 1 1 0  
## 4 Chevrolet 0 0 0 1 0 1 0  
## 5 Porsche 1 0 0 1 1 1 0  
## 6 Acura 1 0 0 0 0 1 0  
## Ford Genesis Smart Dodge Scion Rolls-Royce Cadillac Honda Hyundai  
## 2 0 0 0 0 0 0 0 1 0  
## 4 0 0 0 0 0 0 0 0 0  
## 5 0 0 0 0 0 0 0 1 0  
## 6 0 0 0 0 0 0 0 1 0  
## Volkswagen Mazda Jeep Infiniti Land Rover Kia Mitsubishi BMW FIAT  
## 2 0 0 0 0 0 0 0 0 0  
## 4 0 0 0 0 0 0 0 0 0  
## 5 0 0 0 0 0 0 0 0 0  
## 6 0 0 0 0 0 0 0 0 0  
## Lincoln Lamborghini Jaguar GMC Toyota Nissan Aston Martin Volvo Ferrari  
## 2 0 0 0 0 0 0 1 0 0  
## 4 0 0 0 0 1 0 1 0 0  
## 5 0 0 0 1 0 0 1 1 0  
## 6 0 0 0 0 0 0 1 0 0  
## Chrysler Bentley Tesla Ram Alfa Romeo Subaru Buick Lexus Audi NA  
## 2 0 0 0 0 0 0 0 0 0 0  
## 4 0 0 0 0 0 0 0 1 0 0  
## 5 0 0 1 0 0 0 0 0 1 0  
## 6 0 0 0 0 0 0 0 0 0 0

# Find the cluster that a movie has fallen in.  
# This gives us the row number.  
subset(makes, MakeName == "Honda")

## MakeName McLaren Maserati MINI Chevrolet Porsche Acura Mercedes-Benz  
## 15 Honda 1 0 1 0 0 1 0  
## Ford Genesis Smart Dodge Scion Rolls-Royce Cadillac Honda Hyundai  
## 15 0 0 0 0 0 0 0 1 0  
## Volkswagen Mazda Jeep Infiniti Land Rover Kia Mitsubishi BMW FIAT  
## 15 0 0 0 0 0 0 0 0 0  
## Lincoln Lamborghini Jaguar GMC Toyota Nissan Aston Martin Volvo Ferrari  
## 15 0 0 0 0 1 0 0 0 0  
## Chrysler Bentley Tesla Ram Alfa Romeo Subaru Buick Lexus Audi NA  
## 15 0 0 0 0 0 0 0 0 0 0

# Which clusterGroup is movie n in?  
clusterGroups[15]

## 15   
## 1

cluster7 = subset(makes, clusterGroups == 1)  
head(cluster7)

## MakeName McLaren Maserati MINI Chevrolet Porsche Acura Mercedes-Benz  
## 1 McLaren 1 0 1 0 0 0 0  
## 3 MINI 1 0 1 0 0 0 0  
## 14 Cadillac 1 0 1 0 0 1 0  
## 15 Honda 1 0 1 0 0 1 0  
## 30 Toyota 1 0 1 0 0 0 0  
## 34 Ferrari 1 0 1 0 0 0 0  
## Ford Genesis Smart Dodge Scion Rolls-Royce Cadillac Honda Hyundai  
## 1 0 0 0 0 0 0 0 1 0  
## 3 0 0 0 0 0 0 0 1 0  
## 14 0 0 0 0 0 0 1 1 0  
## 15 0 0 0 0 0 0 0 1 0  
## 30 0 0 0 0 0 0 0 1 0  
## 34 0 0 1 0 0 0 0 1 0  
## Volkswagen Mazda Jeep Infiniti Land Rover Kia Mitsubishi BMW FIAT  
## 1 0 0 0 0 0 0 0 0 0  
## 3 0 0 0 0 0 0 0 0 0  
## 14 0 0 0 0 0 0 0 0 0  
## 15 0 0 0 0 0 0 0 0 0  
## 30 0 0 0 0 0 0 0 0 0  
## 34 0 0 0 0 0 0 0 0 0  
## Lincoln Lamborghini Jaguar GMC Toyota Nissan Aston Martin Volvo Ferrari  
## 1 0 0 0 0 1 0 0 0 0  
## 3 0 0 0 0 1 0 0 0 0  
## 14 0 0 0 0 1 0 0 0 0  
## 15 0 0 0 0 1 0 0 0 0  
## 30 0 0 0 0 1 0 0 0 0  
## 34 0 0 0 0 1 0 0 0 1  
## Chrysler Bentley Tesla Ram Alfa Romeo Subaru Buick Lexus Audi NA  
## 1 0 0 0 0 0 0 0 0 0 0  
## 3 0 0 0 0 0 0 0 0 0 0  
## 14 0 0 0 0 0 0 0 0 0 0  
## 15 0 0 0 0 0 0 0 0 0 0  
## 30 0 0 0 0 0 0 0 0 0 0  
## 34 0 0 0 0 0 0 0 0 0 0

# Find the cluster that a movie has fallen in.  
# This gives us the row number.  
subset(makes, MakeName == "BMW")

## MakeName McLaren Maserati MINI Chevrolet Porsche Acura Mercedes-Benz  
## 24 BMW 1 1 1 0 0 0 0  
## Ford Genesis Smart Dodge Scion Rolls-Royce Cadillac Honda Hyundai  
## 24 1 0 0 0 0 0 0 0 0  
## Volkswagen Mazda Jeep Infiniti Land Rover Kia Mitsubishi BMW FIAT  
## 24 0 0 0 0 0 0 0 1 0  
## Lincoln Lamborghini Jaguar GMC Toyota Nissan Aston Martin Volvo Ferrari  
## 24 0 0 0 0 0 0 0 0 0  
## Chrysler Bentley Tesla Ram Alfa Romeo Subaru Buick Lexus Audi NA  
## 24 0 0 0 0 0 0 0 1 0 0

# Which clusterGroup is movie n in?  
clusterGroups[24]

## 24   
## 7

cluster2 = subset(makes, clusterGroups == 7)  
cluster2

## MakeName McLaren Maserati MINI Chevrolet Porsche Acura Mercedes-Benz  
## 11 Dodge 1 0 1 0 0 0 0  
## 17 Volkswagen 0 0 1 0 0 0 0  
## 20 Infiniti 1 0 1 0 0 0 0  
## 24 BMW 1 1 1 0 0 0 0  
## 27 Lamborghini 1 0 0 0 0 0 0  
## 31 Nissan 1 0 0 0 0 0 1  
## 42 Lexus 1 0 0 0 0 0 0  
## 44 0 0 0 0 0 0 0  
## Ford Genesis Smart Dodge Scion Rolls-Royce Cadillac Honda Hyundai  
## 11 0 0 0 1 0 0 0 0 0  
## 17 0 0 0 0 0 0 0 0 0  
## 20 0 0 0 0 0 0 0 0 0  
## 24 1 0 0 0 0 0 0 0 0  
## 27 0 0 0 0 0 0 0 1 0  
## 31 0 0 0 0 0 0 0 0 0  
## 42 0 0 0 0 0 0 0 0 0  
## 44 0 0 0 0 0 0 0 0 0  
## Volkswagen Mazda Jeep Infiniti Land Rover Kia Mitsubishi BMW FIAT  
## 11 0 0 0 0 0 0 0 0 0  
## 17 1 0 0 0 0 0 0 0 0  
## 20 0 0 0 1 0 0 0 0 0  
## 24 0 0 0 0 0 0 0 1 0  
## 27 0 0 0 0 0 0 0 0 0  
## 31 0 0 0 0 0 0 0 0 0  
## 42 0 0 0 0 0 0 0 0 0  
## 44 0 0 0 0 0 0 0 0 0  
## Lincoln Lamborghini Jaguar GMC Toyota Nissan Aston Martin Volvo Ferrari  
## 11 0 0 0 0 1 0 0 0 0  
## 17 0 0 0 0 1 0 0 0 0  
## 20 0 0 0 0 1 0 0 0 0  
## 24 0 0 0 0 0 0 0 0 0  
## 27 0 1 0 0 1 0 0 0 0  
## 31 0 0 0 0 0 1 0 0 0  
## 42 0 0 0 0 1 0 0 0 0  
## 44 0 0 0 0 0 0 0 0 0  
## Chrysler Bentley Tesla Ram Alfa Romeo Subaru Buick Lexus Audi NA  
## 11 0 0 0 0 0 1 0 0 0 0  
## 17 0 0 0 0 0 0 0 0 0 0  
## 20 0 0 0 0 0 0 0 0 0 0  
## 24 0 0 0 0 0 0 0 1 0 0  
## 27 0 0 0 0 0 0 0 0 0 0  
## 31 0 0 0 0 0 0 0 0 0 0  
## 42 0 0 0 0 0 0 0 1 0 0  
## 44 0 0 0 0 0 0 0 0 0 0

# Find the cluster that a movie has fallen in.  
# This gives us the row number.  
subset(makes, MakeName == "GMC")

## MakeName McLaren Maserati MINI Chevrolet Porsche Acura Mercedes-Benz  
## 29 GMC 1 0 1 0 1 0 0  
## Ford Genesis Smart Dodge Scion Rolls-Royce Cadillac Honda Hyundai  
## 29 0 0 0 0 0 0 0 1 0  
## Volkswagen Mazda Jeep Infiniti Land Rover Kia Mitsubishi BMW FIAT  
## 29 0 0 0 0 0 0 0 0 0  
## Lincoln Lamborghini Jaguar GMC Toyota Nissan Aston Martin Volvo Ferrari  
## 29 0 0 0 1 1 0 0 0 0  
## Chrysler Bentley Tesla Ram Alfa Romeo Subaru Buick Lexus Audi NA  
## 29 0 0 0 0 0 0 0 0 0 0

# Which clusterGroup is movie n in?  
clusterGroups[29]

## 29   
## 10

cluster2 = subset(makes, clusterGroups == 10)  
head(cluster2)

## MakeName McLaren Maserati MINI Chevrolet Porsche Acura Mercedes-Benz  
## 16 Hyundai 1 0 1 0 0 0 0  
## 21 Land Rover 1 0 0 0 1 0 0  
## 25 FIAT 1 0 1 0 0 0 0  
## 29 GMC 1 0 1 0 1 0 0  
## 40 Subaru 1 0 0 0 0 0 0  
## Ford Genesis Smart Dodge Scion Rolls-Royce Cadillac Honda Hyundai  
## 16 0 0 0 0 0 0 0 1 1  
## 21 0 0 0 0 0 0 0 1 0  
## 25 0 0 0 0 0 0 0 1 0  
## 29 0 0 0 0 0 0 0 1 0  
## 40 0 0 0 0 0 0 0 1 0  
## Volkswagen Mazda Jeep Infiniti Land Rover Kia Mitsubishi BMW FIAT  
## 16 0 0 0 0 0 0 0 0 0  
## 21 0 0 0 0 1 0 0 0 0  
## 25 0 0 0 0 0 0 0 0 1  
## 29 0 0 0 0 0 0 0 0 0  
## 40 0 0 0 0 0 0 0 0 0  
## Lincoln Lamborghini Jaguar GMC Toyota Nissan Aston Martin Volvo Ferrari  
## 16 0 0 0 1 1 0 0 0 0  
## 21 0 0 0 1 1 0 0 0 0  
## 25 0 0 0 1 1 0 1 0 0  
## 29 0 0 0 1 1 0 0 0 0  
## 40 0 0 0 1 1 0 0 0 0  
## Chrysler Bentley Tesla Ram Alfa Romeo Subaru Buick Lexus Audi NA  
## 16 0 0 0 0 0 0 0 0 0 0  
## 21 0 0 0 0 0 1 0 0 0 0  
## 25 0 0 0 0 0 0 0 0 1 0  
## 29 0 0 0 0 0 0 0 0 0 0  
## 40 0 0 0 0 0 1 0 0 0 0

# Find the cluster that a movie has fallen in.  
# This gives us the row number.  
subset(makes, MakeName == "Volvo")

## MakeName McLaren Maserati MINI Chevrolet Porsche Acura Mercedes-Benz  
## 33 Volvo 1 1 1 1 1 1 0  
## Ford Genesis Smart Dodge Scion Rolls-Royce Cadillac Honda Hyundai  
## 33 0 1 1 0 0 1 0 1 0  
## Volkswagen Mazda Jeep Infiniti Land Rover Kia Mitsubishi BMW FIAT  
## 33 0 0 0 0 0 1 0 0 0  
## Lincoln Lamborghini Jaguar GMC Toyota Nissan Aston Martin Volvo Ferrari  
## 33 1 0 0 0 1 0 0 1 0  
## Chrysler Bentley Tesla Ram Alfa Romeo Subaru Buick Lexus Audi NA  
## 33 0 0 0 0 0 0 0 0 0 0

# Which clusterGroup is movie n in?  
clusterGroups[33]

## 33   
## 5

cluster2 = subset(makes, clusterGroups == 5)  
head(cluster2)

## MakeName McLaren Maserati MINI Chevrolet Porsche Acura Mercedes-Benz  
## 9 Genesis 0 1 1 1 1 1 0  
## 28 Jaguar 0 1 1 1 1 1 0  
## 33 Volvo 1 1 1 1 1 1 0  
## 39 Alfa Romeo 1 1 1 1 1 1 0  
## Ford Genesis Smart Dodge Scion Rolls-Royce Cadillac Honda Hyundai  
## 9 0 1 1 0 1 1 0 0 0  
## 28 1 1 1 0 1 0 0 1 0  
## 33 0 1 1 0 0 1 0 1 0  
## 39 1 1 1 0 0 1 0 0 0  
## Volkswagen Mazda Jeep Infiniti Land Rover Kia Mitsubishi BMW FIAT  
## 9 0 0 0 0 0 0 0 0 0  
## 28 0 0 0 0 0 0 0 0 0  
## 33 0 0 0 0 0 1 0 0 0  
## 39 0 0 0 0 0 1 0 0 1  
## Lincoln Lamborghini Jaguar GMC Toyota Nissan Aston Martin Volvo Ferrari  
## 9 0 0 0 1 0 0 1 1 0  
## 28 1 0 1 0 0 0 1 1 0  
## 33 1 0 0 0 1 0 0 1 0  
## 39 1 0 0 0 0 0 1 1 0  
## Chrysler Bentley Tesla Ram Alfa Romeo Subaru Buick Lexus Audi NA  
## 9 0 0 0 0 0 0 0 0 0 0  
## 28 0 0 1 0 0 0 0 0 0 0  
## 33 0 0 0 0 0 0 0 0 0 0  
## 39 0 0 0 0 1 0 0 0 0 0

# Find the cluster that a movie has fallen in.  
# This gives us the row number.  
subset(makes, MakeName == "Mercedes-Benz")

## MakeName McLaren Maserati MINI Chevrolet Porsche Acura  
## 7 Mercedes-Benz 1 0 0 0 0 1  
## Mercedes-Benz Ford Genesis Smart Dodge Scion Rolls-Royce Cadillac Honda  
## 7 1 0 0 0 0 0 0 0 1  
## Hyundai Volkswagen Mazda Jeep Infiniti Land Rover Kia Mitsubishi BMW  
## 7 0 0 0 0 0 0 0 0 0  
## FIAT Lincoln Lamborghini Jaguar GMC Toyota Nissan Aston Martin Volvo  
## 7 0 0 0 0 0 1 0 0 0  
## Ferrari Chrysler Bentley Tesla Ram Alfa Romeo Subaru Buick Lexus Audi NA  
## 7 0 0 0 0 0 0 0 0 0 1 0

# Which clusterGroup is movie n in?  
clusterGroups[7]

## 7   
## 3

cluster2 = subset(makes, clusterGroups == 3)  
head(cluster2)

## MakeName McLaren Maserati MINI Chevrolet Porsche Acura  
## 7 Mercedes-Benz 1 0 0 0 0 1  
## 22 Kia 1 0 1 0 0 1  
## 26 Lincoln 1 1 1 1 0 1  
## 32 Aston Martin 1 0 1 0 0 1  
## 37 Tesla 1 0 1 1 0 1  
## 43 Audi 1 0 1 0 0 1  
## Mercedes-Benz Ford Genesis Smart Dodge Scion Rolls-Royce Cadillac Honda  
## 7 1 0 0 0 0 0 0 0 1  
## 22 0 0 0 0 0 0 0 0 1  
## 26 1 0 0 0 0 0 0 0 0  
## 32 0 0 0 0 0 0 0 0 1  
## 37 0 0 0 0 0 0 0 0 1  
## 43 1 0 0 0 0 0 0 0 1  
## Hyundai Volkswagen Mazda Jeep Infiniti Land Rover Kia Mitsubishi BMW  
## 7 0 0 0 0 0 0 0 0 0  
## 22 0 0 0 0 0 0 1 0 0  
## 26 0 0 0 0 0 0 0 0 0  
## 32 0 0 0 0 0 0 0 0 0  
## 37 0 0 0 0 0 0 0 0 0  
## 43 0 0 0 0 0 0 0 0 0  
## FIAT Lincoln Lamborghini Jaguar GMC Toyota Nissan Aston Martin Volvo  
## 7 0 0 0 0 0 1 0 0 0  
## 22 0 0 0 0 0 1 0 1 0  
## 26 0 1 1 0 0 1 0 0 0  
## 32 0 0 0 0 0 1 0 1 0  
## 37 0 0 0 0 0 1 0 1 0  
## 43 1 0 0 0 0 1 0 1 0  
## Ferrari Chrysler Bentley Tesla Ram Alfa Romeo Subaru Buick Lexus Audi  
## 7 0 0 0 0 0 0 0 0 0 1  
## 22 0 0 0 0 0 0 0 0 0 0  
## 26 0 0 0 0 0 0 0 0 0 1  
## 32 0 0 0 0 0 0 0 0 0 1  
## 37 0 0 0 1 0 0 0 0 0 1  
## 43 0 0 0 0 0 0 0 0 0 1  
## NA  
## 7 0  
## 22 0  
## 26 0  
## 32 0  
## 37 0  
## 43 0

# scion, jeep, mazda  
subset(makes, MakeName == "Scion")

## MakeName McLaren Maserati MINI Chevrolet Porsche Acura Mercedes-Benz  
## 12 Scion 1 0 1 0 0 1 0  
## Ford Genesis Smart Dodge Scion Rolls-Royce Cadillac Honda Hyundai  
## 12 0 0 1 0 1 0 0 1 0  
## Volkswagen Mazda Jeep Infiniti Land Rover Kia Mitsubishi BMW FIAT  
## 12 0 1 0 0 0 0 0 0 1  
## Lincoln Lamborghini Jaguar GMC Toyota Nissan Aston Martin Volvo Ferrari  
## 12 0 0 0 0 0 0 0 0 0  
## Chrysler Bentley Tesla Ram Alfa Romeo Subaru Buick Lexus Audi NA  
## 12 0 0 0 0 0 0 0 0 0 0

# Which clusterGroup is movie n in?  
clusterGroups[12]

## 12   
## 8

cluster2 = subset(makes, clusterGroups == 8)  
head(cluster2)

## MakeName McLaren Maserati MINI Chevrolet Porsche Acura Mercedes-Benz  
## 12 Scion 1 0 1 0 0 1 0  
## 18 Mazda 1 0 1 0 0 1 0  
## 36 Bentley 1 1 1 0 1 1 0  
## 41 Buick 0 0 1 0 0 1 1  
## Ford Genesis Smart Dodge Scion Rolls-Royce Cadillac Honda Hyundai  
## 12 0 0 1 0 1 0 0 1 0  
## 18 0 0 1 0 0 0 0 1 0  
## 36 0 0 1 0 0 0 1 1 0  
## 41 0 0 1 0 0 0 0 1 0  
## Volkswagen Mazda Jeep Infiniti Land Rover Kia Mitsubishi BMW FIAT  
## 12 0 1 0 0 0 0 0 0 1  
## 18 0 1 0 0 0 0 0 0 0  
## 36 0 1 0 0 0 0 0 0 0  
## 41 0 0 0 0 0 0 0 0 0  
## Lincoln Lamborghini Jaguar GMC Toyota Nissan Aston Martin Volvo Ferrari  
## 12 0 0 0 0 0 0 0 0 0  
## 18 0 0 0 1 1 0 1 0 0  
## 36 0 0 0 0 1 0 1 0 0  
## 41 0 0 0 1 0 0 1 0 0  
## Chrysler Bentley Tesla Ram Alfa Romeo Subaru Buick Lexus Audi NA  
## 12 0 0 0 0 0 0 0 0 0 0  
## 18 0 0 0 0 0 1 0 0 0 0  
## 36 0 1 0 0 0 0 0 0 1 0  
## 41 0 0 0 0 0 0 1 0 0 0

subset(makes, MakeName == "Jeep")

## MakeName McLaren Maserati MINI Chevrolet Porsche Acura Mercedes-Benz  
## 19 Jeep 0 0 0 0 0 1 0  
## Ford Genesis Smart Dodge Scion Rolls-Royce Cadillac Honda Hyundai  
## 19 0 0 0 0 0 0 0 1 0  
## Volkswagen Mazda Jeep Infiniti Land Rover Kia Mitsubishi BMW FIAT  
## 19 0 0 1 0 0 0 0 0 0  
## Lincoln Lamborghini Jaguar GMC Toyota Nissan Aston Martin Volvo Ferrari  
## 19 0 0 0 0 1 0 0 0 0  
## Chrysler Bentley Tesla Ram Alfa Romeo Subaru Buick Lexus Audi NA  
## 19 0 0 0 0 0 1 0 0 0 0

# Which clusterGroup is movie n in?  
clusterGroups[19]

## 19   
## 11

cluster2 = subset(makes, clusterGroups == 11)  
head(cluster2)

## MakeName McLaren Maserati MINI Chevrolet Porsche Acura Mercedes-Benz  
## 19 Jeep 0 0 0 0 0 1 0  
## 23 Mitsubishi 0 0 0 0 0 0 0  
## Ford Genesis Smart Dodge Scion Rolls-Royce Cadillac Honda Hyundai  
## 19 0 0 0 0 0 0 0 1 0  
## 23 0 0 0 0 0 0 1 1 0  
## Volkswagen Mazda Jeep Infiniti Land Rover Kia Mitsubishi BMW FIAT  
## 19 0 0 1 0 0 0 0 0 0  
## 23 0 0 0 0 0 0 1 0 0  
## Lincoln Lamborghini Jaguar GMC Toyota Nissan Aston Martin Volvo Ferrari  
## 19 0 0 0 0 1 0 0 0 0  
## 23 0 0 0 0 0 0 0 0 0  
## Chrysler Bentley Tesla Ram Alfa Romeo Subaru Buick Lexus Audi NA  
## 19 0 0 0 0 0 1 0 0 0 0  
## 23 0 0 0 0 0 1 0 0 0 0

subset(makes, MakeName == "Mazda")

## MakeName McLaren Maserati MINI Chevrolet Porsche Acura Mercedes-Benz  
## 18 Mazda 1 0 1 0 0 1 0  
## Ford Genesis Smart Dodge Scion Rolls-Royce Cadillac Honda Hyundai  
## 18 0 0 1 0 0 0 0 1 0  
## Volkswagen Mazda Jeep Infiniti Land Rover Kia Mitsubishi BMW FIAT  
## 18 0 1 0 0 0 0 0 0 0  
## Lincoln Lamborghini Jaguar GMC Toyota Nissan Aston Martin Volvo Ferrari  
## 18 0 0 0 1 1 0 1 0 0  
## Chrysler Bentley Tesla Ram Alfa Romeo Subaru Buick Lexus Audi NA  
## 18 0 0 0 0 0 1 0 0 0 0

# Which clusterGroup is movie n in?  
clusterGroups[18]

## 18   
## 8

cluster2 = subset(makes, clusterGroups == 8)  
head(cluster2)

## MakeName McLaren Maserati MINI Chevrolet Porsche Acura Mercedes-Benz  
## 12 Scion 1 0 1 0 0 1 0  
## 18 Mazda 1 0 1 0 0 1 0  
## 36 Bentley 1 1 1 0 1 1 0  
## 41 Buick 0 0 1 0 0 1 1  
## Ford Genesis Smart Dodge Scion Rolls-Royce Cadillac Honda Hyundai  
## 12 0 0 1 0 1 0 0 1 0  
## 18 0 0 1 0 0 0 0 1 0  
## 36 0 0 1 0 0 0 1 1 0  
## 41 0 0 1 0 0 0 0 1 0  
## Volkswagen Mazda Jeep Infiniti Land Rover Kia Mitsubishi BMW FIAT  
## 12 0 1 0 0 0 0 0 0 1  
## 18 0 1 0 0 0 0 0 0 0  
## 36 0 1 0 0 0 0 0 0 0  
## 41 0 0 0 0 0 0 0 0 0  
## Lincoln Lamborghini Jaguar GMC Toyota Nissan Aston Martin Volvo Ferrari  
## 12 0 0 0 0 0 0 0 0 0  
## 18 0 0 0 1 1 0 1 0 0  
## 36 0 0 0 0 1 0 1 0 0  
## 41 0 0 0 1 0 0 1 0 0  
## Chrysler Bentley Tesla Ram Alfa Romeo Subaru Buick Lexus Audi NA  
## 12 0 0 0 0 0 0 0 0 0 0  
## 18 0 0 0 0 0 1 0 0 0 0  
## 36 0 1 0 0 0 0 0 0 1 0  
## 41 0 0 0 0 0 0 1 0 0 0

# Lincoln  
subset(makes, MakeName == "Lincoln")

## MakeName McLaren Maserati MINI Chevrolet Porsche Acura Mercedes-Benz  
## 26 Lincoln 1 1 1 1 0 1 1  
## Ford Genesis Smart Dodge Scion Rolls-Royce Cadillac Honda Hyundai  
## 26 0 0 0 0 0 0 0 0 0  
## Volkswagen Mazda Jeep Infiniti Land Rover Kia Mitsubishi BMW FIAT  
## 26 0 0 0 0 0 0 0 0 0  
## Lincoln Lamborghini Jaguar GMC Toyota Nissan Aston Martin Volvo Ferrari  
## 26 1 1 0 0 1 0 0 0 0  
## Chrysler Bentley Tesla Ram Alfa Romeo Subaru Buick Lexus Audi NA  
## 26 0 0 0 0 0 0 0 0 1 0

# Which clusterGroup is movie n in?  
clusterGroups[26]

## 26   
## 3

cluster2 = subset(makes, clusterGroups == 3)  
head(cluster2)

## MakeName McLaren Maserati MINI Chevrolet Porsche Acura  
## 7 Mercedes-Benz 1 0 0 0 0 1  
## 22 Kia 1 0 1 0 0 1  
## 26 Lincoln 1 1 1 1 0 1  
## 32 Aston Martin 1 0 1 0 0 1  
## 37 Tesla 1 0 1 1 0 1  
## 43 Audi 1 0 1 0 0 1  
## Mercedes-Benz Ford Genesis Smart Dodge Scion Rolls-Royce Cadillac Honda  
## 7 1 0 0 0 0 0 0 0 1  
## 22 0 0 0 0 0 0 0 0 1  
## 26 1 0 0 0 0 0 0 0 0  
## 32 0 0 0 0 0 0 0 0 1  
## 37 0 0 0 0 0 0 0 0 1  
## 43 1 0 0 0 0 0 0 0 1  
## Hyundai Volkswagen Mazda Jeep Infiniti Land Rover Kia Mitsubishi BMW  
## 7 0 0 0 0 0 0 0 0 0  
## 22 0 0 0 0 0 0 1 0 0  
## 26 0 0 0 0 0 0 0 0 0  
## 32 0 0 0 0 0 0 0 0 0  
## 37 0 0 0 0 0 0 0 0 0  
## 43 0 0 0 0 0 0 0 0 0  
## FIAT Lincoln Lamborghini Jaguar GMC Toyota Nissan Aston Martin Volvo  
## 7 0 0 0 0 0 1 0 0 0  
## 22 0 0 0 0 0 1 0 1 0  
## 26 0 1 1 0 0 1 0 0 0  
## 32 0 0 0 0 0 1 0 1 0  
## 37 0 0 0 0 0 1 0 1 0  
## 43 1 0 0 0 0 1 0 1 0  
## Ferrari Chrysler Bentley Tesla Ram Alfa Romeo Subaru Buick Lexus Audi  
## 7 0 0 0 0 0 0 0 0 0 1  
## 22 0 0 0 0 0 0 0 0 0 0  
## 26 0 0 0 0 0 0 0 0 0 1  
## 32 0 0 0 0 0 0 0 0 0 1  
## 37 0 0 0 1 0 0 0 0 0 1  
## 43 0 0 0 0 0 0 0 0 0 1  
## NA  
## 7 0  
## 22 0  
## 26 0  
## 32 0  
## 37 0  
## 43 0

# Lamborghini  
subset(makes, MakeName == "Lamborghini")

## MakeName McLaren Maserati MINI Chevrolet Porsche Acura Mercedes-Benz  
## 27 Lamborghini 1 0 0 0 0 0 0  
## Ford Genesis Smart Dodge Scion Rolls-Royce Cadillac Honda Hyundai  
## 27 0 0 0 0 0 0 0 1 0  
## Volkswagen Mazda Jeep Infiniti Land Rover Kia Mitsubishi BMW FIAT  
## 27 0 0 0 0 0 0 0 0 0  
## Lincoln Lamborghini Jaguar GMC Toyota Nissan Aston Martin Volvo Ferrari  
## 27 0 1 0 0 1 0 0 0 0  
## Chrysler Bentley Tesla Ram Alfa Romeo Subaru Buick Lexus Audi NA  
## 27 0 0 0 0 0 0 0 0 0 0

# Which clusterGroup is movie n in?  
clusterGroups[27]

## 27   
## 7

cluster2 = subset(makes, clusterGroups == 7)  
head(cluster2)

## MakeName McLaren Maserati MINI Chevrolet Porsche Acura Mercedes-Benz  
## 11 Dodge 1 0 1 0 0 0 0  
## 17 Volkswagen 0 0 1 0 0 0 0  
## 20 Infiniti 1 0 1 0 0 0 0  
## 24 BMW 1 1 1 0 0 0 0  
## 27 Lamborghini 1 0 0 0 0 0 0  
## 31 Nissan 1 0 0 0 0 0 1  
## Ford Genesis Smart Dodge Scion Rolls-Royce Cadillac Honda Hyundai  
## 11 0 0 0 1 0 0 0 0 0  
## 17 0 0 0 0 0 0 0 0 0  
## 20 0 0 0 0 0 0 0 0 0  
## 24 1 0 0 0 0 0 0 0 0  
## 27 0 0 0 0 0 0 0 1 0  
## 31 0 0 0 0 0 0 0 0 0  
## Volkswagen Mazda Jeep Infiniti Land Rover Kia Mitsubishi BMW FIAT  
## 11 0 0 0 0 0 0 0 0 0  
## 17 1 0 0 0 0 0 0 0 0  
## 20 0 0 0 1 0 0 0 0 0  
## 24 0 0 0 0 0 0 0 1 0  
## 27 0 0 0 0 0 0 0 0 0  
## 31 0 0 0 0 0 0 0 0 0  
## Lincoln Lamborghini Jaguar GMC Toyota Nissan Aston Martin Volvo Ferrari  
## 11 0 0 0 0 1 0 0 0 0  
## 17 0 0 0 0 1 0 0 0 0  
## 20 0 0 0 0 1 0 0 0 0  
## 24 0 0 0 0 0 0 0 0 0  
## 27 0 1 0 0 1 0 0 0 0  
## 31 0 0 0 0 0 1 0 0 0  
## Chrysler Bentley Tesla Ram Alfa Romeo Subaru Buick Lexus Audi NA  
## 11 0 0 0 0 0 1 0 0 0 0  
## 17 0 0 0 0 0 0 0 0 0 0  
## 20 0 0 0 0 0 0 0 0 0 0  
## 24 0 0 0 0 0 0 0 1 0 0  
## 27 0 0 0 0 0 0 0 0 0 0  
## 31 0 0 0 0 0 0 0 0 0 0

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Cmd+Option+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Cmd+Shift+K* to preview the HTML file).