## Marketing Analytics Prof. Swagato Chatterjee Vinod Gupta School of Management Indian Institute of Technology, Kharagpur Lecture 38: Recommendation Engine and Retail Analytics (Contd.)

Hello, everybody, welcome to Marketing Analytics Course, this is Dr. Swagato Chatterjee from the VGSOM IIT, Kharagpur who is taking this course. We are in week seven, and we are discussing Recommendation Engine. So, in this video and probably in the next video also, we will discuss certain examples of recommendation engine and their applications with real world data.

So, in this particular data set, here, we have movie ratings and where people have given ratings for movies. And I also have genre of the movies and I want to find out how I can recommend certain movies based on this genre and based on this, I would say interest towards those genre where data is being collected from let us say, IMDb or let us say Netflix or something like that.

		Embranat Mildary Connections	
a D 7 Eter		Constant Propert Dataset +	E tie •
		Global Environment • C	1
		Data	
¢.		O movies 10329 obs. of 3 van O ratings 105339 obs. of 4 van	riables Iriables
		Files Plots Packages Help Viewer	
Console Terminal -	-0		
C:/Users/Dell3/Deektop/Week7/Session 4/			
<pre>&gt; setwd("C:/Users/Dell5/Desktop/week7/Session 4") #Tinks &lt;- read.csv("Tinks.csv") &gt; movies &lt;- read.csv("movies.csv",stringsAsFactors=FALSE) &gt; ratings &lt;- read.csv("ratings.csv")</pre>			

(Refer Slide Time: 1:10)



4.0 1217896523

4.0 1217896150



So here, the dataset, if I just read the dataset, the dataset looks like this, there are two datasets which we will merge. One is called movies dataset where the movie ID, the title of the movie, and the genres are written here. And then the ratings dataset, where basically around 1 lakh ratings are there, 1 lakh 5 thousand ratings are there, user ID, movie ID, rating and timestamp at what time point this guy has given the rating. So I will actually find out, so, there are basically if I am not wrong, ratings \$ user ID, if I am trying to find out and then find out the unique of them. So, how many unique users are there?

If I am not wrong, there are 668 unique users, yes. So, there 1 to 668 unique users are there and how many movies? Basically, there 10,329 movies. So, that is what I have. So, now, based on that, I have to find out which movie should be recommended.

(Refer Slide Time: 2:17)









Now, if somebody has already rated a movie, then you will not see the movie anymore, then I have to find out his similar movie which he might like and likeness data, whether he will like or dislike a data, a movie will come from basically the rating that this guy has provided. So, I will ask for a library called recommender lab and ggplot2, recommender lab whatever we did before, the item based and user based collaborative filtering, this guy will do it on its own quickly.

For bigger data, it helps so, recommender lab and ggplot2 is two libraries that I am calling, if you have not installed it, you have to first install it and then and how to install it? You can write install .packages and then the package name or you can go here and install and write the package name here and then that will install the package. So, I have installed the package beforehand and I am just running this thing here. So, the first thing is processing of the data, I will break the data into genres.

So, for that what I am doing is I am creating a genres column, so which is a genre column, only the genres of my dataset and then I am using a library called data.table. And then in with this library, I am breaking the genre with c, there is called string split. So, this string split will actually split the screen based of the genre , 1, that means genre's first column, based on this particular sign. Whenever this sign occurs, it will just split it and then stringAsFactors is equal to false and it will create a dataset.

(Refer	Slide	Time:	4:03	)
--------	-------	-------	------	---

movies	recour =genres2 =genres =	- Environment History Connections	-
		9, 💣 🔒 📅 Import Dutaset + 🧹 🗮 List	. 8
	moviesSgenres	- Q,	
1	Adventure Anigation Children Comedy Fantasy	Data	
2	Adventure)Children/Fantasy	Ogenres 10329 obs. of 1 variable	
3	Comedy/Homance	Ogenres2 10329 obs. of 10 variables	
4	Comedy(Drama)Romance	0 movies 10329 obs. of s variables	
5	Comedy	Vialings 10333 005. 01 4 valiables	
6	Action)Crime(Thriller		
7	Comedy/Romance	Files Plots Packages Help Viewer	-0
	Adventure)Children	er als Plann Plann Plannt • O	
. 9	Action		
10	ActioniAdventure(Thriller	2	
howing 1	to 10 of 10,329 entries		
onsole	Terminal =	-0	
C:/Users/	Dell5/Deuktop/Week7/Semion 4/		
view	ry(data:table) jmessage: 'dat.table'was built under R ve s2 <- as.data.frame(tstrsplitgen type (genres2) (genres2)	tOTthreads). Latest news: r-datatable.com sion 3.6.2 s(.1). (1)*. convertnewe), s=FALSE)	

moviereco.r a		0.1	General			Add	ins +					B Project 0
101217	genres2	× 🔲	genres ×							-0	Environment	History Connections
	Filter								Q		er 🖯 🕾	* Import Dataset • 🔏 📃 List •
c.Adventure.	.Adventure.	.Comedy.	.Comedy.	.Comedy.	Action	c_An	imation	Children.	RomanceDra	maNACrime	Global Envi	rooment • Q
1 Adventure		14	P			A	- 197			- 0	Data	10329 obs. of 1 variable
2 Adventure						Child	nen (				genres2	10329 obs. of 10 variables
S Comedy						noma	nce				Omovies	10329 obs. of 3 variables
Comedy						244					0 ratings	105339 obs. of 4 variables
6 Action						Crime						
7 Comedy						Roma	nce				Files Plots	Packages Help Viewer
8 Adventure						Child	nen					Inn Deport . 0
9 Action						700						
peing 1 to 10 of 10	329 entries											
nsole Termina	н									-0		
library(dat ta.table 1. rning messa ckage 'data genres2 <- view(genres view(genres	a.table) 12.8 usi uge: table' as.data. (2) (3)	ng 2 ti was bu frame(	hreads ilt und tstrspl strings	(see ?) ler R vu it(geni typi AsFacto	etDTth rsion es[,1] e.conve ers=FAL	reads) 3.6.2 , '[1] rt=TRU SE)	. Late	est new	us: r-datat	able.com		
00	X	3		0								
• 9 * •	8.8	01/	Getetli			• Add	ns •					I Project Ø
moviereca.r =	genres2	×								-0	Environment	History Connections
0.1201.2	Filter								Q		C 8 8	Import Dutaset • 🧹 📄 List •
* 1	2	1	4	5	6	7		. 9	10		Global Envi	ionment • Q
1 Adventure	Animation	Children	Cornedy	Fantasy	NA.	104	NA	244	hin	à	Data	10329 obs. of 1 variable
2 Adventure	Children	Fantasy	NA.	-NR	MA.	-NA	204	A4.	\$		o genres2	10329 obs. of 10 variables
3 Comedy	Romance	144									Omovies	10329 obs. of 3 variables
4 Comedy	Chama	Romance	104	And I	100	Tan Inter	144	144	nen h-k		0 ratings	105339 obs. of 4 variables
6 Action	Come	Theiller	ALA.		All I		ALL .	144				
7 Comedy	Romance	ALL.	NA	NR.	NK.	has	NA	AL	hit		Files Plots	Packages Help Viewer
8 Adventure	Children	NA	NA.	NH.	14	NA	744	MA	701			Jimm Deport • 0
9 Action	NA	100	NA	704	NA:	NA (	NI)	NA.	101			
10 Action	Adventure	Thriller	144	Are.	14	104	NA	14	hin	į.		
wing 1 to 10 of 10	1,329 entries											
nsole Termina	ы.									-0		
ckage 'data genres2 <- view(genres colnames(ge view(genres	.table' as.data. (2) () () (2) (2) (2)	was bu frame() :- c(1:	ilt und tstrspl strings 10)	ler R va it(gen type ()	res[,1] , conve rs=FAL	3.6.2 , '[1] rt=TRU SE)	ġ.					
e Calk Bas Pati	Hanne Ball (	na hele (	terk - the Gista th	e)Aprocticas	1.8	• Add	ns •			147		T Project (
	Gaura	on Save	97	-20				- Der	1941 244		Environment	Import Dataset +
moviereco.r ×	a pre-pr	ocess	ng	1000			105034	null	1.37 30		Global Envi	iconnent • Q
movierecor x		lata.fr	ame (mov	resSgei	ires, s	trings	AsFacto	ors=FA	.st)	5	Data	
9 ## Dat 10 genres 11 librar	y(data.t	(able)		trsplin	(genre	s[,1],	'[ ]'	5			O genres	10329 obs. of 1 variable
9 ## Dat 10 genres 11 Tibrar 12 genres	s <- as.e y(data.t :2 <- as.	table) data.f	rame(ts		LYDE.	Conver	E TRUE	~			O genres2	10329 obs. of 10 variables
9 ## Dat 10 genres 11 Tibrar 12 genres 13 14	s <- as.e ry(data.t :2 <- as.	table) .data.f	rame(ts	ringsA	Factor	S=PALS	N				A mAA162	AUJED UDD, OF D VARIAULES
novierecar i 9 ## Dati 10 genres 11 Tibrar 12 genres 13 14 15 colnar	s <- as.t ry(data.t i2 <- as. wes(genre	table) .data.f 152) <-	rame(ts st c%1:10	ringsA: ))	Factor	S=PALS	×.				Oratings	105339 obs. of 4 variables
novierecur x 9 ## Dati 10 genres 11 librar 12 genres 13 14 15 colnar 16 17 genre, 18	s <- as.t ry(data.t :2 <- as. tes(genre ,1ist <-	table) .data.fi is2) <- c(°Act	rame(ts c(1:10 ion", "	Adventi	Factor	Animat	ion", '	"child	ren", "Eastaru"		o ratings	105339 obs. of 4 variables
movierecor 1 9 ## Dat 10 genres 11 librat 12 genres 13 14 15 colnas 16 17 genre, 18 19	s <- as.t ry(data.t s2 <- as. tes(genre ,list <-	table) data.fr ns2) <- c("Act "Com "File	rame(ts st c{1:10 ion", " edy", " n-Noir"	ringsA: ))  'Adventi 'Crime" ', "Hori	Factor "Docum or", "	Animat entary Musica	ion", ' ", "Drv 1", "Mj	"child ama", ystery	ren", "Fantasy", ","Romance"	5	O ratings	105339 obs. of 4 variables
moviercor t 9 #2 Dat 10 genres 11 librar 12 genres 13 14 15 colnas 16 16 17 genre, 19 20 21 22 genre, 23 genre, 23 genre,	s <- as.r ry(data.r s2 <- as. mes(genre ,list < matrix matrix[]	table) .data.fr es2) <- c("Act "File "Sci L,] <-	rame(ts st c(1:10 ion", " edy", " n-Noir" -Fi", " ix(0,10 genre_1	ringsA )) Advent: 'crime" ', "Horn Thrill 0330,18 list #se	Factor "Docum "or", " er", "W #enpt t firs	Animat entary Musica ar", " y matr t row	ion", ' ", "Dra 1", "My Western ix, 10 to genu	"child ama", ystery n") # 1 330=no re lis:	ren", "Fantasy", ","Romance we have 18 of movies	genres i 1, 18=no	O ratings	105339 obs. of 4 variables Packages Help Viewer Imm
moviercor t 9 # Dat 10 genres 11 librar 12 genres 13 14 15 colnas 19 20 21 22 genre, 23 genre, 23 genre, 23 genre, 23 genre, 23 genre, 24 colnas 25 colnas	s <- as.: ry(data.: s2 <- as. es(genre ,list <- matrix - matrix[] es(genre	cable) .data.fr es2) <- c("Act "Com "Fill "Sci L,] <- L,] <-	rame(ts st c(1:10 ion", edy", H-Noir" -Fi", ix(0,10 genre_1 x) <- g	ringsA: )) Adventr "crime" , "Horn "hrille 0330,18 ist #se penre_1	Factor "Docum "or", " #empt st firs ist #se	Animat entary Musica ar", " y matr t row t colu	ion", ' ", "Dra 1", "My Western ix, 10 to gen to gen to gen	"child ana", ' ystery n") # 1 330=no re lis es to 1	ren", "Fantasy", "Romance" we have 18 of movies t genre list	genres i 1, 18=no	O ratings	105339 obs. of 4 variables Packages Help Viewer Dame ⊉Doport+ 0   €
moviercos r 9 ## Dat 10 genres 11 libraz 12 genres 13 14 15 colnas 16 19 20 21 22 genre, 19 20 21 22 genre, 19 20 21 22 genres 19 20 21 20 20 21 20 21 20 21 20 20 21 20 21 20 20 21 20 20 20 21 20 20 20 20 20 20 20 20 20 20	s <- as.: ry(data.: s2 <- as. les(genre ,list <- matrix - matrix[] les(genre ):	table) data.f (************************************	rame(ts st c(1:10 edy", " m-Noir" -Fi", " ix(0,10 genre_1 x) <- g	ringsA Advent: Crime" , "Horr Thrille 0330,18 ist #se penre_1	Făctor "Docum "Docum "or", " " P", "w P ≢empt ist #se	Animat entary Musica ar", " y matr t row t colu	ion", ' ", "Ora 1", "My Western ix, 10 to geni no name	"Child ama", ystery n") # 1 330wno re lis es to (	ren", "Fantasy", ", "Romance" we have 18 of movies- t genre list	genres i 1, 18=no RScript I	Ples Plots	105339 obs. of 4 variables Puckage Help Viewer → Dipport + 0 d
moviencos r 9 ## Dati 10 genresi 11 libraz 12 genresi 14 15 colnari 16 17 genre, 19 20 21 22 genre, 23 genre, 23 genre, 24 colnari 529 (Determination of the second of t	s <- as.: ry(data.: s2 <- as mes(genre ]ist <- matrix - matrix[] res(genre ):	table) data.f (es2) <- c("Act "Com "Fil "Sci (- matr L,] <- L_matri	rame(ts st c(1:10 ion", " edy", " m-Noir" -Fi", " ix(0,10 genre_1 x) <- g	ringsA Adventi "crime" , "Horn Thrill 0330,18 list #se penre_1	Făctor "Docum "or", " r", "w I ≢enpt ist #se	Animat entary Musica ar", "r y matr t row t colu	ion", ' ", "Drw 1", "My Western ix, 10 to genu to genu to name	"Child ana", ' ystery n") # 1 330wno re list es to (	ren", "Fantasy", "."Romance" we have 18 of movies- t genre list	genres i 1, 18=no RScript I	Ples Plots	105339 obs. of 4 variables Puckage Help Viewer
moviercor / 9 ## pat 10 genres 11 librar 12 genres 13 14 15 colnas 16 17 genre, 18 19 20 21 genre, 23 genre, 23 genre, 23 genre, 23 genre, 23 genre, 25 genre, 25 genre, 26 genres, 27 genre, 28 genre, 29 genre, 29 genre, 20 genr	<pre>s &lt;= as.4 ry(data.ts2 &lt;= as. les(genru ]list &lt;= matrix &lt;= matrix[] les(genru b : les(genru b : les(genru b :</pre>	table) data.fr es2) < c("Act "Com "Fill "Sci c- matri: Semion 4/ ing 2 ti was bur	rame(ts st c(1:10 ion", " edy", " m-Noir" -Fi", " ix(0,10 genre_1 x) <- g " hreads ilt und	(see ?	Factor "Docum "Docum or", " #empt t firs ist #se petDTth ursion	Animat entary Musica ar", "I y matr t colu reads) 3.6.2	ion", ' ", "Dru 1", "My western ix, 10: to geni nn name	"Child ama", ystery n") # 1 330=no re lis es to p es to p	ren", "Rantasy", "Romance" have 18 of movies- t genre list es: r-datat	genres i 1, 18=no R Script = able.com	O ratings Files Plots	105339 obs. of 4 variables

) Kinam		
0 - 0	l 🐨 + 1 🕄 - 1 🕄 - 1 A Grits flort dettart at 1 🛛 - 1 Addins +	(S) Project (Norse) -
• movie 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31	<pre>ecu: ecu: Sourcensse Q *: [ + Run ** + Source * 2 penre_list &lt;- c("Action", "Adventure", "Animation", "Children", "Comedy, "Crime", Documentary, "Drama", "Fantasy, "Till+Noir", "Norror", "Musical", "Systey", Romande", "Sci-fel, "Thriller,", "mar", "watern") # we have us penres i genre_matrix &lt;- matrix(0,10330,16) #empty matrix, 10330no of evoluse1, 18=no genre_matrix[]. &lt;- genre_list #set first woo to genre list colnames(genre_matrix) &lt;- genre_list #set column names to genre list #iterate through matrix for (i in 1:nool genres2)) ( for (i in 1:nool genres2)) ( genre_matrix(i-1,genre_matrix(1,] == genres2(i,c)) genre_matrix(i-1,genrat_col) &lt;- 1 </pre>	Environment History Connections
32 33 2085 Console C/Umm, > View > coln > view > genr + + + al > view > view	<pre>} ' ' The function of the second second</pre>	

So, genre2 gets, what it does? It is actually breaking if you check genres if it is adventure, animation, children, comedy and fantasy, so adventure, animation, children, comedy and fantasy so, that is how it has broken and the rest of them are blind. So, we have created and if I just column names of genres 2 if I make it 1 to 10 now, it will be clearly seen that I am actually breaking the genres of each column into separate separate columns. So, these are my various genres.

And if I will find out the unique of these genres in the whatever they are in this 1000, 3000, 10,329 observations of 10 variables, you will basically get 668 sorry, 18 different kinds of genres. So, these are 18 different kind of genres that is coming up. So, that is a genre list that we have and based on this genre list, I will create a matching so, whether what is the code for this particular genre?

# (Refer Slide Time: 5:07)

<pre>province:/ proved to the set of the set</pre>	moviereco.r =		014				· Ad	ins •				Project. (No
<pre>     Prove the set of the se</pre>		genres	ú.								-5	Environment History Connections
<pre>1 2 2 3 4 5 5 6 7 8 9 30 2 domes Colors Ender Son 2 1023 cbs. cf 1 variable 0 domes 100 cc 1 cbs. cf 1 variable 0 domes 10 cc 1 cbs. cf 1 variable 0 domes 10 cc 1 cbs. cf 1 variable 0 domes 10 cc 1 cbs. cf 1 variable 0 domes 10 cc 1 cbs. cf 1 variable 0 domes 10 cc 1 cbs. cf 1 variable 0 domes 10 cc 1 cbs. cf 1 variable 0 domes 10 cc 1 cbs. cf 1 variable 0 domes 10 cc 1 cbs. cf 1 variable 0 domes 10 cc 1 cbs. cf 1 variable 0 domes 10 cc 1 cbs. cf 1 variable 0 domes 10 cc 1 cbs. cf 1 variable 0 domes 1 variable</pre>		Filter								9		💓 🔲 🐡 Import Dataset • 🧳 👘 List •
<pre>1 demon Chain Issue on the set of the s</pre>	* 1	2	, +	4	5	6	7	1		10		Global Environment • Q
2 Admine Date future (a set of a set of	1 6	Annation	Children	Contests	Cintra	14		NA.	NA.		3	Data
<pre>3 Comesy Remark Int At At</pre>	2 Adventure	Children	fantasy	ALA.	105	AA.	104	244	ALA.	AM.		genres 10329 obs. of 1 variable
<pre> • Comey Duma Remarks int of the los of</pre>	3 Corredy	Romance	NA	NA	NA.	NA:	NA.	AM.	NA.	AVA :		genres2 10329 obs. of 10 variables
<pre>S Comedy in in in a in in</pre>	4 Comedy	Drama	Romance	ALA.	NA.	N.S.	NA	NA	NA	NA		Omovies 10329 obs. of 3 variables
<pre>4 Actor Done Toring op: 1 At At</pre>	\$ Comedy	101	NA	NA.	NA.	244	.NA	AVA .	NA.	AM.		oratings 105339 obs. of 4 variables
2 Comedy Remark Date in the late in	6 Action	Crime	Toriller	-	NA.	NA.	NA.	AA.	.84	AVA :		
<pre># Ademove Children in a in an i</pre>	7 Cornedy	Romance	10	NA	144	NA:	NA	NA	244	hiń.		Files Plots Packages Help Viewer
<pre>9 Adds: Ut W W W W W W W W W W W W W W W W W W</pre>	8 Adventure	e Children	NA	AA.	No.	.71A	NA.	14	NA.	A14		an ab 🖉 James 🧏 Export • 🔍 🧹
<pre>B Adden Addentur Thele IN ON IN IN</pre>	9 Action	64	NA	NA:	NA	NA :	:104.7	AA.	NAC:	AVA.		
<pre>weq1501001020remtors mode Tennind {</pre>	10 Action	Adventure	Thriller	AL.	NA.	N6	104		244	IVA		
<pre>mode Turnind Turning Turning Tu</pre>	eing 1 to 10 of 1	0,329 entries										
<pre>WurgentUtilityUtil</pre>	sole Termina										-	
<pre>colnames(genres2) &lt;&lt; c(1:10) fiew(genres2)</pre>	Users/Dellty/Del	ditoo/Week?	Semion 47									
<pre>indec_genres:): i</pre>	olnames(or	enres2) -	- c(1:1	(0)								
Image: International Internationes Internatinternatinte International International International												
<pre>moderaur: for the source of the so</pre>	Contractor Street	Anna Anna A								_		
Image: Source on Save Q / · ( )       - # Kun ** ( ) <th>Cole the Part</th> <th></th> <th>eng Padar Ta</th> <th>Galati</th> <th>etheration</th> <th>- 11</th> <th>B • Ad</th> <th>ins •</th> <th></th> <th>_</th> <th>_</th> <th>scar Broject (W</th>	Cole the Part		eng Padar Ta	Galati	etheration	- 11	B • Ad	ins •		_	_	scar Broject (W
17       genre_list <- c("Action", "Adverture", "Animation", "Children", "Gaussi, "Fantasy", "Coman", "Science", "Comparison of movies-1, 18eno of genre_matrix(], <- genre_list #set first row to genre list	Maria Para	R R	0 A	ek Mey CGalasti	etiestia		B • Ad	ins •			-5	Environment History Connections
<pre>13</pre>	noviereco.r	Source	on Save	Contention	etiesto		⊞•Ad	ins •	Run	*** 0 50		Environment History Connections
<pre>20 "Sci-Fi", "Thriller", "war", "western") # we have 18 genres i 22 genre_matrix (- matrix(0,10330,18) #empty matrix, 10330mo of movies-1, 18ero 32 genre_matrix(1,1) &lt;- genre_list fist fist for to genre list 33 genre_matrix(1,1) &lt;- genre_list fist fist for to genre list 34 fitarate through matrix 35 fitarate through matrix 37 for (i in 1:nron(genres2)) { 38 genre_matrix(1,1,1) == genres2[i,c]) 39 genre_matrix(1,1,1) == genres2[i,c]) 31 genre_matrix(1,1,1) == genres2[i,c]) 32 genre_matrix(1,1,1) == genres2[i,c]) 33 genre_matrix(1,1,1) == genres2[i,c]) 34 fitarate through matrix 35 fitematrix(1,1,2) == genres2[i,c]) 36 genre_matrix(1,1,2) == genres2[i,c]) 37 genre_matrix(1,1,2) == genres2[i,c]) 38 fitematrix(1,1,2) == genres2[i,c]) 39 genre_matrix(1,1,2) == genres2[i,c]) 39 genre_matrix(1,1,2) == genres2[i,c]) 30 genre_matrix(1,1,2) == genres2[i,c]) 31 fitematrix(1,1,2) == genres2[i,c]) 32 fitematrix(1,1,2) == genres2[i,c]) 33 fitematrix(1,1,2) == genres2[i,c]) 34 fitematrix(1,1,2) == genres2[i,c]) 35 fitematrix(1,1,2) == genres2[i,c]) 36 fitematrix(1,1,2) == genres2[i,c]) 37 fitematrix(1,1,2) == genres2[i,c]) 38 fitematrix(1,1,2) == genres2[i,c]) 39 fitematrix(1,1,2) genre_matrix(1,1,2) == genres2[i,c]) 39 fitematrix(1,1,2) genre_matrix(1,1,2) == genres2[i,c]) 30 fitematrix(1,1,2) genre_matrix(1,1,2) == genres2[i,c]) 31 fitematrix(1,1,2) genre_matrix(1,1,2) ge</pre>	noviereco.r	Source	on Save   c("Acti	General Or General On	• [] Advent	ure",	19 • Ad	ins •	-* Run "Childr	en",		Project 04      Indicent History Connections      Project Dulant -
<pre>42 genre_matrix &lt;- matrix(0,10330,18) #empty matrix, 10330=mo of movies=1, 18ero 33 genre_matrix(1,1) &lt;- genre_list #set first row to genre list colname(genres_l) 54 for (ci in 1:nrow(genres2)) { 55 for (ci in 1:nrow(genres2)) { 56 genre_matrix(1,1) genmat_c0] &lt;- 1 31 } 33 } 33 } 33 } 34 /// "Film-koir", "horror", "musical", "Mystery", "Romance", "Sci-Fi", "Thriller", "mar", "musical", "Mystery", "Romance", "sci-Fi", "Thriller", "mar", "musical", "Mystery", "Romance", "sci-Fit", "thriller", "mar", "sci-Fit", "thriller", "mar", "sci-Fit, "thriller", "thriller", "mar", "sci-Fit, "thriller", "mar", "sci-Fit, "thriller", "thr</pre>	moviereco.r =	Source	on Save   c("Acti "Come "Film	General General on", ' dy", '	• i [] Advent Crime	ure", ',"Docu	- Ad	ins •	→ Run "Childr ama", " vsterv"	en", Fantasy", "Romance"	urce • 1 2	Emfragment History Connections Tendersonnent History Connections The Depart Dutate - Connections Connect
<pre>23 genre_matrix[1,] - openre_list set first row to genre list colnames(genre_matrix) ~ openre_list set first row to genre list 25 25 27 for (i in l:ncw(genres2)) { 29 genret_col = which(genre_matrix[1,] == genres2[i,c]) 30 genrem_matrix[i+1,genret_col] ~ 1 31 j 33 } 33 } 34 commod commod commod commod</pre>	17 genre, 18 19 20	Source	on Save   c("Acti "Come "Film "Sci-	on", ' dy", Fi",	• [] • Advent Crime , "Hou Thrill	ure", ,"Docu ror", ler", "	Animat "Animat "Musica war", '	ins • ", "Dr 1", "M Wester	⇒Bun "Childr ama", " ystery" n") # w	•• • • sou en", Fantasy", "Romance" e have 18	urce • I ≥ genres i	Breint N Emfronnent History Connections
<pre>c4 columnes(genre_matrix) &lt; genre_list set column mames to genre list 36 piterate through matrix 37 for (in 1:nrow(genres2)) { 39 genre_matrix[1,1] == genres2[i,c]) 30 genre_matrix[1+1,genmat_col] &lt;- 1 31 } 32 } 41 (bplow) : 42 (bplow) : 43 (bplow) : 44 (column) ************************************</pre>	17 genre, 18 19 20 21 22 genre,	Source list <-	on Save	Q Control on", i dy", i Fi", x(0,10	• [] Advent "Advent ", "Hot "Thrill	ure", ',"Docu ror", ler", "	B - Ad "Animat mentary "Musica war", '	ins • ", "Dr 1", "M Wester	→Run "Childr ama", " ystery" n") # w 330wno	** 1 3 so en", Fantasy", "Romance" e have 18 of movies+	=□ ute • 2 genres 1 1, 18=no	Environment History Connections
<pre>26 Piterate through matrix 27 for (in incom/genres2)) { 28 for (c in 1:ncol/genres2)) { 30 genre_matrix(i-1,genre_tot] &lt; 1 31 ] 32 } 33 / 1 (PpLew) ? web Terminal Unart/Delt/Delto/Delto/Polatos/Wes2/sense / "Film-Nof", "Norror", "Master", "Researce", "Sci-Fi", "thriller", "war", "Wester") ? we have 18 genres in tot riew(genres2) rene_matrix(i,1) &lt;- genre_list #set first row to genre list</pre>	novierecor i movierecor i i 17 genre, 18 19 20 21 22 genre, 23 genre,		on Save   c("Acti "Come "Film "Sci- c- matri L, ] <- 9	on", ' dy", ' - Noir', ' Fi', ' x(0,10	- [] Advent Crime , "Hou Thrill 0330,10 list #	ure", ,"Docu ror", ler", " ) #emp let fir	Adi "Animat mentary "Music: war", ' ty mati st row	ins • '', 'Dr I'', 'M Wester 'ix, 10 to gen	→Run "childr ama", " ystery" n") # w 330wno re ]ist	*** so en", Fantasy", Romance" e have 18 of movies+	urte - ≥ genres i 1, 18⊨no	Brejet 0/      Emvironment History Connections
<pre>22 Tor (1 in 1.nron(genres2)) { 32 genrat_col = which(genre_matrix[1,] == genres2[i,c]) 33 genre_matrix[i-1,genrat_col] &lt;- 1 33 3 33 / 33 / (Option) : 34 1 (Option) : 35 / 35 / 35 / 35 / 35 / 35 / 35 / 35 /</pre>	novierecur i movierecur i 17 genre, 18 19 20 21 22 genre, 23 genre, 24 colnar 25		on Save   c("Acti "Come "Film "Sci- c- matri i,] <- g _matrix	Q 2 on", dy", -Noir Fi", x(0,10 enre_) () <- (	•i [] 'Advent 'crime' ', "Hou 'Thrill 0330,12 list #penre_1	ure", ',"Docu ror", er", " () #emp et fir ist #s	Animat "Animat "Musica War", ' ty mati st row et colu	ins • ion", "Dr l", "M Wester ix, 10 to gen mn nam	→Run "Childr ama", " ystery" n") # w 330wno re list es to g	** 3*50 en", Fantasy", , "Romance" e have 18 of movies+ enre list	genres i 1, 18=no	
<pre>29 genrat_col = which(genre_matrix(1,1) == genres2(i,c)) 31 genre_matrix(1+1,genmat_col) &lt;- 1 33 a) 33 a) 33 a) 33 a) 34 a) 35 a) 35 a) 36 a) 37 a) 38 a) 38 a) 39 a) 30 a) 30</pre>	novierecur i novierecur i 17 genre, 18 19 20 21 22 genre, 23 genre, 23 genre, 24 colnar 25		on Save   c("Acti "Come "Film "Sci- - matrix l,] <= g L_matrix	Q ( ) Q ( ) on", ' idy", i i-Noir' Fi", ' x(0,1( ienre_) ) <- ( ix	<pre>- [] Advent Crime ', "Hor Thrill 0330,18 list # genre_'</pre>	ure", "Docu ror", er", ) #emp et fir ist #s	Animat mentary "Musica War", ' ty math st row et colu	ins • ", "Dr. 1", "M Wester Tix, 10 to gen mn nam	→Run "Childr ama", " ystery" n") # w 330wno re list es to g	** * 50 en", Fantasy", "Romance" e have 18 of movies+ enre list	urce - ≥ genres i 1, 18=no	Emframment History Connections Temporament History Connections Cabout Invisionment - Q Data O genres 10329 obs. of 1 variables O genres 10329 obs. of 1 variables O genres 10329 obs. of 3 variables O movies 10329 obs. of 3 variables O ratings O ratings
<pre>30 genre_matrix[1:1;genre_tor] &lt; 1 32 } 33 genre_matrix[1:1;genret_tor] &lt; 1 32 } 33 genre_matrix[1:1;genret_tor] &lt; 1 33 genret_tor] &lt; 1 33 ge</pre>	Con mo Ami Con movierecor i intervention 17 genre, 18 19 20 21 22 genre, 23 genre, 24 collar 25 fitter 27 for (128 - for (28		on Save   c("Acti "Come "Film "Sci- c- matri i,] <= g _matrix ugh matr row(genr ncol(ge	Q on", -Noir Fi", x(0,1( jenre_) ) <- ( fix es2)) nres2)	<pre>. [] . dvent . dvent . dvent . dvent Hot</pre>	ure", ',"Docu ror", er", " ) #emp et fir ist #s	Adianimati mentary "Musica war", ' ty mati st row et colu	ins • ion", "Dr 1", "M Wester ix, 10 to gen mn nam	⇒Run "Childr ama", " ystery" n") # w 330=no re list es to g	** _* So en", - antasy", "Romance" e have 18 of movies- enre list	penres i 1, 18≈no	Environment History Connections The post Data O genre_matr. Large matrix (185940 elemen O genres 10329 obs. of 1 variables O genres 10329 obs. of 1 variables O movies 10329 obs. of 3 variables O movies 10329 obs. of 3 variables O movies 10329 obs. of 4 variables
<pre>32 }</pre>	Con text for movierecor i information inf		on Save   c("Acti "Come "Film "Sci- c- matri t,] <- g s_matrix row(genr ncol(ge l = whic definition	Q on", i-Noir' Fi", x(0,1( tenre_ c) <- ( ix tes2)) nres2) h(gen	<pre>. [] 'Advent 'Crime' ', 'Hoo 'Theill 0330,18 list # j offe_matri { ) { re_matri </pre>	ure", ',"Docu ror", er", " !) #emp iet fir ist #s	<pre>"Animat mentary "Musicz war", ' ty mat: st row et colu</pre>	ins - ", "Dr ", "Dr "Wester 'ix, 10 to gen mn nam	<pre>*Rus "childr ama", " ystery" n") # w 330wno re list es to g ,c])</pre>	** * 50 en", ** antasy", "Romance" e have 18 of movies- enre list	penres i 1, 18⊨no	
<pre>33 f (%plevm) : Rscyri: abb Terminal ////DHI/Dubtop/Wex0//sembon 4/ // "Sci-Fi", "horror", "Mar", "Western") # we have 18 genres in tot riem(genres) iem(genres) iem(genres) iem(genres) es mer_matrix &lt; matrix(0,10330,18) #empty matrix, 10330emo of movies+1, 18emo of g es mer_matrix(1,1) &lt;- genre_list #set first row to genre list</pre>	Con mon Amil Con Table Con Control Contr		on Save   c("Acti "Come "Film "Sci- c- matri i,] <- g _matrix row(genr ncol(ge = whic -ix[i+1,	Q /	• [] Advent ', "Hor ', "Hor 'Thrill 0330,18 list # genre_1 { }) { re_matr t_col]	<pre>ure", '' , 'Doc', ror", '' er", '' ) #emp et fir ist #s 'ix[1,] &lt;- 1</pre>	<pre>"Animat mentary "Musica War", ' ty mati st row et colu</pre>	ins • ". "Dr 1", "M Wester Tix, 10 to gen mn nam	<pre>"#Rus "Childr ama", ystery" n") # w 330wno re list es to g ,c])</pre>	** s so en", Fantasy", "Romance" e have 18 of movies- enre list	ute - ≷ genres i 1, 18=no	Invicement       History Connections         Invicement       Ital         Ital       Ital
<pre>usb Terminal Trillm-Noir", "Norror", "Musical", "Nystery", "Romance", "sci=#;", "thriller", "war", "westery)"# Romance", riew(genres) riew(genres2) remer_matrix &lt;- matrix(0,10330,18) #empty matrix, 10330=mo of movies+1, 18=mo of g ess ener_matrix[1,] &lt;- genre_list #set first row to genre list</pre>	One was not         One           One         Image: Comparison of the second seco		on Save   c ("Acti "Come "Film "Sci- - - - - - - - - - - - - -	Q Cont on", ' +Noin', ' Fi", ' x(0,10 (eenre_' ;) <- ( ' ix * es2)) ,mres2; hgemat	<pre>ethester identified adventif adventif crime crime</pre>	<pre>cure", ", "Docu ", "Docu ror", " )) #emp set fir ist #s fix[1,]</pre>	"Animat mentary "Musica", ' ty matri st row et colu	dns - "."or 1", "wester ix, 10 to gen mn nam	<pre>"   Run "childr ama", " ystery" n") # w 330wno re list es to g ,c])</pre>	** so en", Fantasy", "Romance" e have 18 of movies- enre list	ute I≷ genres i 1, 18+no	Project Of      Project O
<pre>Unrn(DeBL(Neddog/Weed/Aembin 4/ " "Film-Noir", "Horror", "Musical", "Hystery", "Romance", "sci=Fi", "thriller", "war", "western") # we have 18 genres in tot riew(genres) riew(genres) regre_matrix &lt;- matrix(0,10330,18) #empty matrix, 10330=no of movies=1, 18=no of g reg reg reg reg reg reg reg reg reg re</pre>	One is non-         One is non-           Immovierecor i         Immovierecor i           If         genre,		on Save   CC ("Acti "Con action "Attack of the second "Sci- 	Q Const on", dy", Fi", x(0,1( ienre_' ix ves2)) inrres2; h(genut	<pre>. [] Advent Crime" ", "Hori Thrill 0330,12 ist # penre_1 { }) { re_matri t_col]</pre>	ure", ","Docu ror", "br et fir ist #s fix[1,] <= 1	B - Ad "Animat mentary "Musica war", ' ty mat st row et colu	dns • .", "or .", "or .", "wester .ix, 10 to gen min nam	→ Run "Childr ama", " ama", " amage of the star amage of the star	•• 3• So en", "Romance" e have 18 of movies+ enre list	genres i 1, 18=no	
"Film-Noir", "Norror", "Musical", "Mystery", "Romance", "Sci-Fi", "Thriller", "War", "Western") # we have 18 genres in tot view(genres) view(genres) reme_matrix <- matrix(0,10330,18) #empty matrix, 10330=mo of movies=1, 18=mo of g res perfer_matrix[1,] <- genre_list #set first row to genre list	One was non-           One was non-           One was non-           If           If </td <td></td> <td>on Save c ("Acti "Come "Film "Sci- c- matri t,] &lt;- g matrix ugh matr ncol(ge = whic -ix[i+1,</td> <td>Q Canada on", ' dy", ' H-Noir', ' Fi", ' x(0,1( leane_' Fi", ' x(0,1( leane_' ix es2)) inres2; h(genut</td> <td>+ [] "Advent "crime" ", "Hot Thrill 0330,14 List #1 penre_1 [ ]) { </td> <td>ure", ,"Docu ror",  ) #emp et fir ist #s 'ix[1,] &lt;&lt; 1</td> <td>B - Ad "Animat mentary "Musica war", ' ty mati st row et colu == ger</td> <td>inn", "","Dr "1", "M Wester "ix, 10 to gen mn nam</td> <td>→ Run "Childr ama", " ystery" ") f w 330wno re list sto g es to g .c])</td> <td>** je sov en", fantasy", "Romance" e have 18 of movies+ enre list</td> <td>genres i 1, 18×no R Script :</td> <td>Propert Out      Indexy Connections      Propert Dataset      Connections      Connect</td>		on Save c ("Acti "Come "Film "Sci- c- matri t,] <- g matrix ugh matr ncol(ge = whic -ix[i+1,	Q Canada on", ' dy", ' H-Noir', ' Fi", ' x(0,1( leane_' Fi", ' x(0,1( leane_' ix es2)) inres2; h(genut	+ [] "Advent "crime" ", "Hot Thrill 0330,14 List #1 penre_1 [ ]) { 	ure", ,"Docu ror",  ) #emp et fir ist #s 'ix[1,] << 1	B - Ad "Animat mentary "Musica war", ' ty mati st row et colu == ger	inn", "","Dr "1", "M Wester "ix, 10 to gen mn nam	→ Run "Childr ama", " ystery" ") f w 330wno re list sto g es to g .c])	** je sov en", fantasy", "Romance" e have 18 of movies+ enre list	genres i 1, 18×no R Script :	Propert Out      Indexy Connections      Propert Dataset      Connections      Connect
"Sci-Fi", "Thriller", "war", "western") # we have 18 genres in tot view(genres2) view(genres2) genre_matrix <- matrix(0,10330,18) #empty matrix, 10330=mo of movies=1, 18=mo of g res genre_matrix(1,) <- genre_list #set first row to genre list	Control         Control <t< td=""><td></td><td><pre>het het it on Save c ("Actii "Come "Film "Sci- c- matrii l,] &lt; g matrix high matrix high matrix is (i+1, - white is (i+1, - white)</pre></td><td>Q ( Control control dy, , , , , , , , , , , , , , , , , , ,</td><td>+ [] "Advent "crime" "Hot Thril 0330,12 list # penre_1 { } ) { ) } * crime t_cc]</td><td>ure", ',"Docu ror", 'er", '' )) #emp et fir ist #s fix[1,] &lt;- 1</td><td>"Animat mentary" "Musica Wasar", a' ti st row ti colu == gef</td><td>inn", "","Dr 11", "M Wester Tix, 10 to gen min nam</td><td>→ Run "Childr ama", " ystery" ") f w 330wno ré list sto g (sto g (,c))</td><td>** 3* So en", 5* So Fantasy", "Romance" e have 18 of movies- enre list</td><td>genres i 1, 18eno R Songt s</td><td>Imiroament       History Connections         Imiroament       Interface         Imiroament       Interface</td></t<>		<pre>het het it on Save c ("Actii "Come "Film "Sci- c- matrii l,] &lt; g matrix high matrix high matrix is (i+1, - white is (i+1, - white)</pre>	Q ( Control control dy, , , , , , , , , , , , , , , , , , ,	+ [] "Advent "crime" "Hot Thril 0330,12 list # penre_1 { } ) { ) } * crime t_cc]	ure", ',"Docu ror", 'er", '' )) #emp et fir ist #s fix[1,] <- 1	"Animat mentary" "Musica Wasar", a' ti st row ti colu == gef	inn", "","Dr 11", "M Wester Tix, 10 to gen min nam	→ Run "Childr ama", " ystery" ") f w 330wno ré list sto g (sto g (,c))	** 3* So en", 5* So Fantasy", "Romance" e have 18 of movies- enre list	genres i 1, 18eno R Songt s	Imiroament       History Connections         Imiroament       Interface
view(genres2) view(genres2) genre_matrix <- matrix(0,10330,18) #empty matrix, 10330=no of movies+1, 18=no of g genre_matrix(1,) <- genre_list #set first row to genre list	Construction         Construction           10         10         10           17         genre.         11           18         19         20           20         21         genre.           23         genre.         24           24         colnait         25           26         [Fitter:         27.7 for (:           28.*         for (:         28.*           32         3         :           33         :         :           32         3         :           33         :         :           33         :         :           33         :         :	<pre></pre>	<pre>meg News Te Merce Te Merc</pre>	Q Q Q Q Q Q Q Q Q Q Q Q Q Q	<pre>intertie 'Advent 'Crime' Crime' 'Thril 0330,14 list #1 penre_' { }) { { col}</pre>	ure", "Docu "ror", "Bocu er", " )) #emp ist #s 'ix[1,] ''Musi	"Animat mentary" "Musica ty mati ty mati ty colu == gef	ins • (ion", ") ")", "W wester rix, 10 to gen m nam wres2[i Myster	"*Run "Childm ystery" w 330wno ré list es to g ,c))	** * So en", Fantasy", "Romance" e have 18 of movies+ enre list	genres i 1, 18=no	Propert Out      Iniferry Connection      Propert Dutat      Connection      Propert Dutat      Connection      Connectio
<pre>view(genres2) genre_matrix &lt;- matrix(0,10330,18) #empty matrix, 10330=no of movies+1, 18=no of g res genre_matrix[1,] &lt;- genre_list #set first row to genre list</pre>	Construction         Construction           100         100           117         genre,           18         19           20         22           21         22           22         genre,           23         genre,           24         colnait           25         fritern           26         fritern           27         for           28         for           29         gf           31         3           32         3           51         (Repleve           sold         Terminal           Utert/Delli/Del         Utert/Delli/Del	<pre></pre>	<pre>meg News Te Merce Te Merc</pre>	Galadian Galadian Galadian Galadian Galadian (Galadian (Galadian) Galadian Galadian Galadian Galadian Galadian Galadian (Galadian (Galadian) (Galadian (Galadian) (Ga	<pre>interter "Advent "crime" ", "Hor Thril 0330,14 list #j genre_" { }) { { </pre>	"Musi "Was", "Was", "States "States "States "States "States "States "States "States "States "States "States "States "States "States Sta	<pre>"Animat "Animat "Musica", ' "Musica", ' "war", ' ty mat/ st row tt colu == ger </pre>	ins • (ion", ") " ", "or in, 10 to gen mn nam wester rix, 10 to gen mn nam wester (i Myster rn") #	"*Run "Childm ystery" w 330wno ré list es to g ,c))	** .so of movies- enre list	genres i 1, 18=no R Songt : s in tot	Project Ou Inviconnent History Connections Connect Dataset - Connections Connect - Connections Connect - Connections
penre_matrix <- matrix(0,10330,18) #empty matrix, 10330+no of movies+1, 18+no of g res genre_matrix[1,] <- genre_list #set first row to genre list	Image         Image           17         genre,           18         genre,           20         genre,           21         genre,           23         genre,           24         colnar           5         for           25         genre,           23         genre,           23         genre,           29         gi           31         j           32         j           33         top, for           sobe         Termina           then/Delli/des         top, for	mes has h Source matrix () matrix () matrix () mes (genru tate throng in 1:nn: n 1:	<pre>May Table To Table Tab</pre>	Galantian Galant	<pre></pre>	<pre>''ure", ''Docu ror", 'er", '' )) #emp et fir ist #s ''ix(1,] ''war",</pre>	<pre>"Animat "Animat "Musica", ' "Musica", ' "War", ' ty mat/ st row tt colu == ger ' "west</pre>	ins - (ion", "pr 1", "h wester ix, 10 to gen mo nam res2[i Myster rn") #	"de lan "chilar ystery" ") # w 330wno re list so g es to g (,c)) y","Rom	** ] * So en", Fantasy", "Romance" e have 18 of movies+ enre list enre list ance", e 18 genre	genres i 1, 18=no R Script : S in tot	Invicement History Connection  Control of the second seco
genre_matrix[1,] <- genre_list #set first row to genre list	In a serie of the series of th		<pre>Mew. New To New To</pre>	<pre>v debt debt oon</pre>	<pre>intertion 'Advent Crime' ', "Horo Thrill 0330,14 list # genre_' { } } { ( )) { re_matrice t_col] } prror", ller",</pre>	<pre>''ure", ''Docu ''or", ''er", '' 'ist fir 'ist #s ''ist(1,] ''Musi ''war",</pre>	"Animat mentary "Musica st row et coli == ger cal", ' "weste	ins - '', "or '', "or in, '', "wester ix, 10 to gen mn nam mres2[i Myster, rn") #	"*Run "childr stery" ") f w 330 mo 330 mo story stery" "childr story (c)) y", "Rom we hav	** * so en", "Romance" e have 18 of movies- enre list ance", e 18 genre	genres i 1, 18=no R Script : s in tot	Propert On  Invironment History Connection  Control Data  Propert Control
	riew (genret riew (genret) riew (genret	matrix : matrix	<pre>Medic To Mark To</pre>	<pre>vector ing vector ing vector</pre>	<pre>:: [] 'Advent 'crime' 'crime' 'crime' 'crime' 'crime' 'advent 'crime' '()) { 'crime' 'crime'' 'crime''' 'crime''' 'crime''''''''''''''''''''''''''''''''''</pre>	<pre>ure", ', 'Docu ror", er", er", er" ist #s 'ix(1,) '*war", upty ma</pre>	<pre>"Animat mentary "Musica st row t colu == ger cal", ' "weste trix, 1</pre>	ins - '', ''or '', ''or '', ''or '', ''or '', ''or '', ''or ''', ''or '''', ''or ''''''''''''''''''''''''''''''''''	"*Run "childr ma", " ystery" ") f w 330+mo iss to g ,c]) y", "Rom we hav o of mo	10 Soo en", "artasy", "artasy", "artasy", "artasy", "artasy "artasy have 18 of movies+ enre list enre list ance", e 18 genre uvies+1, 18	rer - 2 genres i 1, 18-no RSorpt : s in tot	Propert Out      Indiversent History Connections      Propert Dutatet      Propert Dutatet      Propert Dutatet      Propert      Data      Openres_attr_atage matrix (185940 element.      Openres_10329 obs. of 1 variables     oratings     10339 obs. of 1 variables     ratings     Totage Help Viewer      Propert O      Propert      Propert O      Propert O      Propert O

ବ 😆 😰 🕈 🛪 🕥 🔅 😢

											1. 0 . 0
0	noviereco.r	genre_	matrix ×						-0	Environment History Connections	-0
	1.131	V Filter						0,		💣 🔛 🌁 Import Dataset + 🧹 🗮 Li	a • 1 🤅
*	Action	Adventure	Animation	Children	Comedy	Crime	Documentary	Drama	Fantasy	Global Environment • Q.	
1	Action	Adventure	Animation	Children	Cornedy	Crime	Documentary	Drama	Fantasy	Data Depre matr. Large matrix (185940 element	
2	0	0	0	0	0	0	0	0	0	genres 10329 obs. of 1 variable	
3	0	0	0	0	0	0	0	0	0	genres2 10329 obs. of 10 variables	
4	0	0	0	0	0	0	0	0	0	O movies 10329 obs. of 3 variables	
5	0	0	0	0	0	0	0	0	0	oratings 105339 obs. of 4 variables	
6	0	0	0	0	0	0	0	0	0	Files Plots Packages Help Viewer	- C
7	0	0	0	0	0	0	0	0	0		
8	0	0	0	0	0	0	0	0	0		
9	.0	0	0	0	0	0	0	0	0 -		
Con	ing 1 to 10 c	of 10,330 entries							-5		
C/1 1 V 9 9 0 0	iew(genr iew(genr enre_mat enre_mat olnames( iew(genr	Desktop/Week7 "Se res) trix <- ma trix[1,] < (genre_matrix)	/Semico 4/ ci-Fi", "Th trix(0,103: - genre_lit rix) <- gen	nriller", 30,18) #em st #set fi nre_list #	"war", "w mpty matri irst row t fset colum	vestern" ix, 1033 to genre in names	) # we have : O=no of movid list to genre list	18 genres es+1, 18: st	in tot •		

So, if this is the first particular movies genre, so then I will get a matrix in such a way such that that matrix will have number of movies is 10,329. So 10,329 rows and 18 columns, 18 columns each has one genre. And the value will be zero if that particular movie falls in that genre, and that data will come from this particular matrix. And if it is not in that genre then 0, if it is in that genre then 1, so, that is what I am going to create now.

So what I am doing is that I am creating see, there is a matrix of 0 with 10,330, that means first row will be the column name and 18 columns. So that is what I am creating, the first row genre matrix 1, the first row is basically the genre list, the column names and then column names of genre matrix is basically the genre list so, that is what I am doing here. So, if you check the genre matrix right now, it is basically all zeros, okay. And then what I am doing is for i = 1 to n row of genres2 that means 10,330 and then c = 1 to n column of genres2, that means 1 to 18, what will I do? If it is matching then 1 otherwise 0.

## (Refer Slide Time: 6:35)

() them	-								
0.0	· · ·		Go to tilerture		· Addins				I Project (None)
0 movies	ecor =							-0	Environment History Connections
41	ALB Da	ource on Save	97.1			- PRU	n 🐂 🗦 So	urce • 2	💣 🔒 🐡 Import Dulaset + 🧹 👘 Eliat + 👩
21 22	genre_matr	ix matri	x(0,10330	.18) #emp	ty matrix	k, 10330≃n	o of movies+	1, 18=no	🚯 Global Environment • 🛛 🔍
23 24	genre_matr colnames(ge	ix[1,] <- g enre_matrix	penre_list c) <- genre	#set fir _list #s	st row to et column	o genre li n names to	genre list	5	Ogenre_matr Large matrix (185940 elemen.
25 26	Fiterate th	hrough matr	ix					_	Ogenres 10329 obs. of 1 variable
27 + 28 +	for (i in )	1:nrow(genr	res2)) {						Ogenres2 10329 obs. of 10 variables
29	genmat,	_col = whic	h(genre_m	trix[1,]	= genre	s2[i,c])			O ratings 105339 obs. of 4 variables
31	genre	metrix(1+1	germat_co						Files Plots Packages Help Viewer
32	2								er sol 2 mm Deport • 0 1
34 35	genre_matr	ix2 <- as.c	ine lata.frame	genre_ma	trix[-1,]	, strings	AsFactors=FA	LSE) #rei	
36 - 37	for (c in )	1:ncol(genr	e_matrix2	) {					
261	(Top Level) 2			R				R Script 1	
Console	Terminal =							-0	
> colna	Dell'I/Deditop/W	<pre>matrix) &lt;-</pre>	genre_list	#set co	lunn nane	is to genr	e list	4	
> View(	genre_matri	ix) h marcix							
> for (	i in l:nro	w(genres2))							1 and
+ 101	enmat_col	= which(gen	re_matrix	[1,] == g	enres2[i,	([s			1 Carles
+ }	penre_matri)	x[1+1,genma	t_col] <-	1					
+ }								- 1	
-	-			-	_	_		3	
🧿 関	8 9	X 🌒 🖟							1.6
The Late Color	The Part Second Pa	lula Delog Polia 1	oh my						Kara Maria
0.0	<b>e</b> • 11	alle la	Gato fileture	tion []	· Addins	•			Project: (None) -
• movier	eco.r ×							0	Environment History Connections
20	C Repeat	surce on Save	h oenre m	reix[1.]	geore	-+Ru	n 👐 🔆 50	urce + R	💣 🖶 📑 Import Dutaset • 🧉 🗮 List • 🐨
30	genre_r	matrix[i+1,	germat_co	] <- 1	yeine	(ar[iic])			Data
31	, '							8	🛛 genre_matr_ Large matrix (185940 elemen_ 🗌
33 34	#convert in	nto datafra	une						Ogenre_retr. 10329 obs. of 18 variables
35 36 +	genre_matr	ix2 <- as.d 1:ncol(genr	ata.frame e_matrix2	(genre_ma	trix[-1,]	, strings	AsFactors	iLSE) #rei	Ogenres 10329 obs. of 1 variable
37	genre_mat	trix2[,c] <	- as.inte	er(genre	_matrix2(	[,c])			O movies 10329 obs. of 3 variables 📃
39			and for						Files Plots Packages Help Viewer
40	years <- at	s.data.fram	e(novies)	itle, st	ringsAsFa	ctors=FAL	SE)		9 9 / / Imm Baport • 0 / 2
42 43 -	library(dat substrRight	ta.table) t <- functi	ion(x, n){						
44 45	substr(x	, nchar(x)-	n+1, nchai	·(x))	_			8	
36:1	(Top Level) 1							R Script 1	
Console	Terminal ×							-0	
> for (	i in 1:nro	w(genres2))	1					3	
+ for	(c in 1:no	col(genres2 which(gen	<pre>)) {     re matrix</pre>	1.1 == 0	enres2[i	(1)			
+ 9	enre_matrix	x[i+1,genma	t_col] <-	1		-17			and the
+ }		102							
> #conv > genre	_matrix2 <-	- as.data.f	rame(genre	_matrix[	-1.], str	ingsAsFac	tors=FALSE)	#remove	
first r	ow, which a	was the gen	ire list					1	
-	-		Con Law	-	_		_	3	SIL
9. 🜒	8 9	X 🎱 🐺	0						A Martin
The Late Color	You Pate James N	un Deug Polie 1	ok ne						
0.0	C. 18	ale a	Ge to the hum		· Addins	•			Project (None)
• movier	ecor = ge	enre_matrix2						-0	Environment History Connections
0.01	I) V. Filter	~					Q,		💣 🔒 📑 Import Dutaset • 🧉 📄 List • 😨
Actio	on Adventur	re Animatio	in Children	Comed	ly Crime	Documen	tary Drama	Fantasy	Data
1 0	1	1	1	1	ŋ	0	0	1	Ogenre_matr Large matrix (185940 elemen
2 0	1	0	1	0	0	0	0	1	Ogenre_matr. 10329 obs. of 18 variables
3 0	0	0	0	1	0	0	0	0	genres 10329 obs. of 1 variable
4 0	0	0	0	1	0	0	1	0	O movies 10329 obs. of 3 variables
6 1	0	0	0	0	1	0	0	0	Files Plots Packages Help Viewer
7 0	0	0	0	1	0	0	0	0	9 9 7 June 3 Deport - 0 1 4
8 0	1	0	1	0	0	0	0	0	
. 9.1	0	0	0	.0	0	0	0	0	
Showing 1 t	to 10 of 10,329 en	tries							
Console	Terminal =							-0	
Ci/Users/C	Delli/Desktop/W	nek?/Senion 4/	3 (0					d.	
+ 9	enmat_col	= which(gen	re_matrix	[1,] == g	enres2[i,	([)			
+ } 9	enre_matri)	x[1+1,genma	t_col] <-	1					
+ } > #conv	ert into da	atafrane							1-11-11-1
> genre	_matrix2 <-	- as.data.f	rame(genre	_matrix[	-1,], str	ingsAsFac	tors=FALSE)	#remove	
> View(	genre_matr	ix2)	e nat					1	
>									
9 0	0 0	x 🙆 😡	0						1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

So, if I just run this thing quickly and then drop the first row, see first row I am dropping here, if I just drop the first row and what do I get in genre matrix2? I am getting genre matrix 2 like this. So, the first movie is action, animation, children, comedy, the second movie is adventure, children and fantasy and so on. Now, this part, you can do it in different way. If you have a better algo for that, you can do it in your way. This part is similar so, I have to create this, this is something that I want to know that which movie and with genre I am creating a mapping for that.

(Refer Slide Time: 7:16)

• 📭 😴 • 🔄 🔄 📥 🖉 🔺 Carlo Hofwatton 👘 🗌 🖼 • Addins •	Prior don
	And and a state of the state of
NUMBER STREET	Consections million and a consections million and a consection and a conse
<pre>29 genmat_col = which(genre_matrix[1,] == genres2[i,c])</pre>	Clobal Engineerant -
<pre>30 genre_matrix(i+1,genmat_col) &lt;= 1</pre>	
32 }	genre_matr. 10329 obs. of 18 variables
33	gen is 10329 obs. of 1 variable
34 #convert into dataframe 25 convert into dataframe 26 converting of a state frame(convertexist) 1 stringsAcEactors.EALER) #co.	Openies 10329 obs. of 3 variables
36+ for (c in 1:ncol(genre_matrix2)) {	Oratings 105339 obs of 4 variables
<pre>37 genre_matrix2[,c] &lt;- as.integer(genre_matrix2[,c])</pre>	values
38 ] #convert from characters to integers	
40 #Create a matrix to search for a movie by genre:	Files Plots Packages Help Viewer
41 years <- as.data.frame(moviesStitle, stringsAsFactors=FALSE) 42 Jibnerg/data_table)	🔿 🛸 🖉 Jaam 🖓 Deport • 🔍 🖌
42 - substrRight <- function(x, n)	
<pre>44 substr(x, nchar(x)-n+1, nchar(x))</pre>	
45 /	
E39 (tip Level) 1 KScript 1	
nsole Terminal	
Guers/Dell3/Desitop/Week7/Semion 4/ 🖗 🚽	
1	
fconvert into dataframe	
<pre>genre_matrix2 &lt;- as.data.frame(genre_matrix[-1,], stringsAsFactors=FALSE) #remove</pre>	
rst row, which was the genre list	
for (c in 1:ncol(genre_matrix2)) {	
<pre>genre_matrix2[,c] &lt;- as.integer(genre_matrix2[,c])</pre>	
} #convert from characters to integers	
a da fan No Jane Ald Nag Mak San Nag	
n dar fan An Joan Kal Hag Mai, ten Ne • Q	Report Dear
Con the Are been boo being here, the new	Project (Harrison and Harrison
Con tes file bees had beeg hele bee new • Q Con tes file bees had beeg hele bee new • Q Con tes file been test test test test test test test t	Environment History Connections mm
con ten fin ten fin fin fin ten     contributive	Prijett Oon     Mitory Cenerctions     mi     District Outset     Lut     dobal Environment     Outset     Outset
the first fir	Project Data      Project
Out the fire here here here here here here here h	Project Dan      Environment History Connections      Project Dan      Project Dan      Project Data      Openre_matr. Large matrix (185940 elemen.
1 de les les les les les les les les les le	Environment History Connections m
r de ten fen ben ben ben hen hen hen noviereur i 1	Prepert Own      Environment History Connections      More Training Connections      Mor
a tes tes her bes her her her her tes a tes tes her bes her her her her tes a tes tes her bes her her her tes a tes tes her bes her her her her tes a tes tes her bes her her her her her her tes tes tes her her her her tes tes tes her	Environment History Connections Control Devices Control
<pre></pre>	Emironment History Connections Emironment History Connections Control Connections Control Control Control Control Control Control Control Control
<pre>e de me he he he he he he he he he e de me he he he he he he he he e e e e e e e e e e e e e e e e e e e</pre>	Environment History Connections m Control Connections m Control Connections m Control Connections m Control Connections m Control Connections m Control Connections m Connections m
<pre> invertion into data frame invertion into data frame if into data data if into data if</pre>	Environment Hintory Cennections Environment Hintory Cennections Constant
<pre> tex ms he bee be bee he bee he bee tex tex he bee he bee he bee he bee tex tex he bee he b</pre>	Project Dan      Environment History Connections      Project Dan
<pre>not no no her her her her her her noverears 33 formert into dataframe 35 genre_matrix2(-c) &lt; as.integer(genre_matrix[-1,], stringsAsFactors=FALSE) #ren 36 for (c in lincol(genre_matrix2)) { 37 genre_matrix2(-c) &lt;-as.integer(genre_matrix2(-c)) 38 f formert from characters to integers 39 dCreate a matrix to search for a movie by genre: 40 forest a matrix to search for a movie by genre: 41 biorry(data.table) 42 substr(k, nchar(x)=n1, nchar(x)) 43 substr(k) &lt;-finction(x, n)( 45 substr(k) = data.frame(substr(substrRight(substrRight(years) moviesStitle', 6 47</pre>	Project dura  Environment History Connections  Project Dura  Content of the second se
<pre> text text text text text text text text</pre>	Environment History Connections Towing and the second sec
<pre>nowiencers nowiencers nowiencers</pre>	Environment Hinkoy Cenerctions Environment Hinkoy Cenerctions Constant - Constant - Constant Constant - Constant - Constant - Constant Constant - Constant
<pre>inverters in the law law inverters invertige inverters invert</pre>	Propert Derr Environment History Connections Control Derregation (1999) Control Der
<pre> to me to An one her her her her her her novercurs  Source on Same Q, +</pre>	Project Dara Environment History Connections Project Dara Project Dar
a de me me be hen hen hen hen hen         a de me hen hen hen hen hen hen         a de me hen hen hen hen hen hen         a de me hen hen hen hen hen hen         a de me hen hen hen hen hen hen         a de me hen hen hen hen hen hen         a de me hen hen hen hen hen hen         a de me hen hen hen hen hen hen         a de me hen hen hen hen hen hen hen         a de me hen hen hen hen hen hen hen         a de me hen hen hen hen hen hen hen hen hen he	Propert Own  Indicate History Connections  Content  General History Connections  General Content  General C
<pre> inserver into dataframe inserver into dataframe</pre>	Project Okr Environment Hinlory Connections Control Oner Connections Control Oner Connection Ogenre_matr. Large matrix (185940 elemen. Ogenre_matr. 10329 obs. of 18 variables Ogenres2 10329 obs. of 1 variables O movies 10329 obs. of 1 variables Denvies 10329 obs. of 3 variables
<pre>tes me Ano See Ball Deep Main New Me newVercus's Source on Same Q A + 1  Addins + 33 4 #Convert into dataframe 35 opene_matrix2 &lt;- as.data.frame(genre_matrix2[-1,], stringsAsFactors=FALSE) Free 36 * For (c: in incol(genre_matrix2)) { 37 opene_matrix2(c) &lt;- as.integers 38 } #Convert into dataframe 40 #Create a matrix to search for a movie by penre: 41 Wears &lt;- as.data.frame(movies/title, stringsAsFactors=FALSE) 42 library(dat.table) 43 substrikingt &lt;- function(x, n) { 44 substrik, nchar(x)) - 45 years &lt;- as.data.frame(substrishtright(substrikinght(years5) moviesStitle', 6 45 search.matrix &lt;- cbind(source) -</pre>	Project Der Environment History Connections Project Der Project Der
<pre></pre>	Project Dire Environment History Connections Project Dura Data O genre_matr. Large matrix (185940 elemen O genre_matr 10329 obs. of 1 variables O genres 10329 obs. of 1 variables O genres 10329 obs. of 3 variables O movies 10329 obs. of 3 variables Data Deta
<pre></pre>	Propert dem Environment History Connections Conta
<pre>tex mu he her her her her her her her her her</pre>	Propert Derr Environment History Connections Chall Inducated - Chall Inducated - Chal
<pre>note not not not have have not note not not not have have not not not not not not not source on Save A data frame (genre_matrix[-1,], stringsAsFactors=FALSE) #ren for (in in incol(genre_matrix2)) {</pre>	Project Der Environment History Connections Project Der Data Ogenne_matr. Large matrix (185940 elemen. Ogenne_matr. Large matrix (185940 elemen. Ogenne_matr. 10329 obs. of 1 variables Ogennes 10329 obs. of 1 variables Ogennes 10329 obs. of 3 variables Ogennes 10329 obs. of 3 variables Provies 10329 obs. of 3 variables



Now, next is for c in 1 to 18 n column genre matrix2 is 18, so, c is 1 to 18 the genre matrix for each column I am changing it to their integer values. So, if I just run this, is doing nothing but changing all values to their integer values. Now, what will I do? I will check, create a search for a movie by genre so, this is an off topic, but this is something that helps me in developing further. So, I am creating a data frame called years which is the movie title. So, years is equal to as data frame movie titles, string as false so, years looks this, so, this is the movie title column basically.

And then I am using library called data table, to what to do? To substring so, I am saying that substring write is a function, which will substring from left to right. So, this is what I am doing substring function and using the substring function from here, I will only take the years this 1995, this 1995, this 1995 I will scrape so, you can do right how much? From right side it is like the right function, if you have used Excel, it is the right function, on the right it is the second to 1, 2, 3, 4, fifth. So, it should start from 2 in that 5. So, from right side that is what I am doing.

So years is equal to you see that 1, 4 basically is something that it is doing, it is subtracting, creating a subset and that years is getting me all this year values of this particular thing. So again, you can do it in Excel if you want, but I am doing it here.

#### (Refer Slide Time: 9:18)

House In Life Code Have Part 5	asser Build Datag Faille Took Help		
0 • 🗣 💣 •	🗑 🖓 🗋 🍐 🖌 🗛 Goto Herbrichten	🔠 • Addins •	Project (No
moviereco.r =		-0	Environment History Connections
4 B B	Source on Save 🛛 🍳 🎢 📲 🚺	🛶 Run 👘 👍 Source + ≷	🤕 🔒 📪 Import Dataset 🔹 🧹 👘 List 🔹
40 #Creat 41 years 42 librar	e a matrix to search for a movi <- as.data.frame(moviesStitle, y(data.table)	e by genre: stringsAsFactors=FALSE)	Global Environment - Q. O genre_matr. 10329 obs. of 18 variables
43 - substr 44 subs	Right <- function(x, n){ tr(x, nchar(x)-n+1, nchar(x))	3	O genres 10329 obs. of 1 variable O genres2 10329 obs. of 10 variables
46 years 47 48 search	<- as.data.frame(substr(substrR	<pre>ight(substrRight(years5'moviesStitle', 6 hstr(movies[.2].1, nchar(movies[.2])-6).</pre>	o movies 10329 obs. of 3 variables O ratings 105339 obs. of 4 variables
49 colnam 50	es(search_matrix) <- c("movield	", "title", "year", genre_list)	Files Plots Packages Help Viewer
51 write. 52 search 53	csv(search_matrix, "search.csv _matrix <- read.csv("search.csv	) ", stringsAsFactors=FALSE)	9 9 / Jam Boot • 0
S4 # Exam S5 subset	ple of search an Action movie p (search_matrix, Action == 1 & y	roduced in 1995: ear == 1995)Stitle	
48.34 (Top Level)	1	R Script a	
Console Terminal	x	-0	
C:/Users/Dell1/Desle	top/Week7/Semion 4/ 🔅		
<pre>&gt; library(dat. &gt; substrRight + substr(x, + } &gt; view(years) &gt; years &lt;- as 1,4)) &gt; view(years)</pre>	a.table) <- function(x, n){ nchar(x)-n+1, nchar(x)) .data.frame(substr(substrRight(	substrRight(years\$'movies\$title', 6),\$),	
Thum Left Code You Pilet 3	num but bing falls for the		
) • 🍕 🐨 •	G G to flottestion	🔠 • Addins •	Enject (No. 1997)
• movierecor =	movies #	-0	Environment History Connections
01017	Filter	Q	💣 🔒 📅 Import Dataset + 🔏 📃 List +
* movield	itte	i genres	Global Environment • Q
1	1 Ry Story (1995)	Adventure Animation Children Comedy(Fantasy	genre_matr. 10329 obs. of 18 variables
2	2 Jumanji (1995)	Adventure Children Fantasy	Ogenres 10329 obs. of 1 variable
1	3 Grumpier Old Men (1995)	Corredy(Romance	o genres2 10329 obs. of 10 variables
4	4 Waiting to Exhale (1995)	Comedy/Dramajllomance	Oratings 105339 obs. of 4 variables
5	5 Father of the Bride Part II (1995)	Comedy	years 10329 obs. of 1 variable
6	6 Heat (1995)	Action(Crime)Thriller	Play Make Backware Male Manual
7	7 Sabrina (1995)	Corredy/Romanor	The ros racage nep viewer
1	8 Tom and Huck (1995)	AdventurejChildren	Export -
,	9 Sudden Death (1995)	Action	
10	10 GoldenEve (1995)	Action(Adventure(Thritter -	
several 1 to 16 of 16	327 Billing		
Console Terminal	K	-0	
<pre>C/User/Delli/Delli &gt; library(data &gt; substrRight + substr(x)</pre>	<pre>top/Week7/Semicn 4/ # a.table) &lt;- function(x, n){     nchar(x)-n+1, nchar(x)) </pre>	2	
+ } > View(years)			

> View(years) > years <- as.data.frame(substr(substrRight(substrRight(years5'moviesStitle', 6),5), 1.4) > view(movies) > view(movies)



moviere	co.r =	search_mat	brix ×				-	6	Environment History Connections	-0
6		Filter				Q,			🐨 🔒 🐨 Import Dataset + 🧹 💿	list - 16
movie	td 1	title	63	year	Action	Adventure	Animation	ŝ	Global Environment • Q.	
1	11	loy Story		1995		0			genre_matr_ 10329 obs. of 18 variable	5
2	2.1	kamanii		1995				1	ogenres 10329 obs. of 1 variable	
	1.0	Grumpier Old Mr	en	1995			1		Omnuies 10329 obs. of 3 variables	
1		Waiting to Exhale		1995					oratings 105339 obs. of 4 variable	5
5	5 1	Father of the Brid	ie Part II	1995		0	)		osearch_mat_ 10329 obs. of 21 variable	s 📄
6	6.)	Heat		1995		1 1	)		Files Plots Packages Help Viewer	-
7	7.1	Sabrina		1995		0	)			
8	8 1	fom and Huck		1995		0	í.			
9	9.9	Sudden Death	_	1995			)	1		
howing 1 to	10 of 10	,329 entries								
Console	Terminal	D					-			
C/Users/D	ell5/Desi	ktop/Week7/Ser	mion 47							
vien()	(ears)							7		
years	<- as	.data.fram	ne(substr(substri	tight(substrRight)	(years\$'m	oviesStitle	, 6),5)	γ.		
,4))										
vienty	(ears)	1							A COLOR	
search	mate	iv ca chin	dimovies[ 1] si	hete/movies[ 2]	nchar(m	nuine[ 21).	6) year			
genre	matri	x2)	infinories[iv]; a	mari (maricalis);	e fine india (an	or real ( + ) -	ey, jean	1		
	asisa	arch matri	ix) <- c("movieId	", "title", "year	", genre	list)				
- colnan										

Next is I will create the same thing with a movie. So I am getting a search matrix, which is binding movies data first column. That means movie ID, next is substring of movies one, two, so the movie name, basically only the movie name, then the years and then the genre matrix. So, when I put this and then put the whole column names, this is the search matrix that I get a movie ID, the title, the year, and then the genre id matrix. So, this is what I have created from this. I save this so that I can use it for later purpose.





So this is how it looks like basically, the search matrix will look like movie ID, title, year and the overall genre of the movies. So, you can create a pivot in this particular matrix or you can create, you can search like this subset, search matrix, action is equal to 1 that means it is a movie action, and years = 1995, what is the title of that? If I find out, so these are the movies, which are, which come out in 1995 under action movies.

So if you want furthermore, let us say, I want action = 1 and comedy = 1 and 1995. So, these are the movies which are action and comedy. So, you can create those kind of search.

Now, what I will do next is I will create a binary rating. So, create a user profile is what is my next objective, so I am creating binary ratings anything 4 and 5 is high, 3, 2, 1 is low. So, that is what I am creating here so, binary rating is equal to ratings and then if the value is greater

than 3 then put it 1, if it is smaller than 3 then - 1. So, that is what I just trying to do. Likes means 1, dislikes means - 1 so, it will run for some time. And let us see how much time it runs.

So, i in 1 to n row binary ratings, binary ratings is 1 lakh. So, it will take quite a bit of time to totally find out so, I will see it is now 28,000. And what I am trying to do here, basically it will go up to 1 lakh. What I am going to do is that sometimes it does not matter if there are multiple gaps let us say 1, 2, 3, 4, 5, multiple levels of ratings, creating similarity becomes very difficult sometimes. So, 1 - 1 is basically a more or less I would say dichotomy. So, if you like 1, if you dislike - 1 and why did I take  $\geq$  4, 4 or 5? Because oftentimes you have seen that the reviews are a little bit positively skewed in nowadays.

So, if it is positively skewed so, 1, 2, 3 is considered to be low and 4 and 5 is considered to be high and we are actually trying to do that we are in 98,000, 105000 will take okay. So, it is over so, I have created a binary rating. So, how does the binary ratings look like?

Image: Constraint of the sectors of the sec	Image: Constraint of the second sec
• usefd         movidid         reling         disettamp           1         1         1277973         0         <	userid         movidal         userid
1 1 10 <sup>10</sup> 1 127707793 2 1 24 -1 127707793 3 1 32 1 127707507 4 1 47 1 127705076 5 1 50 1 127705056 5 1 50 1 127705033 6 1 100 1 127705033 6 1 100 1 127705033 6 1 106 1 127705033 6 1 106 1 127705784 9 1 156 -1 127705784 9 1 156 -1 127705786 Browleg 1b 10 0105/138 entries Sceneb Terminal	1       1       100%       1       122789307         2       1       24       -1       122789307         3       1       32       1       122789307         4       1       122789307       0 binaryrati_ 105339 obs. of 4 variable         5       1       30       1       122789306         6       1       10       1       122789319         6       1       10       1       122789319         6       1       10       1       122789319         7       1       10       4       122789319         8       1       105       -1       122789319         10       1       1       122789319       10         10       1       222789318       10       10         10       1       22789318       10       10         10       1       22789318       10       10         10       1       22789318       10       10         10       1       22789318       10       10         10       1       22789318       10       10 </td
2       1       24       -1       12789507         3       1       32       1       12789506         4       1       07       1       12789506         5       1       50       1       12789507         6       1       10       1       12789503         7       1       150       -1       12789509         7       1       150       -1       12789509         8       1       166       1       22789784         9       1       156       -1       22789784         9       1       156       -1       22789784         9       1       156       -1       22789784         9       1       156       -1       22789784         9       1       104       1       22789784         9       1       104       1       22789784         9       1       104       1       22789784         9       1       104       1       22789784         9       1       104       1       22789784         9       10104/105/104 emtes	2       1       24       -1       12789807         3       1       32       1       12789307         4       1       47       1       12789336         5       1       50       1       10229       05. of 1         6       1       100       1       127893050       0       0       genre_matr.       10329       05. of 1       Variable         7       1       100       1       127893050       0       genres       10329       05. of 1       variable         9       1       161       1       12789304       0       0       genres
3       1       32       1       12/200536         4       1       47       1       12/200536         5       1       50       1       12/200532         6       1       10       1       12/200530         7       1       150       -1       12/20050         8       1       164       1       12/200546         9       1       165       -1       12/200546         9       1       164       1       22/200546         9       1       164       1       22/200546         9       1       164       1       22/200546         9       1       104       1       22/200546         9       1       104       1       22/200546         10       1       20/107155       100       100         10       10       105/107/108       -1       20/200715         10       10       10/5/108       -1       12/2007/306         Showing 10       10       10/5/108       -1       12/2007/306	3       1       32       1       122786346       0
4       1       47       1       122789556         5       1       50       1       122789559         6       1       101       1       122789590         7       1       150       -4       12789959         8       1       161       1       22789786         9       1       165       -1       122789786         10       1       22789786	4     1     47     1     122789535       5     1     50     1     122789533       6     1     110     1     122789540       7     1     150     -4       1     161     1     122789540       9     1     166     1       1     161     1     122789540       9     1     166     1       1     165     4     122789546       10     1     04     -4       10     1     04     -4       10     1     04     -4       10     1     04     -4       10     04     -4       10     04     -4       10     04     -4       10     04     -4       10     04     -4       10     04     -4       10     04     -4       10     04     -4       10     04     -4       10     04     -4       10     104
5       1       50       1       12/1790503         6       1       100       1       12/1790509         7       1       150       -4       12/1790706         8       1       161       1       12/1790706         9       1       165       -4       12/1790716         10       1       20/167/166	5       1       50       1       12/1796523         6       1       10       1       12/1796523         7       1       150       4       12/1795966         8       1       161       1       12/1797135         10       1       20/14       12/1797135       10         10       1       20/4       -1       12/1797135
6       1       10       1       12/179910         7       1       159       -4       12/179910         8       1       161       1       12/1799104         9       1       159       -4       12/1799104         9       1       359       -1       12/1799105         10       1       324       -1       12/1799106         Showing 11b 12 of 105/128 embies	6         1         110         1         12289310           7         3         109         4         12289340           8         1         161         1         122199340           9         1         165         4         122199315           10         1         202199316         1         1022199316
7       1       150       -1       1227095940         0       1       161       1       227095746         9       1       165       -1       227095786         30       1       304       -1       1227095786         Sciencing 1 to 10 of 105/138 embies       -       -	7         1         150         -1         122/39/39/40         Here         Non a rackges         Here         Nona rackges         Here </td
1 164 1 12/19/964     9 1 165 -1 12/19/974     1 165 -1 12/19/9715     10 1 204 -1 12/19/9716     10 10/10/10/10 entes  Consolt Terminal	0         1         161         1         1221989864           9         1         165         -1         1221989105           10         1         304         -1         1221989366
9 1 165 -1 1227897115 10 1 204 -1 122789716 Browing 1b 10 01 105,138 entries Censols Terminal	9         1         165         -1         122/98/115           10         1         201/98/716         -         -
10 1 204 -1 1227/095796	10 1 204 -1 1217855786 -
Showing 1 to 10 of 10/L99 entries	
Console Terminal -	3moreng 1 to 10 of 10%,see entries
	Console Terminal =
C;/Users/Dell1/Desktop/Week/7/Semion 4/ 🔅	C;/Users/Dellt/Desktop/Week//Session 4/
> View(binaryratings)	> View(binaryratings)

### (Refer Slide Time: 12:36)

Name	
9 • 1 🤫   🛫 • 1 🖓 - 0 1 🚔	(B) Project: (Noriet
0 moverecur =	Environment History Connections
()   2)   Source on Save   Q 7 - [   -+ Run   ++	🕒 Source 🔹 💣 🔒 👘 Import Dataset + 🧹 👘 Elikt + 🌀
64	* 🍓 Global Environment • Q
65 - For (i in 1:nrow(binaryratings))(	Data
67 binaryratings[1,3] > 3)(	Obinaryrati_ 105339 obs. of 4 variables
68 )	Ogenre_matr_ Large matrix (185940 elemen_
69- elset	Openre_matr. 10329 obs. of 18 variables
70 binaryratings[1,3] <1	Openres 10329 obs. of 1 variable
72	Ogenres2 10329 obs. of 10 variables
73	
74 # convert binaryratings matrix to the correct format:	Files Plots Packages Help Viewer
75 binaryratings2 <- dcast(binaryratings, movieid-userid, value.var 76	Faring , the state Plane Papert - 0
77 · for (i in 1:ncol(binaryratings2))(	
<pre>78 binaryratings2[which(is.na(binaryratings2[,i]) == TRUE),i] &lt;- ()</pre>	0
R0 *	
651 (lop Level) 1	R Script a
Console Terminal =	-0
C:/Users/Dell3/Deditop/Week7/Session 4/ -/0	
> View(binaryratings)	
5 (	
	A FILL
	Contraction of the second
9 8 9 7 X 9 7 1 0	No. 10 March 1997







Binary ratings look like this, basic thing 1 or - 1 nothing else. So, this code is actually binary code you could have done a if else, simple if else would have taken much lesser time, anyways. So, then what I do is convert this binary rating so, what I will do is I will convert binary matrix with correct format.

So, the format is like this binary ratings2 which I run and it will create a binary ratings2, which is the movie ID. And then correspondingly, there are lots of people 669 variables means 669 observations, all these NS means nobody has given review for that. And if they have given some review either it is positive or negative, 1 means positives, - 1 is negative. So, that is what I have got so, basically 10,325 movies and 669 users, I will got a 0 1 matrix for that. So, it is a large matrix actually and then, so, if by chance, if there is nothing there, then if there is NA, then put a 0 there basically.

So, I am putting 0 for all the values where it is NA and then remove the movie ID column. So, that means I am getting up 10,325 observations, which is the movies, 668 observations, which are the users and 0 1 and - 1 this is the three values that are there, 0 means there was no rating, 1 means he liked it, - 1 means he basically disliked so, that is a binary rating formula that I have created.

#### (Refer Slide Time: 14:14)





Now remove rows that are not rated from movies datasets, there are certain movies which are absolutely not rated by anybody. So, if they are not rated by anybody, why should I use them? So, we are doing that and then movies rows that are not rated for genre matrix2 also so, any movies which has been not rated you have to remove that from genre matrix2 also. That is why I am creating genre matrix3, so, genre matrix2 to 3 there are four movies, which were not rated 10,329 10,225 so, that means there are four movies which were not rated, we removed them.

Similarly movies2 basically have 10,325, therefore movies which were not rated by anybody, we removed them. Fair enough, that is how we are reducing the dataset, so the calculation takes less time. Then what, calculate the dot product of the genre matrix and the ratings

matrix and obtain the user profile. So, understand what is the genre matrix. Genre matrix is 10,325 of 18 genres, whether this movie is belonging to this genre, and then what is my rating matrix? Rating matrix is basically 10,325 of various people giving a rating.

Now, if I want to find out that whether this user is fond of comedy, then what will I do? I have seen that this user, this 668 from this binary matrix, this one, if I just click on this one, let us see user number one. Let it come and then I will say, let us say user number one whether he so, user number one if you say that 10,325 or let us says user number two because the data is showing here, he has seen this movie, he has not liked this movie, he has not liked this movie and if you come down further he has liked this movie. So, some movies user number 2 like, some movies user number 2 did not like.

And then all these movies, like movie number 14, if I go to genre matrix, I know that movie number 14 is basically a, movie number 14, movie number 14 is basically is a drama movie, which he liked. And then it is only a drama probably.

Then, let us say this one, which he did not like, which one was that? That was movie number 3, or movie number 5 he did not like. So, in the genre matrix I know movie number 3 is basically a comedy, movie number 5 is also comedy. And movie number 3 is nothing else. Probably romance also, movie number 5 is not romance. So, movie number 3 and 5 both are comedy which the user number 2 did not like, so from this information I can find out what is that he is net liking, the overall liking of a genre, how would I do that? Which movies he has seen, which genre it is from there, I will find out how much is rated for them, + 1 or - 1.

## (Refer Slide Time: 17:34)



So that is what I am going to do here in the next set of calculations. So, I am creating a result which is 18, 668, why 18? That means 18 rows, 668 columns, each column is one consumer or one user. And I am saying that whether he, basically a sum of 1's and - 1's, whether he has seen a movie in a particular genre, if like liked 1, if he disliked - 1. If the net sum is positive, then overall he is liking, if the net sum is negative then overall he disliked in that genre. So, that is what I am populating here, quickly so, result, I got this result column.

So now, if I just see this matrix carefully, what I get is user number 2 do not, user number 1 likes the first genre most 28, then sixth genre, the eighth genre and probably 16th genre. These are the genres that user number 1 likes, user number 2, probably this one, second genre, ninth genre. Now, if by chance see there are two guys who are, it is user number 1 is a

movie buff, he watches more movie, so the one which one he dislikes, likes a ninth is still much higher than the most liked one for the second user, which is 9 here. So he is 9 and this 9 has two different meaning. For the user 2 this 9 means that he likes the eighth genre most but he does not watch movie much.

On the other hand, this 9 means that he watches movie very much. 9 rating is actually in comparison to 28, 27, 27 they are very small so, it is not his favorite genre, though the overall rating is 9. So, how to check that situation I have to normalize that. So, I am doing that, first of all convert the, so, I am converting them to 0 and 1, if it is smaller than 0, 0 if it is greater than 0, 1. Now I am creating a dataset, which will be used by this recommended engine, that recommender lab library to create the recommendation engine.

(Refer Slide Time: 19:49)

🔸 🎕 🤕 + 🔛 🗐 👛 🔺 Gehefietuut	ion 🔄 🛙 🔠 • Addins •			Project (None)
movierecor = result = binaryrating:2 =	genre, matrix3 ×	- Environm	sent History Connections	
Instruction     Instruction       115     ## assume that users like similarit       116     ## assume that users like similarit       117     # that are closest in similarit       118     # represents a user's preference       129     # use Jaccard Distance to measu       120     # the user-Based Collaborative       121     # the user-Based Collaborative       122     # the user-Based Collaborative       123     # the user-Based Collaborative       124     # Create ratings matrix, mag. use       125     ratingnat < ass.matrix(ratings, use	→ Run → Aun → Aun ar items, and retrieve movies y to a user's profile, which e for an item's feature. re the similarity between user profiles Filtering Approach userId, columns = movield rid-movield, value.var = "rating", ma.m t[1]) #remove userIds	e - 2 C C C C C C C C C C C C C C C C C C	<pre>mm may common proper butset - ngmat Large matrix (6 ngs 10533 0bs. of n mum [1.8, 1:66 ch_mat. 10329 obs. of 2 s 10329 obs. of 1 </pre>	eut • 9 897100 eleme 4 variables 81 11 11 12 2 variables variable
129 # Similarity Calculation Method 130 # Nearest Neighbors: 30 131 ***********************************	: Cosine Similarity	, II R Script 2		
Console Terminal =		=0		
<pre>ttaching package: 'reshape2' the following objects are masked from     dcast, melt &gt; Acreate ratings matrix. Rows = user1 &gt; ratingmat &lt;- dcast(ratings, user1d-m &gt; ratingmat &lt;- as.matrix(ratingmat[,-1 &gt; ]</pre>	"package:data.table": d. Columns = movieId ovieId, value.var = "rating", na.rm=FAL ]) fremove userIds	a)		





So, what I will do is I will create a rating mat, the library called reshape. And rating mat is getting created, rating matrix is a large matrix which looks like this. Let me just put it up here which is nothing but a reshaped version of the dataset that we have created till now. Let it come. So, if you check it carefully it is saying user ID to movie ID and the value is rating, user ID, movie ID and the value is rating, that is what is getting plotted here and the rating matrix all these 1's are basically the user IDs in x axis and y axis is movie id. So, it is taking a little bit higher time. So, the view is not coming properly, I will not focus on that.

And then the method is UBCF, see similarity calculation whether on cosine similarity that is what we are doing, nearest neighbors, remember, there we took five neighbors, here we are taking up to 30 neighbors, the library is recommender lab. The rating matrix is we are converting it to recommender labs parse matrix.

So, these are basically specific to this particular library you have to use this, you have no choice and then the similarity, basically similarity of the users is similarity rating mat 1 to 4, method is cosine which is user so, I am doing a user to users similarity matrix and if I try to find out a similarity score for the first four guys, this is how it looks like. So, you have taken first four, you can take the whole 1 to 668 it will give 1 to 668, I have taken the first four for the simplicity of calculation and this is what I am getting the image, the similarity matrix.

So, the yellowish it is the better, that is the image that we are getting so, this one is coming as to be good we have check the red colors again once more.

Then I can also find out the, compute similarity between the first four movies so, I can do it for the movie by movie similarity also so, for the first four movie if I try to find out the similarity, this is the similarity, see 0.9, 0.95, 0.91 so, the first four movies are very similar probably with each other.

Now, explore the value for ratings so, this is where I am actually trying to find out the ratings, the value of ratings so, vector ratings is equal to as vector, rating mat \$ data sorry, at the rate data so, this is a different format for sparse matrix and unique of vector rating. These are the various vector ratings that you got 5, 4, 3, 4.5, 1.2 and so on. And I am creating a table for that and it will just show me what is the tabular ratings so, vector ratings is these are the how many vector ratings that you have got is something that has been listed here.

#### (Refer Slide Time: 22:56)





Now, vector ratings  $\neq 0$  aI will only take why it is  $\neq$  zero? Because zeros are NA's if you remember, so, I am taking that and making them factor and then if I am plotting them by chance, it is basically a histogram plot that how many different kind of ratings that I am getting here. Then if I explore the viewing behavior with a similar way, I will get, I will not spend time on this, we will just check it, I get a viewing behavior of the top five movies also so, these are the top five movies viewing behavior.

And then if I visualize the matrix, I should visualize this one, the heat map so, the heat map of the first rows and columns so, this is the user rows and this is the item columns. The darker it is, the more it is preference towards that based on the dataset that we have got. Now, what I will do is I will that rating mat if you remember, the rating mat, this is what I am coming back once more the rating mat dataset.

## (Refer Slide Time: 23:54)





This is actually the one that I will use in my UBCF the recommender lab library. So, I am normalizing it first and after normalization this is the values that you are getting, the item columns and the user rows. And all these 1's are basically the heat map, what is a similarity level 2 to - 2, various items and various rows how close it is, how is the chance of seeing this particular thing.

And then I use UBCF, UBCF means? User based collaborative filtering, and nearest number = 30, cosine is equal to, method is equal to cosine, you can change this cosine to correlation and etcetera. So, that is how you get the model and you find out the details of the data. So, the model details of the data says that there are 668 and 10,035 rating matrix of class, real rating matrix. With this many ratings has been used and this have normalized using center of rows.

#### (Refer Slide Time: 25:00)



The Car are find here full here fully here the	
- 🔍 👻 - 🖓 🖓 - 🖓 - 🖉 - Addins -	Project (None)
moverecur x      result x      binarytatings2 x      genre_matrix3 x	Environment History Connections
Source on Save   Q / .   -+ Run   ++   + Source - 2	💣 🖯 📅 Import Dutaset - 🖌 📃 List - 👩
<pre>216 for (i in 1:10)[ 217 recom_result[i] &lt;- as,character(subset(movies, 218 movies/movie1d == as.integer(recom_1) 219 } 220 221 222 # Evaluation; 223 evaluation.scheme &lt;- evaluationscheme(ratingmat, 223 evaluation.scheme &lt;- evaluationscheme(ratingmat, 223 evaluation.scheme &lt; evaluationscheme(ratingmat, 223 evaluation.scheme &lt; evaluation.scheme(ratingmat, 223 eval</pre>	Cobat Inviconment -     recom_resu. chr [1:10, 1] "Eve's Bayou     orecommende. Large Recommender (2 Mb)     result mm [1:18, 1:668] 1 1 1 1 1.     O search_mat. 10329 obs. of 22 variables     O table_views 10325 obs. of 22 variables     O vaers 10329 obs. of 1 variable
2/4 method="cross-validation", 225 k=5, given"3, 226 goodatting=5) #k=5 meaning a 5-fold cro 227 evaluation_results <- evaluate(evaluations.cheme, 228 method="uBCF",	Files Plots Packages Help Viewer
229     n=c(\$,10,20)]       230     eval_results <- getConfusionMatrix(evaluation_r§sults)[[1]]	
Consol Terminal CurveN/Jession 4/ Cytown(DoH)/Dustrop/Timek/Jession 4/ > evaluation_results <- evaluate(evaluation_scheme, + method="UBCF", + nec(5,10,200) UBCF run fold/sample [model time/prediction time] 1 [0sec/4,77sec] 2 [0.01sec/4,9sec] 3 [0.01sec/4,46sec] 5	Heatmap of the top users and movies

Now, if you want to find out a recommendation, the top 10 recommendation. So, for the first person I am taking rating mat 1, if you change this to 2 or 1 to 10, you will get the recommendations for all the 10 guys. So, recommendations for 10, it is the top 10 list for one user is something that is saved here in recom. So, if you check the recom, basically, the items and the ratings and the item levels have been written here.

So, the next part is basically if you want to see the recom, this is the recom list. So recom underscore list is basically the first person, the id of the various movies. And obtain the recommendations based on this recom list if I just tried to find out the names, then recom result will give me the names of the movies for the first guy. So, that is how based on UBCF you can check the code properly, we are doing it.

Now, if you have to evaluate and if I evaluate with n = 1, n = 3, see 1 nearest neighbor, 3 nearest neighbor, let us say 5, these are the three things that we are checking let us say 3, 5, 10 and 20. 5, 10, 20 these are the three guys I will check then how will I run it? I will just run it like this. So, good ratings = 5, bad rating so run this. First rating is taking 5 nearest neighbor, then it will take 10 nearest neighbor and then it will take 20 nearest neighbor and what is the evaluation score and etcetera is something that we will see right now.

(Refer Slide Time: 26:58)



So, if I check that evaluation results, you will get it here so, this is the evaluation result for 5, 10 and 20 corresponding true positive, true negative rates and etcetera are giving here. So, the reference is this particular link, you can get a better discussion about the about this particular thing in this link, I have taken the code from that link and probably the dataset has been first publicly available. And this is how we create the recommendation engine with rating dataset with bigger dataset. In the next video, I will also show you how to create a recommendation engine with a smaller dataset which can still be handle able. Thank you very much. I will see you in the next video.