Assignment 2 Final v2

Sara Monedero

2022-06-28

library (readr)  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

## ✔ ggplot2 3.3.6 ✔ dplyr 1.0.9  
## ✔ tibble 3.1.7 ✔ stringr 1.4.0  
## ✔ tidyr 1.2.0 ✔ forcats 0.5.1  
## ✔ purrr 0.3.4

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(knitr)  
library(bslib)

##   
## Attaching package: 'bslib'

## The following object is masked from 'package:utils':  
##   
## page

#Section 1: Data Description  
#The data set love\_actually\_appearences is based on the famous movie ‘Love Actually’, and what collects is the if all the actors that participated on that movie appear or not on each of the movie’s scenes.   
  
#The information has been collected by two different variables:  
#1. Actors: First we have a list of all the actors that participated on the movie such as: Bill Nighy, Keira Knightley, Hugh Grant,… There’s a total of 14 actors.   
#2. Scenes: The movie has been broken in a total of 71 scenes that together complete the movie. On the data set they are ordered by chronological order.   
  
#With this dataset I want to see who are the actors that appeared in more scenes. I think having that information is an essential starting point that gives a sense about the movie.   
  
#We could later double check this ranking with a secondary data base called love\_actually\_adjencies where we on the rows and columns we have the 14 actors displayed. This dataset collects the relation ship between the characters of the movie, assigning them an index that goes from 0 to 13, being 0 a null relationship and 13 the strongest. Although I don’t have a detailed plan of how the index has been pondered, based on the article https://fivethirtyeight.com/features/the-definitive-analysis-of-love-actually-the-greatest-christmas-movie-of-our-time/ I imagine the index is based on:  
#1. The time those characters share on screen  
#2. The amount of conversation (in minutes) they have on screen   
  
#Format:  
#Both data sets are saved in csv format.   
#Both datasets Appearances and Adjencies are flat files, they have a clear structure where each line of the plain text holds only one record. Records are separated by commas, which matches with the fact they are saved as csv files. We see that the data bases are easy to understand what they are explaining.  
#The databases are not fixed width since they are saved as CSV, and CSV files can’t have a fixed length of their columns.

#Section 2: Data Reading

#Adjacencies Dataset  
url <-"https://raw.githubusercontent.com/fivethirtyeight/data/master/love-actually/love\_actually\_adjacencies.csv"  
rawadjacencies <- read\_csv(url)

## Rows: 14 Columns: 15  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): actors  
## dbl (14): bill\_nighy, keira\_knightley, andrew\_lincoln, hugh\_grant, colin\_fir...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

#To be able to read both csv files I had to use the packages readr with the function readcsv\_(url previously added here).

#Section 3: Clean the Data

#For Adjacencies database  
as.integer(rawadjacencies$bill\_nighy)

## [1] 9 2 2 2 1 1 0 1 1 3 1 1 1 1

adjacencies <- gather(rawadjacencies, key="Other Actor", value="Relationship Index", 2:15)  
as.data.frame(adjacencies)

## actors Other Actor Relationship Index  
## 1 bill\_nighy bill\_nighy 9  
## 2 keira\_knightley bill\_nighy 2  
## 3 andrew\_lincoln bill\_nighy 2  
## 4 hugh\_grant bill\_nighy 2  
## 5 colin\_firth bill\_nighy 1  
## 6 alan\_rickman bill\_nighy 1  
## 7 heike\_makatsch bill\_nighy 0  
## 8 laura\_linney bill\_nighy 1  
## 9 emma\_thompson bill\_nighy 1  
## 10 liam\_neeson bill\_nighy 3  
## 11 kris\_marshall bill\_nighy 1  
## 12 abdul\_salis bill\_nighy 1  
## 13 martin\_freeman bill\_nighy 1  
## 14 rowan\_atkinson bill\_nighy 1  
## 15 bill\_nighy keira\_knightley NA  
## 16 keira\_knightley keira\_knightley 7  
## 17 andrew\_lincoln keira\_knightley 7  
## 18 hugh\_grant keira\_knightley 1  
## 19 colin\_firth keira\_knightley 2  
## 20 alan\_rickman keira\_knightley 1  
## 21 heike\_makatsch keira\_knightley 0  
## 22 laura\_linney keira\_knightley 3  
## 23 emma\_thompson keira\_knightley 1  
## 24 liam\_neeson keira\_knightley 1  
## 25 kris\_marshall keira\_knightley 2  
## 26 abdul\_salis keira\_knightley 2  
## 27 martin\_freeman keira\_knightley 1  
## 28 rowan\_atkinson keira\_knightley 0  
## 29 bill\_nighy andrew\_lincoln NA  
## 30 keira\_knightley andrew\_lincoln NA  
## 31 andrew\_lincoln andrew\_lincoln 8  
## 32 hugh\_grant andrew\_lincoln 1  
## 33 colin\_firth andrew\_lincoln 2  
## 34 alan\_rickman andrew\_lincoln 2  
## 35 heike\_makatsch andrew\_lincoln 1  
## 36 laura\_linney andrew\_lincoln 4  
## 37 emma\_thompson andrew\_lincoln 2  
## 38 liam\_neeson andrew\_lincoln 1  
## 39 kris\_marshall andrew\_lincoln 2  
## 40 abdul\_salis andrew\_lincoln 2  
## 41 martin\_freeman andrew\_lincoln 1  
## 42 rowan\_atkinson andrew\_lincoln 0  
## 43 bill\_nighy hugh\_grant NA  
## 44 keira\_knightley hugh\_grant NA  
## 45 andrew\_lincoln hugh\_grant NA  
## 46 hugh\_grant hugh\_grant 13  
## 47 colin\_firth hugh\_grant 1  
## 48 alan\_rickman hugh\_grant 3  
## 49 heike\_makatsch hugh\_grant 1  
## 50 laura\_linney hugh\_grant 1  
## 51 emma\_thompson hugh\_grant 3  
## 52 liam\_neeson hugh\_grant 2  
## 53 kris\_marshall hugh\_grant 1  
## 54 abdul\_salis hugh\_grant 1  
## 55 martin\_freeman hugh\_grant 2  
## 56 rowan\_atkinson hugh\_grant 0  
## 57 bill\_nighy colin\_firth NA  
## 58 keira\_knightley colin\_firth NA  
## 59 andrew\_lincoln colin\_firth NA  
## 60 hugh\_grant colin\_firth NA  
## 61 colin\_firth colin\_firth 13  
## 62 alan\_rickman colin\_firth 1  
## 63 heike\_makatsch colin\_firth 0  
## 64 laura\_linney colin\_firth 2  
## 65 emma\_thompson colin\_firth 1  
## 66 liam\_neeson colin\_firth 1  
## 67 kris\_marshall colin\_firth 1  
## 68 abdul\_salis colin\_firth 1  
## 69 martin\_freeman colin\_firth 1  
## 70 rowan\_atkinson colin\_firth 0  
## 71 bill\_nighy alan\_rickman NA  
## 72 keira\_knightley alan\_rickman NA  
## 73 andrew\_lincoln alan\_rickman NA  
## 74 hugh\_grant alan\_rickman NA  
## 75 colin\_firth alan\_rickman NA  
## 76 alan\_rickman alan\_rickman 13  
## 77 heike\_makatsch alan\_rickman 7  
## 78 laura\_linney alan\_rickman 4  
## 79 emma\_thompson alan\_rickman 9  
## 80 liam\_neeson alan\_rickman 2  
## 81 kris\_marshall alan\_rickman 1  
## 82 abdul\_salis alan\_rickman 1  
## 83 martin\_freeman alan\_rickman 2  
## 84 rowan\_atkinson alan\_rickman 1  
## 85 bill\_nighy heike\_makatsch NA  
## 86 keira\_knightley heike\_makatsch NA  
## 87 andrew\_lincoln heike\_makatsch NA  
## 88 hugh\_grant heike\_makatsch NA  
## 89 colin\_firth heike\_makatsch NA  
## 90 alan\_rickman heike\_makatsch NA  
## 91 heike\_makatsch heike\_makatsch 9  
## 92 laura\_linney heike\_makatsch 3  
## 93 emma\_thompson heike\_makatsch 3  
## 94 liam\_neeson heike\_makatsch 0  
## 95 kris\_marshall heike\_makatsch 1  
## 96 abdul\_salis heike\_makatsch 0  
## 97 martin\_freeman heike\_makatsch 0  
## 98 rowan\_atkinson heike\_makatsch 0  
## 99 bill\_nighy laura\_linney NA  
## 100 keira\_knightley laura\_linney NA  
## 101 andrew\_lincoln laura\_linney NA  
## 102 hugh\_grant laura\_linney NA  
## 103 colin\_firth laura\_linney NA  
## 104 alan\_rickman laura\_linney NA  
## 105 heike\_makatsch laura\_linney NA  
## 106 laura\_linney laura\_linney 11  
## 107 emma\_thompson laura\_linney 2  
## 108 liam\_neeson laura\_linney 1  
## 109 kris\_marshall laura\_linney 1  
## 110 abdul\_salis laura\_linney 1  
## 111 martin\_freeman laura\_linney 1  
## 112 rowan\_atkinson laura\_linney 0  
## 113 bill\_nighy emma\_thompson NA  
## 114 keira\_knightley emma\_thompson NA  
## 115 andrew\_lincoln emma\_thompson NA  
## 116 hugh\_grant emma\_thompson NA  
## 117 colin\_firth emma\_thompson NA  
## 118 alan\_rickman emma\_thompson NA  
## 119 heike\_makatsch emma\_thompson NA  
## 120 laura\_linney emma\_thompson NA  
## 121 emma\_thompson emma\_thompson 12  
## 122 liam\_neeson emma\_thompson 5  
## 123 kris\_marshall emma\_thompson 1  
## 124 abdul\_salis emma\_thompson 1  
## 125 martin\_freeman emma\_thompson 2  
## 126 rowan\_atkinson emma\_thompson 1  
## 127 bill\_nighy liam\_neeson NA  
## 128 keira\_knightley liam\_neeson NA  
## 129 andrew\_lincoln liam\_neeson NA  
## 130 hugh\_grant liam\_neeson NA  
## 131 colin\_firth liam\_neeson NA  
## 132 alan\_rickman liam\_neeson NA  
## 133 heike\_makatsch liam\_neeson NA  
## 134 laura\_linney liam\_neeson NA  
## 135 emma\_thompson liam\_neeson NA  
## 136 liam\_neeson liam\_neeson 13  
## 137 kris\_marshall liam\_neeson 1  
## 138 abdul\_salis liam\_neeson 1  
## 139 martin\_freeman liam\_neeson 2  
## 140 rowan\_atkinson liam\_neeson 1  
## 141 bill\_nighy kris\_marshall NA  
## 142 keira\_knightley kris\_marshall NA  
## 143 andrew\_lincoln kris\_marshall NA  
## 144 hugh\_grant kris\_marshall NA  
## 145 colin\_firth kris\_marshall NA  
## 146 alan\_rickman kris\_marshall NA  
## 147 heike\_makatsch kris\_marshall NA  
## 148 laura\_linney kris\_marshall NA  
## 149 emma\_thompson kris\_marshall NA  
## 150 liam\_neeson kris\_marshall NA  
## 151 kris\_marshall kris\_marshall 7  
## 152 abdul\_salis kris\_marshall 5  
## 153 martin\_freeman kris\_marshall 1  
## 154 rowan\_atkinson kris\_marshall 0  
## 155 bill\_nighy abdul\_salis NA  
## 156 keira\_knightley abdul\_salis NA  
## 157 andrew\_lincoln abdul\_salis NA  
## 158 hugh\_grant abdul\_salis NA  
## 159 colin\_firth abdul\_salis NA  
## 160 alan\_rickman abdul\_salis NA  
## 161 heike\_makatsch abdul\_salis NA  
## 162 laura\_linney abdul\_salis NA  
## 163 emma\_thompson abdul\_salis NA  
## 164 liam\_neeson abdul\_salis NA  
## 165 kris\_marshall abdul\_salis NA  
## 166 abdul\_salis abdul\_salis 7  
## 167 martin\_freeman abdul\_salis 3  
## 168 rowan\_atkinson abdul\_salis 0  
## 169 bill\_nighy martin\_freeman NA  
## 170 keira\_knightley martin\_freeman NA  
## 171 andrew\_lincoln martin\_freeman NA  
## 172 hugh\_grant martin\_freeman NA  
## 173 colin\_firth martin\_freeman NA  
## 174 alan\_rickman martin\_freeman NA  
## 175 heike\_makatsch martin\_freeman NA  
## 176 laura\_linney martin\_freeman NA  
## 177 emma\_thompson martin\_freeman NA  
## 178 liam\_neeson martin\_freeman NA  
## 179 kris\_marshall martin\_freeman NA  
## 180 abdul\_salis martin\_freeman NA  
## 181 martin\_freeman martin\_freeman 7  
## 182 rowan\_atkinson martin\_freeman 0  
## 183 bill\_nighy rowan\_atkinson NA  
## 184 keira\_knightley rowan\_atkinson NA  
## 185 andrew\_lincoln rowan\_atkinson NA  
## 186 hugh\_grant rowan\_atkinson NA  
## 187 colin\_firth rowan\_atkinson NA  
## 188 alan\_rickman rowan\_atkinson NA  
## 189 heike\_makatsch rowan\_atkinson NA  
## 190 laura\_linney rowan\_atkinson NA  
## 191 emma\_thompson rowan\_atkinson NA  
## 192 liam\_neeson rowan\_atkinson NA  
## 193 kris\_marshall rowan\_atkinson NA  
## 194 abdul\_salis rowan\_atkinson NA  
## 195 martin\_freeman rowan\_atkinson NA  
## 196 rowan\_atkinson rowan\_atkinson 2

#Section 4 #Adjacencies Database #This dataframe has 196 and 3columns. The names of the columns and a brief description of each are in the table below:`

column\_names <- c(ls(adjacencies))  
column\_description <- c("Actor 1 on the relationship", "Actor 2 in the relationship", "Index that shows how strong ancd close is the relationship")  
table.df <- data.frame(column\_names, column\_description)  
knitr::kable(table.df, "simple", col.names = c("Column Name", "Column Description"), align = c("c", "c"))

| Column Name | Column Description |
| --- | --- |
| actors | Actor 1 on the relationship |
| Other Actor | Actor 2 in the relationship |
| Relationship Index | Index that shows how strong ancd close is the relationship |
| adjacencies |  |

#Section 5 #Adjacencies Dataset

data\_pick3 <- select(adjacencies, "actors", "Other Actor", "Relationship Index")

Summarytable<-summary(data\_pick3)   
Summarytable

## actors Other Actor Relationship Index  
## Length:196 Length:196 Min. : 0.00   
## Class :character Class :character 1st Qu.: 1.00   
## Mode :character Mode :character Median : 1.00   
## Mean : 2.61   
## 3rd Qu.: 3.00   
## Max. :13.00   
## NA's :91