week_4_assignment

Adejare Windokun

Friday, September 19, 2014

1. Describe your algorithm for deciding how to compare "best popular" movies between years.

Deciding which movie is best popular is difficult as there are various different attributes to use. In addition there are over 50,000 movies in the database. In order to make the number more manageable, I am going to compare these movies in a time series across decades instead.

The best movies are going to be a combination of movies that had very high ratings by the viewers, and also had the top votes. If not, it is quite possible for a dedicated group of people to vote very highly for a movie, and therefore it would rank very high despite the fact that it was not popular.

Pseudocode 1 Divided up all the movies by decade, those before 1960, will be put in a Pre 1960 bin. 2 Extract all movies that had a rating equal to or higher than 98 percentile in their corresponding decade bin 3 Extract all movies that in addition to ranking at the top 2% also ranked at the top 2% for votes 4 Show a plot of the votes, and ratings versus decades 5 Print out a list of those movies

2. Provide R code that supports your conclusions. - see below
3. Use at least one visualization in support of your conclusion from the ggplot2 package. 4 Use at least one function in support of your conclusion from the plyr package - used the dplyr package instead

Deliver your code, document, and results in R Markdown. MSDA

```r
movies <- read.delim("C:/Users/jare/SkyDrive/WorkDocs/CUNY/607/Week 4/movies.tab", header=TRUE, stringsAsFactors = FALSE)
# will take out columns that are not needed
movies = subset(movies, select = -(7:24))
#install.packages("ggplot2") graphic package, already installed
#install.packages("dplyr")
require (dplyr)
```

```
## Loading required package: dplyr
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##     filter, lag
##
```

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

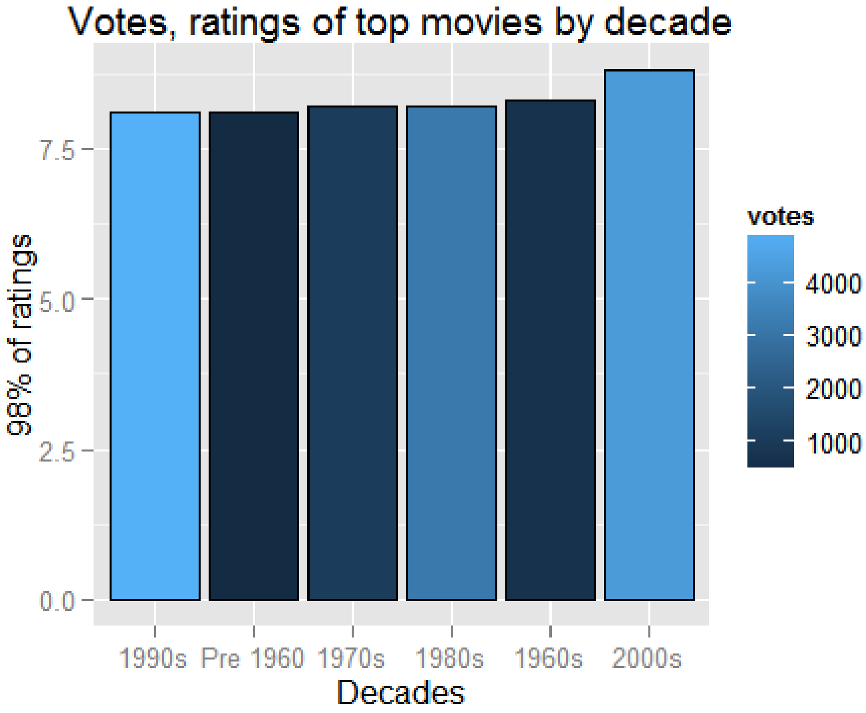
```
require(ggplot2)
```

```
## Loading required package: ggplot2
## 
## Attaching package: 'ggplot2'
## 
## The following object is masked _by_ '.GlobalEnv':
## 
##     movies
```

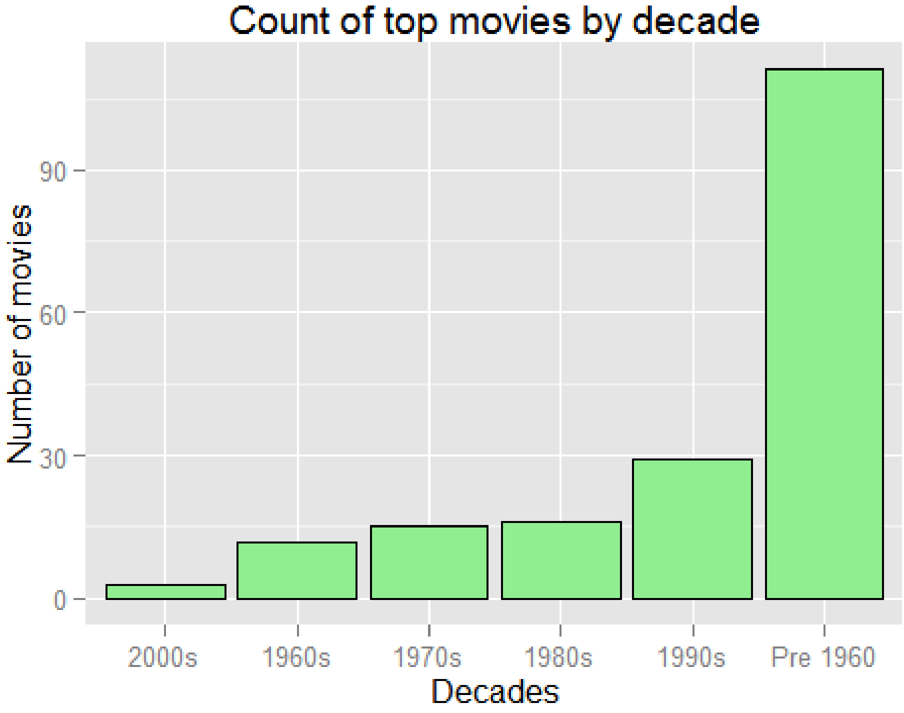
```
# function to convert year to corresponding decade

CreateDecades = function(d){

    if (d >= 2010){

        return("2010s")

    }else if (d >=2000 ) {

        return ("2000s")
    } else if (d >= 1990){

        return ("1990s")

    } else if (d >= 1980){

        return ("1980s")

    }else if (d >=1970 ) {

        return ("1970s")

    } else if (d >= 1960){

        return ("1960s")
    } else {

        return ("Pre 1960")
    }
}

movies = tbl_df(movies)
# will create a column to store the decades
movies$decades = sapply(movies$year, CreateDecades)
```

```r
#convert the newly created column to a factor for use by the aggregation
function
movies$decades = as.factor(movies$decades)

# create a subset of the movies to use with the aggregate function, we are
going to aggregate on rating and votes
submovies = select(movies, decades, rating, votes)

aggsubmovies = aggregate(. ~ decades, data = submovies, FUN = function (x)
c(q = quantile(x, probs = 0.95)))


# will now merge our newly created dataframe with the movie dataframe so that
we can exculde rows we do not want. We will merge on decades which is
contained in both dataframes

newmovies = merge(movies, aggsubmovies, by.x = "decades",by.y =  "decades")


df = filter(newmovies, (votes.x >=  votes.y) & (rating.x >= rating.y))

g = select(df, title, year, decades,rating.x, rating.y, votes.x, votes.y)

#g

summaryg =  g %.%
group_by(decades) %.%
summarize(count = n())
```

Visualization in support of conclusions: 1.



Visualization in support of conclusions: 2.



List of movies per decade that were in the top 2% of both rating and votes:

```
select(g, title, year, decades, rating = rating.x, votes = votes.x)
```

```
##                                                                     title
## 1                                        Buono, il brutto, il cattivo, Il
## 2                                                      Lawrence of Arabia
## 3                                                 C'era una volta il West
## 4                                                   2001: A Space Odyssey
## 5                                                   To Kill a Mockingbird
## 6                                                       Great Escape, The
## 7                                               Manchurian Candidate, The
## 8                                                          Apartment, The
## 9   Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb
## 10                                                                 Psycho
## 11                                                                Yojimbo
## 12                                                Battaglia di Algeri, La
## 13                                                              Chinatown
## 14                                                                   Jaws
## 15                                                                  Alien
## 16                                                    Clockwork Orange, A
## 17                                                         Godfather, The
## 18                                                Godfather: Part II, The
## 19                                        Monty Python and the Holy Grail
## 20                                                            Taxi Driver
## 21                                                              Manhattan
## 22                                                             Annie Hall
## 23                                                         Apocalypse Now
## 24                                                             Sting, The
## 25                                        One Flew Over the Cuckoo's Nest
## 26                                                              Star Wars
## 27                                                                 Patton
## 28                                                                    Ran
## 29                                            Once Upon a Time in America
## 30                                                                 Aliens
## 31                                                           Blade Runner
## 32                                                    Princess Bride, The
## 33                                                         Hotaru no haka
## 34                                                  Nuovo cinema Paradiso
## 35                                                Raiders of the Lost Ark
## 36                                                              Boot, Das
## 37                                                    Fanny och Alexander
## 38                                                                Amadeus
## 39                                                            Raging Bull
## 40                                                           Shining, The
## 41                                                      Full Metal Jacket
## 42                                                      Elephant Man, The
## 43                         Star Wars: Episode V - The Empire Strikes Back
## 44                                                    Straight Story, The
## 45                                                            Toy Story 2
## 46                                                  Trois couleurs: Rouge
## 47                                                             Lola rennt
## 48                                                    American History X
## 49                                                          Forrest Gump
## 50                                            Terminator 2: Judgment Day
## 51                                                   Usual Suspects, The
## 52                                                   Saving Private Ryan
## 53                                                            Fight Club
## 54                                                                 Fargo
## 55                                             Silence of the Lambs, The
## 56                                                         Mononoke-hime
## 57                                                       American Beauty
## 58                                                                Festen
## 59                                                          Pulp Fiction
## 60                                  Wallace & Gromit: The Wrong Trousers
## 61                                                        Reservoir Dogs
## 62                                                            Braveheart
## 63                                                                 Se7en
## 64                                                       Green Mile, The
## 65                                                      Sixth Sense, The
## 66                                                           Matrix, The
## 67                                                            Goodfellas
## 68                                                     L.A. Confidential
## 69                                             Shawshank Redemption, The
## 70                                                            Unforgiven
## 71                                       Wallace & Gromit: A Close Shave
## 72                                                      Schindler's List
## 73                                Lord of the Rings: The Two Towers, The
## 74                    Lord of the Rings: The Fellowship of the Ring, The
## 75                        Lord of the Rings: The Return of the King, The
## 76                                                        Gold Rush, The
## 77                                                     Ballada o soldate
## 78                                          Mr. Smith Goes to Washington
## 79                                                    Great Expectations
## 80                                                    Germania anno zero
## 81                                                 Night at the Opera, A
## 82                                                          12 Angry Men
## 83                                                             Dodsworth
## 84                                               Philadelphia Story, The
## 85                                                        Paths of Glory
## 86                                             Streetcar Named Desire, A
## 87                                                                 Ikiru
## 88                                                   Trouble in Paradise
## 89                                                             Duck Soup
## 90                                                        Music Box, The
## 91                                               Quatre cents coups, Les
## 92                                                        Big Sleep, The
## 93                                                     On the Waterfront
## 94                                        Day the Earth Stood Still, The
## 95                                                               Ben-Hur
## 96                                                                 Laura
## 97                                         Bridge on the River Kwai, The
## 98                                                            Duck Amuck
## 99                                                       Big Parade, The
## 100                                              Meshes of the Afternoon
## 101                                                    Wizard of Oz, The
## 102                                        Sunrise: A Song of Two Humans
## 103                                                   North by Northwest
## 104                                                Bride of Frankenstein
## 105                                                                Faust
## 106                                                   Nuit et brouillard
## 107                                                   African Queen, The
## 108                                                    Lost Weekend, The
## 109                                                                 Cops
## 110                                                          Corbeau, Le
## 111                                                 Grapes of Wrath, The
## 112                                    I Am a Fugitive from a Chain Gang
## 113                                                  Great Dictator, The
## 114                                                        All About Eve
## 115                                                             Midnight
## 116                                                         Safety Last!
## 117                                                            Notorious
## 118                                                           Casablanca
## 119                                                  Maltese Falcon, The
## 120                                                               Harvey
## 121                                                              Vertigo
## 122                                                 Notti di Cabiria, Le
## 123                                          Passion de Jeanne d'Arc, La
## 124                                                          Rear Window
## 125                                                              Rebecca
## 126                                                        Body and Soul
## 127                                                         General, The
## 128                                                         Modern Times
## 129                                             Night of the Hunter, The
## 130                                                         Citizen Kane
## 131                                                          City Lights
## 132                                                                    M
## 133                                                         Sunset Blvd.
## 134                                Nosferatu, eine Symphonie des Grauens
## 135                                                      Brief Encounter
## 136                                                     Bringing Up Baby
## 137                                                   To Be or Not to Be
## 138                                                Bronenosets Potyomkin
## 139                                        Adventures of Robin Hood, The
## 140                                                        Roman Holiday
## 141                                                             Kid, The
## 142                                                              Ukigusa
## 143                                                         Killing, The
## 144                                       All Quiet on the Western Front
## 145                                         Best Years of Our Lives, The
## 146                                                           Strada, La
## 147                                                        Touch of Evil
## 148                                                 It Happened One Night
## 149                                               Enfants du paradis, Les
## 150                                               Voyage dans la lune, Le
## 151                                                         Sherlock, Jr.
## 152                                                     Shadow of a Doubt
## 153                                                       His Girl Friday
## 154                                                  Strangers on a Train
## 155                                  Life and Death of Colonel Blimp, The
## 156                                                            Metropolis
## 157                                                  Shichinin no samurai
## 158                                                    Sullivan's Travels
## 159                                     Kabinett des Doktor Caligari, Das
## 160                                                    Gone with the Wind
## 161                                                       Out of the Past
## 162                                                   Ladri di biciclette
## 163                                                   Grande illusion, La
## 164                                                      Double Indemnity
## 165                                           Witness for the Prosecution
## 166                                                             High Noon
## 167                                              Kind Hearts and Coronets
## 168                                                     Rabbit of Seville
## 169                                                                 Greed
## 170                                                  Sjunde inseglet, Det
## 171                                                      Some Like It Hot
## 172                                                   Singin' in the Rain
## 173                                                   Quai des brumes, Le
## 174                                                        Third Man, The
## 175                                                    What's Opera, Doc?
## 176                                                      Diaboliques, Les
## 177                                                Salaire de la peur, Le
## 178                                                            Umberto D.
## 179                                                         Thin Man, The
## 180                                                  Ox-Bow Incident, The
## 181                                                             Stalag 17
## 182                                                 It's a Wonderful Life
## 183                                                            White Heat
## 184                                           Matter of Life and Death, A
## 185                                     Treasure of the Sierra Madre, The
## 186                                                        Searchers, The
##     year  decades rating  votes
## 1   1966    1960s    8.8  30224
## 2   1962    1960s    8.6  31230
## 3   1968    1960s    8.7  17241
## 4   1968    1960s    8.3  64982
## 5   1962    1960s    8.5  30139
## 6   1963    1960s    8.3  20075
## 7   1962    1960s    8.4  14232
## 8   1960    1960s    8.4  11214
## 9   1964    1960s    8.7  63471
## 10  1960    1960s    8.6  53962
## 11  1961    1960s    8.4   8968
## 12  1965    1960s    8.4   2023
## 13  1974    1970s    8.4  25695
## 14  1975    1970s    8.2  46978
## 15  1979    1970s    8.3  63400
## 16  1971    1970s    8.3  65283
## 17  1972    1970s    9.1 122755
## 18  1974    1970s    8.9  71363
## 19  1975    1970s    8.4  60565
## 20  1976    1970s    8.5  47873
## 21  1979    1970s    8.2  12391
## 22  1977    1970s    8.3  21462
## 23  1979    1970s    8.5  64785
## 24  1973    1970s    8.4  23678
## 25  1975    1970s    8.7  70234
## 26  1977    1970s    8.8 134640
## 27  1970    1970s    8.2  14351
## 28  1985    1980s    8.4  11981
## 29  1984    1980s    8.2  19292
## 30  1986    1980s    8.3  63961
## 31  1982    1980s    8.2  74749
## 32  1987    1980s    8.2  53946
## 33  1988    1980s    8.3   6886
## 34  1989    1980s    8.4  13999
## 35  1981    1980s    8.7  93511
## 36  1981    1980s    8.5  28920
## 37  1982    1980s    8.2   4359
## 38  1984    1980s    8.3  36955
## 39  1980    1980s    8.4  26739
## 40  1980    1980s    8.3  50908
## 41  1987    1980s    8.2  47757
## 42  1980    1980s    8.2  16422
## 43  1980    1980s    8.8 103706
## 44  1999    1990s    8.1  14679
## 45  1999    1990s    8.1  40941
## 46  1994    1990s    8.1  10385
## 47  1998    1990s    8.2  32791
## 48  1998    1990s    8.4  59677
## 49  1994    1990s    8.2  89722
## 50  1991    1990s    8.2  77614
## 51  1995    1990s    8.7 103854
## 52  1998    1990s    8.3 100267
## 53  1999    1990s    8.5 112092
## 54  1996    1990s    8.2  65597
## 55  1991    1990s    8.5  92060
## 56  1997    1990s    8.3  19350
## 57  1999    1990s    8.5 109991
## 58  1998    1990s    8.2  11983
## 59  1994    1990s    8.8 132745
## 60  1993    1990s    8.4  14976
## 61  1992    1990s    8.3  69240
## 62  1995    1990s    8.3  92437
## 63  1995    1990s    8.4  88371
## 64  1999    1990s    8.1  60142
## 65  1999    1990s    8.2  96987
## 66  1999    1990s    8.5 143853
## 67  1990    1990s    8.6  68219
## 68  1997    1990s    8.4  69600
## 69  1994    1990s    9.1 149494
## 70  1992    1990s    8.1  30868
## 71  1995    1990s    8.2   9261
## 72  1993    1990s    8.8  97667
## 73  2002    2000s    8.8 114797
## 74  2001    2000s    8.8 157608
## 75  2003    2000s    9.0 103631
## 76  1925 Pre 1960    8.3   6429
## 77  1959 Pre 1960    8.6    496
## 78  1939 Pre 1960    8.4  11565
## 79  1946 Pre 1960    8.1   2340
## 80  1948 Pre 1960    8.1    516
## 81  1935 Pre 1960    8.1   6008
## 82  1957 Pre 1960    8.7  29278
## 83  1936 Pre 1960    8.2    639
## 84  1940 Pre 1960    8.2  10536
## 85  1957 Pre 1960    8.6  14391
## 86  1951 Pre 1960    8.1   9623
## 87  1952 Pre 1960    8.4   4113
## 88  1932 Pre 1960    8.1   1304
## 89  1933 Pre 1960    8.3   9870
## 90  1932 Pre 1960    8.1    773
## 91  1959 Pre 1960    8.2   6482
## 92  1946 Pre 1960    8.3   9980
## 93  1954 Pre 1960    8.4  12803
## 94  1951 Pre 1960    8.1  10148
## 95  1959 Pre 1960    8.1  21031
## 96  1944 Pre 1960    8.2   4632
## 97  1957 Pre 1960    8.4  22408
## 98  1953 Pre 1960    8.4   1260
## 99  1925 Pre 1960    8.7    435
## 100 1943 Pre 1960    8.1    459
## 101 1939 Pre 1960    8.3  38386
## 102 1927 Pre 1960    8.4   2653
## 103 1959 Pre 1960    8.6  35648
## 104 1935 Pre 1960    8.1   4734
## 105 1926 Pre 1960    8.4    731
## 106 1955 Pre 1960    8.1   1268
## 107 1951 Pre 1960    8.1  13765
## 108 1945 Pre 1960    8.1   2661
## 109 1922 Pre 1960    8.1    542
## 110 1943 Pre 1960    8.3    477
## 111 1940 Pre 1960    8.2   7855
## 112 1932 Pre 1960    8.1   1198
## 113 1940 Pre 1960    8.4  10465
## 114 1950 Pre 1960    8.5  13038
## 115 1939 Pre 1960    8.2    483
## 116 1923 Pre 1960    8.2    757
## 117 1946 Pre 1960    8.3  10637
## 118 1942 Pre 1960    8.8  66030
## 119 1941 Pre 1960    8.4  19444
## 120 1950 Pre 1960    8.1   7879
## 121 1958 Pre 1960    8.5  33875
## 122 1957 Pre 1960    8.1   2286
## 123 1928 Pre 1960    8.4   3466
## 124 1954 Pre 1960    8.7  41035
## 125 1940 Pre 1960    8.3  11403
## 126 1947 Pre 1960    8.1    521
## 127 1927 Pre 1960    8.4   6644
## 128 1936 Pre 1960    8.5  10326
## 129 1955 Pre 1960    8.2   7543
## 130 1941 Pre 1960    8.7  61083
## 131 1931 Pre 1960    8.5   7802
## 132 1931 Pre 1960    8.5  12053
## 133 1950 Pre 1960    8.6  16149
## 134 1922 Pre 1960    8.1   8404
## 135 1945 Pre 1960    8.3   2583
## 136 1938 Pre 1960    8.2   9250
## 137 1942 Pre 1960    8.3   3579
## 138 1925 Pre 1960    8.2   6042
## 139 1938 Pre 1960    8.2   7359
## 140 1953 Pre 1960    8.1   9431
## 141 1921 Pre 1960    8.2   2687
## 142 1959 Pre 1960    8.3    427
## 143 1956 Pre 1960    8.1   6980
## 144 1930 Pre 1960    8.2   6835
## 145 1946 Pre 1960    8.4   6138
## 146 1954 Pre 1960    8.2   5012
## 147 1958 Pre 1960    8.4  11745
## 148 1934 Pre 1960    8.3   7924
## 149 1945 Pre 1960    8.1   3369
## 150 1902 Pre 1960    8.2   1262
## 151 1924 Pre 1960    8.2   1863
## 152 1943 Pre 1960    8.1   5680
## 153 1940 Pre 1960    8.2   6667
## 154 1951 Pre 1960    8.3  10624
## 155 1943 Pre 1960    8.2   1054
## 156 1927 Pre 1960    8.4  11988
## 157 1954 Pre 1960    8.9  32141
## 158 1941 Pre 1960    8.2   2896
## 159 1920 Pre 1960    8.1   3599
## 160 1939 Pre 1960    8.1  28836
## 161 1947 Pre 1960    8.1   2541
## 162 1948 Pre 1960    8.4   7381
## 163 1937 Pre 1960    8.3   5020
## 164 1944 Pre 1960    8.5  11592
## 165 1957 Pre 1960    8.3   5112
## 166 1952 Pre 1960    8.3  12465
## 167 1949 Pre 1960    8.2   3489
## 168 1950 Pre 1960    8.8    484
## 169 1924 Pre 1960    8.3   1539
## 170 1957 Pre 1960    8.4   9622
## 171 1959 Pre 1960    8.4  25369
## 172 1952 Pre 1960    8.5  20585
## 173 1938 Pre 1960    8.5    519
## 174 1949 Pre 1960    8.5  18394
## 175 1957 Pre 1960    8.3   2781
## 176 1955 Pre 1960    8.2   3068
## 177 1953 Pre 1960    8.2   2820
## 178 1952 Pre 1960    8.1   1320
## 179 1934 Pre 1960    8.1   4755
## 180 1943 Pre 1960    8.2   1940
## 181 1953 Pre 1960    8.2   7153
## 182 1946 Pre 1960    8.6  41199
## 183 1949 Pre 1960    8.2   2626
## 184 1946 Pre 1960    8.1   2093
## 185 1948 Pre 1960    8.4  10470
## 186 1956 Pre 1960    8.2  10211
```