Project 7 -- SEYI OGUNMODEDE

Instructor: Dr. Mark Daniel

- The videos posting make me have more interest in the learning being a novice from the scratch.

Collaboration:

Question 1

In [1]:
```
library(data.table)
```

In [2]:
```
titles <- data.frame(fread("/anvil/projects/tdm/data/movies_and_tv/titles.csv"))
```

In [3]:
```
episodes <- data.frame(fread("/anvil/projects/tdm/data/movies_and_tv/episodes.csv"))
```

In [4]:
```
people <- data.frame(fread("/anvil/projects/tdm/data/movies_and_tv/people.csv"))
```

In [5]:
```
ratings <- data.frame(fread("/anvil/projects/tdm/data/movies_and_tv/ratings.csv"))
```

In [6]:
```
options(jupyter.rich_display = F)
```

In [7]:
```
options(repr.matrix.max.cols=50, repr.matrix.max.rows=200)
```

In [8]:
```
head (titles)
# looking up the kind of data frame it is.
```

```
  title_id  type  primary_title          original_title         is_adult
1 tt0000001 short Carmencita             Carmencita             0
2 tt0000002 short Le clown et ses chiens Le clown et ses chiens 0
3 tt0000003 short Pauvre Pierrot         Pauvre Pierrot         0
4 tt0000004 short Un bon bock            Un bon bock            0
5 tt0000005 short Blacksmith Scene       Blacksmith Scene       0
6 tt0000006 short Chinese Opium Den      Chinese Opium Den      0
  premiered ended runtime_minutes genres
1 1894      NA     1              Documentary,Short
2 1892      NA     5              Animation,Short
3 1892      NA     4              Animation,Comedy,Romance
4 1892      NA    12              Animation,Short
5 1893      NA     1              Comedy,Short
6 1894      NA     1              Short
```

In [9]:
```
titles [titles$ title_id == "tt0108778", ]
# to find out the row in the title table that corressponds
# to the whole friends show,look up the title df
# and find out the role of the title id is (equalto the one for friends),
# cwe get the row and all the columns
```

```
       title_id  type     primary_title original_title is_adult premiered ended
106408 tt0108778 tvSeries Friends       Friends        0        1994      2004
       runtime_minutes genres
106408 22              Comedy,Romance
```

In [10]:
```
titles [titles$ title_id == "tt9114286", ]
# personally for black panther movie
```

```
        title_id  type  primary_title
7683192 tt9114286 movie Black Panther: Wakanda Forever
        original_title                is_adult premiered ended runtime_minutes
7683192 Black Panther: Wakanda Forever 0        2022      NA    NA
        genres
7683192 Action,Adventure,Drama
```

In [11]:
```
head (episodes )
```

```
  episode_title_id show_title_id season_number episode_number
1 tt0041951        tt0041038      1             9
2 tt0042816        tt0989125      1             17
3 tt0042889        tt0989125     NA             \\N
4 tt0043426        tt0040051      3             42
5 tt0043631        tt0989125      2             16
6 tt0043693        tt0989125      2             8
```

In [13]:
```
head (people)
```

```
  person_id name            born died
1 nm0000001 Fred Astaire    1899 1987
2 nm0000002 Lauren Bacall   1924 2014
3 nm0000003 Brigitte Bardot 1934   NA
4 nm0000004 John Belushi    1949 1982
5 nm0000005 Ingmar Bergman  1918 2007
6 nm0000006 Ingrid Bergman  1915 1982
```

In [14]:
```
head (ratings)
```

```
  title_id  rating votes
1 tt0000001 5.7    1718
2 tt0000002 6.0     211
3 tt0000003 6.5    1480
4 tt0000004 6.1     124
5 tt0000005 6.2    2283
6 tt0000006 5.1     127
```

In [15]:
```
head(titles$genres)
# the first six category of artistic composition
```

```
[1] "Documentary,Short"        "Animation,Short"
[3] "Animation,Comedy,Romance" "Animation,Short"
[5] "Comedy,Short"             "Short"
```

In [16]:
```
head(titles$genres, n=50)
# the first 50 of them
```

```
 [1] "Documentary,Short"         "Animation,Short"
 [3] "Animation,Comedy,Romance" "Animation,Short"
 [5] "Comedy,Short"              "Short"
 [7] "Short,Sport"               "Documentary,Short"
 [9] "Romance,Short"             "Documentary,Short"
[11] "Documentary,Short"         "Documentary,Short"
[13] "Documentary,Short"         "Comedy,Short"
[15] "Animation,Short"           "Documentary,Short"
[17] "Documentary,Short"         "Short"
[19] "Comedy,Short"              "Documentary,Short,Sport"
[21] "Documentary,Short"         "Documentary,Short"
[23] "News,Short"                "News,Short,Sport"
[25] "Documentary,Short"         "Documentary,Short"
[27] "Documentary,Short"         "Documentary,Short"
[29] "Documentary,Short"         "Documentary,Short"
[31] "Short"                     "Comedy,Documentary,Short"
[33] "Documentary,Short"         "Comedy,Short"
[35] "Drama,Short"               "Short"
[37] "Documentary,Short,Sport"   "Short"
[39] "Documentary,Short"         "Comedy,Documentary,Short"
[41] "Documentary,Short"         "Documentary,Short"
[43] "Short"                     "Short"
[45] "Documentary,Short"         "Documentary,Short"
[47] "Short"                     "Short,Sport"
[49] "Documentary,Short"         "Documentary,Short"
```

In [17]:
```
length(titles$genres)
# there total together with repitition
```

```
[1] 8064259
```

In [18]:
```
length(unique(titles$genres))
# the unique title of the category of artistic composition
```

```
[1] 2283
```

In [19]:
```
head(unique(titles$genres))
# No repitition of titles of the first six category of artistic composition
```

```
[1] "Documentary,Short"         "Animation,Short"
[3] "Animation,Comedy,Romance" "Comedy,Short"
[5] "Short"                     "Short,Sport"
```

In [20]:
```
head(unique(titles$genres), n=50)
# first 50 no repitition like the previous
```

```
 [1] "Documentary,Short"         "Animation,Short"
 [3] "Animation,Comedy,Romance" "Comedy,Short"
 [5] "Short"                     "Short,Sport"
 [7] "Romance,Short"             "Documentary,Short,Sport"
 [9] "News,Short"                "News,Short,Sport"
[11] "Comedy,Documentary,Short" "Drama,Short"
[13] "Fantasy,Short"             "Horror,Short"
[15] "Comedy,Horror,Short"       "Biography,Short"
[17] "Music,Short"               "Documentary,News,Short"
[19] "Fantasy,Horror,Short"      "Short,War"
[21] "Crime,Short"               "Short,Western"
[23] "Comedy,Short,Sport"        "Comedy,Fantasy,Horror"
[25] "Biography,Drama,Short"     "Family,Fantasy,Romance"
[27] "Drama,Short,War"           "Drama,Family,Fantasy"
[29] "Adventure,Fantasy,Horror" "Comedy,Romance,Short"
[31] "Action,Crime,Drama"        "Comedy,Fantasy,Short"
[33] "Animation,Comedy,Fantasy" "Family,Short"
[35] "Drama,History,Short"       "Action,Drama,Short"
[37] "Crime,Drama,Short"         "Fantasy,Romance,Short"
[39] "Drama,Fantasy,Horror"      "Drama,Horror,Short"
[41] "Drama,Fantasy,Short"       "History,Short"
[43] "Action,Adventure,Comedy"  "Family,Fantasy,Short"
[45] "Action,Crime,Short"        "Adventure,Drama,Short"
[47] "Action,Short"              "Comedy,Music,Short"
[49] "Adventure,Fantasy,Short"   "Comedy,Family,Short"
```

In [22]:
```
head(strsplit(unique(titles$genres), ","))
# then you can do string split
# split these unique occurrences of the genres column in the titles table
# split according to a comma
# then the first six
```

```
[[1]]
[1] "Documentary" "Short"

[[2]]
[1] "Animation" "Short"

[[3]]
[1] "Animation" "Comedy"    "Romance"

[[4]]
[1] "Comedy" "Short"

[[5]]
[1] "Short"

[[6]]
[1] "Short" "Sport"
```

In [21]:
```
class(strsplit(unique(titles$genres), ","))
```

```
[1] "list"
```

In [22]:
```
head(unlist(strsplit(unique(titles$genres), ",")))
```

```
[1] "Documentary" "Short"       "Animation"   "Short"       "Animation"
[6] "Comedy"
```

In [23]:
```
head(unlist(strsplit(unique(titles$genres), ",")), n=50)
```

```
 [1] "Documentary" "Short"       "Animation"   "Short"       "Animation"
 [6] "Comedy"      "Romance"     "Comedy"      "Short"       "Short"
[11] "Short"       "Sport"       "Romance"     "Short"       "Documentary"
[16] "Short"       "Sport"       "News"        "Short"       "News"
[21] "Short"       "Sport"       "Comedy"      "Documentary" "Short"
[26] "Drama"       "Short"       "Fantasy"     "Short"       "Horror"
[31] "Short"       "Comedy"      "Horror"      "Short"       "Biography"
[36] "Short"       "Music"       "Short"       "Documentary" "News"
[41] "Short"       "Fantasy"     "Horror"      "Short"       "Short"
[46] "War"         "Crime"       "Short"       "Short"       "Western"
```

In [24]:
```
unique(unlist(strsplit(unique(titles$genres), ",")))
# to check for the unique listed genres in the titles table after unlisted.
```

```
 [1] "Documentary" "Short"       "Animation"   "Comedy"      "Romance"
 [6] "Sport"       "News"        "Drama"       "Fantasy"     "Horror"
[11] "Biography"   "Music"       "War"         "Crime"       "Western"
[16] "Family"      "Adventure"   "Action"      "History"     "Mystery"
[21] "\\N"         "Sci-Fi"      "Musical"     "Thriller"    "Film-Noir"
[26] "Talk-Show"   "Game-Show"   "Reality-TV"  "Adult"
```

In [25]:
```
head(titles$genres)
```

```
[1] "Documentary,Short"        "Animation,Short"
[3] "Animation,Comedy,Romance" "Animation,Short"
[5] "Comedy,Short"             "Short"
```

In [26]:
```
grepl("Comedy",head(titles$genres))
# allow us to extract just the rows in which comedy appear.
```

```
[1] FALSE FALSE  TRUE FALSE  TRUE FALSE
```

In [28]:
```
head(titles[grepl("Comedy", titles$genres), ])
```

```
   title_id  type  primary_title        original_title           is_adult
3  tt0000003 short Pauvre Pierrot       Pauvre Pierrot           0
5  tt0000005 short Blacksmith Scene     Blacksmith Scene         0
14 tt0000014 short The Waterer Watered  L'arroseur arros\303\251 0
19 tt0000019 short The Clown Barber     The Clown Barber         0
32 tt0000033 short Trick Riding         La voltige               0
34 tt0000035 short Watering the Flowers L'arroseur               0
   premiered ended runtime_minutes genres
3  1892      NA     4              Animation,Comedy,Romance
5  1893      NA     1              Comedy,Short
14 1895      NA     1              Comedy,Short
19 1898      NA    NA              Comedy,Short
32 1895      NA     1              Comedy,Documentary,Short
34 1896      NA     1              Comedy,Short
```

In [29]:
```
head(titles[grepl("Comedy", titles$genres), ]$premiered)
# this display their years of premiered
```

```
[1] 1892 1893 1895 1898 1895 1896
```

In [30]:
```
table(titles[grepl("Comedy", titles$genres), ]$premiered)
```

```
 1892  1893  1894  1895  1896  1897  1898  1899  1900  1901  1902  1903  1904
    1     1     5     8    42   112   192   201   239   165   229   429   344
 1905  1906  1907  1908  1909  1910  1911  1912  1913  1914  1915  1916  1917
  366   407   736  1090  1360  1622  1961  2639  3005  3043  2745  2238  1709
 1918  1919  1920  1921  1922  1923  1924  1925  1926  1927  1928  1929  1930
 1232  1102  1325   997   912   766   801   943   993  1036   917   876   808
 1931  1932  1933  1934  1935  1936  1937  1938  1939  1940  1941  1942  1943
  845   739   773   797   766   852   871   883   716   663   691   635   561
 1944  1945  1946  1947  1948  1949  1950  1951  1952  1953  1954  1955  1956
  466   420   566   653   714  1051  1510  1882  1883  2412  2959  3321  3083
 1957  1958  1959  1960  1961  1962  1963  1964  1965  1966  1967  1968  1969
 3074  3196  3204  3680  3953  3957  4337  4356  4963  4777  5111  4895  5572
 1970  1971  1972  1973  1974  1975  1976  1977  1978  1979  1980  1981  1982
 5543  5752  6035  6748  6156  6245  6043  5936  5402  5408  6173  5521  6533
 1983  1984  1985  1986  1987  1988  1989  1990  1991  1992  1993  1994  1995
 6115  6017  6070  6858  8151  7597  9180  8903 11346 10933 11570 12423 14673
 1996  1997  1998  1999  2000  2001  2002  2003  2004  2005  2006  2007  2008
15480 17180 18826 18500 18195 19903 21090 23804 26087 28952 34078 37103 39734
 2009  2010  2011  2012  2013  2014  2015  2016  2017  2018  2019  2020  2021
43095 48825 57892 65305 71963 77034 80199 84227 83916 78393 70457 61237 28693
 2022  2023  2024  2025  2026
  753    39    11     1     3
```

In [31]:
```
sort(table(titles[grepl("Comedy", titles$genres), ]
           $premiered), decreasing = TRUE)
```

```
 2016  2017  2015  2018  2014  2013  2019  2012  2020  2011  2010  2009  2008
84227 83916 80199 78393 77034 71963 70457 65305 61237 57892 48825 43095 39734
 2007  2006  2005  2021  2004  2003  2002  2001  1998  1999  2000  1997  1996
37103 34078 28952 28693 26087 23804 21090 19903 18826 18500 18195 17180 15480
 1995  1994  1993  1991  1992  1989  1990  1987  1988  1986  1973  1982  1975
14673 12423 11570 11346 10933  9180  8903  8151  7597  6858  6748  6533  6245
 1980  1974  1983  1985  1976  1972  1984  1977  1971  1969  1970  1981  1979
 6173  6156  6115  6070  6043  6035  6017  5936  5752  5572  5543  5521  5408
 1978  1967  1965  1968  1966  1964  1963  1962  1961  1960  1955  1959  1958
 5402  5111  4963  4895  4777  4356  4337  3957  3953  3680  3321  3204  3196
 1956  1957  1914  1913  1954  1915  1912  1953  1916  1911  1952  1951  1917
 3083  3074  3043  3005  2959  2745  2639  2412  2238  1961  1883  1882  1709
 1910  1950  1909  1920  1918  1919  1908  1949  1927  1921  1926  1925  1928
 1622  1510  1360  1325  1232  1102  1090  1051  1036   997   993   943   917
 1922  1938  1929  1937  1936  1931  1930  1924  1934  1933  1923  1935  2022
  912   883   876   871   852   845   808   801   797   773   766   766   753
 1932  1907  1939  1948  1941  1940  1947  1942  1946  1943  1944  1903  1945
  739   736   716   714   691   663   653   635   566   561   466   429   420
 1906  1905  1904  1900  1902  1899  1898  1901  1897  1896  2023  2024  1895
  407   366   344   239   229   201   192   165   112    42    39    11     8
 1894  2026  1892  1893  2025
    5     3     1     1     1
```

What are all the different listed genres (in the titles table)? There are 29 of them, though one appears to be erroneous among them.

[1] "Documentary" "Short" "Animation" "Comedy" "Romance"
[6] "Sport" "News" "Drama" "Fantasy" "Horror"
[11] "Biography" "Music" "War" "Crime" "Western"
[16] "Family" "Adventure" "Action" "History" "Mystery"
[21] "\N" "Sci-Fi" "Musical" "Thriller" "Film-Noir"
[26] "Talk-Show" "Game-Show" "Reality-TV" "Adult"

Look at the years column and the genres column. In which year did the most comedies debut?
year 2016

2016 2017 2015 2018 2014 2013 2019 2012 2020 2011 2010 2009 2008 84227 83916 80199 78393 77034 71963 70457 65305 61237 57892 48825 43095 39734

Question 2

In [32]:
```
outcomeDF <- merge(episodes, titles,
                   by.x = "episode_title_id", by.y = "title_id")
```

In [33]:
```
head(outcomeDF)
```

```
  episode_title_id show_title_id season_number episode_number type
1 tt0041951        tt0041038      1             9              tvEpisode
2 tt0042816        tt0989125      1             17             tvEpisode
3 tt0042889        tt0989125     NA             \\N            tvEpisode
4 tt0043426        tt0040051      3             42             tvEpisode
5 tt0043631        tt0989125      2             16             tvEpisode
6 tt0043693        tt0989125      2             8              tvEpisode
  primary_title                      original_title                    is_adult
1 The Tenderfeet                     The Tenderfeet                    0
2 Othello                            Othello                           0
3 The Tragedy of King Richard II/II The Tragedy of King Richard II/II 0
4 Coriolanus                         Coriolanus                        0
5 The Life of King Henry V           The Life of King Henry V          0
6 Julius Caesar                      Julius Caesar                     0
  premiered ended runtime_minutes genres
1 1949      NA     30             Western
2 1950      NA    135             Drama
3 1950      NA    145             Drama
4 1951      NA     60             Drama
5 1951      NA    133             Drama
6 1951      NA     NA             Drama
```

In [34]:
```
outcomeDF[outcomeDF$show_title_id == "tt9114286", ]
# for black panther
```

```
    episode_title_id show_title_id season_number episode_number type
    primary_title original_title is_adult premiered ended runtime_minutes
    genres
```

In [35]:
```
dim(outcomeDF[outcomeDF$show_title_id == "tt9114286", ])
# for black panther
```

```
[1]  0 12
```

In [36]:
```
dim(outcomeDF[outcomeDF$show_title_id == "tt0108778", ])
# for Freinds shows
```

```
[1] 235  12
```

Use the episode_title_id column and the title_id column from the episodes and titles data.frame's (respectively) to merge the two data.frames.

outcomeDF <- merge(episodes, titles, by.x = "episode_title_id", by.y = "title_id")

Question 3

In [37]:
```
expectDF <- outcomeDF[outcomeDF$show_title_id == "tt0108778", ]
# daeling with only show called Friends
```

In [38]:
```
dim (expectDF)
```

```
[1] 235  12
```

In [39]:
```
head(expectDF)
```

```
      episode_title_id show_title_id season_number episode_number type
93207 tt0583431        tt0108778      7             8              tvEpisode
93208 tt0583432        tt0108778     10             9              tvEpisode
93209 tt0583433        tt0108778     10             17             tvEpisode
93210 tt0583435        tt0108778      8             1              tvEpisode
93211 tt0583436        tt0108778     10             1              tvEpisode
93212 tt0583437        tt0108778      5             1              tvEpisode
      primary_title
93207 The One Where Chandler Doesn't Like Dogs
93208 The One with the Birth Mother
93209 The Last One
93210 The One After I Do
93211 The One After Joey and Rachel Kiss
93212 The One After Ross Says Rachel
      original_title                          is_adult premiered ended
93207 The One Where Chandler Doesn't Like Dogs 0        2000      NA
93208 The One with the Birth Mother            0        2004      NA
93209 The Last One                             0        2004      NA
93210 The One After I Do                       0        2001      NA
93211 The One After Joey and Rachel Kiss       0        2003      NA
93212 The One After Ross Says Rachel           0        1998      NA
      runtime_minutes genres
93207 22              Comedy,Romance
93208 22              Comedy,Romance
93209 30              Comedy,Romance
93210 22              Comedy,Romance
93211 22              Comedy,Romance
93212 22              Comedy,Romance
```

In [ ]:
```
# no title of Friends anywhere.
```

In [40]:
```
head(merge(expectDF, titles, by.x = "show_title_id",
           by.y = "title_id"))

# it shows the title_id and Primary titles, Friends
```

```
  show_title_id episode_title_id season_number episode_number type.x
1 tt0108778     tt0583431         7             8              tvEpisode
2 tt0108778     tt0583432        10             9              tvEpisode
3 tt0108778     tt0583433        10             17             tvEpisode
4 tt0108778     tt0583435         8             1              tvEpisode
5 tt0108778     tt0583436        10             1              tvEpisode
6 tt0108778     tt0583437         5             1              tvEpisode
  primary_title.x
1 The One Where Chandler Doesn't Like Dogs
2 The One with the Birth Mother
3 The Last One
4 The One After I Do
5 The One After Joey and Rachel Kiss
6 The One After Ross Says Rachel
  original_title.x                          is_adult.x premiered.x ended.x
1 The One Where Chandler Doesn't Like Dogs  0          2000        NA
2 The One with the Birth Mother             0          2004        NA
3 The Last One                              0          2004        NA
4 The One After I Do                        0          2001        NA
5 The One After Joey and Rachel Kiss        0          2003        NA
6 The One After Ross Says Rachel            0          1998        NA
  runtime_minutes.x genres.x       type.y   primary_title.y original_title.y
1 22                Comedy,Romance tvSeries Friends         Friends
2 22                Comedy,Romance tvSeries Friends         Friends
3 30                Comedy,Romance tvSeries Friends         Friends
4 22                Comedy,Romance tvSeries Friends         Friends
5 22                Comedy,Romance tvSeries Friends         Friends
6 22                Comedy,Romance tvSeries Friends         Friends
  is_adult.y premiered.y ended.y runtime_minutes.y genres.y
1 0          1994        2004    22                Comedy,Romance
2 0          1994        2004    22                Comedy,Romance
3 0          1994        2004    22                Comedy,Romance
4 0          1994        2004    22                Comedy,Romance
5 0          1994        2004    22                Comedy,Romance
6 0          1994        2004    22                Comedy,Romance
```

In [41]:
```
deepDF <- merge(expectDF, titles, by.x = "show_title_id",
                by.y = "title_id")
```

In [42]:
```
head(merge(deepDF, ratings, by.x = "episode_title_id",
           by.y = "title_id"))
```

```
  episode_title_id show_title_id season_number episode_number type.x
1 tt0583431        tt0108778      7             8              tvEpisode
2 tt0583432        tt0108778     10             9              tvEpisode
3 tt0583433        tt0108778     10             17             tvEpisode
4 tt0583435        tt0108778      8             1              tvEpisode
5 tt0583436        tt0108778     10             1              tvEpisode
6 tt0583437        tt0108778      5             1              tvEpisode
  primary_title.x
1 The One Where Chandler Doesn't Like Dogs
2 The One with the Birth Mother
3 The Last One
4 The One After I Do
5 The One After Joey and Rachel Kiss
6 The One After Ross Says Rachel
  original_title.x                          is_adult.x premiered.x ended.x
1 The One Where Chandler Doesn't Like Dogs  0          2000        NA
2 The One with the Birth Mother             0          2004        NA
3 The Last One                              0          2004        NA
4 The One After I Do                        0          2001        NA
5 The One After Joey and Rachel Kiss        0          2003        NA
6 The One After Ross Says Rachel            0          1998        NA
  runtime_minutes.x genres.x       type.y   primary_title.y original_title.y
1 22                Comedy,Romance tvSeries Friends         Friends
2 22                Comedy,Romance tvSeries Friends         Friends
3 30                Comedy,Romance tvSeries Friends         Friends
4 22                Comedy,Romance tvSeries Friends         Friends
5 22                Comedy,Romance tvSeries Friends         Friends
6 22                Comedy,Romance tvSeries Friends         Friends
  is_adult.y premiered.y ended.y runtime_minutes.y genres.y       rating votes
1 0          1994        2004    22                Comedy,Romance 8.2     3380
2 0          1994        2004    22                Comedy,Romance 8.6     3489
3 0          1994        2004    22                Comedy,Romance 9.7    13064
4 0          1994        2004    22                Comedy,Romance 8.7     3679
5 0          1994        2004    22                Comedy,Romance 8.5     3784
6 0          1994        2004    22                Comedy,Romance 8.9     4354
```

In [43]:
```
newDF <- merge(deepDF, ratings, by.x = "episode_title_id",
               by.y = "title_id")
```

In [45]:
```
head(merge(newDF, ratings, by.x = "show_title_id",
           by.y = "title_id"))
# this gave ratings of episodes and of friends together.
```

```
  show_title_id episode_title_id season_number episode_number type.x
1 tt0108778     tt0583431         7             8              tvEpisode
2 tt0108778     tt0583432        10             9              tvEpisode
3 tt0108778     tt0583433        10             17             tvEpisode
4 tt0108778     tt0583435         8             1              tvEpisode
5 tt0108778     tt0583436        10             1              tvEpisode
6 tt0108778     tt0583437         5             1              tvEpisode
  primary_title.x
1 The One Where Chandler Doesn't Like Dogs
2 The One with the Birth Mother
3 The Last One
4 The One After I Do
5 The One After Joey and Rachel Kiss
6 The One After Ross Says Rachel
  original_title.x                          is_adult.x premiered.x ended.x
1 The One Where Chandler Doesn't Like Dogs  0          2000        NA
2 The One with the Birth Mother             0          2004        NA
3 The Last One                              0          2004        NA
4 The One After I Do                        0          2001        NA
5 The One After Joey and Rachel Kiss        0          2003        NA
6 The One After Ross Says Rachel            0          1998        NA
  runtime_minutes.x genres.x       type.y   primary_title.y original_title.y
1 22                Comedy,Romance tvSeries Friends         Friends
2 22                Comedy,Romance tvSeries Friends         Friends
3 30                Comedy,Romance tvSeries Friends         Friends
4 22                Comedy,Romance tvSeries Friends         Friends
5 22                Comedy,Romance tvSeries Friends         Friends
6 22                Comedy,Romance tvSeries Friends         Friends
  is_adult.y premiered.y ended.y runtime_minutes.y genres.y       rating.x
1 0          1994        2004    22                Comedy,Romance 8.2
2 0          1994        2004    22                Comedy,Romance 8.6
3 0          1994        2004    22                Comedy,Romance 9.7
4 0          1994        2004    22                Comedy,Romance 8.7
5 0          1994        2004    22                Comedy,Romance 8.5
6 0          1994        2004    22                Comedy,Romance 8.9
  votes.x rating.y votes.y
1  3380   8.9      869810
2  3489   8.9      869810
3 13064   8.9      869810
4  3679   8.9      869810
5  3784   8.9      869810
6  4354   8.9      869810
```

In [44]:

```
head(merge(newDF, ratings, by.x = "show_title_id",
           by.y = "title_id")[ ,c("primary_title.x",
                                  "primary_title.y",
                                  "rating.x", "rating.y")])
```

```
  primary_title.x                           primary_title.y rating.x rating.y
1 The One Where Chandler Doesn't Like Dogs  Friends         8.2      8.9
2 The One with the Birth Mother             Friends         8.6      8.9
3 The Last One                              Friends         9.7      8.9
4 The One After I Do                        Friends         8.7      8.9
5 The One After Joey and Rachel Kiss        Friends         8.5      8.9
6 The One After Ross Says Rachel            Friends         8.9      8.9
```

In [45]:
```
friendsDF <- merge(newDF, ratings, by.x = "show_title_id",
                   by.y = "title_id")[ ,c("primary_title.x",
                                          "primary_title.y", "rating.x",
                                          "rating.y")]
```

In [46]:
```
dim(friendsDF)
```

```
[1] 235   4
```

In [47]:
```
head(friendsDF)
```

```
  primary_title.x                          primary_title.y rating.x rating.y
1 The One Where Chandler Doesn't Like Dogs Friends         8.2      8.9
2 The One with the Birth Mother            Friends         8.6      8.9
3 The Last One                             Friends         9.7      8.9
4 The One After I Do                       Friends         8.7      8.9
5 The One After Joey and Rachel Kiss       Friends         8.5      8.9
6 The One After Ross Says Rachel           Friends         8.9      8.9
```

In [48]:
```
names(friendsDF) <- c("episode_title", "show_title",
                      "episode_rating", "show_rating")
```

In [49]:
```
head(friendsDF)
```

```
  episode_title                            show_title episode_rating
1 The One Where Chandler Doesn't Like Dogs Friends    8.2
2 The One with the Birth Mother            Friends    8.6
3 The Last One                             Friends    9.7
4 The One After I Do                       Friends    8.7
5 The One After Joey and Rachel Kiss       Friends    8.5
6 The One After Ross Says Rachel           Friends    8.9
  show_rating
1 8.9
2 8.9
3 8.9
4 8.9
5 8.9
6 8.9
```

In [50]:
```
head(friendsDF$episode_rating)
```

```
[1] 8.2 8.6 9.7 8.7 8.5 8.9
```

In [51]:
```
order(friendsDF$episode_rating)
# this get the analogous rows in the order
```

```
  [1] 185  94 229 231  85 105 164 169 170  44  59 104 106 107 215  14  39 100
 [19] 132 134 158 161 178 186 188 196 197 214 227  29  31  34  43  62  72  87
 [37]  88  90  97 103 111 115 116 135 139 151 152 155 163 168 184 191 212 218
 [55] 220 221 224   1  30  36  46  52  61  68  75  81  86  91  92 101 114 119
 [73] 121 122 124 127 140 143 147 148 150 159 165 171 172 200 210 211 213 216
 [91]  17  22  27  28  42  49  55  64  84  93 113 145 153 154 156 181 183 194
[109] 198 204 206  16  37  47  83  89  95  96 112 176 182 201 230   5  10  11
[127]  15  18  20  23  24  26  56  66  71  79  80 102 108 118 125 126 129 136
[145] 138 142 162 166 177 179 180 192 195 226 228   2   8  12  19  25  32  54
[163]  58  60  67  73 110 141 160 175 193 207 208   4   7  13  45  48  74 123
[181] 128 133 144 149 157 189 217 222   9  40  53  65  69  77 120 131 203 232
[199]   6  38  41  57  98 109 117 190 199  33  35  50  63  76 146 174 223  70
[217]  78  82 130 137 173 187 205 233  51  99 219 234 209 225 235 202 167   3
[235]  21
```

In [52]:
```
tail(friendsDF[order(friendsDF$episode_rating), ], n=7)
# it gives all of the columns organized
# look for the tail and the , last 7
# these are the highest rated episodes of friends
# across all ten seasons of friends
```

```
    episode_title                     show_title episode_rating show_rating
209 The One with the Rumor            Friends    9.3            8.9
225 The One with the Videotape        Friends    9.3            8.9
235 The One with the Proposal         Friends    9.3            8.9
202 The One with the Prom Video       Friends    9.4            8.9
167 The One with the Embryos          Friends    9.5            8.9
3   The Last One                      Friends    9.7            8.9
21  The One Where Everybody Finds Out Friends    9.7            8.9
```

Show the top 5 rows of your final data.frame that contain the top 5 rated episodes. episode_title show_title episode_rating show_rating 209 The One with the Rumor Friends 9.3 8.9

225 The One with the Videotape Friends 9.3 8.9

235 The One with the Proposal Friends 9.3 8.9

202 The One with the Prom Video Friends 9.4 8.9

167 The One with the Embryos Friends 9.5 8.9

3 The Last One Friends 9.7 8.9

21 The One Where Everybody Finds Out Friends 9.7 8.9

Question 4

In [53]:
```
outcomeDF <- merge(episodes, titles, by.x = "episode_title_id",
                   by.y = "title_id")
```

In [54]:
```
expectDF <- outcomeDF[outcomeDF$show_title_id == "tt0108778", ]
```

In [55]:
```
deepDF <- merge(expectDF, titles, by.x = "show_title_id", by.y = "title_id")
```

In [56]:
```
newDF <- merge(deepDF, ratings, by.x = "episode_title_id", by.y = "title_id")
```

In [57]:
```
friendsDF <- merge(newDF, ratings, by.x = "show_title_id", by.y = "title_id")
```

In [58]:
```
friendsDF <- merge(newDF, ratings, by.x = "show_title_id", by.y = "title_id")
[ ,c("primary_title.x", "primary_title.y", "rating.x",
     "rating.y", "season_number")]
```

In [59]:
```
names(friendsDF) <- c("episode_title", "show_title", "episode_rating",
                      "show_rating", "season_number")
```

In [60]:
```
head(friendsDF)
```

```
  episode_title                             show_title episode_rating
1 The One Where Chandler Doesn't Like Dogs Friends    8.2
2 The One with the Birth Mother           Friends    8.6
3 The Last One                            Friends    9.7
4 The One After I Do                      Friends    8.7
5 The One After Joey and Rachel Kiss      Friends    8.5
6 The One After Ross Says Rachel          Friends    8.9
  show_rating season_number
1 8.9          7
2 8.9         10
3 8.9         10
4 8.9          8
5 8.9         10
6 8.9          5
```

In [61]:
```
friendsDF[ (friendsDF$episode_rating>9)& (friendsDF$season_number ==5), ]
```

```
    episode_title                      show_title episode_rating show_rating
21  The One Where Everybody Finds Out  Friends    9.7            8.9
78  The One with All the Resolutions   Friends    9.1            8.9
130 The One with Ross's Sandwich       Friends    9.1            8.9
219 The One with All the Thanksgivings Friends    9.2            8.9
233 The One in Vegas: Part 2           Friends    9.1            8.9
    season_number
21  5
78  5
130 5
219 5
233 5
```

In [62]:
```
subset( friendsDF, (episode_rating>9)& (season_number ==5))
```

```
    episode_title                      show_title episode_rating show_rating
21  The One Where Everybody Finds Out  Friends    9.7            8.9
78  The One with All the Resolutions   Friends    9.1            8.9
130 The One with Ross's Sandwich       Friends    9.1            8.9
219 The One with All the Thanksgivings Friends    9.2            8.9
233 The One in Vegas: Part 2           Friends    9.1            8.9
    season_number
21  5
78  5
130 5
219 5
233 5
```

Use regular old indexing to find all episodes of friends with an episode_rating greater than 9 and season_number of exactly 5.

Repeat the process, but this time use the subset function instead.

Question 5

In [70]:
```
season_number = 5
```

In [71]:
```
friendsDF[friendsDF$episode_rating>9 &
          friendsDF$season_number ==season_number, ]
```

```
    episode_title                     show_title episode_rating show_rating
21  The One Where Everybody Finds Out  Friends    9.7            8.9
78  The One with All the Resolutions   Friends    9.1            8.9
130 The One with Ross's Sandwich       Friends    9.1            8.9
219 The One with All the Thanksgivings Friends    9.2            8.9
233 The One in Vegas: Part 2           Friends    9.1            8.9
    season_number
21  5
78  5
130 5
219 5
233 5
```

In [72]:
```
subset(friendsDF, episode_rating> 9 & season_number == season_number)
```

```
    episode_title                                     show_title
3   The Last One                                      Friends
21  The One Where Everybody Finds Out                 Friends
51  The One Where Ross Got High                       Friends
70  The One with the Morning After                    Friends
78  The One with All the Resolutions                  Friends
82  The One with Chandler in a Box                    Friends
99  The One with Monica and Chandler's Wedding: Part 2 Friends
130 The One with Ross's Sandwich                      Friends
137 The One with Unagi                                Friends
167 The One with the Embryos                          Friends
173 The One with the Flashback                        Friends
187 The One with the Jellyfish                        Friends
202 The One with the Prom Video                       Friends
205 The One with the Red Sweater                      Friends
209 The One with the Rumor                            Friends
219 The One with All the Thanksgivings                Friends
225 The One with the Videotape                        Friends
233 The One in Vegas: Part 2                          Friends
234 The One with Ross's Wedding: Part 2               Friends
235 The One with the Proposal                         Friends
    episode_rating show_rating season_number
3   9.7            8.9         10
21  9.7            8.9          5
51  9.2            8.9          6
70  9.1            8.9          3
78  9.1            8.9          5
82  9.1            8.9          4
99  9.2            8.9          7
130 9.1            8.9          5
137 9.1            8.9          6
167 9.5            8.9          4
173 9.1            8.9          3
187 9.1            8.9          4
202 9.4            8.9          2
205 9.1            8.9          8
209 9.3            8.9          8
219 9.2            8.9          5
225 9.3            8.9          8
233 9.1            8.9          5
234 9.2            8.9          4
235 9.3            8.9          6
```

In [73]:
```
season_number = 6
```

In [74]:
```
friendsDF[friendsDF$episode_rating>9 &
          friendsDF$season_number ==season_number, ]
```

```
    episode_title               show_title episode_rating show_rating
51  The One Where Ross Got High Friends    9.2            8.9
137 The One with Unagi          Friends    9.1            8.9
235 The One with the Proposal   Friends    9.3            8.9
    season_number
51  6
137 6
235 6
```

In [75]:
```
subset(friendsDF, episode_rating> 9 &
       season_number == season_number)
```

```
    episode_title                                       show_title
3   The Last One                                        Friends
21  The One Where Everybody Finds Out                   Friends
51  The One Where Ross Got High                         Friends
70  The One with the Morning After                      Friends
78  The One with All the Resolutions                    Friends
82  The One with Chandler in a Box                      Friends
99  The One with Monica and Chandler's Wedding: Part 2 Friends
130 The One with Ross's Sandwich                        Friends
137 The One with Unagi                                  Friends
167 The One with the Embryos                            Friends
173 The One with the Flashback                          Friends
187 The One with the Jellyfish                          Friends
202 The One with the Prom Video                         Friends
205 The One with the Red Sweater                        Friends
209 The One with the Rumor                              Friends
219 The One with All the Thanksgivings                  Friends
225 The One with the Videotape                          Friends
233 The One in Vegas: Part 2                            Friends
234 The One with Ross's Wedding: Part 2                 Friends
235 The One with the Proposal                           Friends
    episode_rating show_rating season_number
3   9.7            8.9         10
21  9.7            8.9          5
51  9.2            8.9          6
70  9.1            8.9          3
78  9.1            8.9          5
82  9.1            8.9          4
99  9.2            8.9          7
130 9.1            8.9          5
137 9.1            8.9          6
167 9.5            8.9          4
173 9.1            8.9          3
187 9.1            8.9          4
202 9.4            8.9          2
205 9.1            8.9          8
209 9.3            8.9          8
219 9.2            8.9          5
225 9.3            8.9          8
233 9.1            8.9          5
234 9.2            8.9          4
235 9.3            8.9          6
```

Read that provided article and do your best to explain why subset gets a different result than our example that uses regular indexing.

By defaults subset removes NA rows

Subset is useful interactive shortcut for subsetting data frame,

instead of repeating the names of dataframe many times, you save some typing.

majorly it avoid repitition of output that are thesame.

Pledge

By submitting this work I hereby pledge that this is my own, personal work. I've acknowledged in the designated place at the top of this file all sources that I used to complete said work, including but not limited to: online resources, books, and electronic communications. I've noted all collaboration with fellow students and/or TA's. I did not copy or plagiarize another's work.

> As a Boilermaker pursuing academic excellence, I pledge to be honest and true in all that I do. Accountable together – We are Purdue.