will-v-denzel raw

Chris Sarno

12/10/2020

```r
library(stringr)
library(rvest)
```

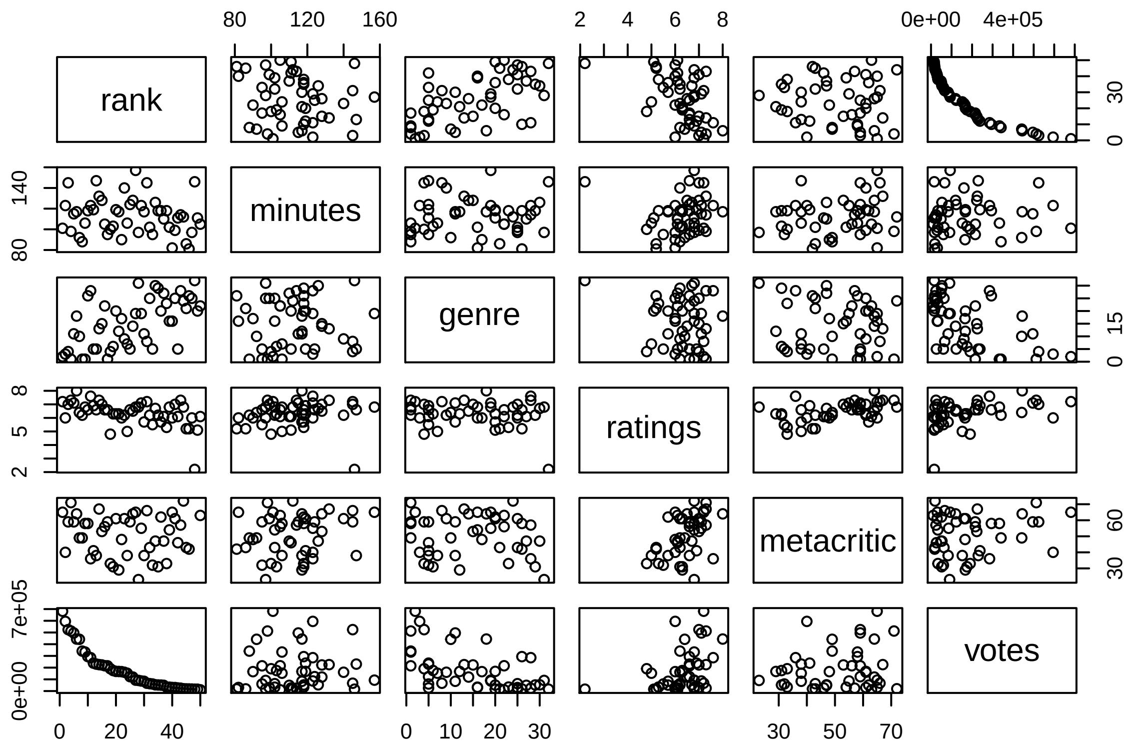
```
## Loading required package: xml2
```

```r
grabFilmInfoFromFilmsPage = function(page)
    {
    # 50 elements
    # # title = id = rank = year = rating = minutes = genre = votes = metascore = desc = millions

    movies = page %>%
        html_nodes(".mode-detail");



    pagecount = length(movies);

    result = data.frame(            matrix(ncol = 11,nrow = pagecount) );
    # a matrix-type form with lots of NA values ...

    colnames(result) = c("rank", "title", "ttid", "year", "rated", "minutes", "genre", "ratings", "meta


    for(i in 1:pagecount)
        {
        movie = movies[i];

        rank = movie %>%
            html_node(".lister-item-index") %>%
            html_text() %>%
            as.numeric();
        result$rank[i] = rank;

        title = movie %>%
            html_node(".lister-item-header a") %>%
            html_text();
        result$title[i] = title;

        ttid = movie %>%
            html_node(".lister-item-header a") %>%
            html_attr("href");

            temp = strsplit(ttid,"/",fixed=T);
```

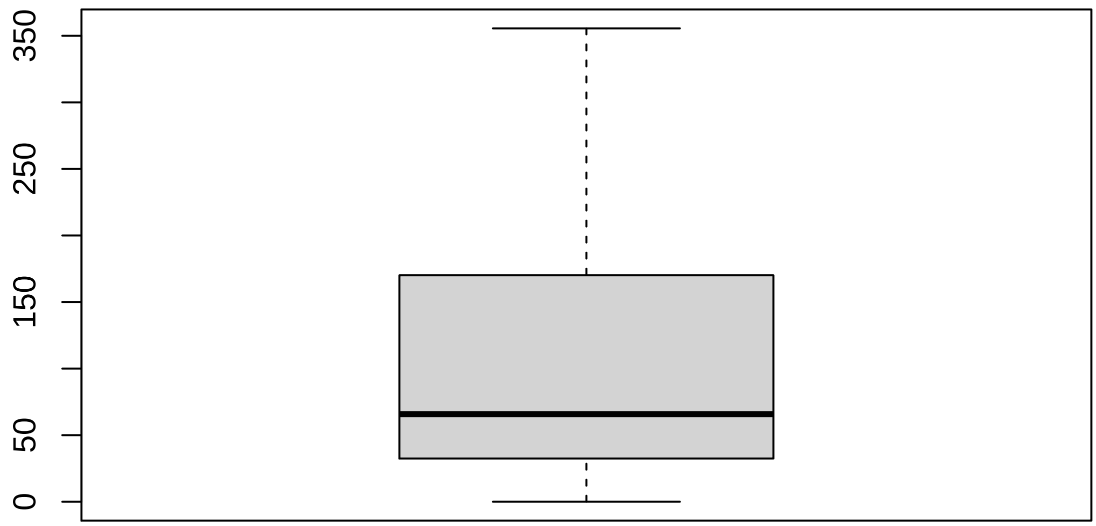
```
    ttid = temp[[1]][3];
    result$ttid[i] = ttid;

    year = movie %>%
        html_node(".lister-item-year") %>%
        html_text();
    year = cleanupYear(year);
    result$year[i] = year;

    rated = movie %>%
        html_node(".certificate") %>%
        html_text();
    result$rated[i] = rated;

    minutes = movie %>%
        html_node(".runtime") %>%
        html_text();
    minutes = cleanupMinutes(minutes);
    result$minutes[i] = minutes;

    genre = movie %>%
        html_node(".genre") %>%
        html_text();
    genre = str_trim(genre);
    result$genre[i] = genre;

    ratings = movie %>%
        html_node("div .rating-list") %>%
        html_attr("title");
            temp = strsplit(ratings,"/",fixed=T);
            temp = gsub("Users rated this","",temp[[1]][1],fixed=T);
            temp = str_trim(temp);
    ratings = as.numeric(temp);
    result$ratings[i] = ratings;

    metacritic = movie %>%
        html_node(".ratings-metascore span") %>%
        html_text();
    metacritic = as.numeric(str_trim(metacritic));
    result$metacritic[i] = metacritic;

    info = movie %>%
        html_nodes(".lister-item-content p span") %>%
        html_text();

    votes = as.numeric(gsub(",","",info[8],fixed=T));
    result$votes[i] = votes;

    millions = cleanupMillions(info[11]);
    result$millions[i] = millions;
    }

result;
```

```
}

cleanupMillions = function(millions)
    {
    millions = gsub('$','',millions, fixed=T);
    millions = gsub('M','',millions, fixed=T);

    millions = as.numeric(millions);
    millions;
    }

cleanupMinutes = function(minutes)
    {
    minutes = gsub('min','',minutes, fixed=T);

    minutes = as.numeric(minutes);
    minutes;
    }

cleanupYear = function(year)
    {
    year = gsub('(','',year, fixed=T);
    year = gsub(')','',year, fixed=T);
    year = gsub('I','',year, fixed=T);
    year = as.numeric(year);
    year;
    }

grabNameFromFilmsPage = function(page)
    {
    name = page %>%
        html_node(".header") %>%
        html_text();

        name = gsub("Most Rated Feature Films With","",name,fixed=T);
        name = str_trim(name);

    name;
    }


grabFilmCountFromFilmsPage = function(page)
    {
    totalcount = page %>%
        html_nodes(".desc") %>%
        html_text();

        temp = strsplit(totalcount,"of",fixed=T);
        temp2 = strsplit(temp[[1]][2],"titles", fixed=T);

        totalcount = str_trim(temp2[[1]][1]);
        totalcount = as.numeric(totalcount);
```

```
        temp2 = strsplit(temp[[1]][1],"to", fixed=T);

        pagecount = str_trim(temp2[[1]][2]);
        pagecount = as.numeric(pagecount);

    result = list();

    result$totalcount = totalcount;
    result$pagecount = pagecount;

    result;
}

grabFilmsForPerson = function(nmid)
    {
    url = paste("https://www.imdb.com/filmosearch/?explore=title_type&role=",nmid,"&ref_=filmo_ref_typ&s

    page1 = read_html(url);
    result = list();
    ## useful for other data purposes
    result$nmid = nmid;

    ## name of person
    result$name = grabNameFromFilmsPage(page1);
    result$countfilms = grabFilmCountFromFilmsPage(page1);

    result$movies.50 = grabFilmInfoFromFilmsPage(page1);

    result;
}
   nmid = "nm0000226";
    will = grabFilmsForPerson(nmid);
    plot(will$movies.50[,c(1,6,7:10)]);
```




```
boxplot(will$movies.50$millions);
```



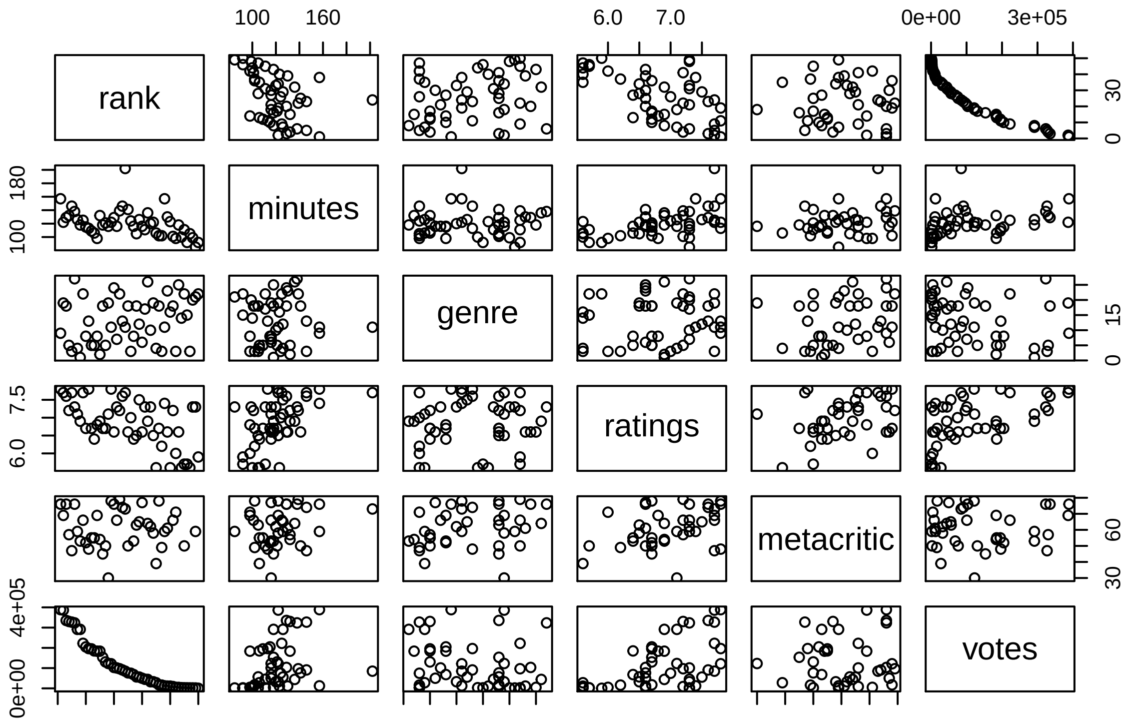

```r
widx =  which.max(will$movies.50$millions);
will$movies.50[widx,];
```

```
##    rank   title       ttid year rated minutes                      genre ratings
## 15   15 Aladdin tt6139732 2019    PG     128 Adventure, Family, Fantasy       7
##    metacritic  votes millions
## 15         53 222928   355.56
```

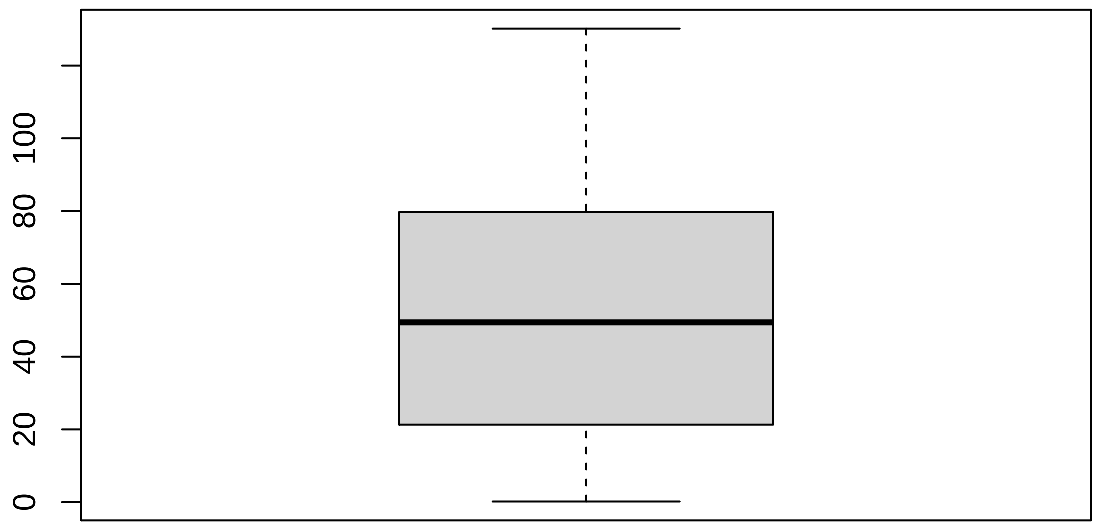
```r
summary(will$movies.50$year);  # bad boys for life ... did data change?
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1993    2001    2006    2007    2014    2020
```

```r
nmid = "nm0000243";
denzel = grabFilmsForPerson(nmid);
plot(denzel$movies.50[,c(1,6,7:10)]);
```




```r
boxplot(denzel$movies.50$millions);
```



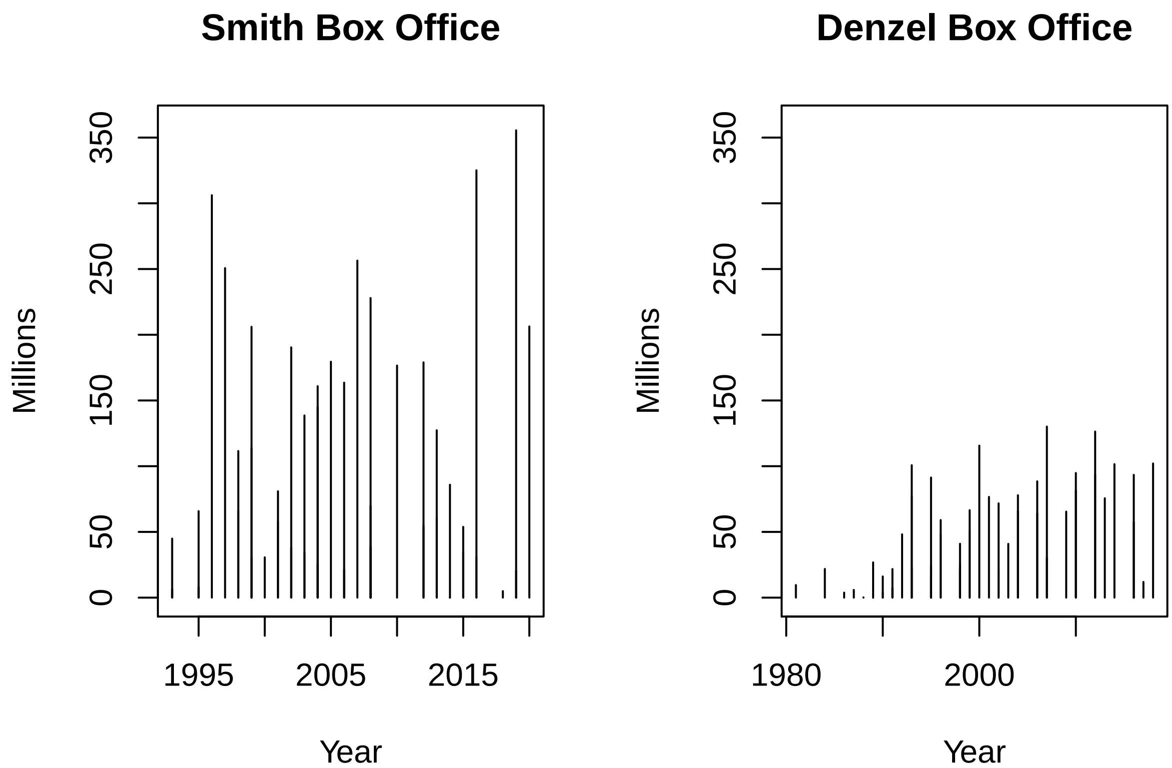

```
    didx =  which.max(denzel$movies.50$millions);
denzel$movies.50[didx,];
```

```
##   rank             title      ttid year rated minutes                   genre
## 1    1 American Gangster tt0765429 2007     R     157 Biography, Crime, Drama
##   ratings metacritic  votes millions
## 1     7.8         76 388731   130.16
```

```
    summary(denzel$movies.50$year);
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1981    1993    1999    2000    2008    2018
```

```
par(mfrow=c(1,2));
plot(will$movies.50$year,will$movies.50$millions, main="Smith Box Office", xlab="Year",ylim=c(0,360)
plot(denzel$movies.50$year,denzel$movies.50$millions, main="Denzel Box Office", xlab="Year",ylim=c(0
```



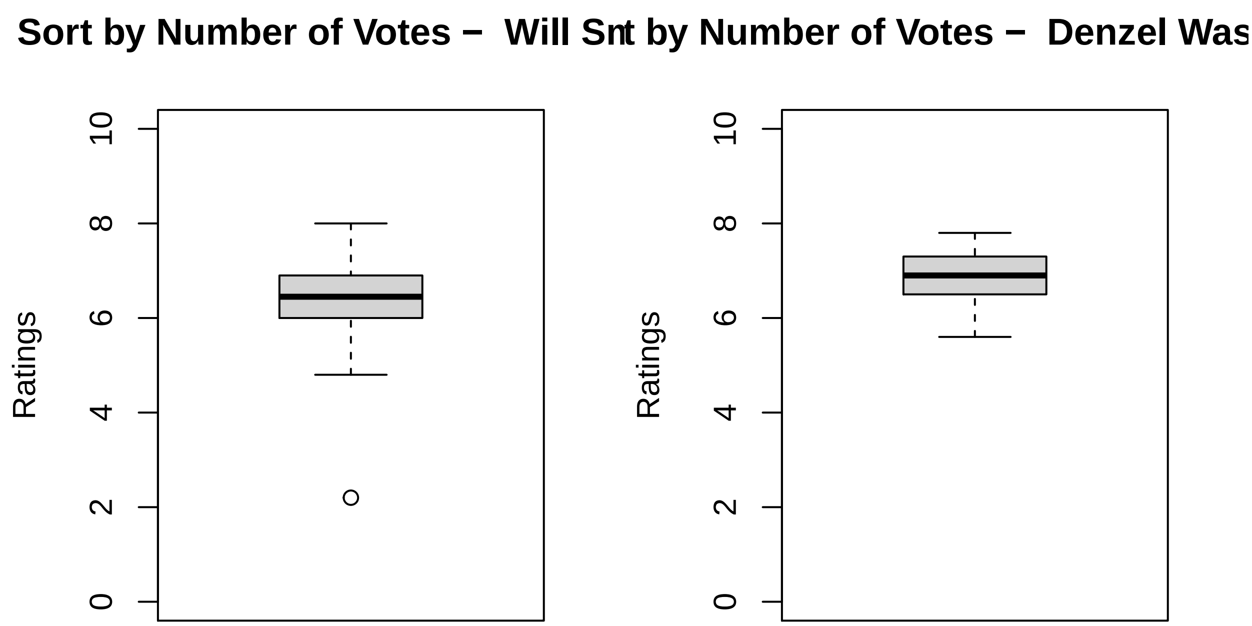

```
summary(will$movies.50$millions)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.02   32.42   65.81  103.42  170.08  355.56       3
```

```
summary(denzel$movies.50$millions)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.19   21.43   49.42   52.12   78.82  130.16       2
```

```
# In the dollars department, there is no contest. Will Smith's movies have higher lows,
# higher highs, a higher average, and higher quartiles. Unequivocally, Will Smith's movies
# are a bigger success at the box office than Denzel Washington's.

    par(mfrow=c(1,2));
    boxplot(will$movies.50$ratings, main=will$name, ylim=c(0,10), ylab="Ratings" )
    boxplot(denzel$movies.50$ratings, main=denzel$name, ylim=c(0,10), ylab="Ratings")
```



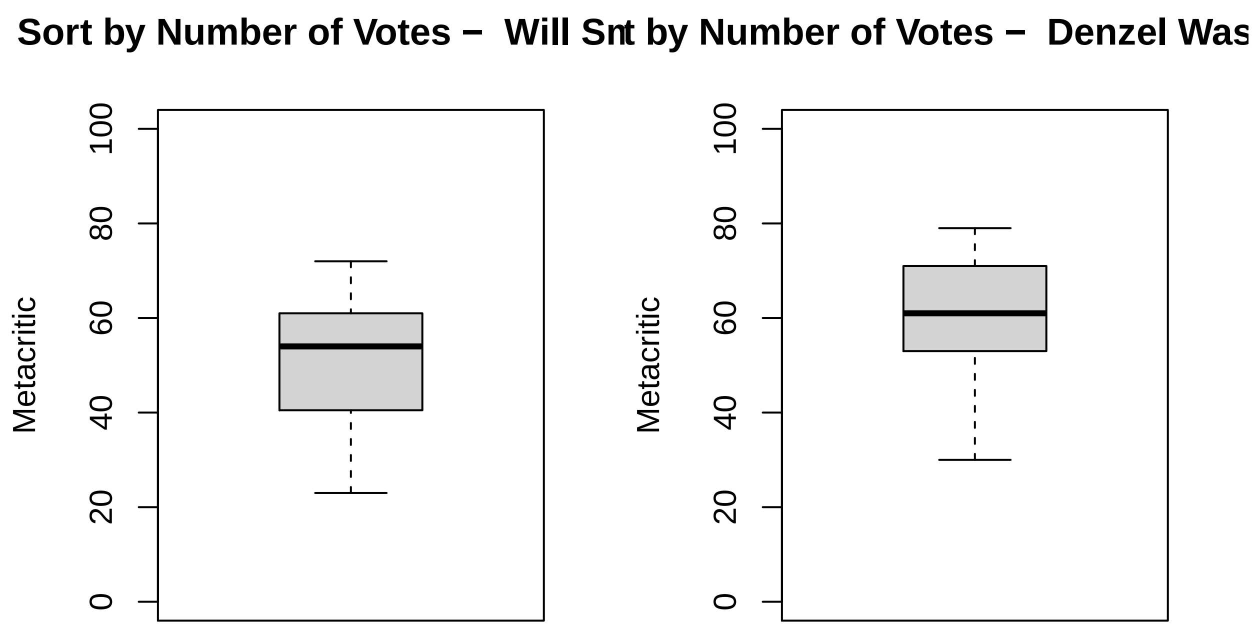

```r
summary(will$movies.50$ratings)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.200   6.000   6.450   6.326   6.875   8.000
```

```r
summary(denzel$movies.50$ratings)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   5.600   6.525   6.900   6.852   7.300   7.800
```

```r
# Viewer ratings tell a different story than box office earnings. While Denzel's highest
# rated movies are unable to crack a rating of 8/10 like Will Smith's highest rated movie,
# his average movie rating, quartile ratings, and lowest rating are all higher. On average,
# viewers enjoy Denzel Washington's movies more than Will Smith's.

par(mfrow=c(1,2));
boxplot(will$movies.50$metacritic, main=will$name, ylim=c(0,100), ylab="Metacritic" )
boxplot(denzel$movies.50$metacritic, main=denzel$name, ylim=c(0,100), ylab="Metacritic")
```




```r
  par(mfrow=c(1,1))
summary(will$movies.50$metacritic)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   23.00   40.50   54.00   50.87   61.00   72.00       3
```

```r
summary(denzel$movies.50$metacritic)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   30.00   53.00   61.00   61.15   71.00   79.00       9
```

```r
# In general, critics seem to agree with the viewers. Their average rating of Denzel
# Washington's movies are higher, as well as the minumum and maximum ratings. There is a
# large area underneith the average in Will Smith's boxplot, indicating that his worse
# movies greatly outnumber his good movies, and thus bring down his metacritic average.
```

```r
library(devtools);
```

```
## Loading required package: usethis
```

```r
library(humanVerseWSU);

path.github = "https://raw.githubusercontent.com/MonteShaffer/humanVerseWSU/master/";

include.me = paste0(path.github, "misc/functions-nlp.R");
source_url( include.me );
```

```
## SHA-1 hash of file is 704afa69d52215d315cb5f59cdc020b0bbfd0b13
```

```
## Warning: package 'tm' was built under R version 4.0.3

## Loading required package: NLP

## Warning: package 'NLP' was built under R version 4.0.3

## Warning: package 'SentimentAnalysis' was built under R version 4.0.3

##
## Attaching package: 'SentimentAnalysis'

## The following object is masked from 'package:base':
##
##     write
```

```r
include.me = paste0(path.github, "misc/functions-nlp-str.R");
source_url( include.me );
```

```
## SHA-1 hash of file is 6bdb234fa84eea995969dc29d6ff2a78f3982131
```

```r
include.me = paste0(path.github, "misc/functions-nlp-stack.R");
source_url( include.me );
```

```
## SHA-1 hash of file is 034efbce0405954198545f8798e119b77a4809c9
```

```r
include.me = paste0(path.github, "misc/functions-nlp-pos.R");
source_url( include.me );
```

```
## SHA-1 hash of file is d8c8cf01c8ead1b6d4228891aa52bac77084a6e7

## Warning: package 'openNLP' was built under R version 4.0.3
```

```r
include.me = paste0(path.github, "humanVerseWSU/R/functions-encryption.R");
source_url( include.me );
```

```
## SHA-1 hash of file is da71dde620bed33db055778b752eefb476f7bf6b
```

```r
path.to.nascent = "C:/Users/Alexander Nevsky/Dropbox/WSU-419/Fall 2020/__student_access__/unit_02_confi

folder.nlp = "nlp/";
path.to.nlp = paste0(path.to.nascent, folder.nlp);


###### UPDATES TO dataframe subset function ######
# inflation adjustments for NA ... and improvements on subsetting
include.me = paste0(path.github, "humanVerseWSU/R/functions-dataframe.R");
source_url( include.me );
```

```
## SHA-1 hash of file is 1149cbf3e865f692b50d4d1983e6364dc56ce62d
```

```r
include.me = paste0(path.github, "humanVerseWSU/R/functions-inflation.R");
source_url( include.me );
```

```
## SHA-1 hash of file is b6d29327e3fe030ca132b135f4a89b6fc6a61a66
```

```r
library(imdb);
imdb::loadDataIMDB();
names(imdb.data);
```

```
##  [1] "all.movies.creatives"         "all.movies.companies"
##  [3] "all.movies.actors.characters" "all.actors.rank"
##  [5] "all.actors.movies"            "all.actors.info"
```

```
##  [7] "moviecount.byyear"            "actors"
##  [9] "glue"                         "headliners"
## [11] "movies"                       "movies.df"
```

```
humanVerseWSU::loadInflationData();
will.search = IMDB.searchPersonName("Will* Smith*");

will.search;
```

```
##            nmid             name                             roles  born.when
## 226   nm0000226       Will Smith Music Department,Actor,Producer  1968-9-25
## 41738 nm0810342    William Smith Actor,Stunts,Miscellaneous Crew  1933-3-24
## 41739 nm0810349    William Smith                             Actor       <NA>
## 41744 nm0810461 William Smithers                             Actor  1927-7-10
## 58036 nm2405238     Willow Smith       Soundtrack,Actress,Writer 2000-10-31
##                              born.where died.when died.where starmeter.rank
## 226   Philadelphia, Pennsylvania, USA      <NA>       <NA>            500
## 41738         Columbia, Missouri, USA      <NA>       <NA>           5000
## 41739                            <NA>      <NA>       <NA>             NA
## 41744         Richmond, Virginia, USA      <NA>       <NA>             NA
## 58036    Los Angeles, California, USA      <NA>       <NA>             NA
##       starmeter.delta
## 226                67
## 41738            1386
## 41739         -191098
## 41744             627
## 58036           -1525
##
## 226    Willard Carroll "Will" Smith, Jr. (born September 25, 1968) is an American actor, comedian, p
## 41738 William Smith is probably best known for his portrayal as "Falconetti" in Rich Man, Poor Man (
## 41739
## 41744 Although character actor William (or Bill) Smithers is not recognizable perhaps by name, the f
## 58036
```

```
will.nmid = will.search$nmid[1];
will.nmid = "nm0000226";
will.movies = IMDB.getMoviesForPerson(will.nmid);
will.movies;
```

```
##           ttid      nmid rank year                                  title
## 39   tt0480249 nm0000226    1 2007                             I Am Legend
## 65   tt1386697 nm0000226    2 2016                           Suicide Squad
## 6    tt0116629 nm0000226    3 1996                        Independence Day
## 7    tt0119654 nm0000226    4 1997                            Men in Black
## 26   tt0343818 nm0000226    5 2004                               I, Robot
## 35   tt0454921 nm0000226    6 2006                The Pursuit of Happyness
## 34   tt0448157 nm0000226    7 2008                                 Hancock
## 12   tt0120912 nm0000226    8 2002                         Men in Black II
## 66   tt1409024 nm0000226    9 2012                          Men in Black 3
## 31   tt0386588 nm0000226   10 2005                                   Hitch
## 41   tt0814314 nm0000226   11 2008                            Seven Pounds
## 4    tt0112442 nm0000226   12 1995                                Bad Boys
## 17   tt0172156 nm0000226   13 2003                             Bad Boys II
## 8    tt0120660 nm0000226   14 1998                      Enemy of the State
## 104  tt6139732 nm0000226   15 2019                                 Aladdin
```

```
## 13   tt0145660 nm0000226   16 1999   Austin Powers: The Spy Who Shagged Me
## 90   tt2381941 nm0000226   17 2015                                   Focus
## 80   tt1815862 nm0000226   18 2013                             After Earth
## 74   tt1596350 nm0000226   19 2012                          This Means War
## 57   tt1229340 nm0000226   20 2013       Anchorman 2: The Legend Continues
## 100  tt5519340 nm0000226   21 2017                                  Bright
## 23   tt0307453 nm0000226   22 2004                              Shark Tale
## 54   tt1155076 nm0000226   23 2010                          The Karate Kid
## 11   tt0120891 nm0000226   24 1999                          Wild Wild West
## 10   tt0120783 nm0000226   25 1998                         The Parent Trap
## 70   tt1502397 nm0000226   26 2020                       Bad Boys For Life
## 19   tt0248667 nm0000226   27 2001                                     Ali
## 97   tt4682786 nm0000226   28 2016                       Collateral Beauty
## 93   tt3322364 nm0000226   29 2015                              Concussion
## 47   tt1025100 nm0000226   30 2019                              Gemini Man
## 5    tt0114558 nm0000226   31 1995                            Strange Days
## 22   tt0300051 nm0000226   32 2004                             Jersey Girl
## 21   tt0284490 nm0000226   33 2002                                Showtime
## 14   tt0146984 nm0000226   34 2000            The Legend of Bagger Vance
## 83   tt1837709 nm0000226   35 2014                           Winter's Tale
## 25   tt0338466 nm0000226   36 2003                            Stuck on You
## 44   tt0947802 nm0000226   37 2008                         Lakeview Terrace
## 82   tt1823664 nm0000226   38 2014                                   Annie
## 20   tt0268397 nm0000226   39 2001              Jimmy Neutron: Boy Genius
## 50   tt1082886 nm0000226   40 2008                            The Wackness
## 101  tt5814534 nm0000226   41 2019                       Spies in Disguise
## 94   tt3721964 nm0000226   42 2018                                  Gringo
## 33   tt0416212 nm0000226   43 2008                 The Secret Life of Bees
## 3    tt0108149 nm0000226   44 1993              Six Degrees of Separation
## 24   tt0328099 nm0000226   45 2003                    Malibu's Most Wanted
## 15   tt0167427 nm0000226   46 1999                               Superstar
## 37   tt0466839 nm0000226   47 2007            I Could Never Be Your Woman
## 2    tt0107478 nm0000226   48 1993                        Made in America
## 106  tt7255568 nm0000226   49 2019                    Student of the Year 2
## 38   tt0466856 nm0000226   50 2006                                     ATL
## 30   tt0384504 nm0000226   51 2004                             Saving Face
## 108  tt7440432 nm0000226   52 2018                                  Quincy
## 1    tt0105810 nm0000226   53 1992                 Where the Day Takes You
## 29   tt0380277 nm0000226   54 2004                             The Cookout
## 53   tt1109477 nm0000226   55 2008                      The Human Contract
## 64   tt1383607 nm0000226   56 2008    Ramadan Mabrouk Abul-Alamein Hamouda
## 46   tt1015971 nm0000226   57 2009   The Boys: The Sherman Brothers' Story
## 95   tt4209900 nm0000226   58 2015                           Fresh Dressed
## 16   tt0169376 nm0000226   59 1998                     Welcome to Hollywood
## 9    tt0120707 nm0000226   60 1998                                    Ride
## 51  tt10883124 nm0000226   61 2019                                    Dads
## 58   tt1230214 nm0000226   62 2009                                Stuntmen
## 28   tt0379487 nm0000226   63 2004                         The Seat Filler
## 89   tt2350432 nm0000226   64 2012 Free Angela and All Political Prisoners
## 103  tt5903926 nm0000226   65 2018                                Sprinter
## 67   tt1409772 nm0000226   66 2010                            A Man's Story
## 27   tt0375663 nm0000226   67 2003                           A Closer Walk
## 73   tt1584918 nm0000226   68 2014                               Murder101
## 86   tt2074488 nm0000226   69 2005              Will Smith: Live in Concert
```

```
## 99   tt5230508 nm0000226   70 2005         AMV Hell 3: The Motion Picture
## 32   tt0414587 nm0000226   71 2005               There's a God on the Mic
## 18   tt0205314 nm0000226   72 1998                         Pesel Ha'Zahav
## 71   tt1546668 nm0000226   73 2007        The 100 Best Black Movies (Ever)
## 102  tt5839150 nm0000226   74 2009                 The Muhammad Ali Story
## 98   tt4919268 nm0000226   75   NA                             Bad Boys 4
## 76   tt1674782 nm0000226   76   NA                       The Karate Kid 2
## 111  tt9620288 nm0000226   77 2021                           King Richard
## 105  tt6598238 nm0000226   78 2020                         Life in a Year
## 110  tt7820302 nm0000226   79   NA                               Bright 2
## 52  tt11012318 nm0000226   80   NA                            The Council
## 62  tt12915962 nm0000226   81   NA           Planes, Trains & Automobiles
## 69   tt1498780 nm0000226   82   NA                              Hancock 2
## 60  tt12530246 nm0000226   83   NA                           Emancipation
## 72   tt1578276 nm0000226   84   NA                  Uptown Saturday Night
## 91   tt2776710 nm0000226   85   NA                             Brilliance
## 55   tt1181790 nm0000226   86   NA              The Billionaire's Vinegar
## 96   tt4463004 nm0000226   87   NA                                 Bounty
## 61  tt12608018 nm0000226   88   NA                                 Bounce
## 107  tt7331400 nm0000226   89   NA                                   1921
## 78   tt1692497 nm0000226   90   NA                 The Redemption of Cain
## 40   tt0812268 nm0000226   91   NA                       The Last Pharaoh
## 49  tt10344550 nm0000226   92   NA           Untitled Solvan Naim Project
## 68   tt1470831 nm0000226   93   NA                    The City That Sailed
## 63   tt1368421 nm0000226   94   NA                             Angelology
## 92   tt2846190 nm0000226   95   NA                  La escribana de uraba
## 42   tt0911023 nm0000226   96   NA                My Wife Hates Your Wife
## 36   tt0464265 nm0000226   97   NA                             Time Share
## 75   tt1609486 nm0000226   98   NA           Harold and the Purple Crayon
## 81   tt1817789 nm0000226   99   NA                        Wheeler Dealers
## 77   tt1691439 nm0000226  100   NA   Untitled Dr. S. Allen Counter Project
## 85   tt1879098 nm0000226  101   NA                    What Would Kenny Do?
## 88   tt2246761 nm0000226  102   NA                               Interpol
## 59   tt1235550 nm0000226  103   NA                  Welcome to the Sticks
## 79   tt1714201 nm0000226  104   NA      Monster Witness Relocation Program
## 43   tt0942916 nm0000226  105   NA                       Sisters of Mercy
## 48   tt1028582 nm0000226  106   NA                           Newton's Law
## 45   tt0979959 nm0000226  107   NA                                 Cooked
## 109  tt7737796 nm0000226  108   NA                           The Prospect
## 56   tt1206884 nm0000226  109   NA                           The Long Run
## 87   tt2168954 nm0000226  110   NA                       Ghost Graduation
## 84   tt1852816 nm0000226  111   NA                                    Joe
##                            genre    rated minutes ratings metacritic  votes
## 39      Action, Adventure, Drama    PG-13     101     7.2         65 675160
## 65    Action, Adventure, Fantasy    PG-13     123     6.0         40 588064
## 6      Action, Adventure, Sci-Fi    PG-13     145     7.0         59 520592
## 7      Action, Adventure, Comedy    PG-13      98     7.3         71 507597
## 26         Action, Drama, Sci-Fi    PG-13     115     7.1         59 491474
## 35              Biography, Drama    PG-13     117     8.0         64 438121
## 34               Action, Fantasy    PG-13      92     6.4         49 437887
## 12     Action, Adventure, Comedy    PG-13      88     6.2         49 337458
## 66     Action, Adventure, Comedy    PG-13     106     6.8         58 329948
## 31               Comedy, Romance    PG-13     118     6.6         58 291991
## 41                         Drama    PG-13     123     7.6         36 282568
```

```
## 4          Action, Comedy, Crime         R      119      6.9           41 235766
## 17         Action, Comedy, Crime         R      147      6.6           38 228329
## 8               Action, Thriller         R      132      7.3           67 224019
## 104   Adventure, Family, Fantasy        PG      128      7.0           53 216257
## 13     Action, Adventure, Comedy     PG-13       95      6.6           59 214064
## 90          Comedy, Crime, Drama         R      105      6.6           56 212789
## 80     Action, Adventure, Sci-Fi     PG-13      100      4.8           33 189337
## 74       Action, Comedy, Romance     PG-13      103      6.3           31 174205
## 57                        Comedy     PG-13      119      6.3           61 165909
## 100    Action, Fantasy, Thriller     TV-MA      117      6.3           29 164913
## 23  Animation, Adventure, Comedy        PG       90      6.0           48 164132
## 54         Action, Drama, Family        PG      140      6.2           61 158842
## 11        Action, Comedy, Sci-Fi     PG-13      106      5.0           38 151132
## 10      Adventure, Comedy, Drama        PG      128      6.5           64 117780
## 70         Action, Comedy, Crime         R      124      6.6           59 112826
## 19       Biography, Drama, Sport         R      157      6.8           65  92508
## 97                Drama, Romance     PG-13       97      6.8           23  89138
## 93       Biography, Drama, Sport     PG-13      123      7.1           55  84750
## 47         Action, Drama, Sci-Fi     PG-13      117      5.7           38  76145
## 5           Action, Crime, Drama         R      145      7.2           66  66102
## 22        Comedy, Drama, Romance     PG-13      102      6.2           43  62700
## 21         Action, Comedy, Crime     PG-13       95      5.5           32  59984
## 14         Drama, Fantasy, Sport     PG-13      126      6.7           47  52831
## 83       Drama, Fantasy, Mystery     PG-13      118      6.2           31  51469
## 25                        Comedy     PG-13      118      5.7           62  50682
## 44        Crime, Drama, Thriller     PG-13      110      6.1           47  48231
## 82         Comedy, Drama, Family        PG      118      5.3           33  33448
## 20  Animation, Action, Adventure         G       82      6.0           65  30195
## 50        Comedy, Drama, Romance         R       99      7.0           61  29701
## 101 Animation, Action, Adventure        PG      102      6.8           54  28764
## 94         Action, Comedy, Crime         R      111      6.1           46  25794
## 33                         Drama     PG-13      114      7.3           57  23962
## 3         Comedy, Drama, Mystery         R      112      6.8           72  19473
## 24                 Comedy, Crime     PG-13       86      5.2           43  18356
## 15               Comedy, Romance     PG-13       81      5.2           42  17781
## 37        Comedy, Drama, Romance     PG-13       97      6.0           NA  16855
## 2                         Comedy     PG-13      111      5.1           NA  15447
## 106        Drama, Romance, Sport Not Rated      146      2.3           NA  15200
## 38          Comedy, Crime, Drama     PG-13      105      6.1           63  10580
## 30        Comedy, Drama, Romance         R       91      7.4           65  10016
## 108       Documentary, Biography     TV-MA      124      7.6           60   3359
## 1         Crime, Drama, Thriller         R      105      6.6           NA   3112
## 29                        Comedy     PG-13       97      3.7           15   2741
## 53                         Drama         R      103      5.1           NA   1970
## 64                Comedy, Family      <NA>      105      6.4           NA     NA
## 46                   Documentary        PG      101      7.8           78    853
## 95   Documentary, History, Music Not Rated       90      6.6           69    840
## 16                        Comedy         R       89      4.9           NA    715
## 9                         Comedy         R       90      4.6           NA    683
## 51   Documentary, Comedy, Family     TV-14       87      6.3           59    668
## 58                        Comedy         R       90      5.1           NA    627
## 28               Comedy, Romance     PG-13       90      5.5           NA    607
## 89                   Documentary Not Rated      102      6.9           73    565
## 103                 Drama, Sport      <NA>      114      6.1           NA     NA
```

```
## 67                  Documentary Not Rated      98      6.2          49    126
## 27                  Documentary      <NA>      NA      8.6          NA     NA
## 73                     Thriller     PG-13      90      4.4          NA     41
## 86           Documentary, Music Not Rated      52      5.7          NA     34
## 99                       Comedy      <NA>      68      7.5          NA     NA
## 32                  Documentary      <NA>      NA      5.6          NA     NA
## 18                  Documentary      <NA>      55      5.2          NA     NA
## 71                  Documentary      <NA>      NA      5.0          NA     NA
## 102                 Documentary      <NA>      52      5.7          NA     NA
## 98    Action, Adventure, Comedy      <NA>      NA       NA          NA     NA
## 76                Action, Drama      <NA>      NA       NA          NA     NA
## 111                       Drama      <NA>      NA       NA          NA     NA
## 105              Drama, Romance     PG-13      NA       NA          NA     NA
## 110    Action, Adventure, Crime      <NA>      NA       NA          NA     NA
## 52                    Biography      <NA>      NA       NA          NA     NA
## 62                       Comedy      <NA>      NA       NA          NA     NA
## 69              Action, Fantasy      <NA>      NA       NA          NA     NA
## 60             Action, Thriller      <NA>      NA       NA          NA     NA
## 72                       Comedy      <NA>      NA       NA          NA     NA
## 91             Sci-Fi, Thriller      <NA>      NA       NA          NA     NA
## 55                        Drama      <NA>      NA       NA          NA     NA
## 96                       Action      <NA>      NA       NA          NA     NA
## 61                        Drama      <NA>      NA       NA          NA     NA
## 107                       Drama      <NA>      NA       NA          NA     NA
## 78                        Drama      <NA>      NA       NA          NA     NA
## 40                        Drama      <NA>      NA       NA          NA     NA
## 49                      Musical      <NA>      NA       NA          NA     NA
## 68               Drama, Fantasy      <NA>      NA       NA          NA     NA
## 63                        Drama      <NA>      NA       NA          NA     NA
## 92                        Drama      <NA>      NA       NA          NA     NA
## 42                       Comedy      <NA>      NA       NA          NA     NA
## 36                       Comedy      <NA>      NA       NA          NA     NA
## 75                    Animation      <NA>      NA       NA          NA     NA
## 81                        Drama      <NA>      NA       NA          NA     NA
## 77                        Drama      <NA>      NA       NA          NA     NA
## 85                       Comedy      <NA>      NA       NA          NA     NA
## 88                       Comedy      <NA>      NA       NA          NA     NA
## 59                       Comedy      <NA>      NA       NA          NA     NA
## 79                       Sci-Fi      <NA>      NA       NA          NA     NA
## 43        Comedy, Drama, Romance      <NA>      NA       NA          NA     NA
## 48    Adventure, Comedy, Family      <NA>      NA       NA          NA     NA
## 45                        Drama      <NA>      NA       NA          NA     NA
## 109                        <NA>      <NA>      NA       NA          NA     NA
## 56                        Drama      <NA>      NA       NA          NA     NA
## 87              Comedy, Fantasy      <NA>      NA       NA          NA     NA
## 84                        Drama      <NA>      NA       NA          NA     NA
##     millions
## 39    256.39
## 65    325.10
## 6     306.17
## 7     250.69
## 26    144.80
## 35    163.57
## 34    227.95
```

```
## 12    190.42
## 66    179.02
## 31    179.50
## 41     69.95
## 4      65.81
## 17    138.61
## 8     111.55
## 104   355.56
## 13    206.04
## 90     53.86
## 80     60.52
## 74     54.76
## 57    127.35
## 100       NA
## 23    160.86
## 54    176.59
## 11    113.81
## 10     66.31
## 70    204.42
## 19     58.20
## 97     31.02
## 93     34.54
## 47     20.55
## 5       7.92
## 22     25.27
## 21     38.08
## 14     30.70
## 83      0.02
## 25     33.83
## 44     39.26
## 82     85.91
## 20     80.94
## 50      2.08
## 101       NA
## 94      4.97
## 33     37.77
## 3       6.41
## 24     34.31
## 15     30.63
## 37        NA
## 2      44.94
## 106     0.78
## 38     21.16
## 30      1.19
## 108       NA
## 1       0.39
## 29     11.54
## 53        NA
## 64        NA
## 46      0.05
## 95        NA
## 16        NA
## 9       5.48
## 51        NA
```

```
## 58         NA
## 28         NA
## 89         NA
## 103        NA
## 67         NA
## 27         NA
## 73         NA
## 86         NA
## 99         NA
## 32         NA
## 18         NA
## 71         NA
## 102        NA
## 98         NA
## 76         NA
## 111        NA
## 105        NA
## 110        NA
## 52         NA
## 62         NA
## 69         NA
## 60         NA
## 72         NA
## 91         NA
## 55         NA
## 96         NA
## 61         NA
## 107        NA
## 78         NA
## 40         NA
## 49         NA
## 68         NA
## 63         NA
## 92         NA
## 42         NA
## 36         NA
## 75         NA
## 81         NA
## 77         NA
## 85         NA
## 88         NA
## 59         NA
## 79         NA
## 43         NA
## 48         NA
## 45         NA
## 109        NA
## 56         NA
## 87         NA
## 84         NA
##
## 39
## 65
## 6
```

```
## 7
## 26
## 35
## 34
## 12
## 66
## 31
## 41
## 4
## 17
## 8
## 104
## 13
## 90
## 80
## 74
## 57
## 100
## 23
## 54
## 11
## 10
## 70
## 19
## 97
## 93
## 47
## 5
## 22
## 21
## 14
## 83
## 25
## 44
## 82
## 20
## 50                                      It's the summer of 1994, and the streets of New York are puls
## 101
## 94
## 33
## 3
## 24
## 15
## 37
## 2
## 106
## 38
## 30
## 108
## 1                                                                                   A group of te
## 29
## 53
## 64
## 46                                                                          Their music is unforg
```

```
## 95                                                                                           Fresh Dressed chron
## 16
## 9
## 51  Director <a href="/name/nm0397171">Bryce Dallas Howard</a> teams up with her father, <a href="/na
## 58
## 28
## 89
## 103
## 67
## 27                                                                                                  This documen
## 73
## 86                                                                                                 Will Smith: L
## 99                                                                                                  An enormousl
## 32
## 18
## 71
## 102
## 98
## 76
## 111
## 105
## 110
## 52
## 62
## 69
## 60
## 72
## 91
## 55
## 96
## 61
## 107                                                                                                        Green
## 78
## 40
## 49
## 68
## 63                                                                                                  When God cre
## 92
## 42
## 36
## 75
## 81
## 77
## 85
## 88
## 59
## 79
## 43
## 48
## 45
## 109
## 56
## 87
## 84
```

```r
will.movies = standardizeDollarsInDataFrame(will.movies, 2000, "millions", "year", "millions2000");
will.movies = sortDataFrameByNumericColumns(will.movies, "millions2000", "DESC");
will.movies$rank.money = 1:nrow(will.movies);
will.rank = subsetDataFrame(imdb.data$all.movies.actors.characters, "nmid", "==", will.nmid);
will.rank;
```

```
##              ttid actor.rank      nmid
## 349353  tt0105810          6 nm0000226
## 353669  tt0107478          3 nm0000226
## 355578  tt0108149          2 nm0000226
## 365501  tt0112442         11 nm0000226
## 376889  tt0116629          1 nm0000226
## 387158  tt0119654          2 nm0000226
## 391606  tt0120660          1 nm0000226
## 393444  tt0120891          1 nm0000226
## 393576  tt0120912          2 nm0000226
## 413566  tt0146984          1 nm0000226
## 434016  tt0172156          2 nm0000226
## 460734  tt0205314         11 nm0000226
## 486656  tt0248667          1 nm0000226
## 521274  tt0307453          1 nm0000226
## 540127  tt0343818          1 nm0000226
## 562065  tt0386588          1 nm0000226
## 586225  tt0448157          1 nm0000226
## 588916  tt0454921          1 nm0000226
## 599482  tt0480249          1 nm0000226
## 613408  tt0814314          1 nm0000226
## 633149  tt1025100          1 nm0000226
## 680397  tt1386697          1 nm0000226
## 682251  tt1409024          1 nm0000226
## 689408  tt1502397          1 nm0000226
## 693711  tt1546668          9 nm0000226
## 715818  tt1815862          2 nm0000226
## 750095  tt2381941          1 nm0000226
## 778362  tt3322364          1 nm0000226
## 809470  tt4682786          1 nm0000226
## 825451  tt5519340          1 nm0000226
## 831014  tt5814534          6 nm0000226
## 837383  tt6139732          1 nm0000226
## 859453  tt7820302          1 nm0000226
## 876732  tt9620288          1 nm0000226
## 642982 tt10883124         13 nm0000226
##                                   character
## 349353                                Manny
## 353669                     Tea Cake Walters
## 355578                                 Paul
## 365501                          Mike Lowrey
## 376889                  Capt. Steven Hiller
## 387158                                  Jay
## 391606                  Robert Clayton Dean
## 393444                           James West
## 393576                                  Jay
## 413566                         Bagger Vance
## 434016                Detective Mike Lowrey
```

```
## 460734                                     Self
## 486656 Cassius Clay /               Muhammad Ali
## 521274                  Oscar            (voice)
## 540127                               Del Spooner
## 562065                                     Hitch
## 586225                              John Hancock
## 588916                             Chris Gardner
## 599482                            Robert Neville
## 613408                                       Ben
## 633149       Henry Brogan /               Junior
## 680397                                  Deadshot
## 682251                                   Agent J
## 689408                                      Mike
## 693711                                      Self
## 715818                              Cypher Raige
## 750095                                     Nicky
## 778362                        Dr. Bennet Omalu
## 809470                                    Howard
## 825451                                Daryl Ward
## 831014                  Lance            (voice)
## 837383             Genie /               Mariner
## 859453                                Daryl Ward
## 876732                         Richard Williams
## 642982                                      Self
```

```r
will.n = nrow(will.movies);
will.ttids = will.movies$ttid;
length( intersect(will.ttids, imdb.data$movies$popular50$ttid) );
```

```
## [1] 19
```

```r
mean(will.rank$actor.rank)
```

```
## [1] 2.628571
```