Házi feladat - Számtech 1.

Abari Kálmán

2020. 09. 13.

# Beolvasás

metacritic <- read.table(file = "adat/metacritic\_games.csv", sep=",", dec=".", header = T, comment.char = "", quote = "\"")

# Sorok és oszlopok száma

str(metacritic)

## 'data.frame': 5699 obs. of 15 variables:  
## $ game : chr "Portal 2" "The Elder Scrolls V: Skyrim" "The Legend of Zelda: Ocarina of Time 3D" "Batman: Arkham City" ...  
## $ platform : chr "PC" "PC" "3DS" "PC" ...  
## $ developer : chr "Valve Software" "Bethesda Game Studios" "GREZZO" "Rocksteady Studios" ...  
## $ genre : chr "Action" "Role-Playing" "Miscellaneous" "Action Adventure" ...  
## $ number\_players : chr "" "No Online Multiplayer" "No Online Multiplayer" "" ...  
## $ rating : chr "E10+" "M" "E10+" "T" ...  
## $ release\_date : chr "Apr 18, 2011" "Nov 10, 2011" "Jun 19, 2011" "Nov 21, 2011" ...  
## $ positive\_critics: int 51 32 84 27 81 52 31 62 14 60 ...  
## $ neutral\_critics : int 1 0 1 0 1 0 0 0 0 1 ...  
## $ negative\_critics: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ positive\_users : int 1700 1616 283 240 251 520 34 187 33 1061 ...  
## $ neutral\_users : int 107 322 20 34 39 112 4 44 8 191 ...  
## $ negative\_users : int 19 451 5 27 11 78 2 29 10 271 ...  
## $ metascore : int 95 94 94 91 90 90 90 90 89 89 ...  
## $ user\_score : int 90 82 90 87 84 85 83 84 71 76 ...

dim(metacritic)

## [1] 5699 15

# Statisztikai mutatók kiírása az adatbázisra

# install.packages("psych")  
# install.packages("DescTools")  
library(psych)  
library(DescTools)

##   
## Attaching package: 'DescTools'

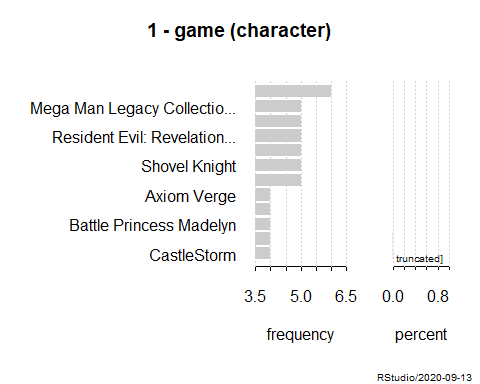
## The following objects are masked from 'package:psych':  
##   
## AUC, ICC, SD

options(scipen = 8)  
describe(metacritic)

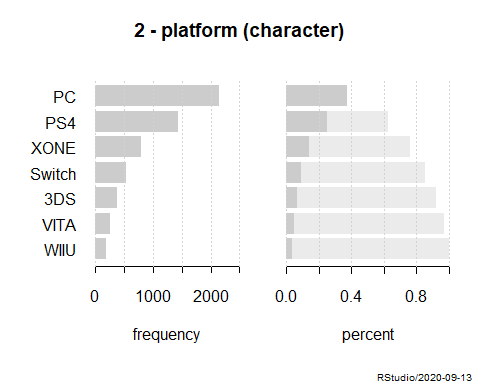
## vars n mean sd median trimmed mad min max  
## game\* 1 5699 2030.36 1157.51 2037 2033.39 1485.57 1 4018  
## platform\* 2 5699 3.32 1.82 3 3.11 1.48 1 7  
## developer\* 3 5699 1074.10 639.16 1058 1075.35 870.29 1 2133  
## genre\* 4 5699 19.25 17.38 7 17.70 4.45 1 54  
## number\_players\* 5 5699 9.88 9.46 8 8.65 10.38 1 36  
## rating\* 6 5699 4.12 2.13 4 4.15 1.48 1 7  
## release\_date\* 7 5699 846.97 466.60 889 854.75 594.52 1 1602  
## positive\_critics 8 5699 13.35 16.20 8 9.96 8.90 0 118  
## neutral\_critics 9 5699 7.96 7.70 6 6.64 4.45 0 65  
## negative\_critics 10 5699 1.15 2.93 0 0.46 0.00 0 51  
## positive\_users 11 5699 28.75 103.69 5 9.91 5.93 0 2762  
## neutral\_users 12 5699 8.64 24.71 2 3.57 2.97 0 506  
## negative\_users 13 5699 14.37 84.94 1 3.13 1.48 0 2998  
## metascore 14 5699 71.86 11.21 73 72.86 10.38 11 97  
## user\_score 15 5699 67.14 13.58 70 68.70 10.38 1 94  
## range skew kurtosis se  
## game\* 4017 -0.02 -1.20 15.33  
## platform\* 6 1.00 -0.22 0.02  
## developer\* 2132 0.02 -1.31 8.47  
## genre\* 53 0.51 -1.52 0.23  
## number\_players\* 35 0.63 -0.54 0.13  
## rating\* 6 -0.11 -1.15 0.03  
## release\_date\* 1601 -0.13 -1.23 6.18  
## positive\_critics 118 2.36 6.49 0.21  
## neutral\_critics 65 2.13 6.28 0.10  
## negative\_critics 51 5.89 55.30 0.04  
## positive\_users 2762 12.11 216.14 1.37  
## neutral\_users 506 8.50 109.21 0.33  
## negative\_users 2998 18.03 442.06 1.13  
## metascore 86 -1.01 1.64 0.15  
## user\_score 93 -1.21 1.79 0.18

Desc(metacritic)

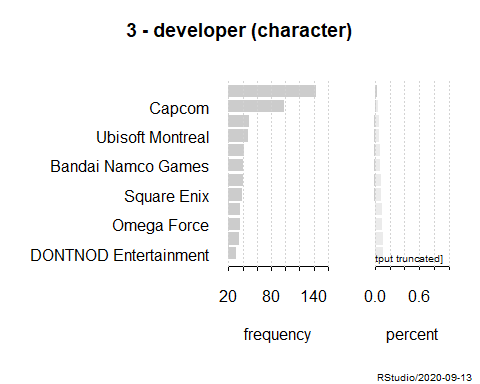
## ------------------------------------------------------------------------------   
## Describe metacritic (data.frame):  
##   
## data frame: 5699 obs. of 15 variables  
## 5699 complete cases (100.0%)  
##   
## Nr ColName Class NAs Levels  
## 1 game character .   
## 2 platform character .   
## 3 developer character .   
## 4 genre character .   
## 5 number\_players character .   
## 6 rating character .   
## 7 release\_date character .   
## 8 positive\_critics integer .   
## 9 neutral\_critics integer .   
## 10 negative\_critics integer .   
## 11 positive\_users integer .   
## 12 neutral\_users integer .   
## 13 negative\_users integer .   
## 14 metascore integer .   
## 15 user\_score integer .   
##   
##   
## ------------------------------------------------------------------------------   
## 1 - game (character)  
##   
## length n NAs unique levels dupes  
## 5'699 5'699 0 4'018 4'018 y  
## 100.0% 0.0%   
##   
## level freq perc cumfreq cumperc  
## 1 Child of Light 6 0.1% 6 0.1%  
## 2 Mega Man Legacy Collection 5 0.1% 11 0.2%  
## 3 Resident Evil: Revelations 5 0.1% 16 0.3%  
## 4 Resident Evil: Revelations 2 5 0.1% 21 0.4%  
## 5 Shantae and the Pirate's Curse 5 0.1% 26 0.5%  
## 6 Shovel Knight 5 0.1% 31 0.5%  
## 7 SteamWorld Dig 5 0.1% 36 0.6%  
## 8 Axiom Verge 4 0.1% 40 0.7%  
## 9 Battle Chasers: Nightwar 4 0.1% 44 0.8%  
## 10 Battle Princess Madelyn 4 0.1% 48 0.8%  
## 11 Bridge Constructor Portal 4 0.1% 52 0.9%  
## 12 CastleStorm 4 0.1% 56 1.0%  
## ... etc.  
## [list output truncated]



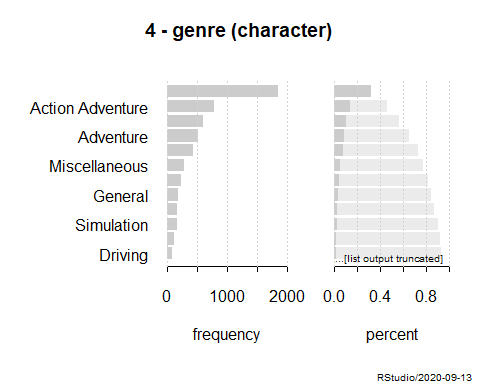
## ------------------------------------------------------------------------------   
## 2 - platform (character)  
##   
## length n NAs unique levels dupes  
## 5'699 5'699 0 7 7 y  
## 100.0% 0.0%   
##   
## level freq perc cumfreq cumperc  
## 1 PC 2'138 37.5% 2'138 37.5%  
## 2 PS4 1'429 25.1% 3'567 62.6%  
## 3 XONE 784 13.8% 4'351 76.3%  
## 4 Switch 526 9.2% 4'877 85.6%  
## 5 3DS 383 6.7% 5'260 92.3%  
## 6 VITA 254 4.5% 5'514 96.8%  
## 7 WIIU 185 3.2% 5'699 100.0%



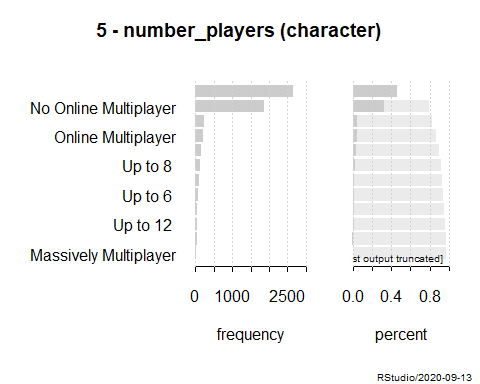
## ------------------------------------------------------------------------------   
## 3 - developer (character)  
##   
## length n NAs unique levels dupes  
## 5'699 5'699 0 2'133 2'133 y  
## 100.0% 0.0%   
##   
## level freq perc cumfreq cumperc  
## 1 Telltale Games 143 2.5% 143 2.5%  
## 2 Capcom 98 1.7% 241 4.2%  
## 3 Nintendo 49 0.9% 290 5.1%  
## 4 Ubisoft Montreal 47 0.8% 337 5.9%  
## 5 TT Games 42 0.7% 379 6.7%  
## 6 Bandai Namco Games 41 0.7% 420 7.4%  
## 7 EA Sports 40 0.7% 460 8.1%  
## 8 Square Enix 39 0.7% 499 8.8%  
## 9 Codemasters 36 0.6% 535 9.4%  
## 10 Omega Force 36 0.6% 571 10.0%  
## 11 Ubisoft 35 0.6% 606 10.6%  
## 12 DONTNOD Entertainment 30 0.5% 636 11.2%  
## ... etc.  
## [list output truncated]



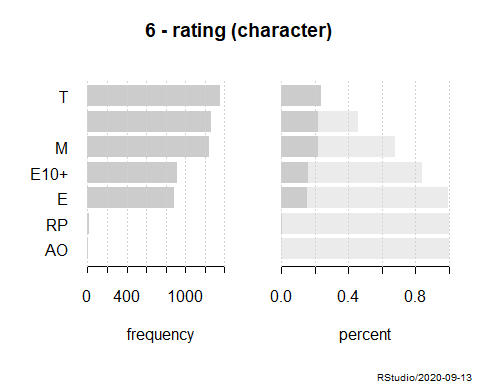
## ------------------------------------------------------------------------------   
## 4 - genre (character)  
##   
## length n NAs unique levels dupes  
## 5'699 5'699 0 54 54 y  
## 100.0% 0.0%   
##   
## level freq perc cumfreq cumperc  
## 1 Action 1'846 32.4% 1'846 32.4%  
## 2 Action Adventure 776 13.6% 2'622 46.0%  
## 3 Role-Playing 596 10.5% 3'218 56.5%  
## 4 Adventure 505 8.9% 3'723 65.3%  
## 5 Strategy 431 7.6% 4'154 72.9%  
## 6 Miscellaneous 282 4.9% 4'436 77.8%  
## 7 Sports 227 4.0% 4'663 81.8%  
## 8 General 173 3.0% 4'836 84.9%  
## 9 Racing 152 2.7% 4'988 87.5%  
## 10 Simulation 152 2.7% 5'140 90.2%  
## 11 Puzzle 108 1.9% 5'248 92.1%  
## 12 Driving 81 1.4% 5'329 93.5%  
## ... etc.  
## [list output truncated]



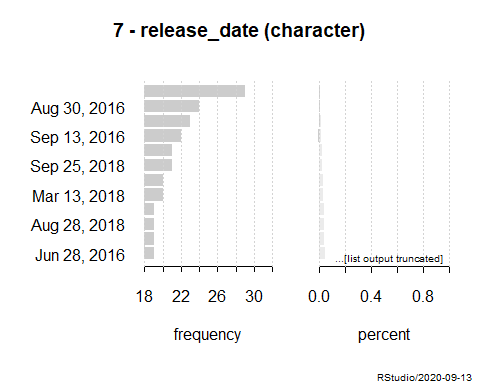
## ------------------------------------------------------------------------------   
## 5 - number\_players (character)  
##   
## length n NAs unique levels dupes  
## 5'699 5'699 0 36 36 y  
## 100.0% 0.0%   
##   
## level freq perc cumfreq cumperc  
## 1 2'640 46.3% 2'640 46.3%  
## 2 No Online Multiplayer 1'869 32.8% 4'509 79.1%  
## 3 Up to 4 222 3.9% 4'731 83.0%  
## 4 Online Multiplayer 218 3.8% 4'949 86.8%  
## 5 2 158 2.8% 5'107 89.6%  
## 6 Up to 8 123 2.2% 5'230 91.8%  
## 7 1 Player 93 1.6% 5'323 93.4%  
## 8 Up to 6 62 1.1% 5'385 94.5%  
## 9 Up to 10 46 0.8% 5'431 95.3%  
## 10 Up to 12 45 0.8% 5'476 96.1%  
## 11 Up to 16 35 0.6% 5'511 96.7%  
## 12 Massively Multiplayer 19 0.3% 5'530 97.0%  
## ... etc.  
## [list output truncated]



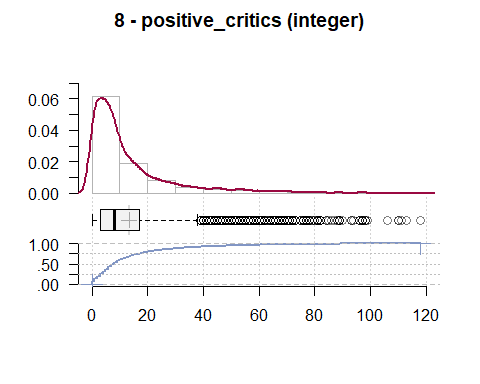
## ------------------------------------------------------------------------------   
## 6 - rating (character)  
##   
## length n NAs unique levels dupes  
## 5'699 5'699 0 7 7 y  
## 100.0% 0.0%   
##   
## level freq perc cumfreq cumperc  
## 1 T 1'360 23.9% 1'360 23.9%  
## 2 1'266 22.2% 2'626 46.1%  
## 3 M 1'247 21.9% 3'873 68.0%  
## 4 E10+ 915 16.1% 4'788 84.0%  
## 5 E 890 15.6% 5'678 99.6%  
## 6 RP 19 0.3% 5'697 100.0%  
## 7 AO 2 0.0% 5'699 100.0%



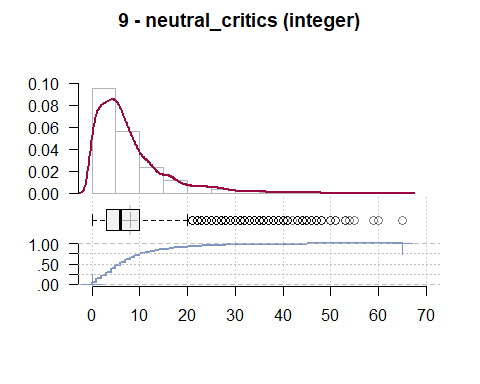
## ------------------------------------------------------------------------------   
## 7 - release\_date (character)  
##   
## length n NAs unique levels dupes  
## 5'699 5'699 0 1'602 1'602 y  
## 100.0% 0.0%   
##   
## level freq perc cumfreq cumperc  
## 1 Nov 18, 2012 29 0.5% 29 0.5%  
## 2 Aug 30, 2016 24 0.4% 53 0.9%  
## 3 Oct 20, 2015 23 0.4% 76 1.3%  
## 4 Sep 13, 2016 22 0.4% 98 1.7%  
## 5 Dec 4, 2018 21 0.4% 119 2.1%  
## 6 Sep 25, 2018 21 0.4% 140 2.5%  
## 7 Apr 18, 2017 20 0.4% 160 2.8%  
## 8 Mar 13, 2018 20 0.4% 180 3.2%  
## 9 Apr 26, 2016 19 0.3% 199 3.5%  
## 10 Aug 28, 2018 19 0.3% 218 3.8%  
## 11 Aug 29, 2017 19 0.3% 237 4.2%  
## 12 Jun 28, 2016 19 0.3% 256 4.5%  
## ... etc.  
## [list output truncated]



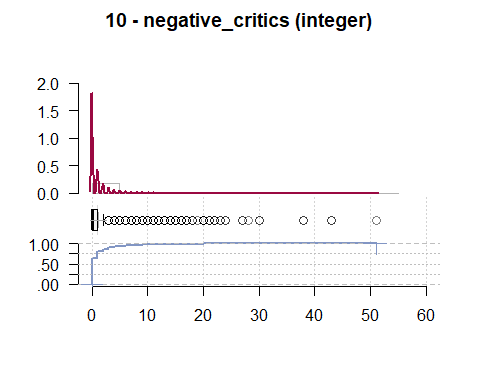
## ------------------------------------------------------------------------------   
## 8 - positive\_critics (integer)  
##   
## length n NAs unique 0s mean meanCI'  
## 5'699 5'699 0 100 416 13.35 12.93  
## 100.0% 0.0% 7.3% 13.77  
##   
## .05 .10 .25 median .75 .90 .95  
## 0.00 1.00 3.00 8.00 17.00 34.00 50.00  
##   
## range sd vcoef mad IQR skew kurt  
## 118.00 16.20 1.21 8.90 14.00 2.36 6.49  
##   
## lowest : 0 (416), 1 (358), 2 (342), 3 (362), 4 (321)  
## highest: 106, 110 (2), 111, 113, 118  
##   
## heap(?): remarkable frequency (7.3%) for the mode(s) (= 0)  
##   
## ' 95%-CI (classic)



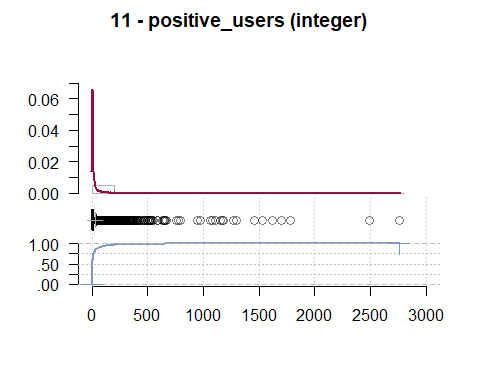
## ------------------------------------------------------------------------------   
## 9 - neutral\_critics (integer)  
##   
## length n NAs unique 0s mean meanCI'  
## 5'699 5'699 0 57 340 7.96 7.76  
## 100.0% 0.0% 6.0% 8.16  
##   
## .05 .10 .25 median .75 .90 .95  
## 0.00 1.00 3.00 6.00 10.00 17.00 24.00  
##   
## range sd vcoef mad IQR skew kurt  
## 65.00 7.70 0.97 4.45 7.00 2.13 6.28  
##   
## lowest : 0 (340), 1 (442), 2 (467), 3 (463), 4 (490)  
## highest: 54, 55, 59, 60, 65  
##   
## heap(?): remarkable frequency (8.7%) for the mode(s) (= 5)  
##   
## ' 95%-CI (classic)



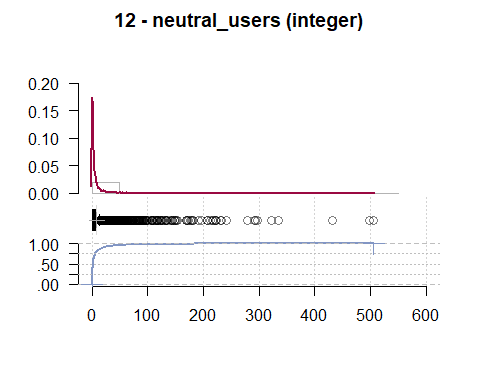
## ------------------------------------------------------------------------------   
## 10 - negative\_critics (integer)  
##   
## length n NAs unique 0s mean meanCI'  
## 5'699 5'699 0 31 3'710 1.15 1.07  
## 100.0% 0.0% 65.1% 1.22  
##   
## .05 .10 .25 median .75 .90 .95  
## 0.00 0.00 0.00 0.00 1.00 3.00 6.00  
##   
## range sd vcoef mad IQR skew kurt  
## 51.00 2.93 2.55 0.00 1.00 5.89 55.30  
##   
## lowest : 0 (3'710), 1 (887), 2 (377), 3 (195), 4 (117)  
## highest: 28, 30 (2), 38 (2), 43 (2), 51  
##   
## heap(?): remarkable frequency (65.1%) for the mode(s) (= 0)  
##   
## ' 95%-CI (classic)



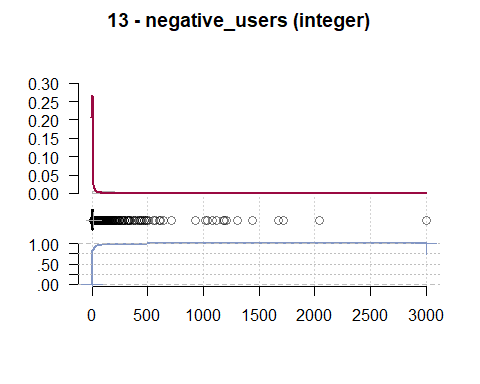
## ------------------------------------------------------------------------------   
## 11 - positive\_users (integer)  
##   
## length n NAs unique 0s mean meanCI'  
## 5'699 5'699 0 318 704 28.75 26.06  
## 100.0% 0.0% 12.4% 31.44  
##   
## .05 .10 .25 median .75 .90 .95  
## 0.00 0.00 2.00 5.00 17.00 61.00 124.10  
##   
## range sd vcoef mad IQR skew kurt  
## 2'762.00 103.69 3.61 5.93 15.00 12.11 216.14  
##   
## lowest : 0 (704), 1 (677), 2 (587), 3 (454), 4 (329)  
## highest: 1'616, 1'700, 1'784, 2'488, 2'762  
##   
## heap(?): remarkable frequency (12.4%) for the mode(s) (= 0)  
##   
## ' 95%-CI (classic)



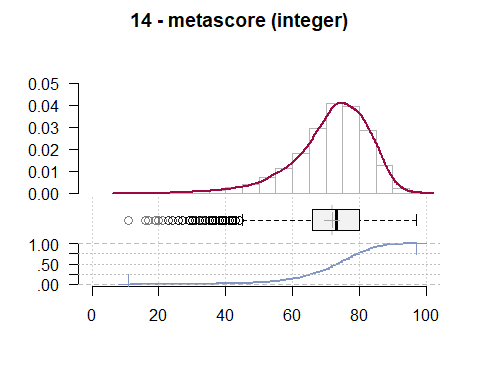
## ------------------------------------------------------------------------------   
## 12 - neutral\_users (integer)  
##   
## length n NAs unique 0s mean meanCI'  
## 5'699 5'699 0 154 1'366 8.64 8.00  
## 100.0% 0.0% 24.0% 9.28  
##   
## .05 .10 .25 median .75 .90 .95  
## 0.00 0.00 1.00 2.00 6.00 20.00 38.00  
##   
## range sd vcoef mad IQR skew kurt  
## 506.00 24.71 2.86 2.97 5.00 8.50 109.21  
##   
## lowest : 0 (1'366), 1 (1'054), 2 (710), 3 (460), 4 (311)  
## highest: 322, 334, 431, 498, 506  
##   
## heap(?): remarkable frequency (24.0%) for the mode(s) (= 0)  
##   
## ' 95%-CI (classic)



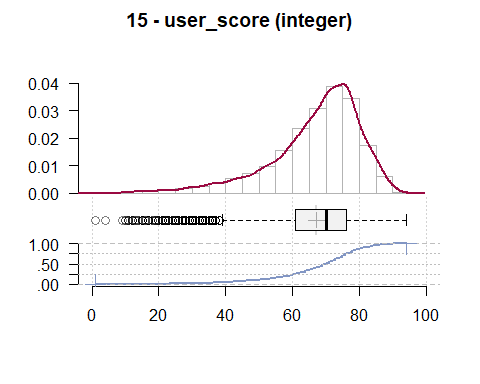
## ------------------------------------------------------------------------------   
## 13 - negative\_users (integer)  
##   
## length n NAs unique 0s mean meanCI'  
## 5'699 5'699 0 214 2'019 14.37 12.17  
## 100.0% 0.0% 35.4% 16.58  
##   
## .05 .10 .25 median .75 .90 .95  
## 0.00 0.00 0.00 1.00 5.00 22.00 49.00  
##   
## range sd vcoef mad IQR skew kurt  
## 2'998.00 84.94 5.91 1.48 5.00 18.03 442.06  
##   
## lowest : 0 (2'019), 1 (929), 2 (575), 3 (332), 4 (248)  
## highest: 1'441, 1'674, 1'720, 2'044, 2'998  
##   
## heap(?): remarkable frequency (35.4%) for the mode(s) (= 0)  
##   
## ' 95%-CI (classic)



## ------------------------------------------------------------------------------   
## 14 - metascore (integer)  
##   
## length n NAs unique 0s mean meanCI'  
## 5'699 5'699 0 79 0 71.86 71.57  
## 100.0% 0.0% 0.0% 72.15  
##   
## .05 .10 .25 median .75 .90 .95  
## 51.00 57.00 66.00 73.00 80.00 84.00 87.00  
##   
## range sd vcoef mad IQR skew kurt  
## 86.00 11.21 0.16 10.38 14.00 -1.01 1.64  
##   
## lowest : 11, 16, 17, 19, 20  
## highest: 93 (11), 94 (5), 95 (3), 96 (2), 97 (6)  
##   
## ' 95%-CI (classic)



## ------------------------------------------------------------------------------   
## 15 - user\_score (integer)  
##   
## length n NAs unique 0s mean meanCI'  
## 5'699 5'699 0 87 0 67.14 66.79  
## 100.0% 0.0% 0.0% 67.49  
##   
## .05 .10 .25 median .75 .90 .95  
## 40.00 49.00 61.00 70.00 76.00 81.00 84.00  
##   
## range sd vcoef mad IQR skew kurt  
## 93.00 13.58 0.20 10.38 15.00 -1.21 1.79  
##   
## lowest : 1, 4, 9, 10 (2), 11 (3)  
## highest: 89 (26), 90 (14), 91 (7), 92 (3), 94  
##   
## ' 95%-CI (classic)



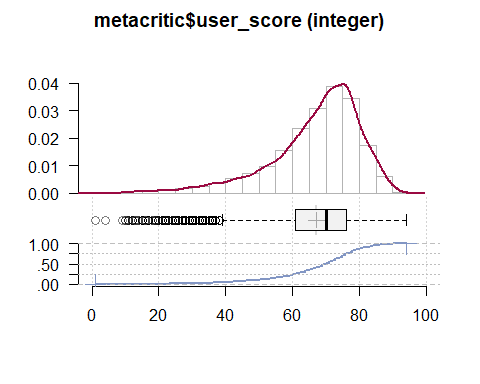
# Egy numerikus oszlopot

describe(metacritic$user\_score)

## vars n mean sd median trimmed mad min max range skew kurtosis se  
## X1 1 5699 67.14 13.58 70 68.7 10.38 1 94 93 -1.21 1.79 0.18

Desc(metacritic$user\_score)

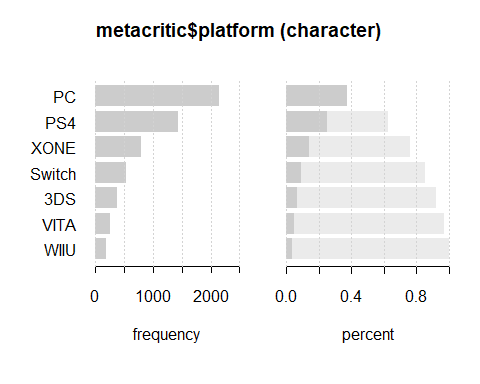
## ------------------------------------------------------------------------------   
## metacritic$user\_score (integer)  
##   
## length n NAs unique 0s mean meanCI'  
## 5'699 5'699 0 87 0 67.14 66.79  
## 100.0% 0.0% 0.0% 67.49  
##   
## .05 .10 .25 median .75 .90 .95  
## 40.00 49.00 61.00 70.00 76.00 81.00 84.00  
##   
## range sd vcoef mad IQR skew kurt  
## 93.00 13.58 0.20 10.38 15.00 -1.21 1.79  
##   
## lowest : 1, 4, 9, 10 (2), 11 (3)  
## highest: 89 (26), 90 (14), 91 (7), 92 (3), 94  
##   
## ' 95%-CI (classic)



# Egy kategorikus oszlop

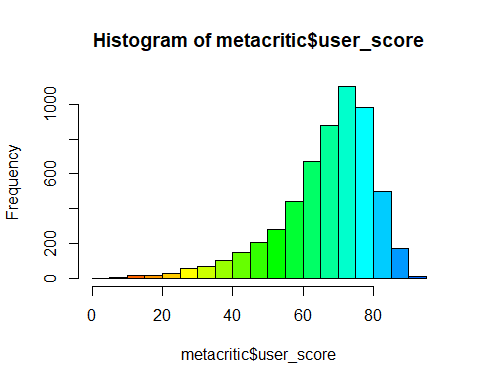
Desc(metacritic$platform)

## ------------------------------------------------------------------------------   
## metacritic$platform (character)  
##   
## length n NAs unique levels dupes  
## 5'699 5'699 0 7 7 y  
## 100.0% 0.0%   
##   
## level freq perc cumfreq cumperc  
## 1 PC 2'138 37.5% 2'138 37.5%  
## 2 PS4 1'429 25.1% 3'567 62.6%  
## 3 XONE 784 13.8% 4'351 76.3%  
## 4 Switch 526 9.2% 4'877 85.6%  
## 5 3DS 383 6.7% 5'260 92.3%  
## 6 VITA 254 4.5% 5'514 96.8%  
## 7 WIIU 185 3.2% 5'699 100.0%



# Két színes ábra

hist(metacritic$user\_score, col=rainbow(30))



boxplot(user\_score~metacritic$platform, data=metacritic, col=topo.colors(7))

