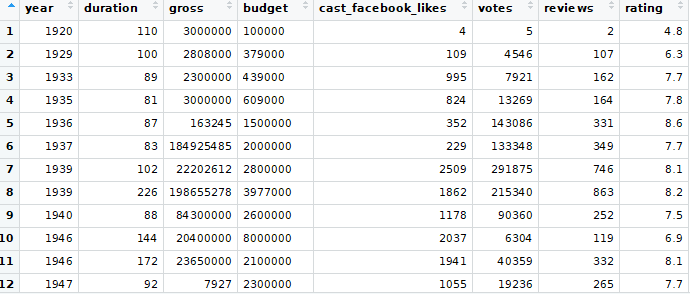
> library(corrplot)

> library(ggplot2)

> library(qgraph)

> data\_frame <- read.csv(file="c:/movies.csv", header=TRUE, sep=",")

> View(data\_frame)  


#перші 12 рядків із 2961

> data\_numeric <- data\_frame[-c(1,2,3)]

> View(data\_numeric)

> #порахуємо ранговий коеф. кореляції Спірмана для числових данних датасету

> corrtest<-corr.test(data\_numeric, method = "spearman")

> corrtest$p #рівень значущості; неформально вказує на ймовірність отримати

year duration gross budget

year 0.000000e+00 3.869423e-01 4.117653e-03 1.150728e-24

duration 1.934712e-01 0.000000e+00 6.584929e-61 5.302368e-82

gross 1.372551e-03 5.487441e-62 0.000000e+00 0.000000e+00

budget 1.438410e-25 3.534912e-83 0.000000e+00 0.000000e+00

cast\_facebook\_likes 7.960954e-19 1.192593e-42 1.794431e-66 2.894696e-93

votes 1.557291e-13 1.101546e-96 0.000000e+00 1.006947e-190

reviews 6.181475e-35 2.589783e-91 6.421917e-239 4.488093e-136

rating 1.509214e-06 8.886976e-113 9.973062e-33 3.434117e-01

cast\_facebook\_likes votes reviews rating

year 4.776572e-18 7.786457e-13 6.181475e-34 6.036855e-06

duration 1.311852e-41 1.982783e-95 4.143653e-90 1.688525e-111

gross 2.332760e-65 0.000000e+00 1.605479e-237 8.975756e-32

budget 4.920983e-92 2.315978e-189 9.873804e-135 3.869423e-01

cast\_facebook\_likes 0.000000e+00 2.085425e-117 2.785498e-75 4.064933e-18

votes 1.042713e-118 0.000000e+00 0.000000e+00 1.949985e-198

reviews 1.989641e-76 0.000000e+00 0.000000e+00 4.492521e-127

rating 5.807047e-19 8.124938e-200 2.139296e-128 0.000000e+00

> #такі результати на вибірці такого розміру й розподілу за умови істинності нульової

> #гіпотези; тобто чим він менше тим більша ймовірність хибності нульової гіпотези

> #тут нульова гіпотеза, що величини некорельовані

> #при р<0.05 відхиляюємо нульову гіпотезу і можемо вважати, що величини корельовані

> corrtest$r #власне пораховані коефіцієнти кореляції Спірмана

year duration gross budget

year 1.00000000 -0.02390408 0.05878784 0.19039548

duration -0.02390408 1.00000000 0.29846650 0.34431068

gross 0.05878784 0.29846650 1.00000000 0.64877136

budget 0.19039548 0.34431068 0.64877136 1.00000000

cast\_facebook\_likes 0.16182691 0.24772722 0.30886610 0.36365807

votes 0.13509876 0.36992324 0.70638138 0.50418839

reviews 0.22382050 0.36001853 0.55502695 0.43368381

rating -0.08825957 0.39768411 0.21644277 -0.01741749

cast\_facebook\_likes votes reviews rating

year 0.1618269 0.1350988 0.2238205 -0.08825957

duration 0.2477272 0.3699232 0.3600185 0.39768411

gross 0.3088661 0.7063814 0.5550269 0.21644277

budget 0.3636581 0.5041884 0.4336838 -0.01741749

cast\_facebook\_likes 1.0000000 0.4072760 0.3305527 0.16245944

votes 0.4072760 1.0000000 0.8691398 0.51446676

reviews 0.3305527 0.8691398 1.0000000 0.42229836

rating 0.1624594 0.5144668 0.4222984 1.00000000

> flattenCorrMatrix <- function(cormat, pmat) {

+ ut <- upper.tri(cormat)

+ data.frame(

+ row = rownames(cormat)[row(cormat)[ut]],

+ column = rownames(cormat)[col(cormat)[ut]],

+ cor =(cormat)[ut],

+ p = pmat[ut]

+ )

+ }

> flattenCorrMatrix(corrtest$r,corrtest$p) #просто вивели все в одну матрицю, для наочності

row column cor p

1 year duration -0.02390408 3.869423e-01

2 year gross 0.05878784 4.117653e-03

3 duration gross 0.29846650 6.584929e-61

4 year budget 0.19039548 1.150728e-24

5 duration budget 0.34431068 5.302368e-82

6 gross budget 0.64877136 0.000000e+00

7 year cast\_facebook\_likes 0.16182691 4.776572e-18

8 duration cast\_facebook\_likes 0.24772722 1.311852e-41

9 gross cast\_facebook\_likes 0.30886610 2.332760e-65

10 budget cast\_facebook\_likes 0.36365807 4.920983e-92

11 year votes 0.13509876 7.786457e-13

12 duration votes 0.36992324 1.982783e-95

13 gross votes 0.70638138 0.000000e+00

14 budget votes 0.50418839 2.315978e-189

15 cast\_facebook\_likes votes 0.40727597 2.085425e-117

16 year reviews 0.22382050 6.181475e-34

17 duration reviews 0.36001853 4.143653e-90

18 gross reviews 0.55502695 1.605479e-237

19 budget reviews 0.43368381 9.873804e-135

20 cast\_facebook\_likes reviews 0.33055269 2.785498e-75

21 votes reviews 0.86913980 0.000000e+00

22 year rating -0.08825957 6.036855e-06

23 duration rating 0.39768411 1.688525e-111

24 gross rating 0.21644277 8.975756e-32

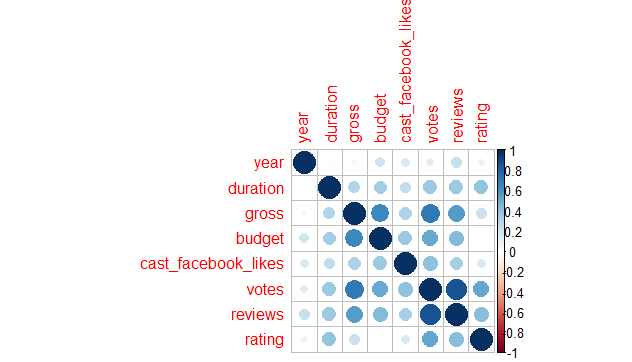
25 budget rating -0.01741749 3.869423e-01

26 cast\_facebook\_likes rating 0.16245944 4.064933e-18

27 votes rating 0.51446676 1.949985e-198

28 reviews rating 0.42229836 4.492521e-127  
#бачимо, що не можемо зробити висновок про наявність кореляції між роком випуску та тривалістю фільму і бюджетом та рейтингом, інші змінні корельовані з відповідним

коефіцієнтом кореляції Спірмана

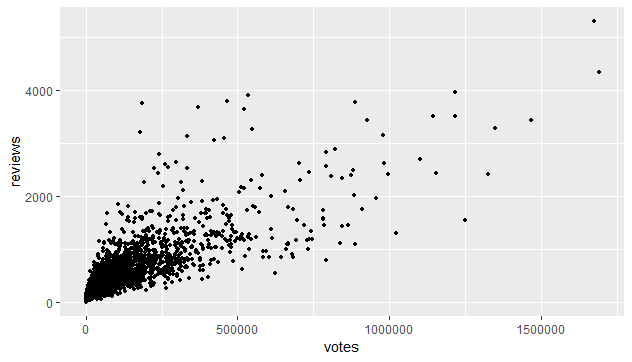
> corrplot(corrtest$r) #будуємо корелограму  


> #бачимо, що однією з найсильніших кореляцій є між відгуками та голосами (р=0)

> #зобразимо їх на графіку

> ggplot(data\_numeric, aes(x=votes, y=reviews))+

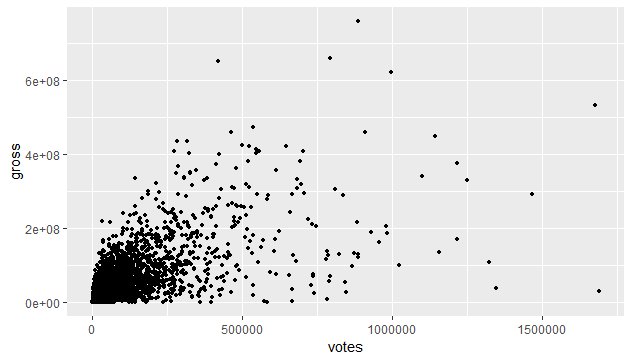
+ geom\_point(size=1)



> #зобразимо також взаємозв'язок між голосами та прибутком фільму (p=0)

> ggplot(data\_numeric, aes(x=votes, y=gross))+

+ geom\_point(size=1)



> #зобразимо кореляційну мережу

> qgraph(corrtest$r)

