Проект

По

Сттистика и емпирични методи

Изготвили:

Александра Яновска, 61753

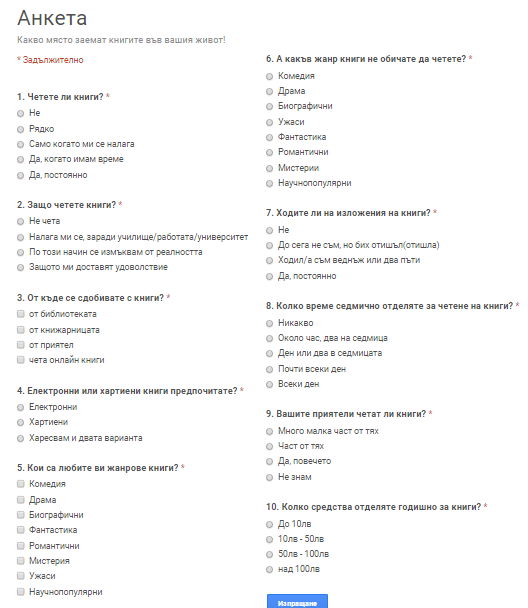
Калина Стоянова, 61755

СИ, 3ти курс, 5та група

1. Анкета

Подготвихме анкета на тема „Какво място заемат книгите във вашия живот“, включваща 10 въпроса.

Линк към анкетата: <https://docs.google.com/forms/d/1prXvFyzT2p9wdLMfVh9nIZS3YesUtPIaqV1iECRtOfA/viewform?usp=send_form>



1. Отговори

На линк-а по-долу се намира таблица с отговори на 40 души.

Линк към отговорите: <https://drive.google.com/file/d/0B12nDlQN8OF_bllpaTI1clBDM2s/view>

1. Въвеждане и изобразяване на данните

* 1.Четете ли книги?

reading = c("Да, постоянно", "Да, постоянно", …);

Функцията table ни показва съотношението между различните отговори на въпроса „Четете ли книги?“. Като можем да забележим, че само 2ма от 40 души или едва 5% не четат. Най-голям дял са хората,четящи само през свободното си време.

> q1 <- table(books$reading)

> q1

Да, когато имам време Да, постоянно

18 9

Не Рядко

2 9

Само когато ми се налага

2

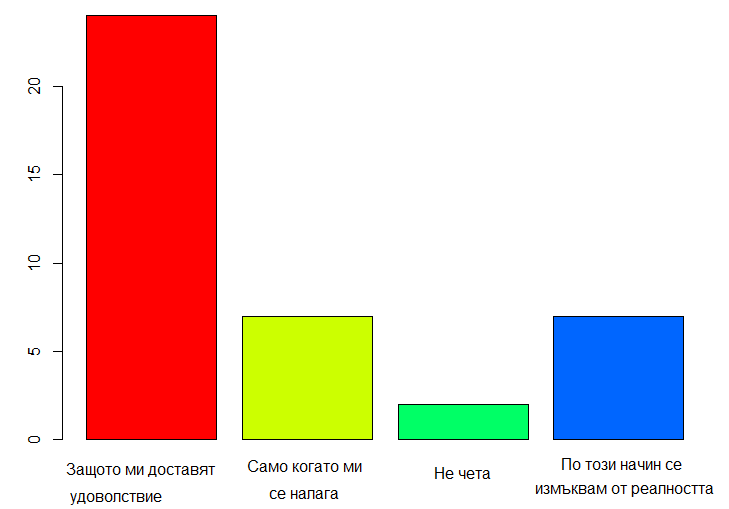
> pie(q1, col=rainbow(5))

Ето и същите резултати нагледно(получени с piechart).



* 2.Защо четете книги?

whyYouRead = c("Защото ми доставят удоволствие", …);



Както за предния въпрос, така и тук използваме table, за да изведем резултатите от въпроса „Защо четете книги?“. С функцията barplot() онагледяваме резултатите.

На графиката ясно се вижда, че хората четат, защото изпитват удоволствие от това.

* 3. От къде се сдобивате с книги?

Въвеждане на данните:  
Тъй като всеки анкетиран има право да изберем повече от едно място от където се сдобива с книги, имаме наличие на повече отговори от броя на анкетираните. Поради тази причина ще въведем данните в матрица.

> placeMatrix = matrix(

+ c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,

+ 0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,

+ 0,0,0,0,0,0,1,1,1,1,1,1,1,0,0,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,

+ 0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,0,0,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,1,1,1

+ ),

+ nrow = 4,

+ ncol=40,

+ byrow=TRUE)

> dimnames(placeMatrix) = list( c("lybrary", "bookstore", "friend", "online"))

След въвеждането на данните в матрица можем да разберем кой е най предпочитаният начин за сдобиване на книга.

> rowSums(placeMatrix)

lybrary bookstore friend online

20 26 18 13

Резултатите показват, че повечето хора предпочитат да си купят книга от книжарницата, отколкото да си я наемат от библиотеката. Само една трета от хората четат онлайн книги, но в резултатите от следващия въпрос, ще видим дали все пак това е предпочитаният им начин на четене.

* 4.Електронни или хартиени книги предпочитате?

eBooksOrPaper = c("Хартиени", "Хартиени", …);

При този въпрос можем да забележим, че макар и технологиите да се развиват толкова бързо, никои не предпочита електронните книги пред мирисът на старата хартиена книга.



* 5. Кои са любите ви жанрове книги?

Отново, поради възможноста за повече от един верен отговор, данните са въведени в матрица.

genreMatrix = matrix(

c(1,1,1,1,0,1,1,1,1,1,0,0,0,0,0,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,

0,1,1,0,0,0,0,1,0,0,0,0,1,0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,

0,0,1,1,0,0,1,0,0,0,0,1,0,0,0,0,0,0,1,0,1,1,1,1,0,0,1,1,0,1,1,1,0,0,0,0,1,1,1,0,

0,0,0,1,0,0,0,0,0,1,0,0,1,0,0,0,0,0,1,0,1,1,0,0,0,1,1,1,1,0,0,0,0,1,1,1,1,1,1,1,

0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,

0,0,0,0,1,1,1,1,1,1,1,1,1,0,0,0,0,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,

0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,

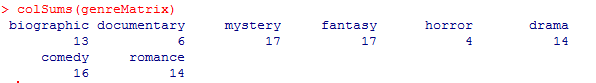
0,0,0,1,0,0,0,0,1,1,1,1,1,0,0,0,0,1,1,1,0,1,0,0,1,0,0,0,1,0,0,0,1,0,0,0,0,0,0,1),

nrow = 40,

ncol=8)

dimnames(genreMatrix) = list( c(), c("biographic", "documentary", "mystery", "fantasy", "horror", "drama", "comedy", "romance"))

С командата colSums виждаме колко гласа е получила всяка категория(жанр). По този начин разбираме кой са любимите жанрове на анкетираните.



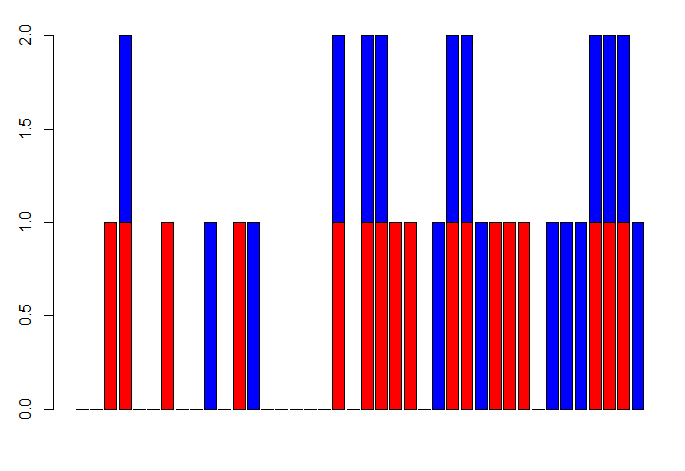
В случая можем ясно да видим, че най-предпочитаните жанрове книги сред нашите анкетирани са мистерията и фантастиката, след което се нареждат комедията, романите, драмата и биографичните книги.

Следва да направим няколко опита за това дали някой съвместими жанрове, всъщност се препокриват и при нашите читатели.

* + След като мистерията и фантастиката сас най-предпочитаните жанрове книги, можем да прочерим колко от хората, които са ги избрали за любими, харесват и двата жанра. За целта ще използваме функциите на матрицата, като знаем как точно са подредени жанровете.

> a=GenreMatrix[c(3,4),]

> barplot(a, col=c("red", "blue"))

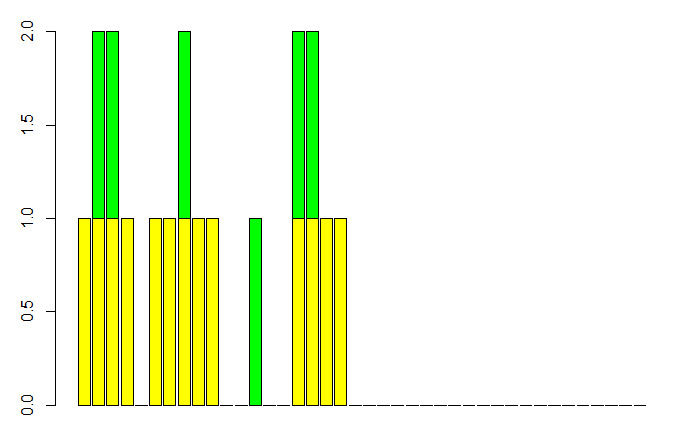


Всеки от двата жанра е заявен 17 пъти, но само 9 от всички 25 харесват и двата.

* + - Като следващ опит ще проверим взаимодействието между любителите на биографичните книги и тези на научно-популярните

>a=GenreMatrix[c(1,2),]

>barplot(a, col=c("yellow", "green"))

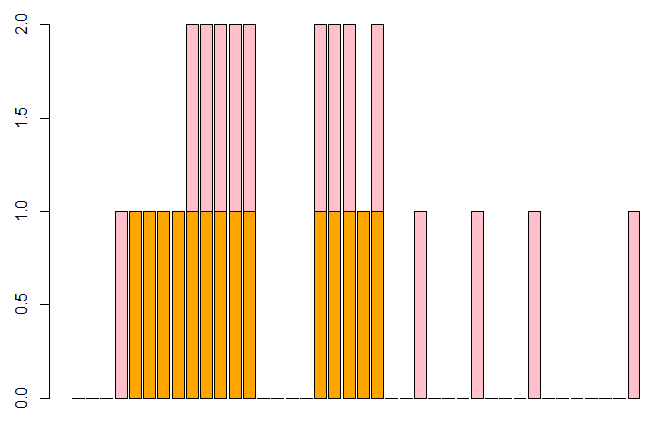


Ясно виждаме, че почти всички от хората които фаресват научно-популярни книги, харесват и биографичните. За съжаление не можем да твърдим обратното.

* Следва да проверим зависимоста межди романите и драмата.

>a=GenreMatrix[c(6,8),]

>barplot(a, col=c("orange", "pink"))

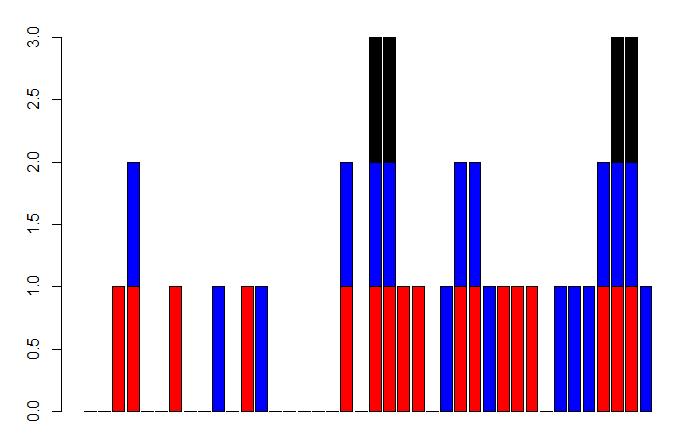


Виждаме, че връзката тук не е много силна, половината отговорите съвпадат.

* Последно решихме да добавим към графата на мистериите и фантастиката най- малко популярният жанр сред нашите читатели, а именно ужаса

>a=GenreMatrix[c(3,4,5),]

>barplot(a, col=c("red", "blue","black"))

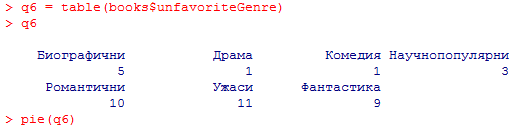


Установихме, че всички хора посочили ужаси, като свой любим жанр, също са посочили и фантастиката и мистерията като свои любими жанрове.

* 6. А какъв жанр книги не обичате да четете?

unfavoriteGenre = c("Ужаси", "Ужаси", "Фантастика", …);

Както на по горните въпроси, така и тук решихме да използваме pie(), за да видим, кой са жанровете книги, който нашите читатели харесватнай-малко.

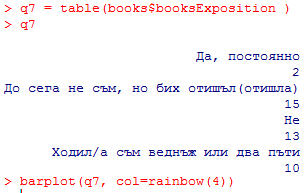




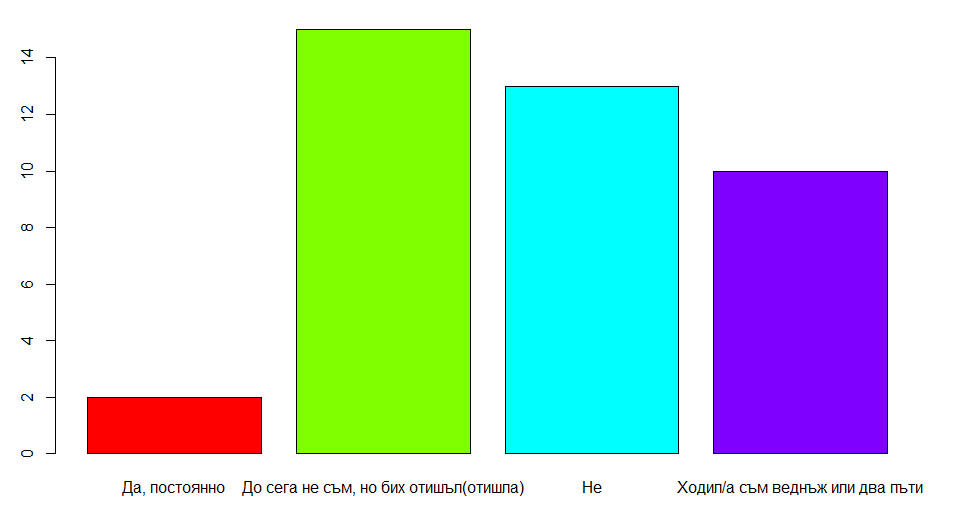
* 7. Ходите ли на изложения на книги?

От първия въпрос разбрахме, че от всички 40 анкетирани, само 5% не четат книги. За това решихме да се поинтересуваме, дали нашите читатели ходят на изложения на книги.

booksExposition = c( "До сега не съм, но бих отишъл(отишла)", "Ходил/а съм веднъж или два пъти", …);



За съжаление 70% от анкетираните никога не са били на изложение на книга, но голяма част от тях, имат желание да посетят такова. Проучването показва,че едва 5% от анкетираните са любители на изложенията и доста чест ги посещават.



* 8. Колко време седмично отделяте за четене на книги?

timeSpend = c("Почти всеки ден", "Почти всеки ден", …);

За съжаление тази статистика показва, че хората, които не отделят никакво време за четене на книги, са повече от тези които четат всеки ден.

> q8 <- table(books$timeSpend)

> q8

Всеки ден Почти всеки ден Ден или два в седмицата

2 7 8

Никакво Около час, два на седмица

8 15

> pie(q8)



* 9. Вашите приятели четат ли книги?

readingFriends = c("Част от тях", "Да, повечето”, …);

В резултатите получени от този въпрос, можем да ясно да разберем, че 98% от анкетираните, знаят дали техните приятели четът книги.

> q9 <-table(books$readingFriends)

> q9

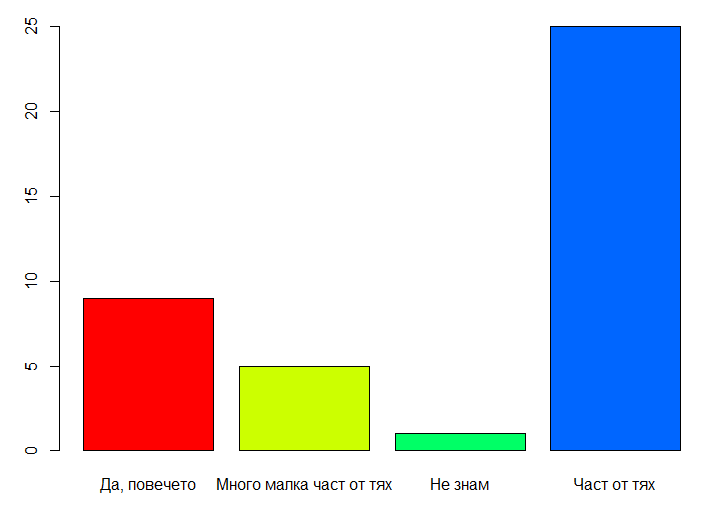
Да, повечето Много малка част от тях

9 5

Не знам Част от тях

1 25

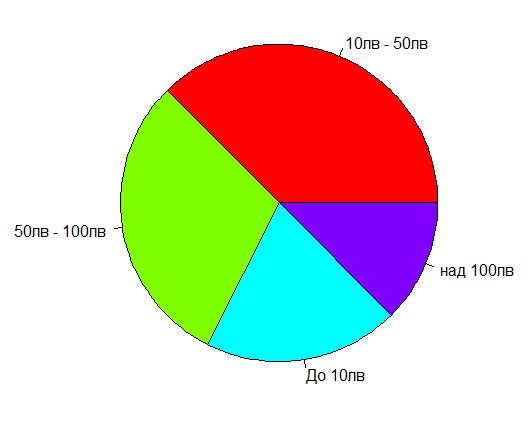
> barplot(q9, col=rainbow(5))

****

* 10.Колко средства отделяте за книги годишно?

moneySpend = c("над 100лв", "50лв - 100лв", …);

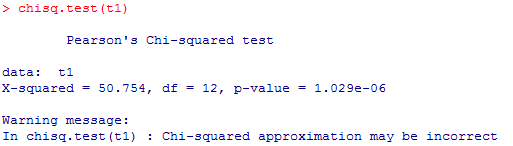
В резултатите от този въпрос забелязвам, че хората, все пак отделят пари за книги.



Поставяме данните от 8 от въпросите(тези с по един отговор) в един data frame

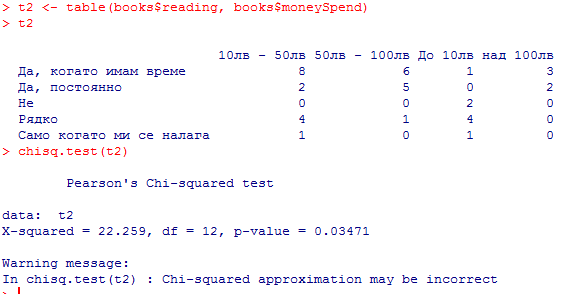
books = data.frame(reading, whyYouRead, eBooksOrPaper, unfavoriteGenre, booksExposition, timeSpend, readingFriends, moneySpend);

1. Изследване на данните

* Съвместно изследване на въпросите: „Четете ли книги?“ и „Защо четете книги?“. Поставяме резултатите от първите 2 въпроса в нашата анкета в таблицата t1. Проверяваме тяхната зависимост с chisq.test().
* 

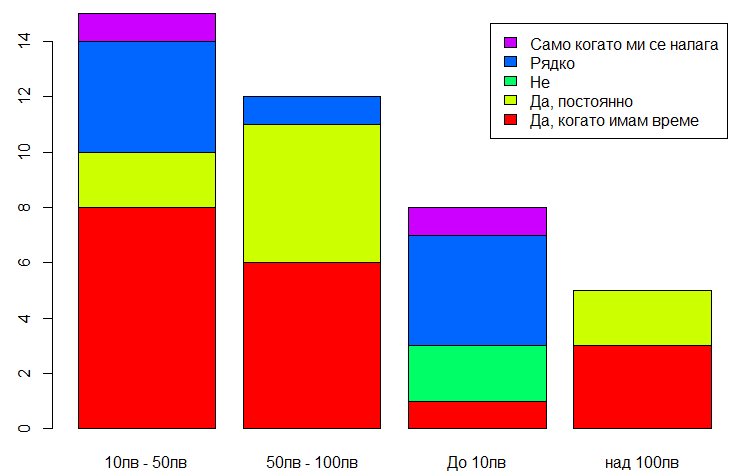
Резултатите от теста показват,че няма ясна зависимост между тях.

* Съвместно изследване на въпросите : „Колко средства отделяте годишно за книги?“ и „Четете ли книги?“

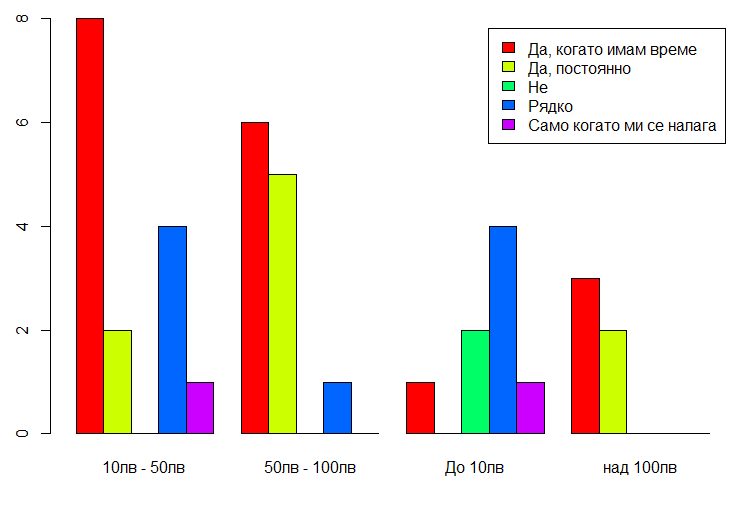
Проверяваме дали между двата въпроса има зависимост, като отново използваме chisq.test(). 

Резултатите ни показват, че този път има зависимост, защото p-valuе <0,5. Нагледно можем да се уверим в това като използваме следната функция:

barplot(t2, col = rainbow(5),legend.text = c( "Да, когато имам време","Да, постоянно","Не", "Рядко", "Само когато ми се налага"));

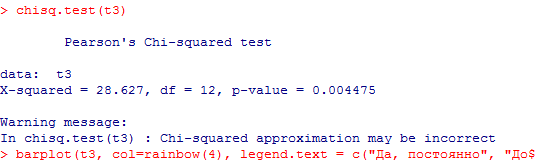


barplot(t2, beside = T, col = rainbow(5), legend.text = c( "Да, когато имам време","Да, постоянно","Не", "Рядко", "Само когато ми се налага"))

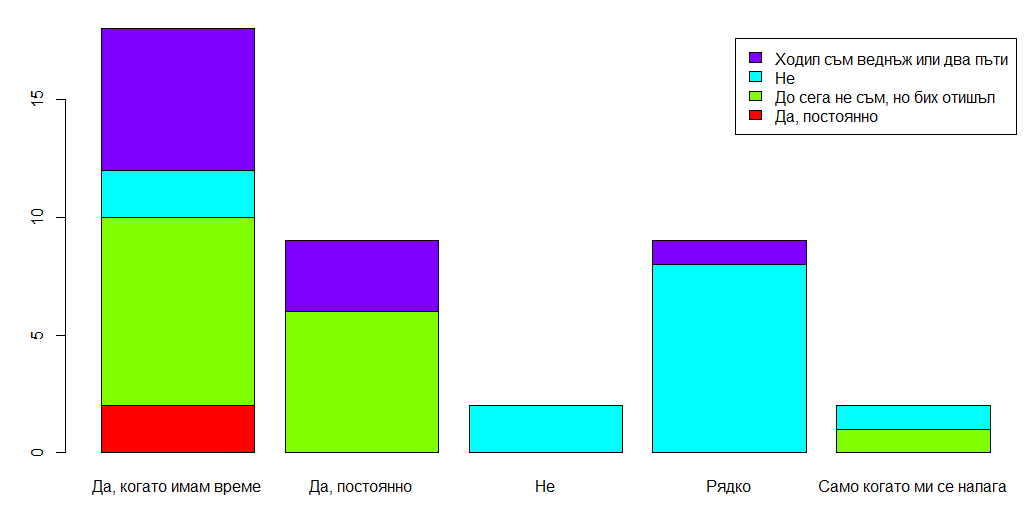


* Съвместно изследване на въпросите: „Четете ли книги?“ и „Ходите ли на изложения на книги?“

Решихме да проверим, каква е връзката между това дали хората четат книги и дали посещават изложения. Проверихме зависимостта им и се оказа, че имат такава.

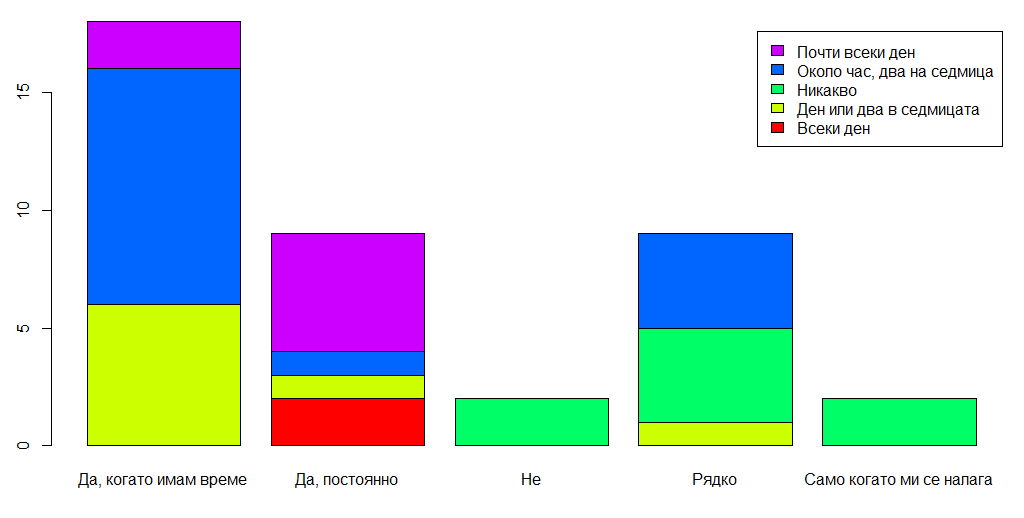


На графиката можем да забележим, че хората, които не обичат да четат, никога не са ходили на изложенияна книги. Обратно на очакванията ни, хората,които постояно посещават изложения на книги, четат само когато намерят време за това.



* Съвместно изследване на въпросите: „Четете ли книги?“ и „Колко време седмично отделяте за четене на книги?“

Тук резултатите от chisq.teat() и графиката, изобщо не могат да ни изненадат. В теста, получихме много ниска стойност за p-value = 0.0001015 < 0,05.



1. Заключение

След обработка и изследване на данните, получени от това проучване, смеем да твърдим, че независимо от бързо развиващите се технологии, хората не са обърнали гръб на книгите.