



СОФИЙСКИ УНИВЕРСИТЕТ "СВ. КЛИМЕНТ ОХРИДСКИ"  
ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА

Курсов проект  
ПО  
Статистика и емпирични методи  
Практикум

Изготвил:  
Радина Нунева, фн 71957, II курс  
специалност Информационни системи

София, 2021 г.

# 1. Тема и данни

Анкетираха съм 55 човека с цел да разбера каква част от живота им са кучетата. Линк към анкета:

[https://docs.google.com/forms/d/e/1FAIpQLSeFAAtOyeDu2691KiEz4kfKPsOZLDfKAiU7sAzA\\_joJCvCyF3A/closedform](https://docs.google.com/forms/d/e/1FAIpQLSeFAAtOyeDu2691KiEz4kfKPsOZLDfKAiU7sAzA_joJCvCyF3A/closedform)

Въпросите присъстващи в анкетата са:

- Имате ли куче?
- Колко кучета бихте искали да имате?
- Кучета или котки?
- За вас кучето е:
- Любим цвят куче:
- Колко пари месечно бихте отделили/отделяте за кучето си? (в лв)
- Кое е най-подходящото място за отглеждане на куче?
- Бихте ли се сблизихи с човек, който не обича кучета?
- Имали ли сте куче до сега?

## 2. Въвеждане и анализ на едномерна променлива

### 2.1. Въпрос: Имате ли куче

★ Въвеждане на данните

```
has_dog <- c("Не, но искам",  
             "Не, но искам",  
             "Не, но искам",  
             "Не, но искам",  
             "Да",  
             "Не, но искам",  
             "Не и не искам",  
             "Да",  
             "Да",  
             "Не, но искам",  
             "Не и не искам",  
             "Не, но искам",  
             "Да",  
             "Не и не искам",  
             "Не, но искам",  
             "Да",  
             "Да",  
             "Да",  
             "Да",  
             "Не, но искам",
```

```

"Не, но искам",
"Не, но искам",
"Не, но искам",
"Да",
"Да",
"Да",
"Да",
"Да",
"Да",
"Не, но искам",
"Не, но искам",
"Не, но искам",
"Не и не искам",
"Да",
"Не и не искам",
"Не и не искам",
"Не, но искам",
"Не и не искам",
"Не, но искам",
"Не, но искам",
"Да",
"Не и не искам",
"Да",
"Да",
"Да",
"Не, но искам",
"Да",
"Да",
"Не и не искам",
"Да",
"Не, но искам",
"Не, но искам",
"Да",
"Не и не искам")

```

## ★ Анализ

Използвам функцията `table()`, защото при категорийните променливи честотата се вижда най-добре чрез таблици.

```
table_has_dog <- table(has_dog)
```

```
table_has_dog
```

```
has_dog
```

Да	Не и не искам	Не, но искам
23	10	22

С функцията `prop.table()` изобразявам процентното разпределение.

```
prop_table_has_dog <- prop.table(table_has_dog)
```

```
prop_table_has_dog
```

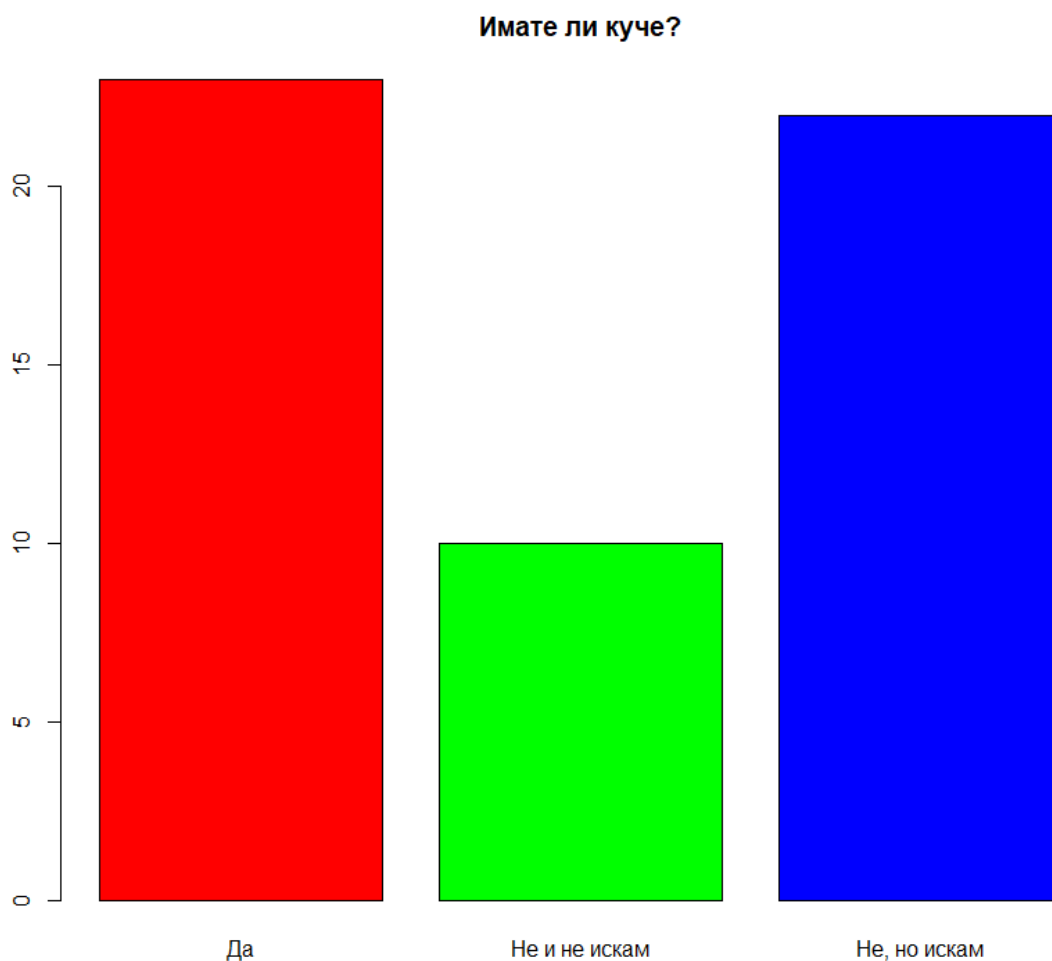
```
has_dog
```

```
Да Не и не искам Не, но искам
```

```
0.4181818 0.1818182 0.4000000
```

Чрез `barplot` показвам частичното разпределение на категориите променливи.

```
barplot(height = table_has_dog, col = rainbow(3), main = "Имате ли куче?")
```



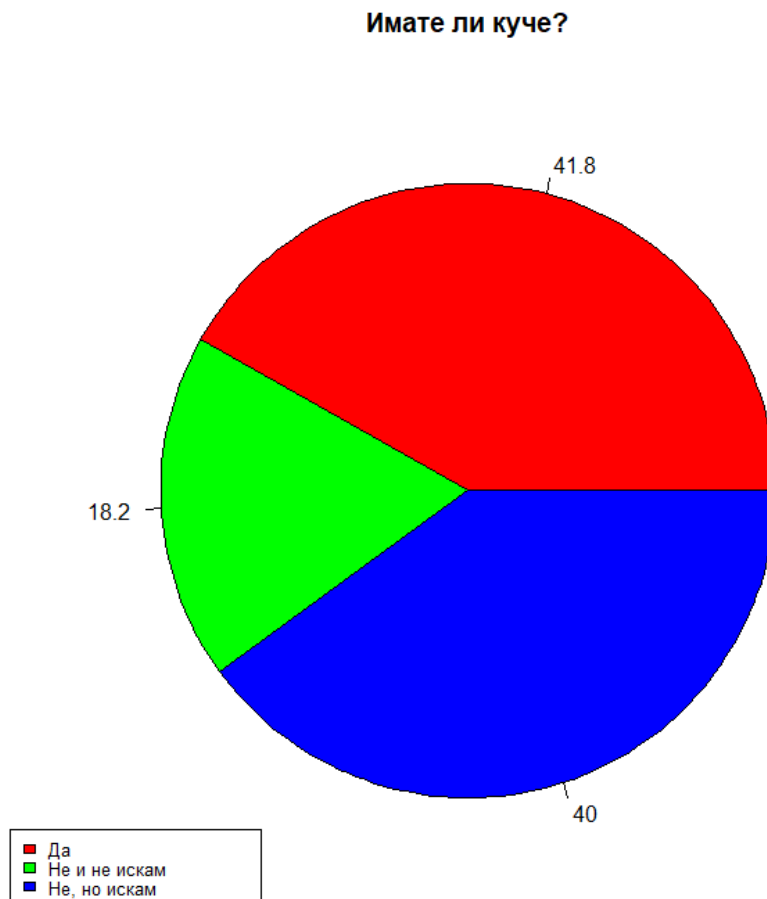
На графиката се вижда, че броят на хората които нямат и не искат куче е най-малък - 10. Останалите 2 отговора имат сходен брой гласове.

За процентното разпределение на данните използвам pie chart

```
piepercent_has_dog<- round(100*table_has_dog/sum(table_has_dog), 1)
```

```
pie(table_has_dog, labels = piepercent_has_dog, main = "Имате ли куче?", col =  
rainbow(length(table_has_dog)))
```

```
legend(x = "bottomleft", legend = c("Да", "Не и не искам", "Не, но искам"), cex =  
0.8, fill = rainbow(length(table_has_dog)))
```



## 2.2. Въпрос: Колко кучета бихте искали да имате?

★ Въвеждане на данните

```
dogs_count <- c(2,
```

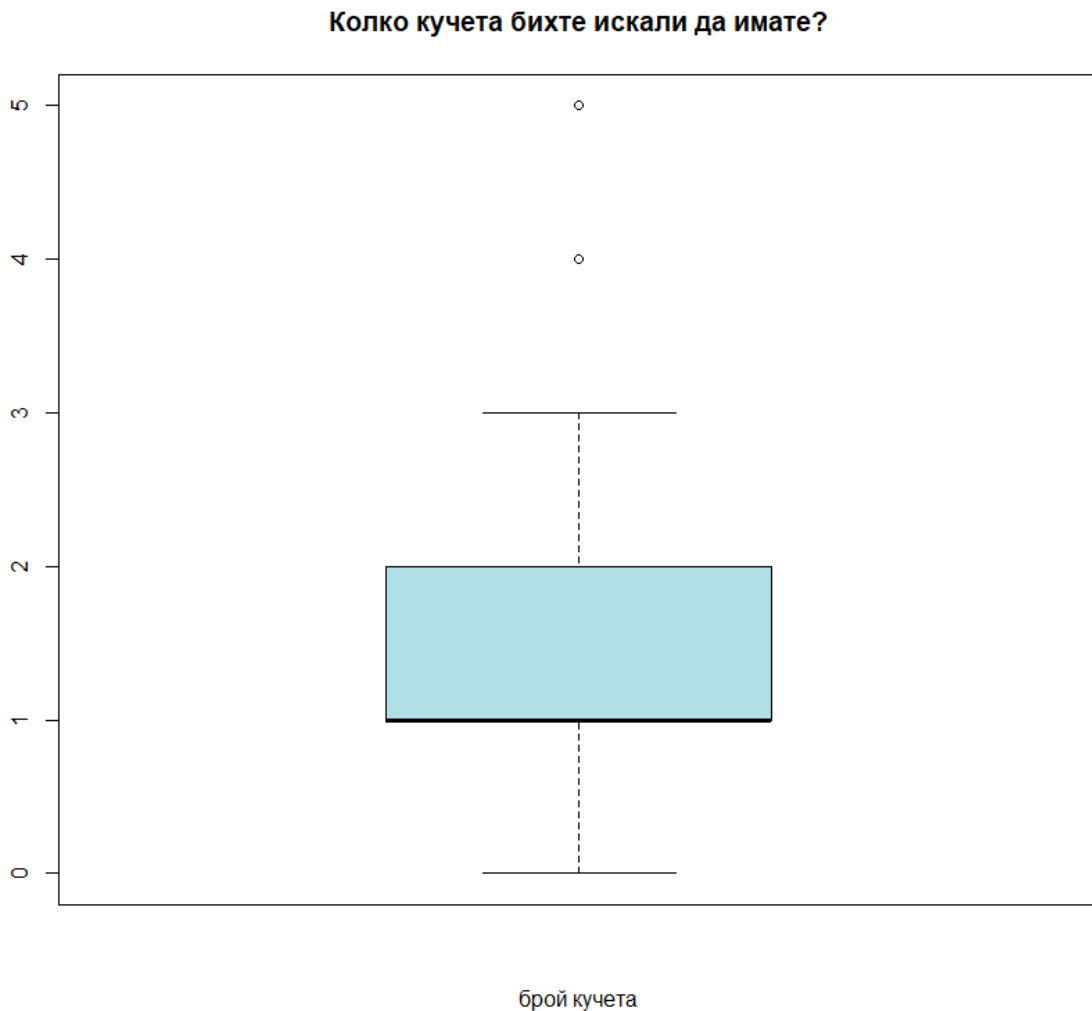
```
2,  
5,  
1,  
2,  
1,  
1,
```

2,  
3,  
1,  
1,  
0,  
1,  
1,  
0,  
2,  
2,  
2,  
2,  
1,  
1,  
1,  
1,  
1,  
1,  
1,  
2,  
2,  
2,  
1,  
2,  
1,  
3,  
1,  
0,  
1,  
0,  
0,  
1,  
0,  
3,  
1,  
1,  
0,  
3,  
1,  
2,  
2,  
5,  
1,  
0,  
1,  
4,  
1,  
2,  
0)

## ★ Анализ

Използвам boxplot за откриване на потенциални outlier-и.

```
boxplot(dogs_count, col = "powderblue", main = "Колко кучета бихте искали да имате?", xlab = "брой кучета")
```



Има открити потенциални outlier-и.

Намирам модата (най-често срещаната стойност във вектора):

```
modeFunction <- function(x) {  
  res_table <- table(x)  
  return(names(res_table)[res_table == max(res_table)])  
}  
modeFunction(dogs_count)  
[1] "1"
```

=> Повечето хора биха искали да имат 1 куче

Намирам медианата (средна стойност):

```
medianFunction <- function(x) {  
  x_sorted <- sort(x)  
  nn <- length(x_sorted)  
  if(nn %% 2 == 0) {  
    return(mean(x_sorted[nn/2 + c(0, 1)]))  
  } else {  
    return(x_sorted[round(nn/2 + 0.25)])  
  }  
}  
medianFunction(dogs_count)  
[1] 1
```

Използвам summary - описателна статистика за центъра на разпределението. Тя показва минималната стойност, 1 квартил, 2 квартил (медиана), 3 квартил и максималната стойност.

```
summary(dogs_count)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	1.000	1.000	1.455	2.000	5.000

Намирам дисперсията (вариацията):

```
var(dogs_count)  
[1] 1.289562
```

Намирам стандартното отклонение, което е оценка на вариацията, която показва колко се отклоняват наблюденията от очакването. То е производно на вариацията и е равно на корен квадратен от дисперсията.

```
sd(dogs_count)  
[1] 1.135589
```

Използвам range() за да намеря обхвата на интервала от най-ниската до най-високата стойност.

```
rangeFunction <- function(x) {  
  max(x) - min(x)  
}  
rangeFunction(dogs_count) #range  
[1] 5
```



Намирам IQR - interquartile range на данните, който е равен на трети квантил минус първи квантил.

```
IQR(dogs_count)
```

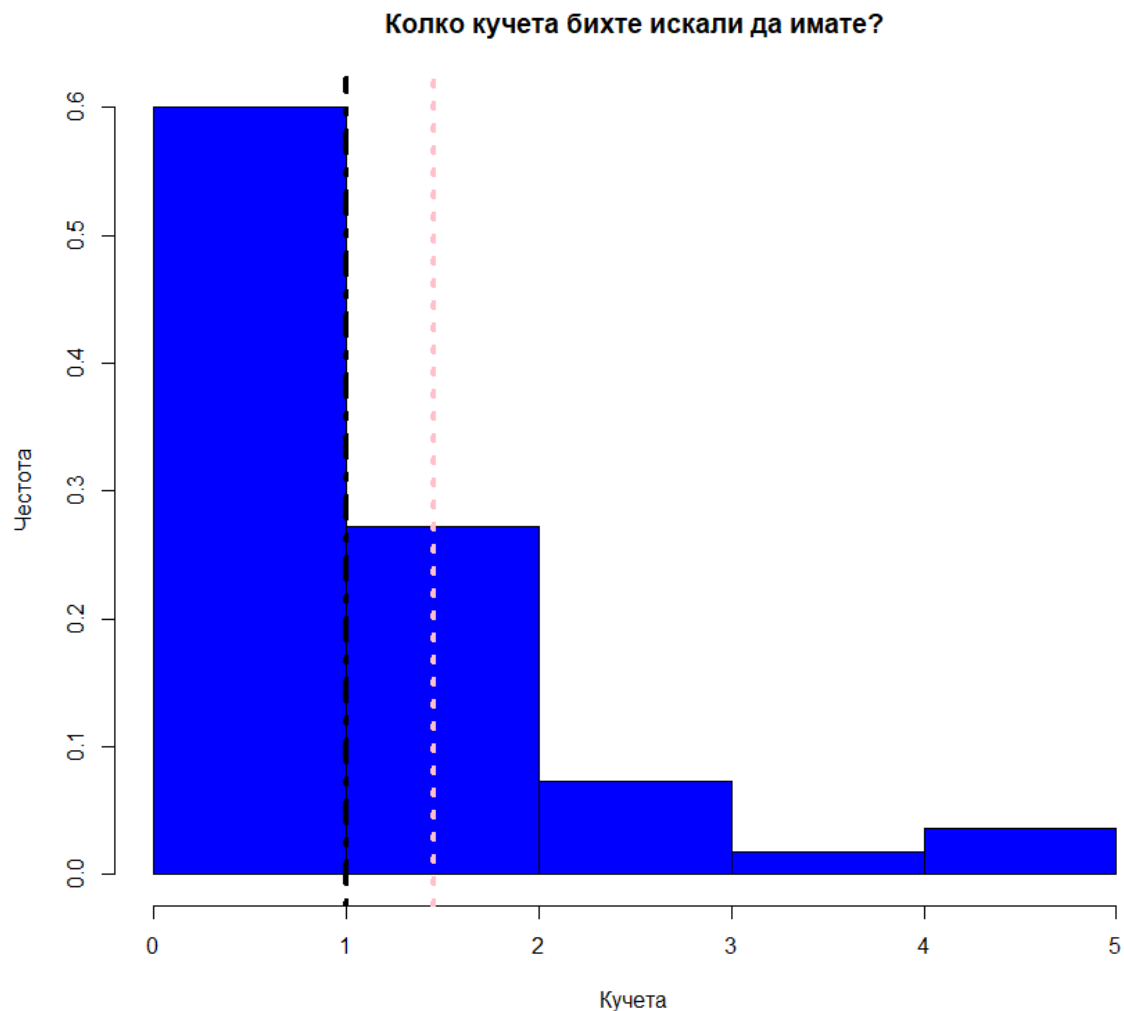
```
[1] 1
```

Използвам хистограма за разпределението на непрекъснатите променливи и черна вертикална линия показваща мястото на средната стойност и розова - медианата.

```
hist(dogs_count, main = "Колко кучета бихте искали да имате?", xlab = "Кучета",  
ylab = "Честота", col = "blue1", prob = T)
```

```
abline(v = mean(dogs_count), lwd = 4, lty = 3, col = "pink")
```

```
abline(v = median(dogs_count), lwd = 4, lty = 4, col = "black")
```



## 2.3. Въпрос: Кучета или котки?

★ Въвеждане на данните

```
dogs_or_cats <- c("Кучета",
```

```
  "Кучета",
```

```
  "И двете",
```

```
  "И двете",
```

```
  "И двете",
```

```
  "Кучета",
```

```
  "Котки",
```

```
  "И двете",
```

```
  "И двете",
```

```
  "И двете",
```

```
  "Кучета",
```

```
  "И двете",
```

```
  "Кучета",
```

```
  "Кучета",
```

```
  "Котки",
```

```
  "Кучета",
```

```
  "Кучета",
```

```
  "И двете",
```

```
  "И двете",
```

```
  "Кучета",
```

```
  "Котки",
```

```
  "Кучета",
```

```
  "Кучета",
```

```
  "Кучета",
```

```
  "Кучета",
```

```
  "Кучета",
```

```
  "Кучета",
```

```
  "Кучета",
```

```
  "Кучета",
```

```
  "И двете",
```

```
  "Кучета",
```

```
  "Кучета",
```

```
  "И двете",
```

```
  "Котки",
```

```
  "Кучета",
```

```
  "Котки",
```

```
  "Кучета",
```

```
  "Котки",
```

```
  "Котки",
```

```
  "И двете",
```

```
  "Кучета",
```

```
  "Кучета",
```

```
  "Кучета",
```

```
  "Кучета",
```

```
  "Кучета",
```

```
"Кучета",  
"Кучета",  
"Кучета",  
"Кучета",  
"Котки",  
"Кучета",  
"Кучета",  
"И двете",  
"И двете",  
"Котки")
```

### ★ Анализ

Отново използвам функцията `table()`, защото при категориите променливи честотата се вижда най-добре чрез таблици.

```
table_dogs_or_cats <- table(dogs_or_cats)  
table_dogs_or_cats
```

```
dogs_or_cats  
И двете  Котки  Кучета  
14      9    32
```

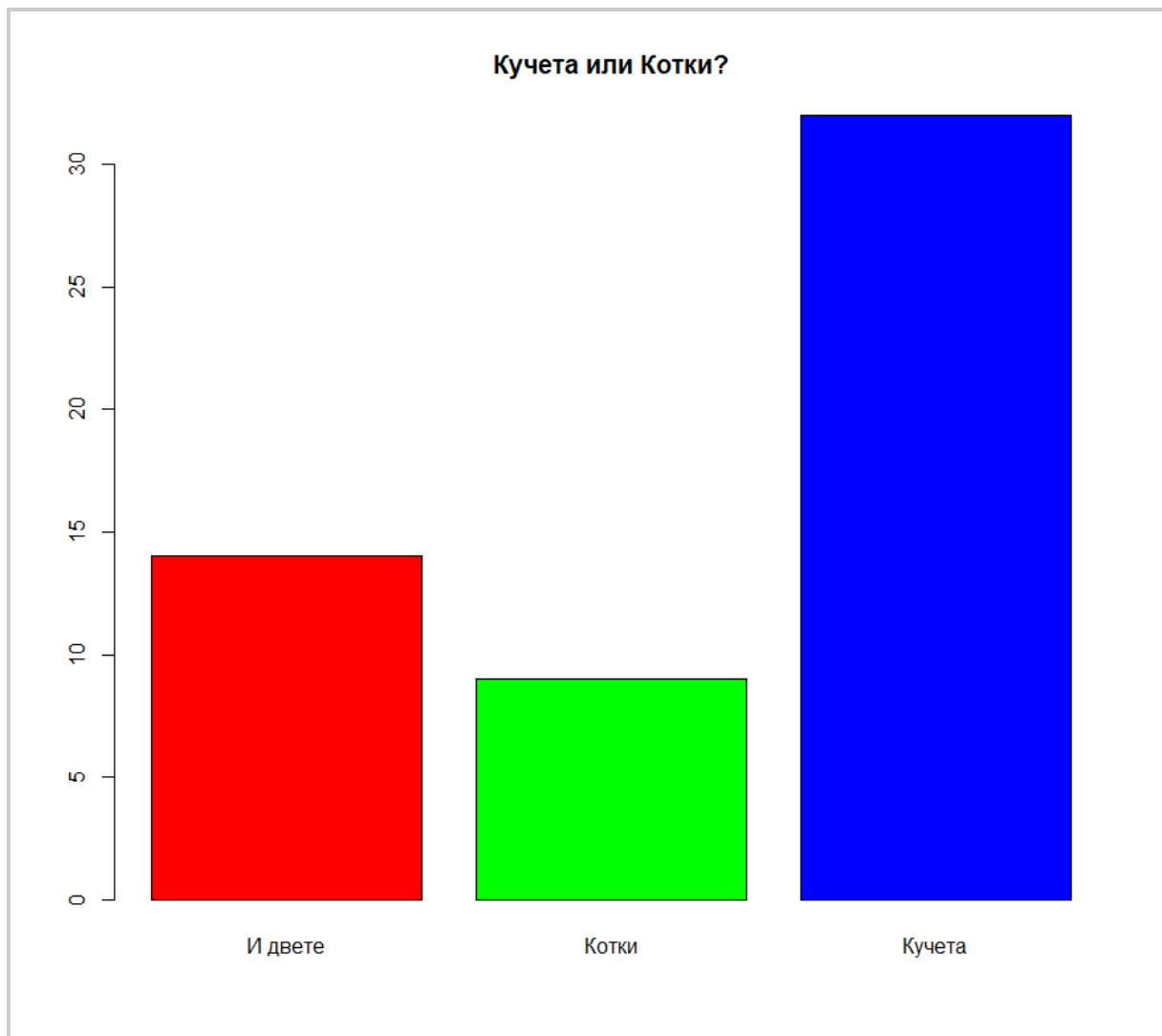
С функцията `prop.table()` изобразявам процентното разпределение.

```
prop_table_dogs_or_cats <- prop.table(table_dogs_or_cats)  
prop_table_dogs_or_cats
```

```
dogs_or_cats  
И двете  Котки  Кучета  
0.2545455 0.1636364 0.5818182
```

Чрез barplot показвам частичното разпределение на категориите променливи.

```
barplot(height = table_dogs_or_cats, col = rainbow(3), main = "Кучета или  
Котки?")
```



На графиката се вижда, че броят на хората които предпочитат кучета е най-голям - над 30. Анкетираните, които предпочитат котки са най-малко.

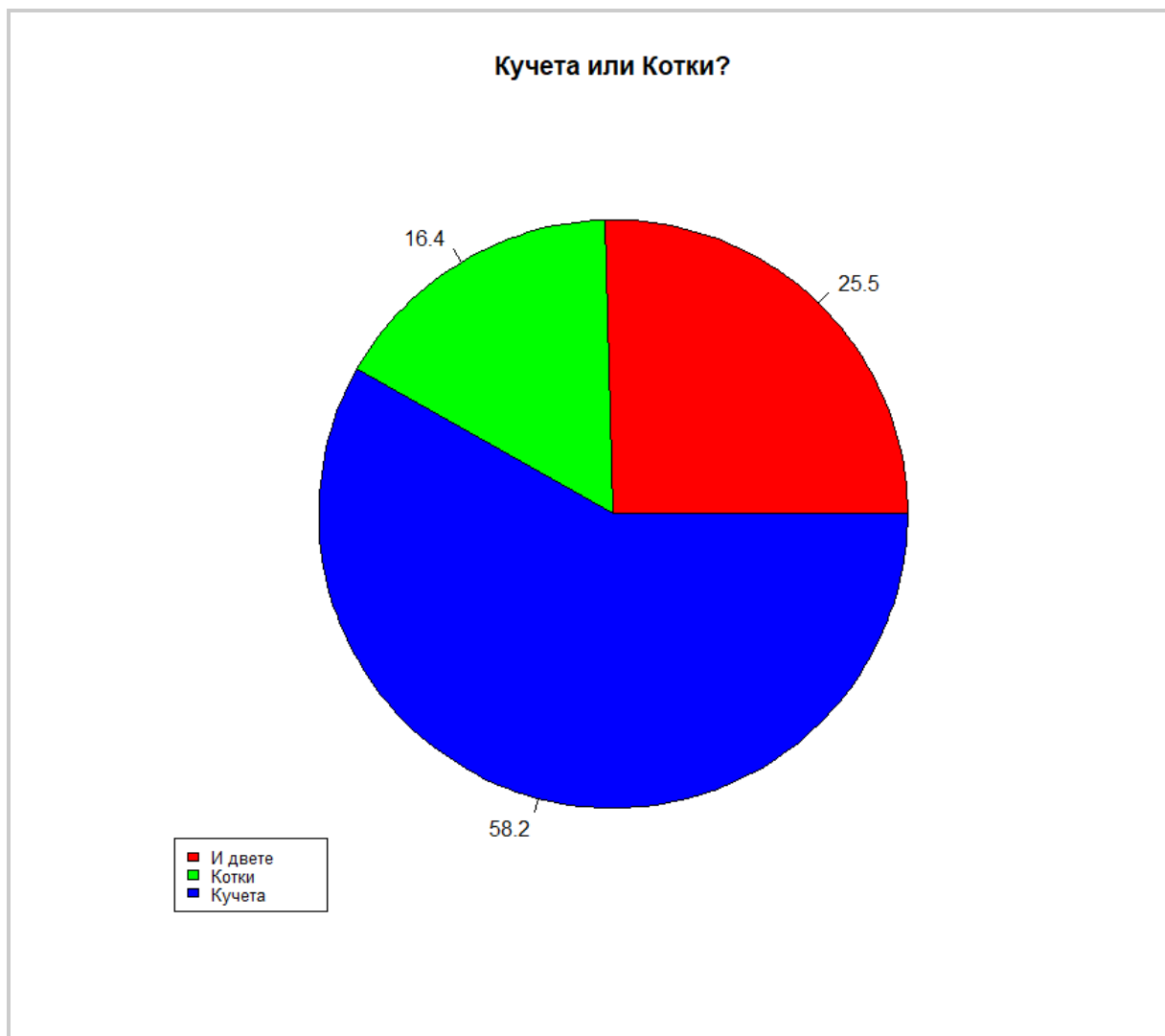
За процентното разпределение на данните използвам pie chart

```
piepercent_dogs_or_cats<-
```

```
round(100*table_dogs_or_cats/sum(table_dogs_or_cats), 1)
```

```
pie(table_dogs_or_cats, labels = piepercent_dogs_or_cats, main = "Кучета или  
Котки?", col = rainbow(length(table_dogs_or_cats)))
```

```
legend(x = "bottomleft", legend = c("И двете", "Котки", "Кучета"), cex = 0.8,  
fill = rainbow(length(table_dogs_or_cats)))
```



## 2.4. Въпрос: За Вас кучето е:

★ Въвеждане на данните

```
dog_is = c("Най-добър приятел",  
  "Най-добър приятел",  
  "Като малко дете",  
  "Най-добър приятел", "Като малко дете",  
  "Най-добър приятел", "Като малко дете",  
  "Най-добър приятел",  
  "Най-добър приятел",  
  "Най-добър приятел",  
  "Най-добър приятел",  
  "Най-добър приятел", "Като малко дете",  
  "Обикновено животно",  
  "Като малко дете",  
  "Най-добър приятел", "Като малко дете",  
  "Като малко дете",  
  "Най-добър приятел", "Като малко дете",  
  "Най-добър приятел", "Като малко дете",  
  "Най-добър приятел", "Като малко дете",  
  "Като малко дете",  
  "Най-добър приятел",  
  "Най-добър приятел",  
  "Като малко дете", "Обикновено животно",  
  "Най-добър приятел",  
  "Най-добър приятел",  
  "Най-добър приятел", "Като малко дете",  
  "Най-добър приятел", "Като малко дете",  
  "Най-добър приятел", "Като малко дете",  
  "Най-добър приятел", "Като малко дете",  
  "Като малко дете",  
  "Най-добър приятел",  
  "Като малко дете",  
  "Най-добър приятел", "Като малко дете",  
  "Най-добър приятел",  
  "Като малко дете",  
  "Като малко дете",  
  "Обикновено животно",  
  "Обикновено животно",  
  "Най-добър приятел", "Като малко дете",  
  "Обикновено животно",  
  "Най-добър приятел",  
  "Като малко дете",  
  "Най-добър приятел", "Като малко дете",  
  "Като малко дете",  
  "Като малко дете",  
  "Като малко дете",  
  "Като малко дете",
```

```
"Като малко дете",  
"Най-добър приятел",  
"Като малко дете",  
"Най-добър приятел", "Като малко дете",  
"Обикновено животно",  
"Най-добър приятел",  
"Обикновено животно",  
"Най-добър приятел", "Обикновено животно",  
"Най-добър приятел", "Като малко дете",  
"Обикновено животно")
```

### ★ Анализ

Тук използвам функцията `table()`, защото при категориите променливи честотата се вижда най-добре чрез таблици.

```
table_dog_is <- table(dog_is)  
table_dog_is
```

```
dog_is  
Като малко дете Най-добър приятел Обикновено животно  
30 33 9
```

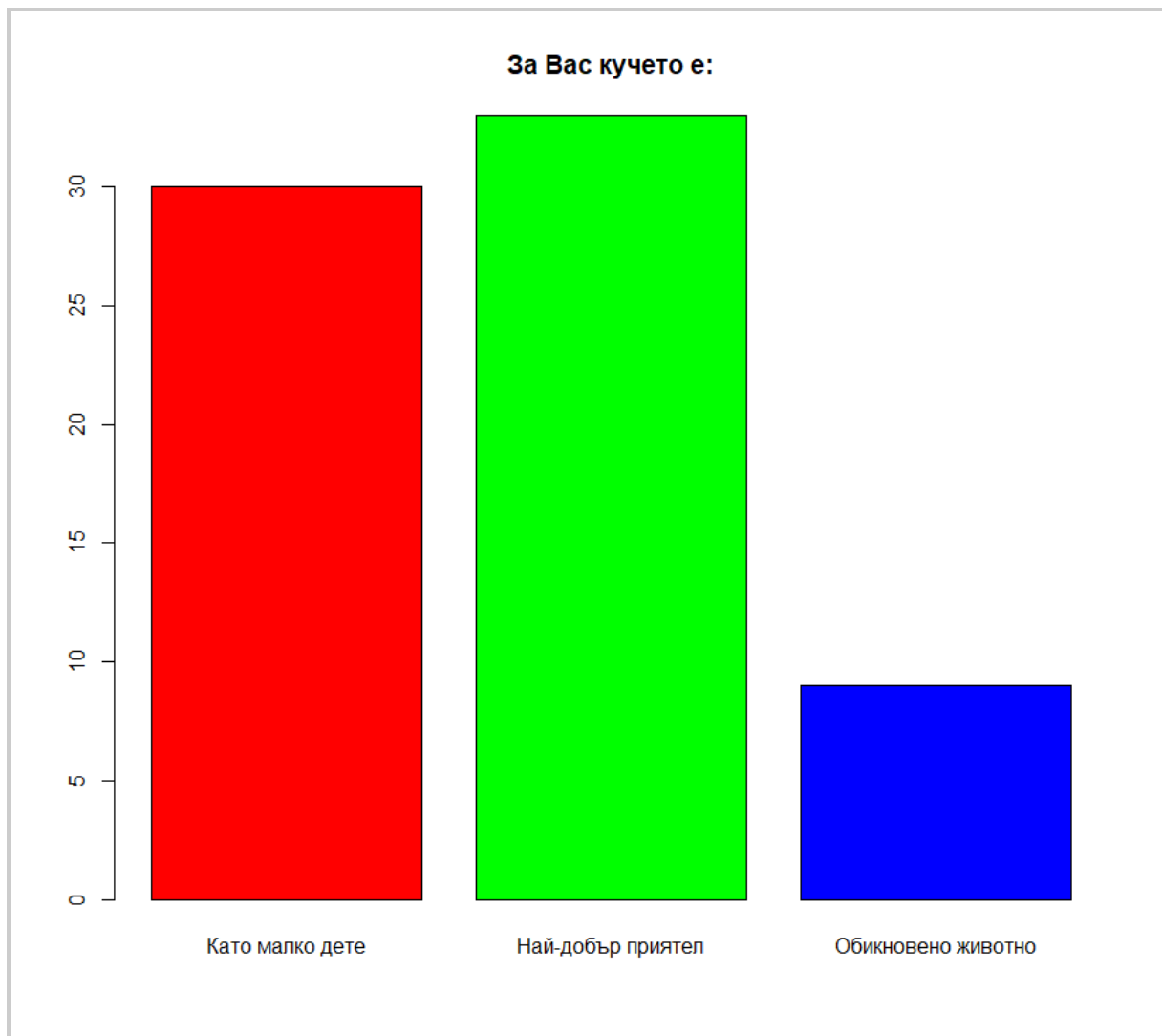
С функцията `prop.table()` изобразявам процентното разпределение.

```
prop_table_dog_is <- prop.table(table_dog_is)  
prop_table_dog_is
```

```
Като малко дете Най-добър приятел Обикновено животно  
0.4166667 0.4583333 0.1250000
```

Чрез barplot показвам частичното разпределение на категориите променливи.

```
barplot(height = table_dog_is, col = rainbow(3), main = "За Вас кучето е:")
```



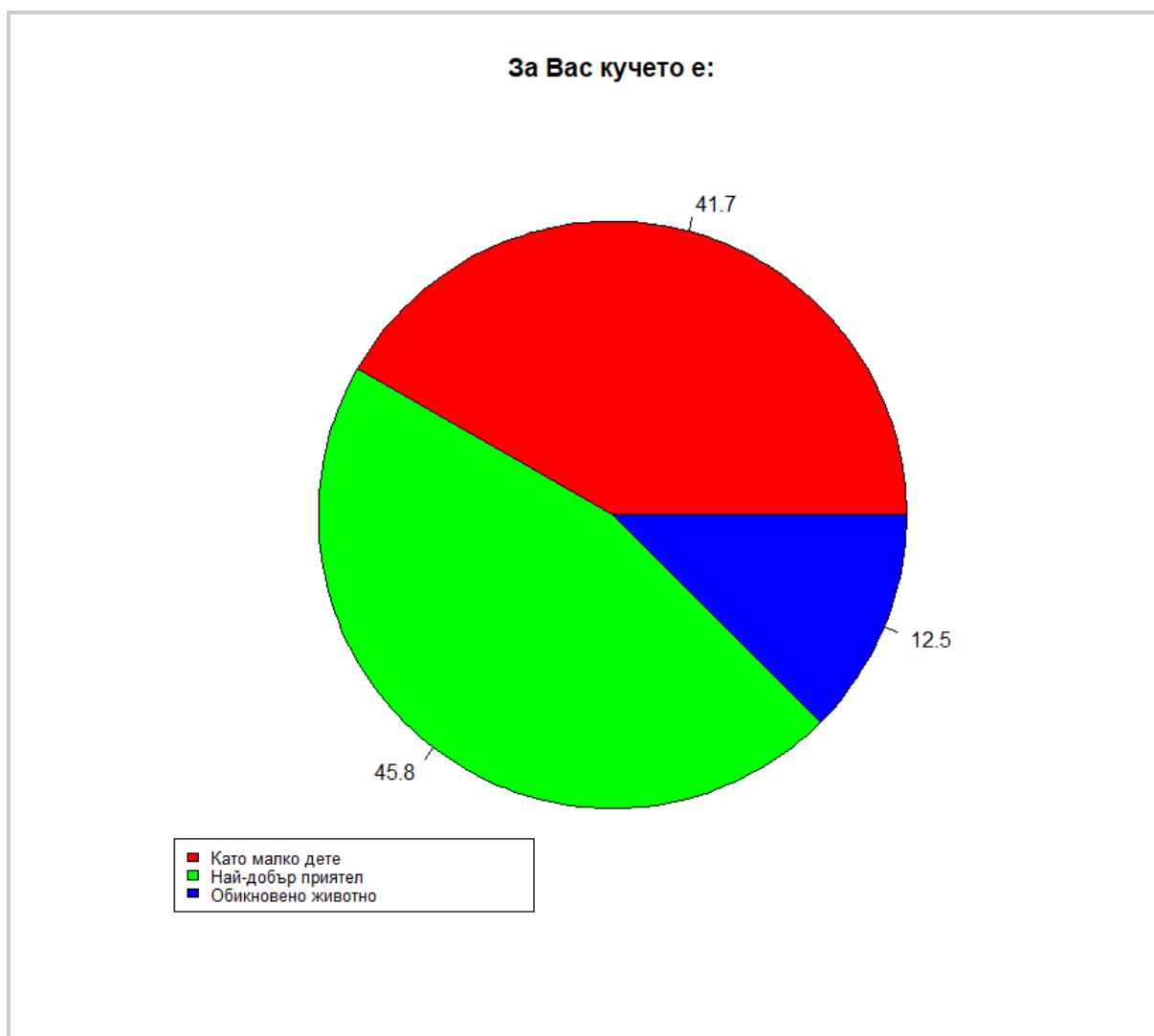
На графиката се вижда, че броят на хората, които кучетата за най-добър приятел е най-голям - над 30. Анкетираните, които смятат, че е просто обикновено животно са най-малко.



За процентното разпределение на данните използвам pie chart  
`piepercent_dog_is<- round(100*table_dog_is/sum(table_dog_is), 1)`

`pie(table_dog_is, labels = piepercent_dog_is, main = "За Вас кучето е:", col = rainbow(length(table_dog_is)))`

`legend(x = "bottomleft", legend = c("Като малко дете", "Най-добър приятел", "Обикновено животно"), cex = 0.8, fill = rainbow(length(table_dog_is)))`



## 2.5. Въпрос: Любим цвят куче:

### ★ Въвеждане на данните

Тук демонстрирам втори начин за въвеждане на информация:

```
fav_dog_color <- c(rep("бял", 15), rep("златист", 16), rep("кафяв", 18), rep("сив", 10), rep("черен", 22), rep("шарен", 14))
```

### ★ Анализ

Използвам функцията table(), защото при категорийните променливи честотата се вижда най-добре чрез таблици.

```
table_fav_dog_color <- table(fav_dog_color)
table_fav_dog_color
```

```
fav_dog_color
бял златист кафяв сив черен шарен
15  16   18  10  22  14
```

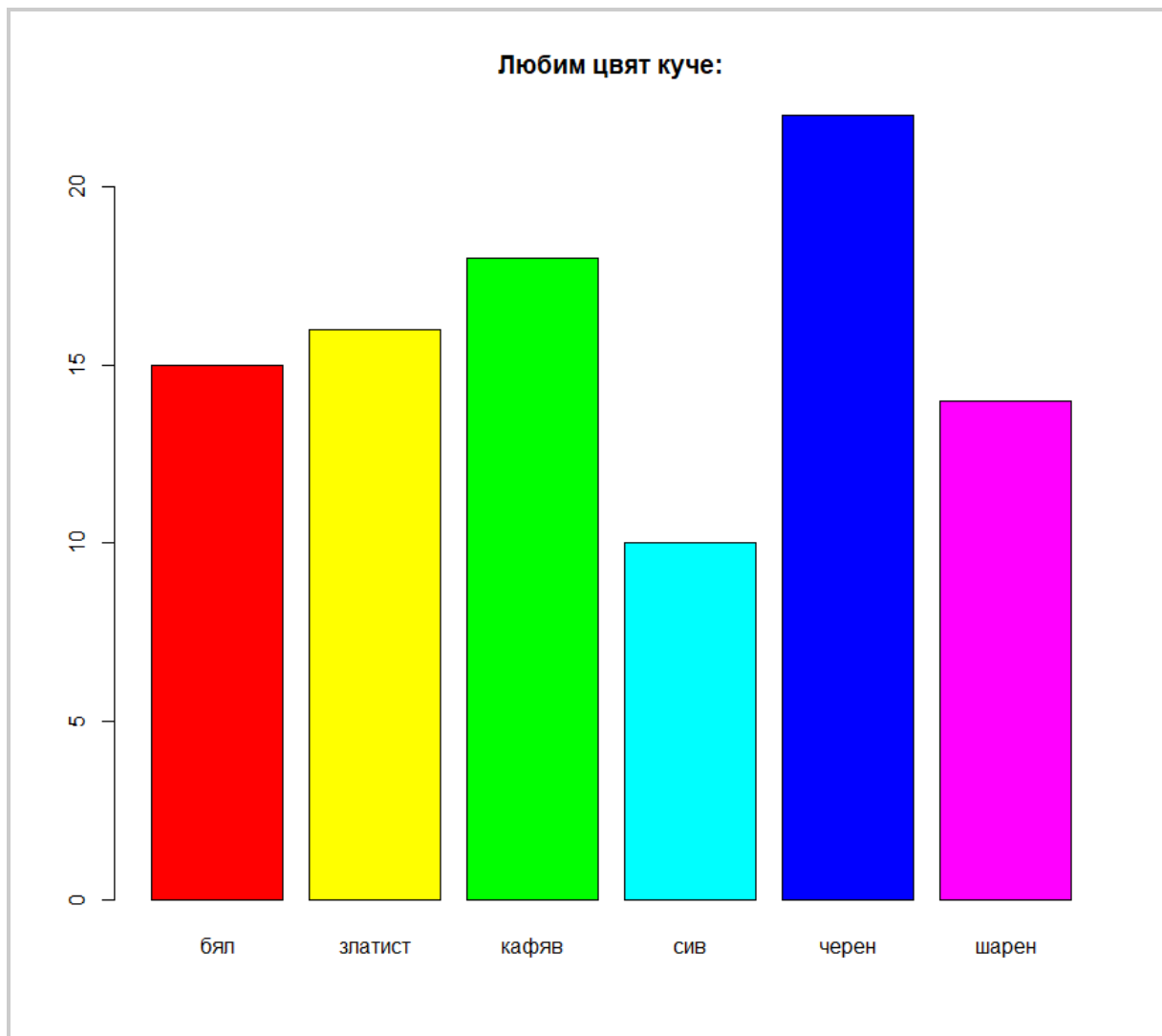
С функцията prop.table() изобразявам процентното разпределение.

```
prop_table_fav_dog_color <- prop.table(table_fav_dog_color)
prop_table_fav_dog_color
```

```
fav_dog_color
бял златист кафяв сив черен шарен
0.1578947 0.1684211 0.1894737 0.1052632 0.2315789 0.1473684
```

Чрез barplot показвам частичното разпределение на категориите променливи.

```
barplot(height = table_fav_dog_color, col =  
rainbow(length(table_fav_dog_color)), main = "Любим цвят куче:")
```



Може да се забележи, че хората, които предпочитат черни кучета са най-много. Анкетираните, които харесват сиви кучета са най-малко.

За процентното разпределение на данните използвам pie chart

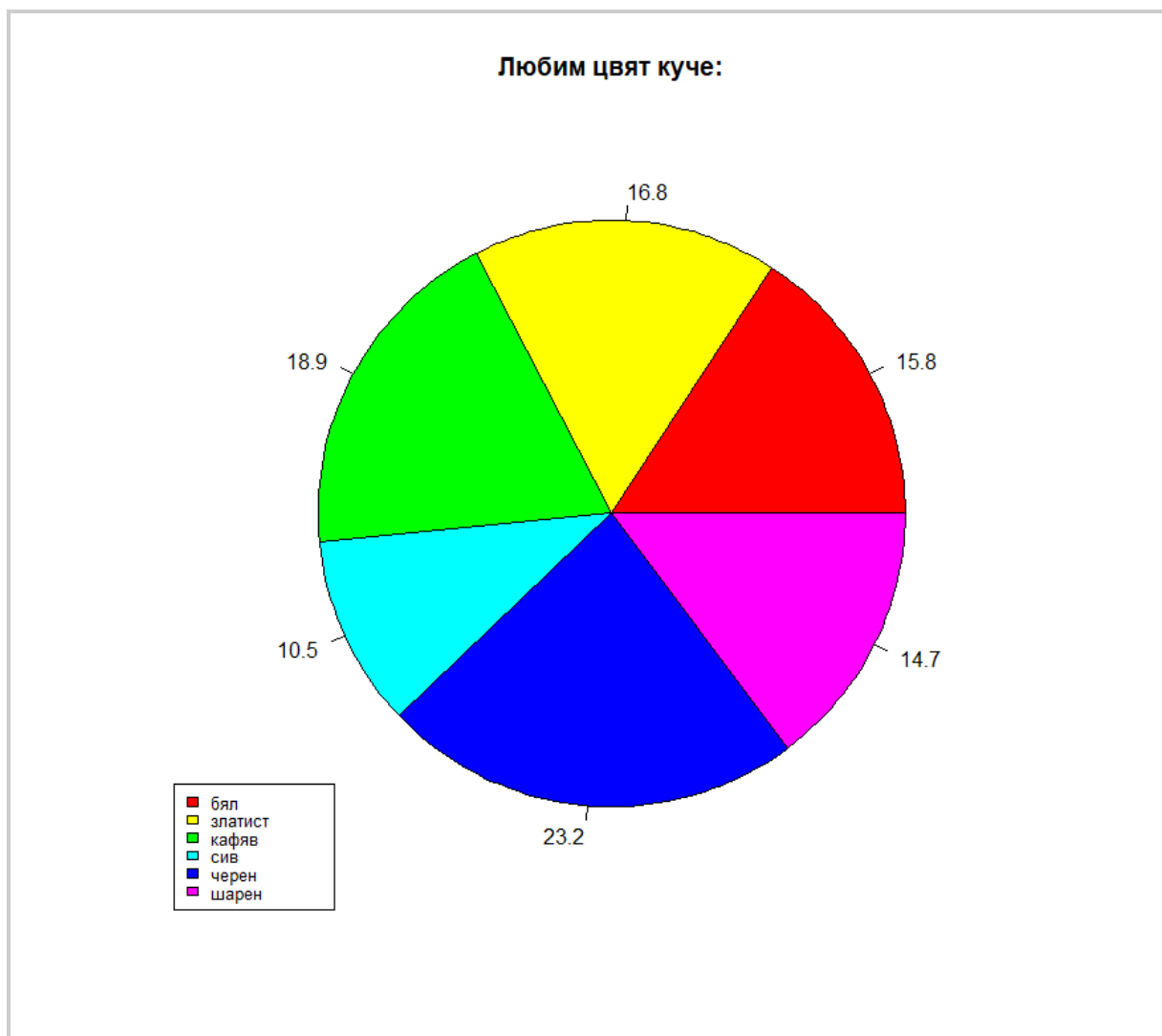
```
piepercent_fav_dog_color<-
```

```
round(100*table_fav_dog_color/sum(table_fav_dog_color), 1)
```

```
pie(table_fav_dog_color, labels = piepercent_fav_dog_color, main = "Любим цвят  
куче:", col = rainbow(length(table_fav_dog_color)))
```

```
legend(x = "bottomleft", legend = c("бял", "златист", "кафяв", "сив", "черен",  
"шарен"), cex = 0.8,
```

```
fill = rainbow(length(table_fav_dog_color)))
```



2.6. Въпрос: Колко пари месечно бихте  
отделили/отделяте за кучето си? (в лв)

★ Въвеждане на данните

```
monthly_expenses <- c(1000,  
100,  
300,  
70,  
1000,  
100,  
200,  
500,  
300,  
250,  
0,  
300,  
50,  
50,  
300,  
100,  
100,  
200,  
40,  
500,  
70,  
200,  
50,  
50,  
200,  
100,  
100,  
200,  
100,  
1000,  
400,  
1200,  
50,  
0,  
30,  
0,  
0,  
200,  
0,  
900,
```

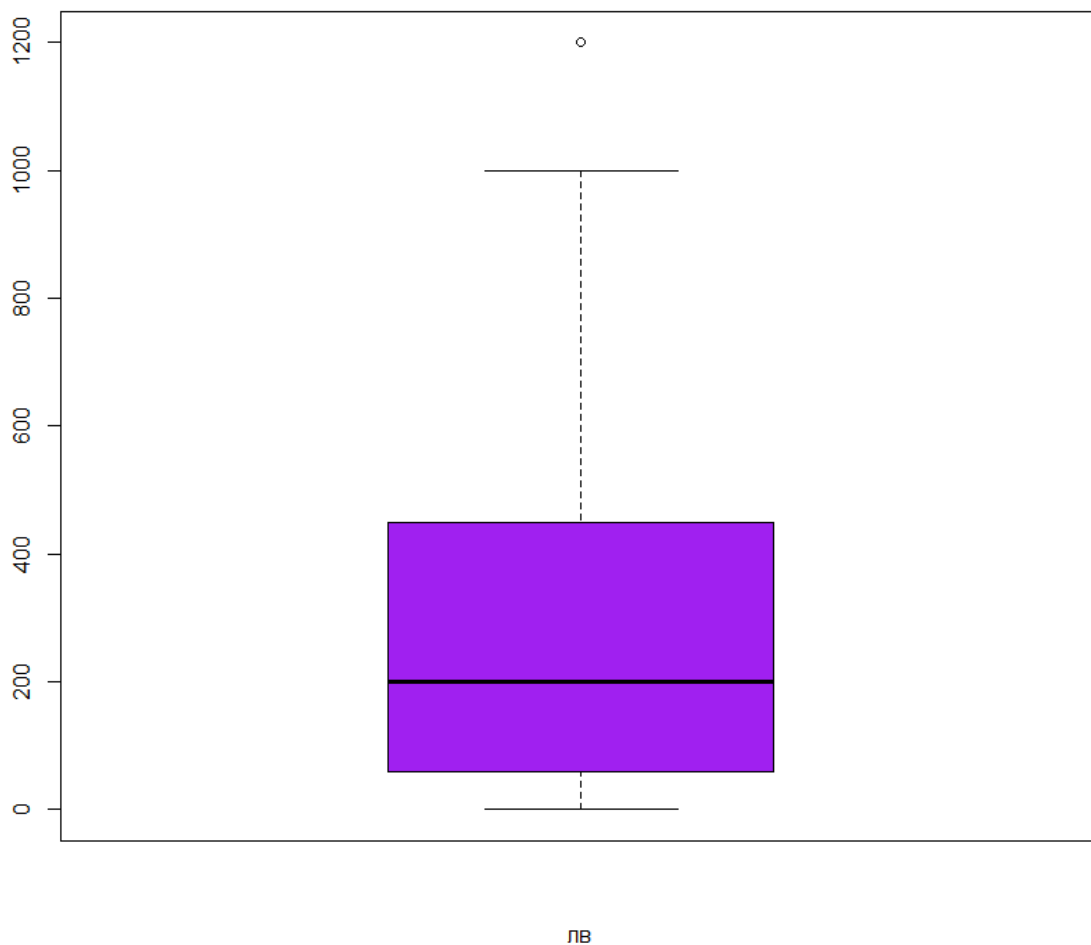
```
900,  
100,  
100,  
200,  
100,  
600,  
900,  
800,  
100,  
0,  
750,  
800,  
400,  
1000,  
0)
```

### ★ Анализ

Използвам boxplot за откриване на потенциални outlier-и.

```
boxplot(monthly_expenses, col = "purple", main = "Колко пари месечно бихте  
отделили/отделяте за кучето си?", xlab = "лв") #outlier
```

Колко пари месечно бихте отделили/отделяте за кучето си?



Има открит потенциален outlier.

Намирам модата (най-често срещаната стойност във вектора):

```
modeFunction <- function(x) {  
  res_table <- table(x)  
  return(names(res_table)[res_table == max(res_table)])  
}  
modeFunction(monthly_expenses) #мода
```

[1] "100"

=> Повечето хора биха отделили 100 лв.

Намирам медианата (средна стойност):

```
medianFunction <- function(x) {
```

```

x_sorted <- sort(x)
nn <- length(x_sorted)
if(nn %% 2 == 0) {
  return(mean(x_sorted[nn/2 + c(0, 1)]))
} else {
  return(x_sorted[round(nn/2 + 0.25)])
}
}
medianFunction(monthly_expenses) #медиана

```

```
[1] 200
```

Използвам summary - описателна статистика за центъра на разпределението. Тя показва минималната стойност, 1 квартил, 2 квартил (медиана), 3 квартил и максималната стойност.

```
summary(monthly_expenses) #квартили
```

```

Min. 1st Qu. Median Mean 3rd Qu. Max.
0.0   60.0   200.0  310.2  450.0 1200.0

```

Намирам дисперсията (вариацията):

```
var(monthly_expenses)
```

```
[1] 119224
```

Намирам стандартното отклонение, което е оценка на вариацията, която показва колко се отклоняват наблюденията от очакването. То е производно на вариацията и е равно на корен квадратен от дисперсията.

```
sd(monthly_expenses) #sd
```

```
[1] 345.2883
```

Използвам range() за да намеря обхвата на интервала от най-ниската до най-високата стойност.

```
rangeFunction <- function(x) {
```

```
  max(x) - min(x)
```

```
}
```

```
rangeFunction(dogs_count) #range
```



```
[1] 1200
```

Намирам IQR - interquartile range на данните, който е равен на трети квартил минус първи квартил.

```
IQR(monthly_expenses) #IQR
```

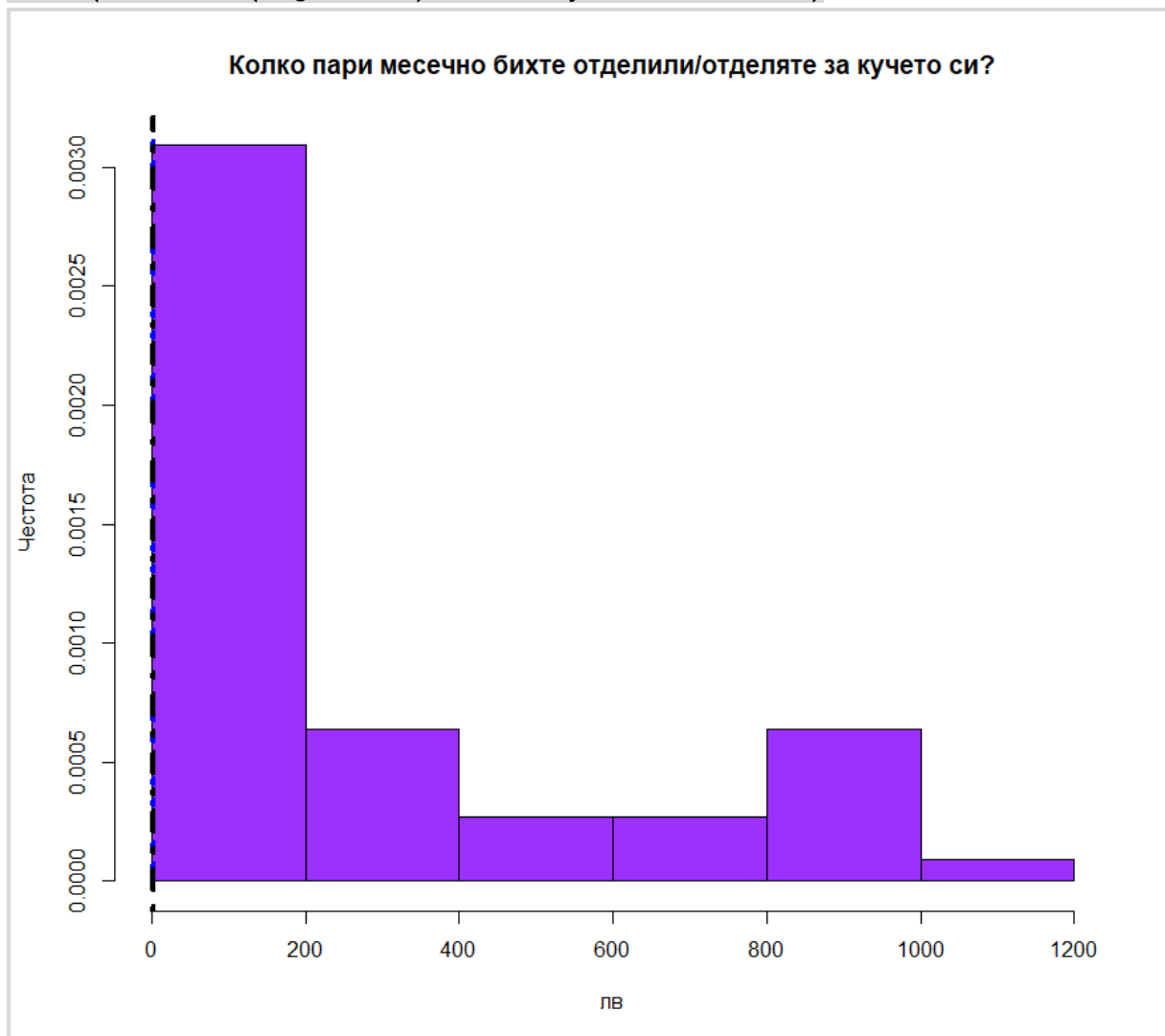
```
[1] 390
```

Използвам хистограма за разпределението на непрекъснатите променливи и синя вертикална линия показваща мястото на средната стойност и черна-медианата.

```
hist(monthly_expenses, main = "Колко пари месечно бихте отделили/отделяте за кучето си?", xlab = "лв", ylab = "Честота", col = "purple1", prob = T)
```

```
abline(v = mean(dogs_count), lwd = 4, lty = 3, col = "blue")
```

```
abline(v = median(dogs_count), lwd = 4, lty = 4, col = "black")
```



## 2.7. Въпрос: Кое е най-подходящото място за отглеждане на куче?

★ Въвеждане на данните

```
most_suitable_place <- c("Къща", "Двор",  
    "Двор",  
    "Къща",  
    "Къща",  
    "Къща", "Двор",  
    "Къща",  
    "Двор",  
    "Къща",  
    "Къща", "Двор",  
    "Апартамент", "Къща",  
    "Къща", "Двор",  
    "Къща",  
    "Двор",  
    "Къща",  
    "Двор",  
    "Двор",  
    "Двор",  
    "Двор",  
    "Апартамент", "Къща", "Двор",  
    "Къща",  
    "Апартамент", "Къща", "Двор",  
    "Къща", "Двор",  
    "Двор",  
    "Къща",  
    "Двор",  
    "Двор",  
    "Апартамент", "Къща",  
    "Къща", "Двор",  
    "Двор",  
    "Апартамент", "Къща", "Двор",  
    "Къща", "Двор",  
    "Апартамент", "Къща", "Двор",  
    "Двор",  
    "Двор",  
    "Къща",  
    "Двор",  
    "Къща",  
    "Къща", "Двор",  
    "Двор",  
    "Къща",  
    "Къща",  
    "Къща", "Двор",  
    "Двор",  
    "Къща",  
    "Къща",  
    "Къща", "Двор",  
    "Къща", "Двор",
```

```

"Къща",
"Апартамент",
"Апартамент", "Къща",
"Къща",
"Апартамент",
"Къща", "Двор",
"Двор",
"Къща",
"Апартамент",
"Къща",
"Къща",
"Двор")

```

## ★ Анализ

Използвам функцията `table()`, защото при категорийните променливи честотата се вижда най-добре чрез таблици.

```

table_most_suitable_place <- table(most_suitable_place)
table_most_suitable_place

```

```

most_suitable_place
Апартамент  Двор  Къща
      10      32      35

```

С функцията `prop.table()` изобразявам процентното разпределение.

```

prop_table_most_suitable_place <- prop.table(table_most_suitable_place)
prop_table_most_suitable_place

```

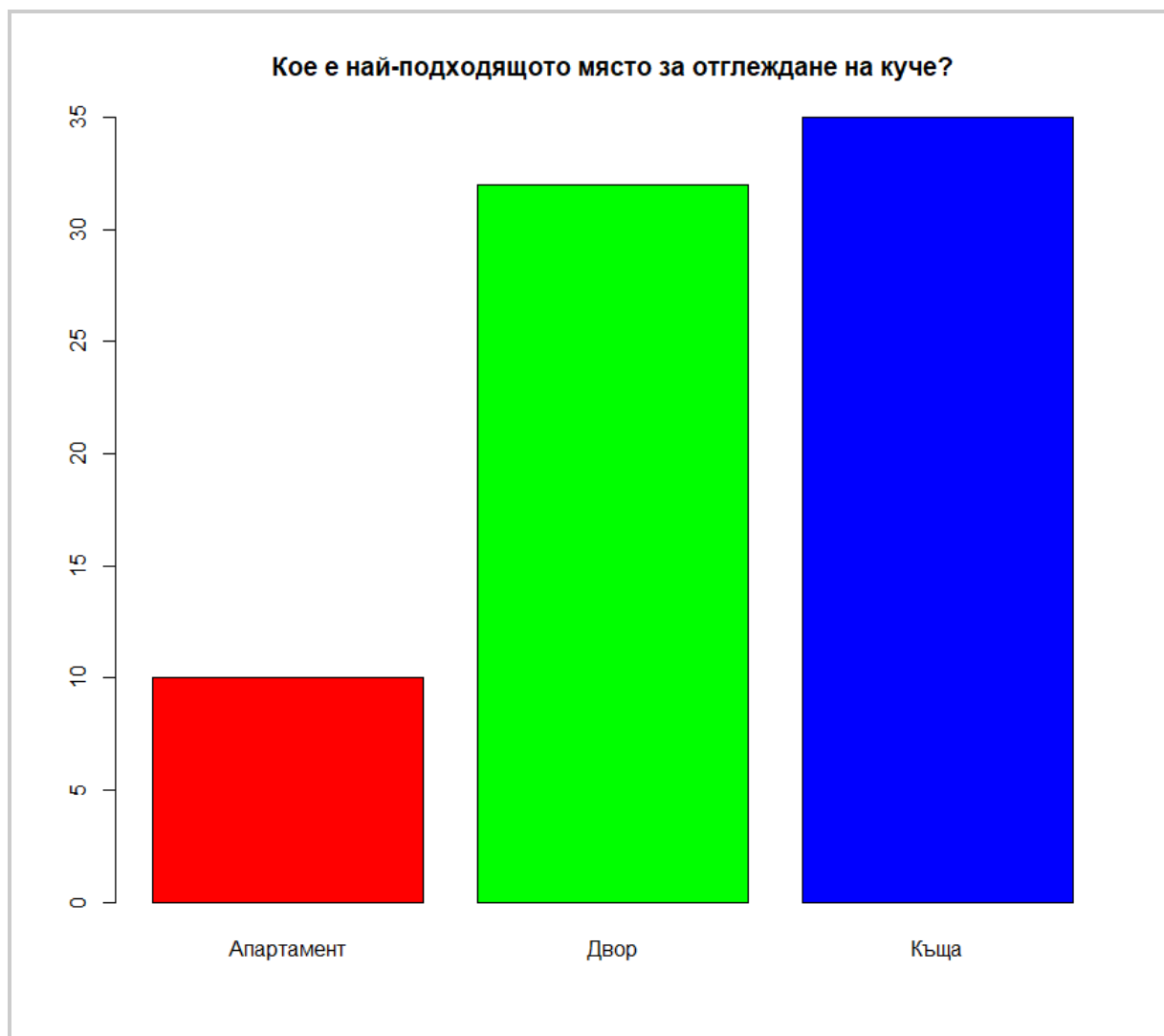
```

most_suitable_place
Апартамент  Двор  Къща
0.1298701 0.4155844 0.4545455

```

Чрез barplot показвам частичното разпределение на категориите променливи.

```
barplot(height = table_most_suitable_place, col =  
rainbow(length(table_most_suitable_place)), main = "Кое е най-подходящото място  
за отглеждане на куче?")
```



Може да се забележи, че анкетираните, които считат апартамента за най-подходящо място за отглеждане на куче, са най-малко.

За процентното разпределение на данните използвам pie chart

```
piepercent_most_suitable_place<-
```

```
round(100*table_fav_dog_color/sum(table_most_suitable_place), 1)
```

```
pie(table_most_suitable_place, labels = piepercent_most_suitable_place, main =
```

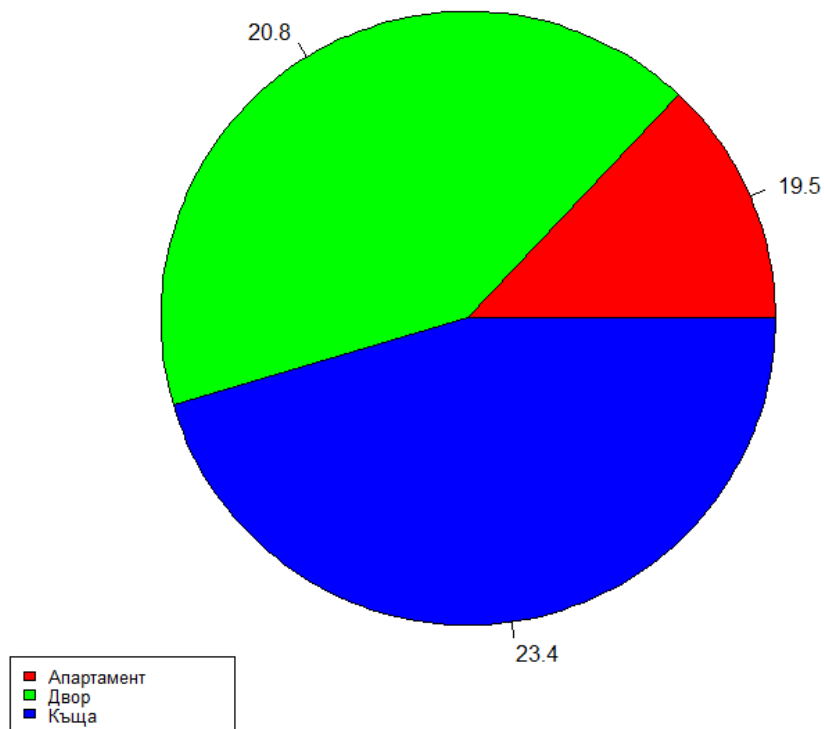
```
"Кое е най-подходящото място за отглеждане на куче?", col =
```

```
rainbow(length(table_most_suitable_place)))
```

```
legend(x = "bottomleft", legend = c("Апартамент", "Двор", "Къща"), cex = 0.8,
```

```
fill = rainbow(length(table_most_suitable_place)))
```

Кое е най-подходящото място за отглеждане на куче?



2.8. Въпрос: Бихте ли се сближили с човек, който не обича кучета?

★ Въвеждане на данните

```
befriend_non_dog_lover <- c(rep("Зависи от човека", 30), rep("Да", 13), rep("Не", 12))
```

## ★ Анализ

Използвам функцията `table()`, защото при категорийните променливи честотата се вижда най-добре чрез таблици.

```
table_befriend_non_dog_lover <- table(befriend_non_dog_lover)
table_befriend_non_dog_lover
```

```
befriend_non_dog_lover
      Да Зависи от човека      Не
      13          30          12
```

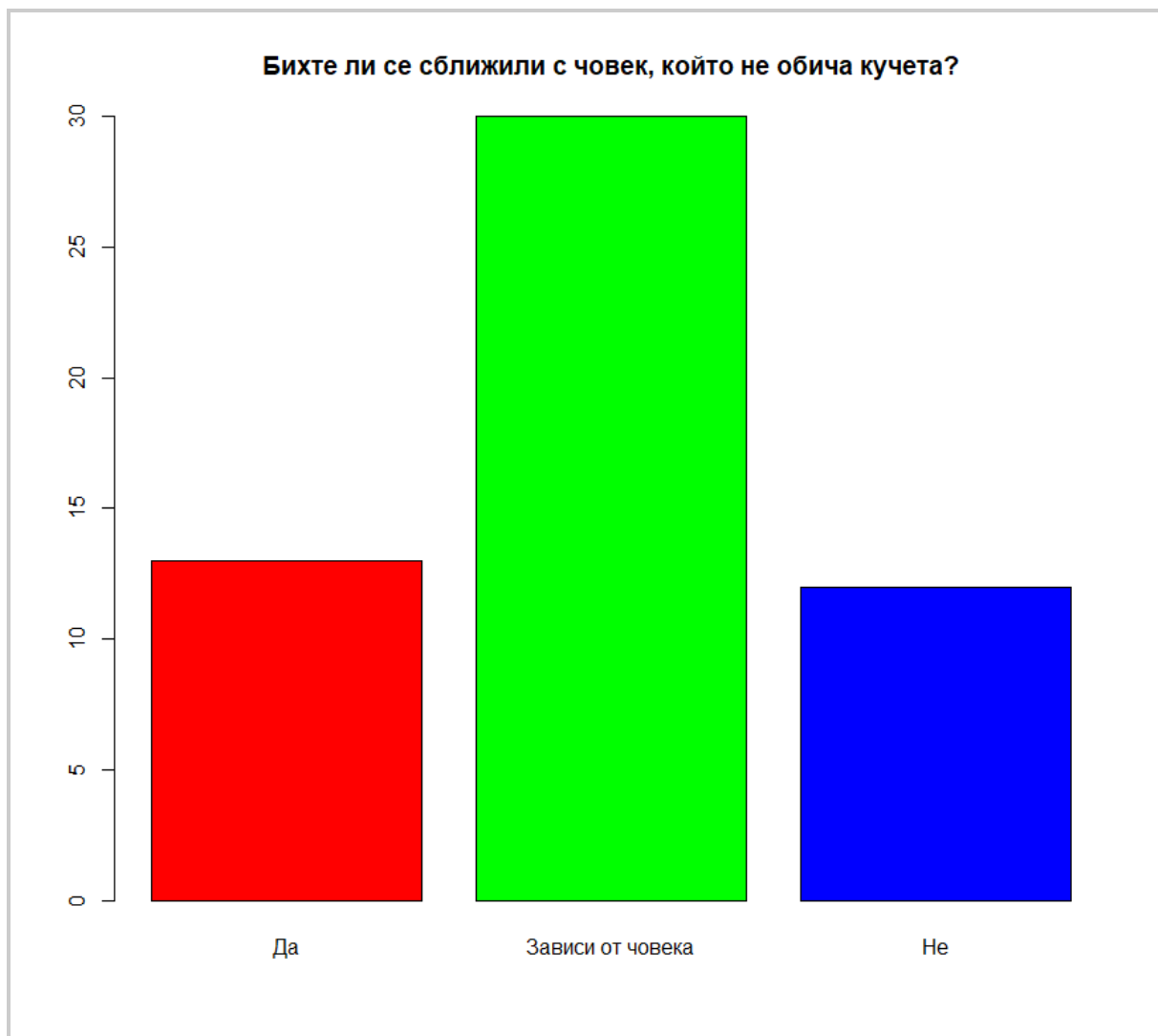
С функцията `prop.table()` изобразявам процентното разпределение.

```
prop_table_befriend_non_dog_lover <-
prop.table(table_befriend_non_dog_lover)
prop_table_befriend_non_dog_lover
```

```
befriend_non_dog_lover
      Да      Зависи от човека      Не
0.2363636  0.5454545  0.2181818
```

Чрез `barplot` показвам частичното разпределение на категорийните променливи.

```
barplot(height = table_befriend_non_dog_lover, col =
rainbow(length(table_befriend_non_dog_lover)), main = "Бихте ли се сближили с
човек, който не обича кучета?")
```



Може да се забележи, че анкетираните, които биха се сприятелили с човек, който не обича кучета, в зависимост от самия него, са най-много - 30.

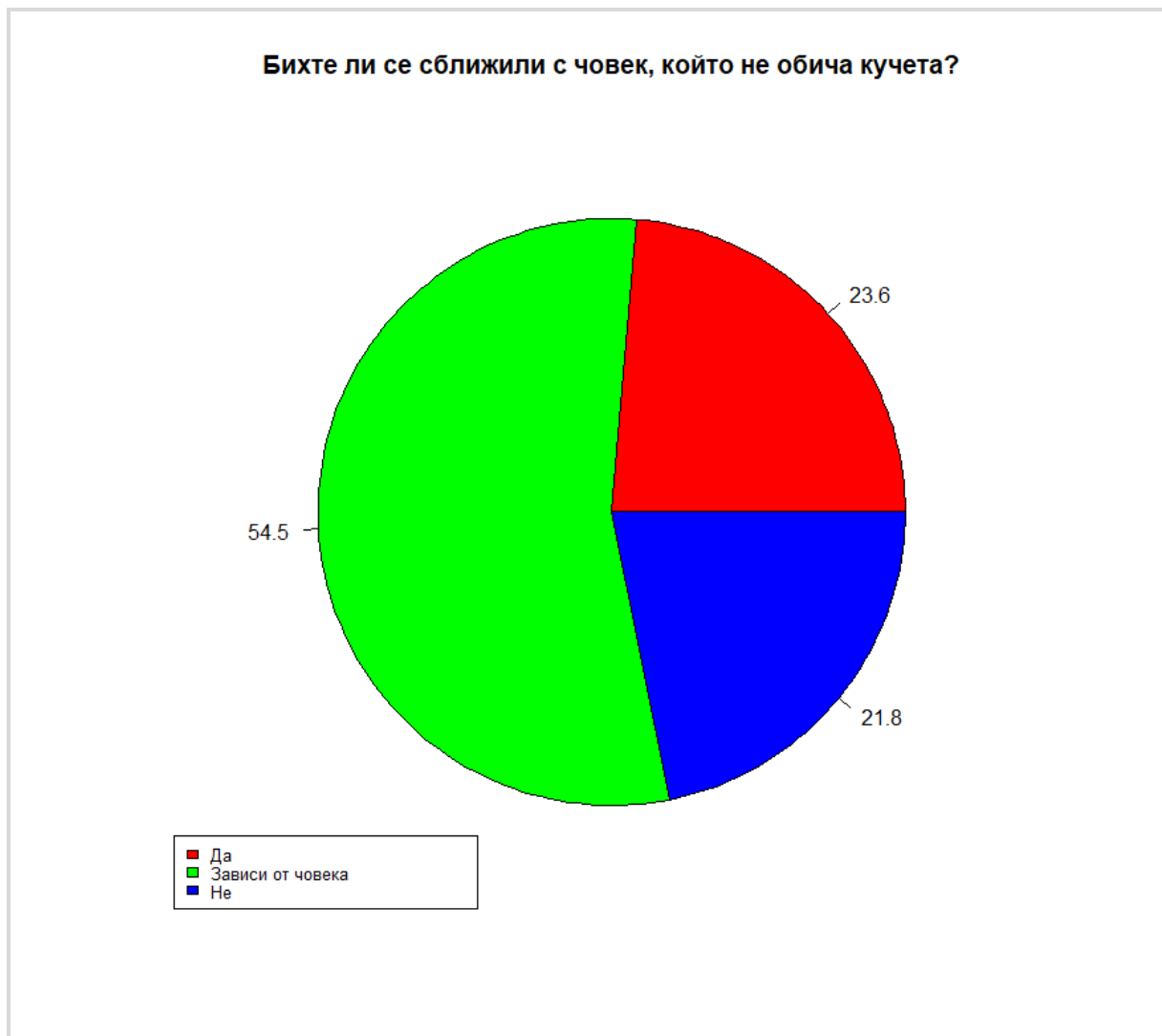
За процентното разпределение на данните използвам pie chart

```
piepercent_befriend_non_dog_lover<-  
round(100*table_befriend_non_dog_lover/sum(table_befriend_non_dog_lover), 1)
```



```
pie(table_befriend_non_dog_lover, labels = piepercent_befriend_non_dog_lover,
main = "Бихте ли се сближили с човек, който не обича кучета?", col =
rainbow(length(table_befriend_non_dog_lover)))
```

```
legend(x = "bottomleft", legend = c("Да", "Зависи от човека", "Не"), cex = 0.8,
fill = rainbow(length(table_befriend_non_dog_lover)))
```



## 2.9. Въпрос: Имали ли сте куче до сега?

★ Въвеждане на данните

```
has_had_dog <- c(rep("Да", 35), rep("Не", 20))
```

## ★ Анализ

Използвам функцията `table()`, защото при категориите променливи честотата се вижда най-добре чрез таблици.

```
table_has_had_dog <- table(has_had_dog)
table_has_had_dog
```

```
has_had_dog
Да Не
35 20
```

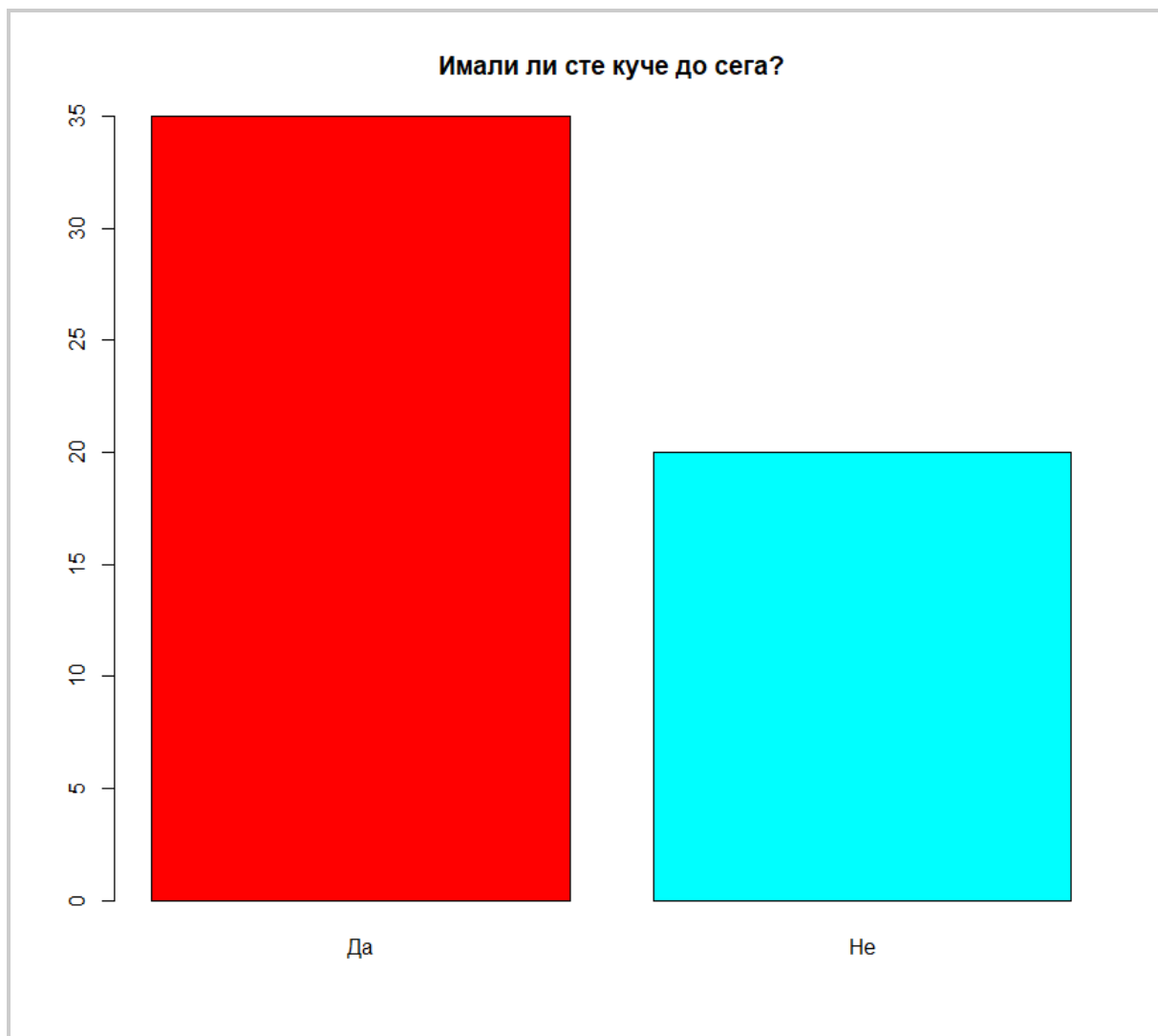
С функцията `prop.table()` изобразявам процентното разпределение.

```
prop_table_has_had_dog <- prop.table(table_has_had_dog)
prop_table_has_had_dog
```

```
has_had_dog
Да      Не
0.6363636 0.3636364
```

Чрез `barplot` показвам частичното разпределение на категориите променливи.

```
barplot(height = table_has_had_dog, col =
rainbow(length(table_has_had_dog)), main = "Имали ли сте куче до сега? ")
```



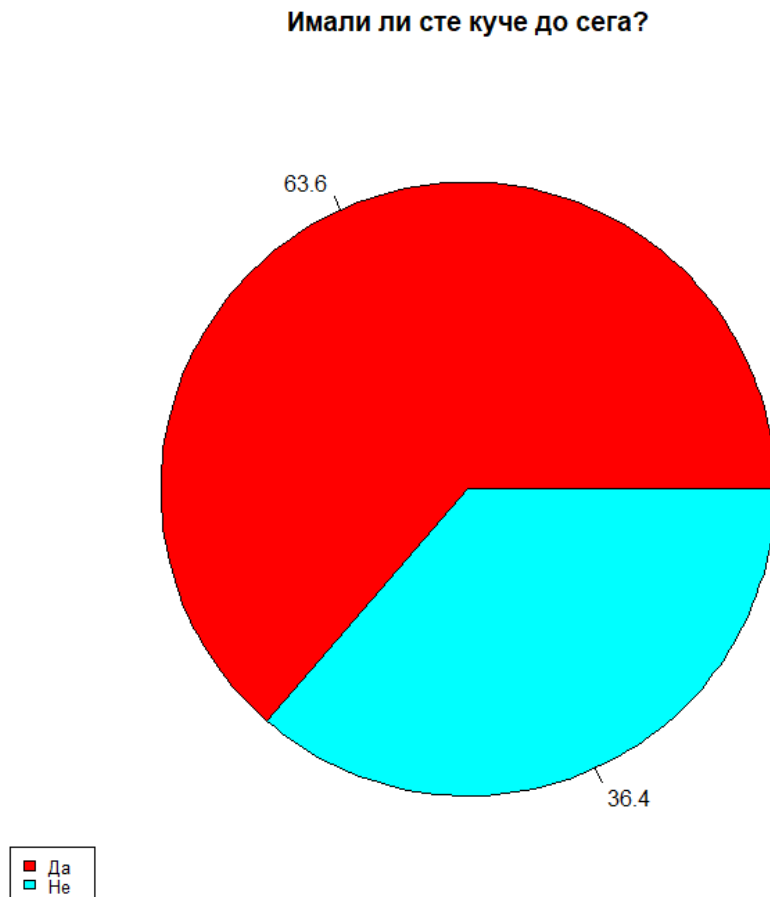
Може да се забележи, че мнозинството от анкетираните са имали куче.

За процентното разпределение на данните използвам pie chart

```
piepercent_has_had_dog<-  
round(100*table_has_had_dog/sum(table_has_had_dog), 1)
```

```
pie(table_has_had_dog, labels = piepercent_has_had_dog, main = "Имали ли сте  
куче до сега? ", col = rainbow(length(table_has_had_dog)))
```

```
legend(x = "bottomleft", legend = c("Да", "Не"), cex = 0.8,  
fill = rainbow(length(table_has_had_dog)))
```



### 3. Многомерен анализ

Обединявам досега създадените вектори в data frame

```
dogs_data <- data.frame(has_dog, dogs_count, dogs_or_cats, monthly_expenses,  
  befriend_non_dog_lover, has_had_dog)
```

### ❖ Числова vs числова

Анализирам връзката между `dogs_count` (броят кучета, които хората биха искали да имат) и `monthly_expenses` (разходите, които биха давали месечно). Използвам `plot` за графично представяне на връзката.

```
plot(dogs_data$dogs_count, dogs_data$monthly_expenses)
```

### ❖ Категорийна vs числова