

**Софийски университет „Св. Климент Охридски“**

*Факултет по математика и информатика*

*Специалност: “Информационни системи”*

*Курс: 2, Група: 1*

*Дисциплина: “Статистика и емпирични методи - практикум”*

# **КУРСОВ ПРОЕКТ**

**Тема: Жените Шофьори**

***Изготвил:***

**Боян Дончев (ФН: 72029)**

***Преподавател:***

**Елза Иванова**

**София 2022**

## 1. Въведение

С моят проект направих проучване, на тема „жените шофьори“ с цел да анализираме общественото мнение на темата. Кой хора смятат, че мъжете са по добри шофьори, и кои мислят обратното. Изготвих анкета с 8 въпроса, на която отговориха 70 човека.

<https://forms.gle/QvZbRNDfdMaGPqd58>

## 2. Импортиране на данните от резултатите

Отговорите от анкетата са извадени в тази таблица.

[https://docs.google.com/spreadsheets/d/1LRcD01Ng\\_YZ3GnDD4wPL8AMTdWfMnqnTCPuvfPAMRiY/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1LRcD01Ng_YZ3GnDD4wPL8AMTdWfMnqnTCPuvfPAMRiY/edit?usp=sharing)

Таблицата се експортира в .csv файл, който лесно се чете в R Studio.

```
1 library(UsingR)
2
3 data <- read.csv("../R/Project/data.csv")
4
5 View(data)
```

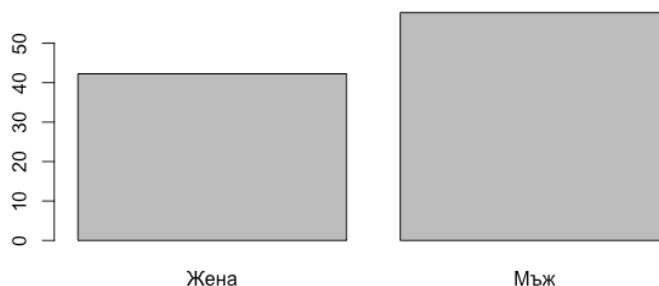
data										Open with Numbers	
Календарь	Временной пояс	Имя пользователя	На каком языке	Какой язык (или перевод) ПТТ	Какой язык (или перевод) ПТТ	Сторона	Какой язык (или перевод) ПТТ	Сторона	Какой язык (или перевод) ПТТ	Какой язык (или перевод) ПТТ	Какой язык (или перевод) ПТТ
04.06.2022 14:48:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 14:49:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 14:50:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 14:51:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 14:52:37	Мск	Ja	20	1	Ja	Ja	Ja	Ja	Ja	Ja	1.0
04.06.2022 14:53:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 14:54:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 14:55:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 14:56:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 14:57:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 14:58:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 14:59:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:00:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:01:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:02:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:03:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:04:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:05:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:06:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:07:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:08:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:09:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:10:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:11:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:12:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:13:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:14:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:15:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:16:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:17:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:18:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:19:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:20:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:21:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:22:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:23:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:24:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5
04.06.2022 15:25:37	Мск	Ja	20	0	Ja	Ja	Ja	Ja	Ja	Ja	0.5

### 3. Анализ на въпросите

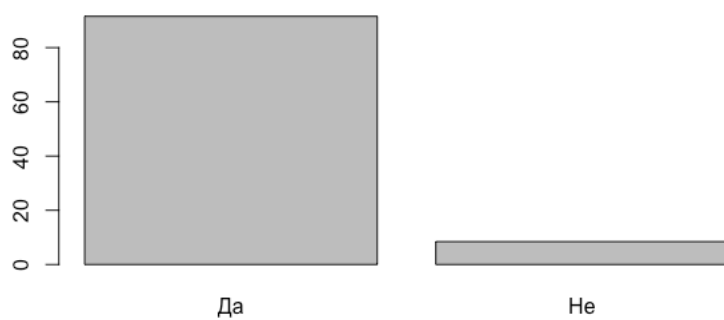
#### barplot анализ

```
1 barplot(height =  
  prop.table(table(data$Вашият.пол))*100)  
2 barplot(height =  
  prop.table(table(data$Имате.ли.шофьорска.книжка.))*100  
3 )  
4 barplot(height =  
  prop.table(table(data$На.колко.години.сте.))*100)  
5 barplot(height =  
  prop.table(table(data$Колко.пъти.сте.претърпявали.ПТП.  
  зад.волана.))*100)  
6 barplot(height =  
  prop.table(table(data$Според.вас.средностатистическият  
  .мъж.по.добър.шофьор.ли.е.от.средностатистическата.жен  
  а.))*100)  
7 barplot(height =  
  prop.table(table(data$Съгласни.ли.сте.с.факта..че.мъже  
  те.са.по.безотговорни.на.пътя..а.жените.са.по.дезориен  
  тирани.))*100)  
8 barplot(height =  
  prop.table(table(data$Според.вас.мъжете.или.жените.зад  
  .волана.са.по.опасни.за.останалите.участници.в.движени  
  ето.))*100)  
9 barplot(height =  
  prop.table(table(data$На.колко.промила.трябва.да.е.еди  
  н.мъж..за.да.кара.по.зле.от.жена.))*100)
```

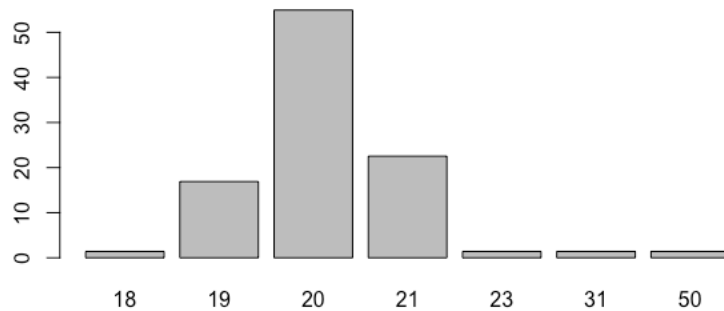
- 
- Вашият пол?



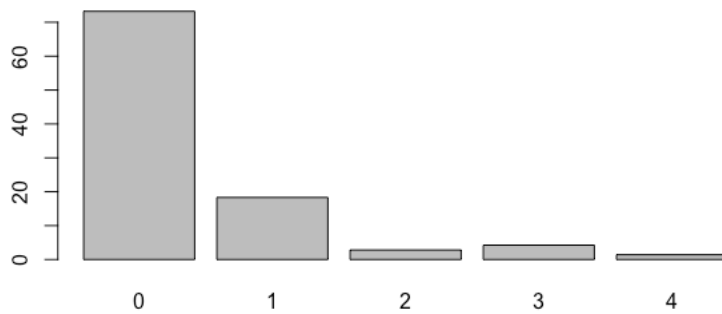
- Имате ли шофьорска книжка?



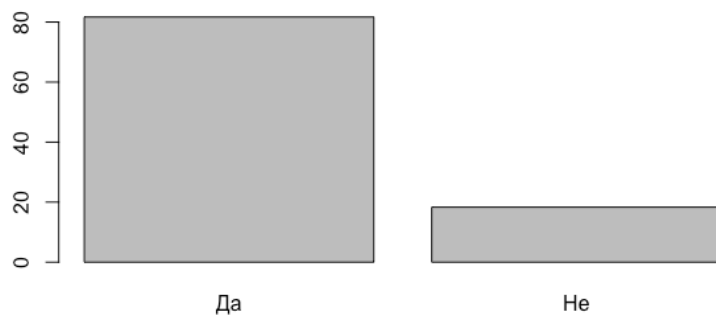
- На колко години сте?



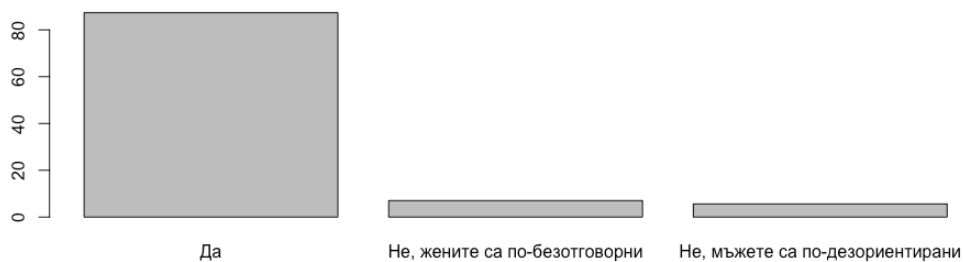
- Колко пъти сте претърпявали ПТП зад волана?



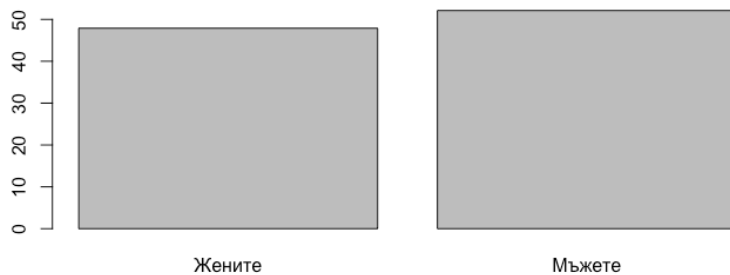
- Според вас средностатистическият мъж по-добър шофьор ли е от средностатистическата жена?



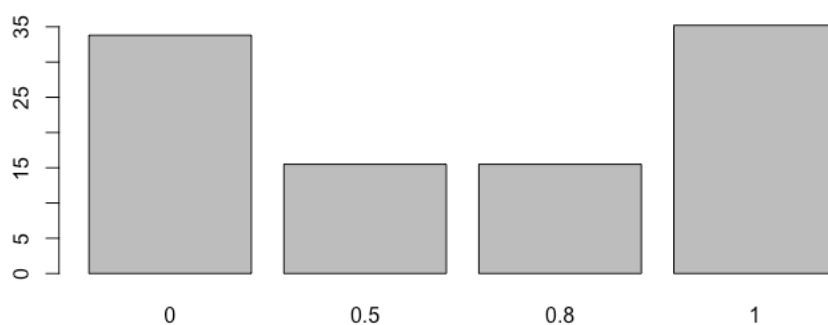
- Съгласни ли сте с факта, че мъжете са по-безотговорни на пътя, а жените са по-дезориентирани?



- Според вас мъжете или жените зад волана са по опасни за останалите участници в движението?



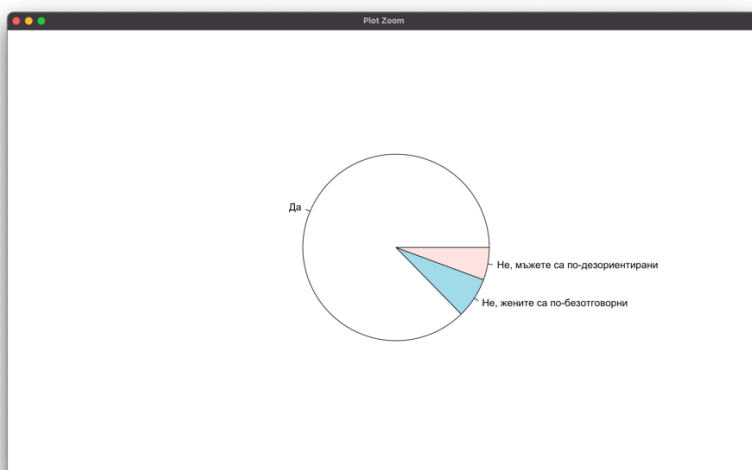
- На колко промила трябва да е един мъж, за да кара по зле от жена?



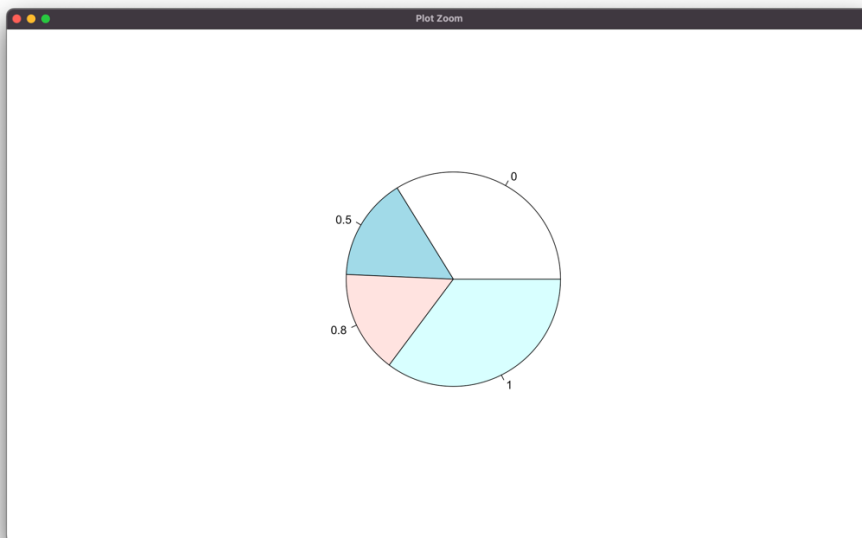
### piechart анализ

```
1 pie(table(data$На.колко.промила.трябва.да.е.един.мъж..
за.да.кара.по.зле.от.жена.))
2
3 pie(table(data$Съгласни.ли.сте.с.факта..че.мъжете.са.п
о.безотговорни.на.пътя..а.жените.са.по.дезориентирани.
))
```

- Съгласни ли сте с факта, че мъжете са по-безотговорни на пътя, а жените са по-дезориентирани?



- На колко промила трябва да е един мъж, за да кара по зле от жена?



### векторен анализ

```
1 mode <- function(vec) {  
2   uniqv <- unique(vec)  
3   uniqv[which.max(tabulate(match(vec, uniqv)))]  
4 }  
5  
6 age <- sort(c(data$На.колко.години.сте.))
```

```
1 summary(age)  
2  
3 mean(age) #Средна стойност  
4 median(age) #Медиана  
5 mode(age) #Мода  
6 var(age) #Дисперсия  
7 sd(age) #Стандартно отклонение
```

```

> summary(age)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 18.00  20.00   20.00   20.65  21.00   50.00

>
> mean(age) #Средна стойност
[1] 20.64789
> median(age) #Медиана
[1] 20
> mode(age) #Мода
[1] 20
> var(age) #Дисперсия
[1] 14.74567
> sd(age) #Стандартно отклонение
[1] 3.84001

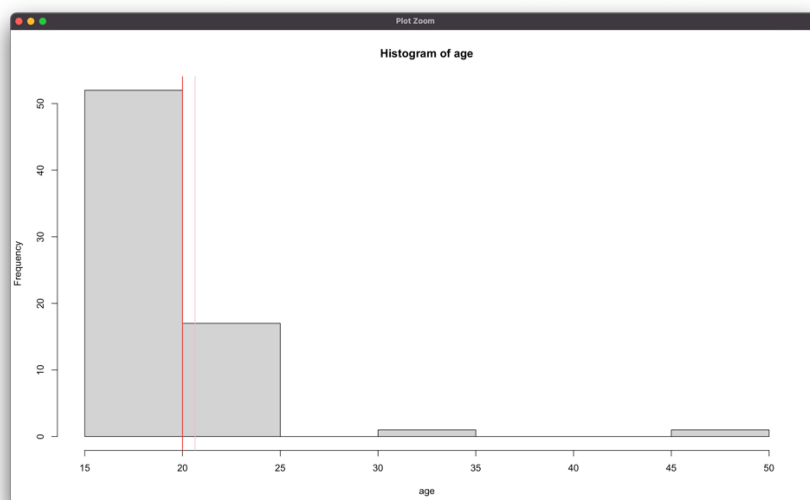
```

### хистограмен анализ

```

1 hist(age)
2 abline(v = mean(age), col = "pink")
3 abline(v = median(age), col = "red")

```





## outliers анализ

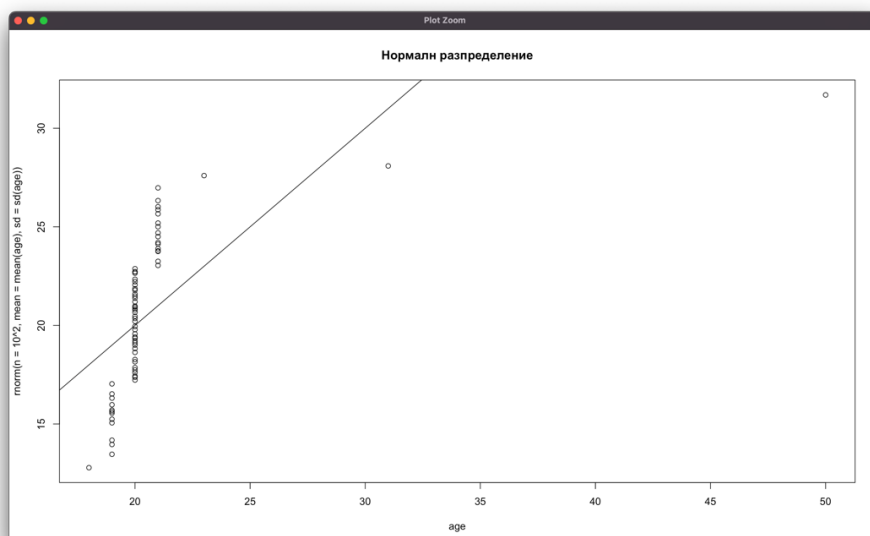
```
1 boxplot(age)
```



Намираме 4 outliers.

## Проверка за нормално разпределение

```
1 qqplot(age, rnorm(n = 10^2, mean = mean(age), sd =  
sd(age)), main = "Нормалн разпределение")  
2 abline(a = 0, b = 1)
```



## 4. Корелационен анализ

На колко години сте?

+

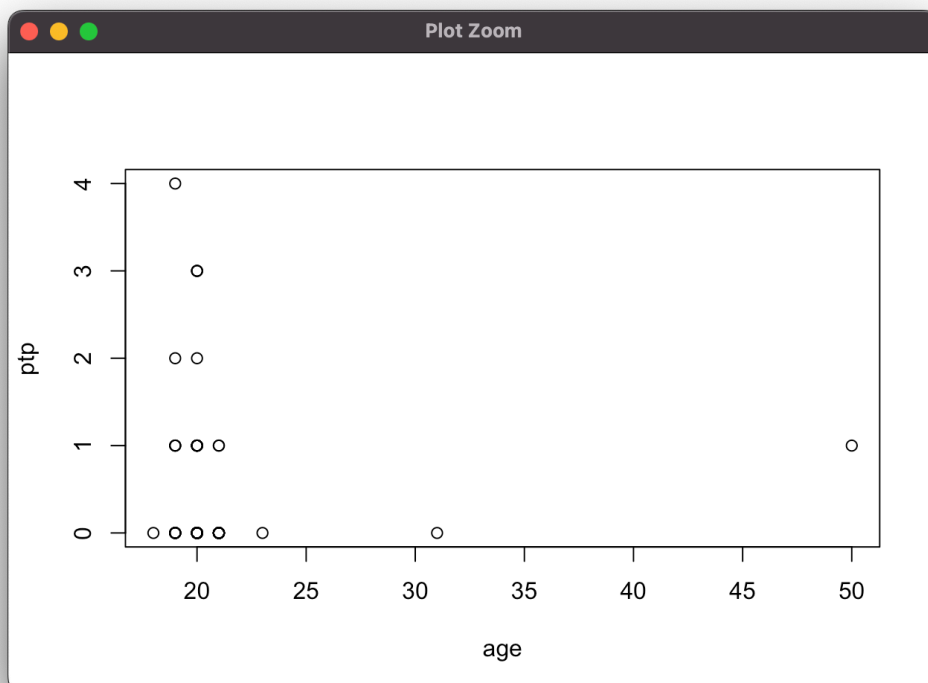
Колко пъти сте претърпявали ПТП зад волана?



```
1 age <- c(data$На.колко.години.сте.)  
2 ptp <-  
  c(data$Колко.пъти.сте.претърпявали.ПТП.зад.волана.)
```



```
1 rho <- round(cor(age, ptp), 3) #коефициент на  
2 корелация  
3 par(mfrow = c(1, 1))  
  plot(age, ptp)
```



```

1 cor(age, ptp)
2 cor.test(age, ptp, method = "spearman")

```

```

> cor(age, ptp)
[1] 0.01548178
> cor.test(age, ptp, method = "spearman")

Spearman's rank correlation rho

data:  age and ptp
S = 70027, p-value = 0.1463
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.1741696

Warning message:
In cor.test.default(age, ptp, method = "spearman") :
  Cannot compute exact p-value with ties

```

На колко години сте?

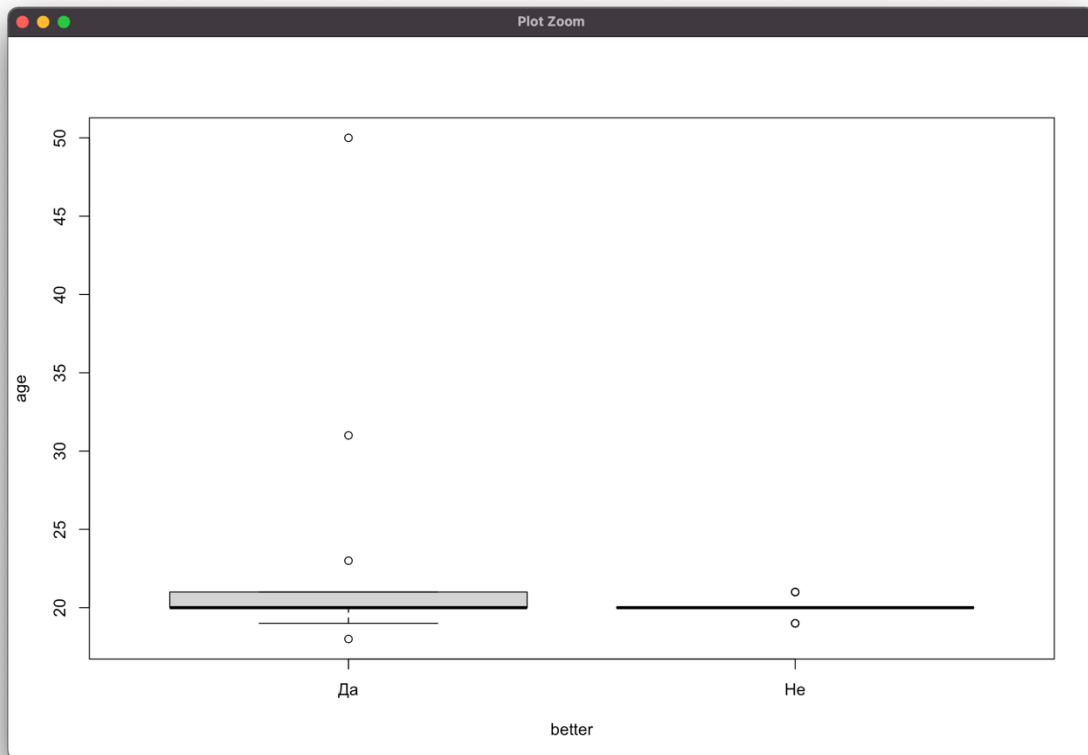
+

Според вас средностатистическият мъж по-добър шофьор ли е от средностатистическата жена?

```

1 better <-
  с(data$Според.вас.средностатистическият.мъж.по.добър.ш
    офьор.ли.е.от.средностатистическата.жена.)
2
3 boxplot(age~better)

```

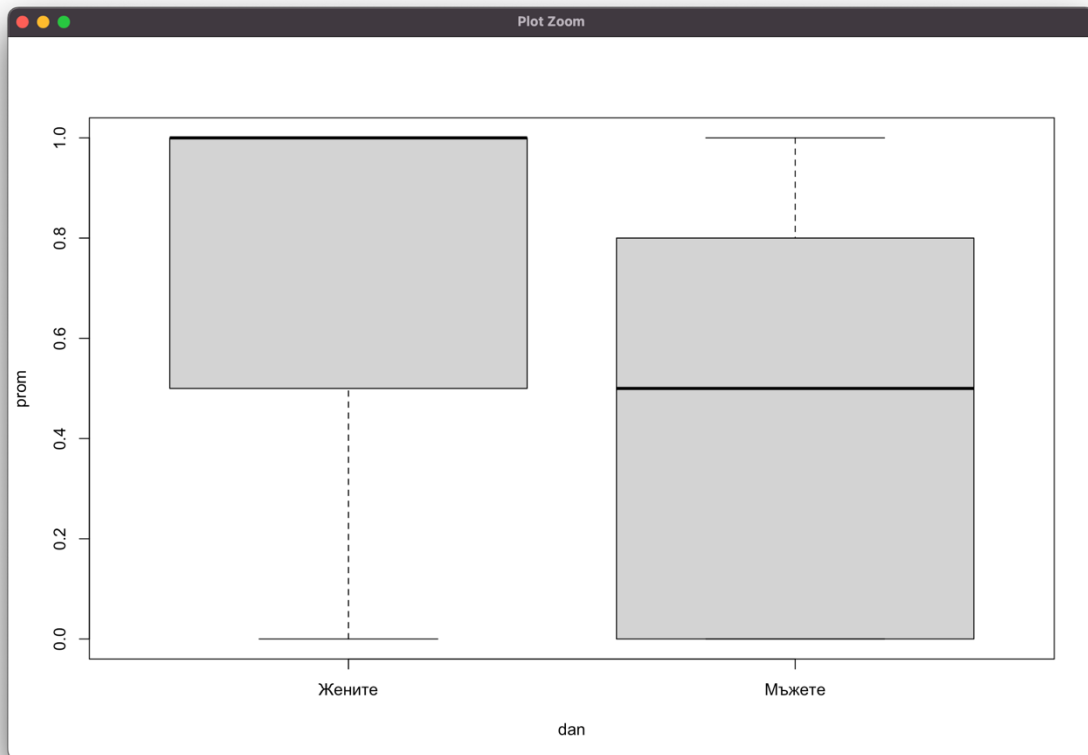


Според вас мъжете или жените зад волана са по опасни за останалите участници в движението?

+

На колко промила трябва да е един мъж, за да кара по зле от жена?

```
1 dan <-
  с(data$Според.вас.мъжете.или.жените.зад.волана.са.по.о
    пасни.за.останалите.участници.в.движението.)
2 prom <-
  с(data$На.колко.промила.трябва.да.е.един.мъж..за.да.ка
    ра.по.зле.от.жена.)
3 boxplot(prom~dan)
```



```
> bartlett.test(prom~dan)

Bartlett test of homogeneity of variances

data:  prom by dan
Bartlett's K-squared = 0.25987, df = 1, p-value = 0.6102
```

```
> kruskal.test(prom~dan)

Kruskal-Wallis rank sum test

data:  prom by dan
Kruskal-Wallis chi-squared = 12.238, df = 1, p-value = 0.0004683
```

## **5. Заключение**

С помощта на множество тестове и инструменти за анализ, можем да кажем следното за резултатите от анкетата ни:

Хората които смятат, че жените са по опасни на пътя, смятат и че мъж трябва да е пиан, за да кара по зле от жена.

Повечето хора смятат, че мъжете са по добри шофьори от жените.