

SAP - Projektni zadatak

Case study *Analiza preferencija mladih ljudi*: Deskriptivna statistika, vizualizacija podataka, statističko zaključivanje, linearna regresija i ANOVA

Domagoj Marinello, Sven Skender, Ana Skukan, Matea Vasilj

1/13/2022

Case study: *Analiza preferencija mladih ljudi*

Interesi mladih ljudi (glazba, filmovi itd.), zdrave navike, odabir načina života i obrasci potrošnje vrlo su važni za različite industrije, kao i donositelje demografskih, poreznih ili mirovinskih politika svake zemlje. Upravo su te teme bile predmet istraživanja provedenog u Slovačkoj nad mladim osobama između 15. i 30. godine života.

Shodno tome, obradili smo slijedeća istraživačka pitanja koja će biti detaljizirana u nastavku ove bilježnice:

1. Razlikuju li se izraženi strahovi ispitanih žena i muškaraca?
2. Možemo li predvidjeti obrazac potrošnje ovisno o žanru glazbe kojeg ispitanik preferira?
3. Možemo li temeljem danih varijabli predvidjeti dob ispitanika?
- 4.

Podaci za analizu sadržani su u datoteci opis pitanja.csv koja se sastoji od informacija prikupljenih upitnikom koji je prezentiran gore navedenoj skupini ljudi. Podatci se sastoje od osobnih informacija ispitanika koje, među ostalim, uključuju dob, visinu, težinu i spol. Preferencije ispitanika prikupljene su kroz kategorije glazbenog i filmskog ukusa, hobija, strahova, (ne)zdravih navika, osobina ličnosti i drugih. Skup podataka sastoji od više pitanja koja pružaju numeričke podatke, primjerice ocjenu preferencije ispitanika na skali od 1 do 5 te od kategorijskih podataka, primjerice, spol.

Prije pregleda odgovora, pogledajmo postavljena pitanja s pojašnjenjima koja se nalaze u datoteci odgovori.csv:

```
odgovori = read.csv("./data/odgovori.csv")
odgovori
```

```
##                                original
## 1                    I enjoy listening to music.
## 2                                I prefer.
## 3                    Dance, Disco, Funk
## 4                                Folk music
## 5                                Country
## 6                                Classical
## 7                                Musicals
## 8                                Pop
## 9                                Rock
## 10                   Metal, Hard rock
## 11                                Punk
## 12                   Hip hop, Rap
## 13                   Reggae, Ska
## 14                   Swing, Jazz
## 15                   Rock n Roll
```

## 16	Alternative music
## 17	Latin
## 18	Techno, Trance
## 19	Opera
## 20	I really enjoy watching movies.
## 21	Horror movies
## 22	Thriller movies
## 23	Comedies
## 24	Romantic movies
## 25	Sci-fi movies
## 26	War movies
## 27	Tales
## 28	Cartoons
## 29	Documentaries
## 30	Western movies
## 31	Action movies
## 32	History
## 33	Psychology
## 34	Politics
## 35	Mathematics
## 36	Physics
## 37	Internet
## 38	PC Software, Hardware
## 39	Economy, Management
## 40	Biology
## 41	Chemistry
## 42	Poetry reading
## 43	Geography
## 44	Foreign languages
## 45	Medicine
## 46	Law
## 47	Cars
## 48	Art
## 49	Religion
## 50	Outdoor activities
## 51	Dancing
## 52	Playing musical instruments
## 53	Poetry writing
## 54	Sport and leisure activities
## 55	Sport at competitive level
## 56	Gardening
## 57	Celebrity lifestyle
## 58	Shopping
## 59	Science and technology
## 60	Theatre
## 61	Socializing
## 62	Adrenaline sports
## 63	Pets
## 64	Flying
## 65	Thunder, lightning
## 66	Darkness
## 67	Heights
## 68	Spiders
## 69	Snakes

70 Rats, mice
 ## 71 Ageing
 ## 72 Dangerous dogs
 ## 73 Public speaking
 ## 74 Smoking habits
 ## 75 Drinking
 ## 76 I live a very healthy lifestyle.
 ## 77 I take notice of what goes on around me.
 ## 78 I try to do tasks as soon as possible and not leave them until last minute.
 ## 79 I always make a list so I don't forget anything.
 ## 80 I often study or work even in my spare time.
 ## 81 I look at things from all different angles before I go ahead.
 ## 82 I believe that bad people will suffer one day and good people will be rewarded.
 ## 83 I am reliable at work and always complete all tasks given to me.
 ## 84 I always keep my promises.
 ## 85 I can fall for someone very quickly and then completely lose interest.
 ## 86 I would rather have lots of friends than lots of money.
 ## 87 I always try to be the funniest one.
 ## 88 I can be two faced sometimes.
 ## 89 I damaged things in the past when angry.
 ## 90 I take my time to make decisions.
 ## 91 I always try to vote in elections.
 ## 92 I often think about and regret the decisions I make.
 ## 93 I can tell if people listen to me or not when I talk to them.
 ## 94 I am a hypochondriac.
 ## 95 I am emphatetic person.
 ## 96 I eat because I have to. I don't enjoy food and eat as fast as I can.
 ## 97 I try to give as much as I can to other people at Christmas.
 ## 98 I don't like seeing animals suffering.
 ## 99 I look after things I have borrowed from others.
 ## 100 I feel lonely in life.
 ## 101 I used to cheat at school.
 ## 102 I worry about my health.
 ## 103 I wish I could change the past because of the things I have done.
 ## 104 I believe in God.
 ## 105 I always have good dreams.
 ## 106 I always give to charity.
 ## 107 I have lots of friends.
 ## 108 Timekeeping.
 ## 109 Do you lie to others?
 ## 110 I am very patient.
 ## 111 I can quickly adapt to a new environment.
 ## 112 My moods change quickly.
 ## 113 I am well mannered and I look after my appearance.
 ## 114 I enjoy meeting new people.
 ## 115 I always let other people know about my achievements.
 ## 116 I think carefully before answering any important letters.
 ## 117 I enjoy childrens' company.
 ## 118 I am not afraid to give my opinion if I feel strongly about something.
 ## 119 I can get angry very easily.
 ## 120 I always make sure I connect with the right people.
 ## 121 I have to be well prepared before public speaking.
 ## 122 I will find a fault in myself if people don't like me.
 ## 123 I cry when I feel down or things don't go the right way.

```

## 124             I am 100% happy with my life.
## 125             I am always full of life and energy.
## 126             I prefer big dangerous dogs to smaller, calmer dogs.
## 127             I believe all my personality traits are positive.
## 128             If I find something the doesn't belong to me I will hand it in.
## 129             I find it very difficult to get up in the morning.
## 130             I have many different hobbies and interests.
## 131             I always listen to my parents' advice.
## 132             I enjoy taking part in surveys.
## 133             How much time do you spend online?
## 134             I save all the money I can.
## 135             I enjoy going to large shopping centres.
## 136             I prefer branded clothing to non branded.
## 137             I spend a lot of money on partying and socializing.
## 138             I spend a lot of money on my appearance.
## 139             I spend a lot of money on gadgets.
## 140             I will hapilly pay more money for good, quality or healthy food.
## 141             Age
## 142             Height
## 143             Weight
## 144             How many siblings do you have?
## 145             Gender
## 146             I am
## 147             Highest education achieved
## 148             I am the only child
## 149             I spent most of my childhood in a
## 150             I lived most of my childhood in a
##             short
## 1             Music
## 2             Slow songs or fast songs
## 3             Dance
## 4             Folk
## 5             Country
## 6             Classical music
## 7             Musical
## 8             Pop
## 9             Rock
## 10            Metal or Hardrock
## 11            Punk
## 12            Hiphop, Rap
## 13            Reggae, Ska
## 14            Swing, Jazz
## 15            Rock n roll
## 16            Alternative
## 17            Latino
## 18            Techno, Trance
## 19            Opera
## 20            Movies
## 21            Horror
## 22            Thriller
## 23            Comedy
## 24            Romantic
## 25            Sci-fi
## 26            War

```

## 27	Fantasy/Fairy tales
## 28	Animated
## 29	Documentary
## 30	Western
## 31	Action
## 32	History
## 33	Psychology
## 34	Politics
## 35	Mathematics
## 36	Physics
## 37	Internet
## 38	PC
## 39	Economy Management
## 40	Biology
## 41	Chemistry
## 42	Reading
## 43	Geography
## 44	Foreign languages
## 45	Medicine
## 46	Law
## 47	Cars
## 48	Art exhibitions
## 49	Religion
## 50	Countryside, outdoors
## 51	Dancing
## 52	Musical instruments
## 53	Writing
## 54	Passive sport
## 55	Active sport
## 56	Gardening
## 57	Celebrities
## 58	Shopping
## 59	Science and technology
## 60	Theatre
## 61	Fun with friends
## 62	Adrenaline sports
## 63	Pets
## 64	Flying
## 65	Storm
## 66	Darkness
## 67	Heights
## 68	Spiders
## 69	Snakes
## 70	Rats
## 71	Ageing
## 72	Dangerous dogs
## 73	Fear of public speaking
## 74	Smoking
## 75	Alcohol
## 76	Healthy eating
## 77	Daily events
## 78	Prioritising workload
## 79	Writing notes
## 80	Workaholism

## 81	Thinking ahead
## 82	Final judgement
## 83	Reliability
## 84	Keeping promises
## 85	Loss of interest
## 86	Friends versus money
## 87	Funniness
## 88	Fake
## 89	Criminal damage
## 90	Decision making
## 91	Elections
## 92	Self-criticism
## 93	Judgment calls
## 94	Hypochondria
## 95	Empathy
## 96	Eating to survive
## 97	Giving
## 98	Compassion to animals
## 99	Borrowed stuff
## 100	Loneliness
## 101	Cheating in school
## 102	Health
## 103	Changing the past
## 104	God
## 105	Dreams
## 106	Charity
## 107	Number of friends
## 108	Punctuality
## 109	Lying
## 110	Waiting
## 111	New environment
## 112	Mood swings
## 113	Appearance and gestures
## 114	Socializing
## 115	Achievements
## 116	Responding to a serious letter
## 117	Children
## 118	Assertiveness
## 119	Getting angry
## 120	Knowing the right people
## 121	Public speaking
## 122	Unpopularity
## 123	Life struggles
## 124	Happiness in life
## 125	Energy levels
## 126	Small - big dogs
## 127	Personality
## 128	Finding lost valuables
## 129	Getting up
## 130	Interests or hobbies
## 131	Parents' advice
## 132	Questionnaires or polls
## 133	Internet usage
## 134	Finances

```
## 135          Shopping centres
## 136          Branded clothing
## 137      Entertainment spending
## 138          Spending on looks
## 139          Spending on gadgets
## 140      Spending on healthy eating
## 141              Age
## 142              Height
## 143              Weight
## 144          Number of siblings
## 145              Gender
## 146          Left - right handed
## 147              Education
## 148              Only child
## 149          Village - town
## 150      House - block of flats
```

Sada, kada smo saznali o kakvim se pitanjima radi, možemo pogledati kako su ona kodirana, kakve smo odgovore uspjeli prikupiti i koliko ih je uopće:

```
pitanja = read.csv("../data/opis pitanja.csv")
head(pitanja)
```

```
##      Music Slow.songs.or.fast.songs Dance Folk Country Classical.music Musical Pop
## 1      5              3      2      1      2              2      1      5
## 2      4              4      2      1      1              1      2      3
## 3      5              5      2      2      3              4      5      3
## 4      5              3      2      1      1              1      1      2
## 5      5              3      4      3      2              4      3      5
## 6      5              3      2      3      2              3      3      2
##      Rock Metal.or.Hardrock Punk Hiphop..Rap Reggae..Ska Swing..Jazz Rock.n.roll
## 1      5              1      1              1              1              3
## 2      5              4      4              1              3              4
## 3      5              3      4              1              4              5
## 4      2              1      4              2              2              2
## 5      3              1      2              5              3              1
## 6      5              5      3              4              3              4
##      Alternative Latino Techno..Trance Opera Movies Horror Thriller Comedy
## 1              1      1              1      1      5      4      2      5
## 2              4      2              1      1      5      2      2      4
## 3              5      5              1      3      5      3      4      4
## 4              5      1              2      1      5      4      4      3
## 5              2      4              2      2      5      4      4      5
## 6              5      3              1      3      5      5      5      5
##      Romantic Sci.fi War Fantasy.Fairy.tales Animated Documentary Western Action
## 1              4      4      1              5      5      3      1      2
## 2              3      4      1              3      5      4      1      4
## 3              2      4      2              5      5      2      2      1
## 4              3      4      3              1      2      5      1      2
## 5              2      3      3              4      4      3      1      4
## 6              2      3      3              4      3      3      2      4
##      History Psychology Politics Mathematics Physics Internet PC
## 1              1              5      1              3      3      5      3
## 2              1              3      4              5      2      4      4
## 3              1              2      1              5      2      4      2
```

## 4	4	4	5	4	1	3	1			
## 5	3	2	3	2	2	2	2			
## 6	5	3	4	2	3	4	4			
##	Economy.Management	Biology	Chemistry	Reading	Geography	Foreign.languages				
## 1		5	3	3	3		5			
## 2		5	1	1	4	4	5			
## 3		4	1	1	5	2	5			
## 4		2	3	3	5	4	4			
## 5		2	3	3	5	2	3			
## 6		1	4	4	3	3	4			
##	Medicine	Law	Cars	Art.exhibitions	Religion	Countryside..outdoors	Dancing			
## 1		3	1	1	1		5 3			
## 2		1	2	2	2	1	1 1			
## 3		2	3	1	5	5	5 5			
## 4		2	5	1	5	4	1 1			
## 5		3	2	3	1	4	4 1			
## 6		4	3	5	2	2	5 1			
##	Musical.instruments	Writing	Passive.sport	Active.sport	Gardening	Celebrities				
## 1		3	2	1	5	5	1			
## 2		1	1	1	1	1	2			
## 3		5	5	5	2	1	1			
## 4		1	3	1	1	1	2			
## 5		3	1	3	1	4	3			
## 6		5	1	5	4	2	1			
##	Shopping	Science.and.technology	Theatre	Fun.with.friends	Adrenaline.sports					
## 1		4	4	2	5		4			
## 2		3	3	2	4		2			
## 3		4	2	5	5		5			
## 4		4	3	1	2		1			
## 5		3	3	2	4		2			
## 6		2	3	1	3		3			
##	Pets	Flying	Storm	Darkness	Heights	Spiders	Snakes	Rats	Ageing	Dangerous.dogs
## 1	4	1	1	1	1	1	5	3	1	3
## 2	5	1	1	1	2	1	1	1	3	1
## 3	5	1	1	1	1	1	1	1	1	1
## 4	1	2	1	1	3	5	5	5	4	5
## 5	1	1	2	1	1	1	1	2	2	4
## 6	2	3	2	2	2	1	2	2	1	1
##	Fear.of.public.speaking				Smoking		Alcohol		Healthy.eating	
## 1				2	never smoked		drink a lot			4
## 2				4	never smoked		drink a lot			3
## 3				2	tried smoking		drink a lot			3
## 4				5	former smoker		drink a lot			3
## 5				3	tried smoking	social	drinker			4
## 6				3	never smoked		never			2
##	Daily.events	Prioritising.workload	Writing.notes	Workaholism	Thinking.ahead					
## 1		2		2		5		4		2
## 2		3		2		4		5		4
## 3		1		2		5		3		5
## 4		4		4		4		5		3
## 5		3		1		2		3		5
## 6		2		2		3		3		3
##	Final.judgement	Reliability	Keeping.promises	Loss.of.interest						
## 1		5		4		4				1

## 2	1	4	4	3
## 3	3	4	5	1
## 4	1	3	4	5
## 5	5	5	4	2
## 6	1	3	4	3
##	Friends.versus.money	Funniness	Fake Criminal.damage	Decision.making Elections
## 1	3	5	1	3 4
## 2	4	3	2	1 2 5
## 3	5	2	4	1 3 5
## 4	2	1	1	5 5 5
## 5	3	3	2	1 3 5
## 6	2	3	1	4 2 5
##	Self.criticism	Judgment.calls	Hypochondria	Empathy Eating.to.survive Giving
## 1	1	3	1	3 1 4
## 2	4	4	1	2 1 2
## 3	4	4	1	5 5 5
## 4	5	4	3	3 1 1
## 5	5	5	1	3 1 3
## 6	4	4	1	4 2 3
##	Compassion.to.animals	Borrowed.stuff	Loneliness	Cheating.in.school Health
## 1	5	4	3	2 1
## 2	4	3	2	4 4
## 3	4	2	5	3 2
## 4	2	5	5	5 1
## 5	3	4	3	5 3
## 6	5	5	2	4 3
##	Changing.the.past	God Dreams	Charity	Number.of.friends
## 1	1 1	4	2	3
## 2	4 1	3	1	3
## 3	5 5	1	3	3
## 4	5 4	3	3	1
## 5	4 5	3	3	3
## 6	3 3	3	2	3
##	Punctuality		Lying	Waiting New.environment
## 1	i am always on time		never	3 4
## 2	i am often early		sometimes	3 4
## 3	i am often running late		sometimes	2 3
## 4	i am often early only to avoid hurting someone			1 1
## 5	i am always on time		everytime it suits me	3 4
## 6	i am often early only to avoid hurting someone			3 4
##	Mood.swings	Appearance.and.gestures	Socializing	Achievements
## 1	3	4	3	4
## 2	4	4	4	2
## 3	4	3	5	3
## 4	5	3	1	3
## 5	2	3	3	3
## 6	3	3	4	2
##	Responding.to.a.serious.letter	Children	Assertiveness	Getting.angry
## 1		3	5	1 1
## 2		4	2	2 5
## 3		4	4	3 4
## 4		3	2	5 5
## 5		3	5	4 2
## 6		2	3	4 3

##	Knowing.the.right.people	Public.speaking	Unpopularity	Life.struggles		
## 1	3	5	5	1		
## 2	4	4	4	1		
## 3	3	2	4	4		
## 4	4	5	3	3		
## 5	3	5	5	2		
## 6	4	4	4	3		
##	Happiness.in.life	Energy.levels	Small...big.dogs	Personality		
## 1	4	5	1	4		
## 2	4	3	5	3		
## 3	4	4	3	3		
## 4	2	2	1	2		
## 5	3	5	3	3		
## 6	3	4	4	3		
##	Finding.lost.valuables	Getting.up	Interests.or.hobbies	Parents..advice		
## 1	3	2	3	4		
## 2	4	5	3	2		
## 3	3	4	5	3		
## 4	1	1	NA	2		
## 5	2	4	3	3		
## 6	3	3	5	3		
##	Questionnaires.or.polls	Internet.usage	Finances	Shopping.centres		
## 1	3 few hours a day	3	4			
## 2	3 few hours a day	3	4			
## 3	1 few hours a day	2	4			
## 4	4 most of the day	2	4			
## 5	3 few hours a day	4	3			
## 6	4 few hours a day	2	3			
##	Branded.clothing	Entertainment.spending	Spending.on.looks	Spending.on.gadgets		
## 1	5	3	3	1		
## 2	1	4	2	5		
## 3	1	4	3	4		
## 4	3	3	4	4		
## 5	4	3	3	2		
## 6	3	3	1	4		
##	Spending.on.healthy.eating	Age	Height	Weight	Number.of.siblings	Gender
## 1	3	20	163	48	1	female
## 2	2	19	163	58	2	female
## 3	2	20	176	67	2	female
## 4	1	22	172	59	1	female
## 5	4	20	170	59	1	female
## 6	4	20	186	77	1	male
##	Left...right.handed	Education	Only.child	Village...town		
## 1	right handed	college/bachelor degree	no	village		
## 2	right handed	college/bachelor degree	no	city		
## 3	right handed	secondary school	no	city		
## 4	right handed	college/bachelor degree	yes	city		
## 5	right handed	secondary school	no	village		
## 6	right handed	secondary school	no	city		
##	House...block.of.flats					
## 1	block of flats					
## 2	block of flats					
## 3	block of flats					
## 4	house/bungalow					

```
## 5      house/bungalow
## 6      block of flats
```

Dimenzije dataseta:

```
dim(pitanja) # broj redaka, broj stupaca (broj primjera, broj varijabli)
```

```
## [1] 1010 150
```

Pomoću summary-ja računamo statistike i doznajemo tipove podataka:

```
summary(pitanja)
```

```
##      Music      Slow.songs.or.fast.songs      Dance      Folk
##  Min.   :1.000   Min.   :1.000             Min.   :1.000   Min.   :1.000
## 1st Qu.:5.000   1st Qu.:3.000             1st Qu.:2.000   1st Qu.:1.000
##  Median :5.000   Median :3.000             Median :3.000   Median :2.000
##  Mean   :4.732   Mean   :3.328             Mean   :3.113   Mean   :2.289
## 3rd Qu.:5.000   3rd Qu.:4.000             3rd Qu.:4.000   3rd Qu.:3.000
##  Max.   :5.000   Max.   :5.000             Max.   :5.000   Max.   :5.000
## NA's    :3      NA's    :2              NA's    :4      NA's    :5
##      Country      Classical.music      Musical      Pop
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
## 1st Qu.:1.000   1st Qu.:2.000   1st Qu.:2.000   1st Qu.:3.000
##  Median :2.000   Median :3.000   Median :3.000   Median :4.000
##  Mean   :2.123   Mean   :2.956   Mean   :2.762   Mean   :3.472
## 3rd Qu.:3.000   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
## NA's    :5      NA's    :7      NA's    :2      NA's    :3
##      Rock      Metal.or.Hardrock      Punk      Hiphop..Rap
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
## 1st Qu.:3.000   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:2.000
##  Median :4.000   Median :2.000   Median :2.000   Median :3.000
##  Mean   :3.762   Mean   :2.361   Mean   :2.456   Mean   :2.911
## 3rd Qu.:5.000   3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:4.000
##  Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
## NA's    :6      NA's    :3      NA's    :8      NA's    :4
##      Reggae..Ska      Swing..Jazz      Rock.n.roll      Alternative      Latino
##  Min.   :1.00   Min.   :1.00   Min.   :1.000   Min.   :1.000   Min.   :1.000
## 1st Qu.:2.00   1st Qu.:2.00   1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000
##  Median :3.00   Median :3.00   Median :3.000   Median :3.000   Median :3.000
##  Mean   :2.77   Mean   :2.76   Mean   :3.142   Mean   :2.829   Mean   :2.842
## 3rd Qu.:4.00   3rd Qu.:4.00   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :5.00   Max.   :5.00   Max.   :5.000   Max.   :5.000   Max.   :5.000
## NA's    :7      NA's    :6      NA's    :7      NA's    :7      NA's    :8
##      Techno..Trance      Opera      Movies      Horror      Thriller
##  Min.   :1.000   Min.   :1.00   Min.   :1.000   Min.   :1.000   Min.   :1.000
## 1st Qu.:1.000   1st Qu.:1.00   1st Qu.:4.000   1st Qu.:1.000   1st Qu.:3.000
##  Median :2.000   Median :2.00   Median :5.000   Median :3.000   Median :4.000
##  Mean   :2.339   Mean   :2.14   Mean   :4.614   Mean   :2.794   Mean   :3.384
## 3rd Qu.:3.000   3rd Qu.:3.00   3rd Qu.:5.000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :5.000   Max.   :5.00   Max.   :5.000   Max.   :5.000   Max.   :5.000
## NA's    :7      NA's    :1      NA's    :6      NA's    :2      NA's    :1
##      Comedy      Romantic      Sci.fi      War
##  Min.   :1.000   Min.   :1.00   Min.   :1.000   Min.   :1.000
## 1st Qu.:4.000   1st Qu.:3.00   1st Qu.:2.000   1st Qu.:2.000
##  Median :5.000   Median :4.00   Median :3.000   Median :3.000
```

##	Mean	:4.495	Mean	:3.49	Mean	:3.113	Mean	:3.156
##	3rd Qu.	:5.000	3rd Qu.	:5.00	3rd Qu.	:4.000	3rd Qu.	:4.000
##	Max.	:5.000	Max.	:5.00	Max.	:5.000	Max.	:5.000
##	NA's	:3	NA's	:3	NA's	:2	NA's	:2
##	Fantasy.Fairy.tales		Animated		Documentary		Western	
##	Min.	:1.00	Min.	:1.000	Min.	:1.000	Min.	:1.000
##	1st Qu.	:3.00	1st Qu.	:3.000	1st Qu.	:3.000	1st Qu.	:1.000
##	Median	:4.00	Median	:4.000	Median	:4.000	Median	:2.000
##	Mean	:3.75	Mean	:3.788	Mean	:3.644	Mean	:2.126
##	3rd Qu.	:5.00	3rd Qu.	:5.000	3rd Qu.	:5.000	3rd Qu.	:3.000
##	Max.	:5.00	Max.	:5.000	Max.	:5.000	Max.	:5.000
##	NA's	:3	NA's	:3	NA's	:8	NA's	:4
##	Action		History		Psychology		Politics	
##	Min.	:1.000	Min.	:1.000	Min.	:1.000	Min.	:1.000
##	1st Qu.	:3.000	1st Qu.	:2.000	1st Qu.	:2.000	1st Qu.	:1.000
##	Median	:4.000	Median	:3.000	Median	:3.000	Median	:2.000
##	Mean	:3.537	Mean	:3.207	Mean	:3.138	Mean	:2.596
##	3rd Qu.	:5.000	3rd Qu.	:4.000	3rd Qu.	:4.000	3rd Qu.	:4.000
##	Max.	:5.000	Max.	:5.000	Max.	:5.000	Max.	:5.000
##	NA's	:2	NA's	:2	NA's	:5	NA's	:1
##	Mathematics		Physics		Internet		PC	
##	Min.	:1.000	Min.	:1.000	Min.	:1.000	Min.	:1.000
##	1st Qu.	:1.000	1st Qu.	:1.000	1st Qu.	:4.000	1st Qu.	:2.000
##	Median	:2.000	Median	:2.000	Median	:4.000	Median	:3.000
##	Mean	:2.335	Mean	:2.065	Mean	:4.176	Mean	:3.136
##	3rd Qu.	:3.000	3rd Qu.	:3.000	3rd Qu.	:5.000	3rd Qu.	:4.000
##	Max.	:5.000	Max.	:5.000	Max.	:5.000	Max.	:5.000
##	NA's	:3	NA's	:3	NA's	:4	NA's	:6
##	Economy.Management		Biology		Chemistry		Reading	
##	Min.	:1.000	Min.	:1.000	Min.	:1.000	Min.	:1.000
##	1st Qu.	:1.000	1st Qu.	:2.000	1st Qu.	:1.000	1st Qu.	:2.000
##	Median	:2.000	Median	:2.000	Median	:2.000	Median	:3.000
##	Mean	:2.644	Mean	:2.665	Mean	:2.165	Mean	:3.159
##	3rd Qu.	:4.000	3rd Qu.	:4.000	3rd Qu.	:3.000	3rd Qu.	:5.000
##	Max.	:5.000	Max.	:5.000	Max.	:5.000	Max.	:5.000
##	NA's	:5	NA's	:6	NA's	:10	NA's	:6
##	Geography		Foreign.languages		Medicine		Law	
##	Min.	:1.000	Min.	:1.000	Min.	:1.000	Min.	:1.000
##	1st Qu.	:2.000	1st Qu.	:3.000	1st Qu.	:1.000	1st Qu.	:1.000
##	Median	:3.000	Median	:4.000	Median	:2.000	Median	:2.000
##	Mean	:3.083	Mean	:3.778	Mean	:2.516	Mean	:2.257
##	3rd Qu.	:4.000	3rd Qu.	:5.000	3rd Qu.	:3.000	3rd Qu.	:3.000
##	Max.	:5.000	Max.	:5.000	Max.	:5.000	Max.	:5.000
##	NA's	:9	NA's	:5	NA's	:5	NA's	:1
##	Cars		Art.exhibitions		Religion		Countryside..outdoors	
##	Min.	:1.000	Min.	:1.00	Min.	:1.000	Min.	:1.000
##	1st Qu.	:1.000	1st Qu.	:1.00	1st Qu.	:1.000	1st Qu.	:3.000
##	Median	:3.000	Median	:2.00	Median	:2.000	Median	:4.000
##	Mean	:2.687	Mean	:2.59	Mean	:2.273	Mean	:3.687
##	3rd Qu.	:4.000	3rd Qu.	:4.00	3rd Qu.	:3.000	3rd Qu.	:5.000
##	Max.	:5.000	Max.	:5.00	Max.	:5.000	Max.	:5.000
##	NA's	:4	NA's	:6	NA's	:3	NA's	:7
##	Dancing		Musical.instruments		Writing		Passive.sport	
##	Min.	:1.000	Min.	:1.000	Min.	:1.000	Min.	:1.000

## 1st Qu.:1.000	1st Qu.:1.000	1st Qu.:1.000	1st Qu.:2.000
## Median :2.000	Median :2.000	Median :1.000	Median :3.000
## Mean :2.462	Mean :2.324	Mean :1.901	Mean :3.388
## 3rd Qu.:4.000	3rd Qu.:4.000	3rd Qu.:3.000	3rd Qu.:5.000
## Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000
## NA's :3	NA's :1	NA's :6	NA's :15
## Active.sport	Gardening	Celebrities	Shopping
## Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
## 1st Qu.:2.000	1st Qu.:1.000	1st Qu.:1.000	1st Qu.:2.000
## Median :3.000	Median :1.000	Median :2.000	Median :3.000
## Mean :3.291	Mean :1.907	Mean :2.362	Mean :3.277
## 3rd Qu.:5.000	3rd Qu.:3.000	3rd Qu.:3.000	3rd Qu.:4.000
## Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000
## NA's :4	NA's :7	NA's :2	NA's :2
## Science.and.technology	Theatre	Fun.with.friends	Adrenaline.sports
## Min. :1.000	Min. :1.000	Min. :2.000	Min. :1.000
## 1st Qu.:2.000	1st Qu.:2.000	1st Qu.:4.000	1st Qu.:2.000
## Median :3.000	Median :3.000	Median :5.000	Median :3.000
## Mean :3.234	Mean :3.025	Mean :4.558	Mean :2.948
## 3rd Qu.:4.000	3rd Qu.:4.000	3rd Qu.:5.000	3rd Qu.:4.000
## Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000
## NA's :6	NA's :8	NA's :4	NA's :3
## Pets	Flying	Storm	Darkness
## Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
## 1st Qu.:2.000	1st Qu.:1.000	1st Qu.:1.000	1st Qu.:1.000
## Median :4.000	Median :2.000	Median :2.000	Median :2.000
## Mean :3.335	Mean :2.062	Mean :1.973	Mean :2.251
## 3rd Qu.:5.000	3rd Qu.:3.000	3rd Qu.:3.000	3rd Qu.:3.000
## Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000
## NA's :4	NA's :3	NA's :1	NA's :2
## Heights	Spiders	Snakes	Rats
## Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
## 1st Qu.:2.000	1st Qu.:1.000	1st Qu.:2.000	1st Qu.:1.000
## Median :2.000	Median :3.000	Median :3.000	Median :2.000
## Mean :2.616	Mean :2.826	Mean :3.028	Mean :2.409
## 3rd Qu.:4.000	3rd Qu.:4.000	3rd Qu.:4.000	3rd Qu.:3.000
## Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000
## NA's :3	NA's :5		NA's :3
## Ageing	Dangerous.dogs	Fear.of.public.speaking	Smoking
## Min. :1.000	Min. :1.000	Min. :1.000	Length:1010
## 1st Qu.:1.000	1st Qu.:2.000	1st Qu.:2.000	Class :character
## Median :2.000	Median :3.000	Median :3.000	Mode :character
## Mean :2.581	Mean :3.043	Mean :2.804	
## 3rd Qu.:4.000	3rd Qu.:4.000	3rd Qu.:4.000	
## Max. :5.000	Max. :5.000	Max. :5.000	
## NA's :1	NA's :1	NA's :1	
## Alcohol	Healthy.eating	Daily.events	Prioritising.workload
## Length:1010	Min. :1.000	Min. :1.000	Min. :1.000
## Class :character	1st Qu.:3.000	1st Qu.:2.000	1st Qu.:2.000
## Mode :character	Median :3.000	Median :3.000	Median :3.000
##	Mean :3.032	Mean :3.075	Mean :2.646
##	3rd Qu.:4.000	3rd Qu.:4.000	3rd Qu.:3.000
##	Max. :5.000	Max. :5.000	Max. :5.000
##	NA's :3	NA's :7	NA's :5

## Writing.notes	Workaholism	Thinking.ahead	Final.judgement
## Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
## 1st Qu.:2.000	1st Qu.:2.000	1st Qu.:3.000	1st Qu.:1.000
## Median :3.000	Median :3.000	Median :3.000	Median :3.000
## Mean :3.083	Mean :2.996	Mean :3.414	Mean :2.649
## 3rd Qu.:4.000	3rd Qu.:4.000	3rd Qu.:4.000	3rd Qu.:4.000
## Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000
## NA's :3	NA's :5	NA's :3	NA's :7
## Reliability	Keeping.promises	Loss.of.interest	Friends.versus.money
## Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
## 1st Qu.:3.000	1st Qu.:3.000	1st Qu.:2.000	1st Qu.:3.000
## Median :4.000	Median :4.000	Median :3.000	Median :4.000
## Mean :3.859	Mean :3.987	Mean :2.709	Mean :3.779
## 3rd Qu.:5.000	3rd Qu.:5.000	3rd Qu.:4.000	3rd Qu.:5.000
## Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000
## NA's :4	NA's :1	NA's :4	NA's :6
## Funniness	Fake	Criminal.damage	Decision.making
## Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
## 1st Qu.:3.000	1st Qu.:1.000	1st Qu.:1.000	1st Qu.:2.000
## Median :3.000	Median :2.000	Median :2.000	Median :3.000
## Mean :3.293	Mean :2.131	Mean :2.604	Mean :3.198
## 3rd Qu.:4.000	3rd Qu.:3.000	3rd Qu.:4.000	3rd Qu.:4.000
## Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000
## NA's :4	NA's :1	NA's :7	NA's :4
## Elections	Self.criticism	Judgment.calls	Hypochondria
## Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
## 1st Qu.:2.000	1st Qu.:3.000	1st Qu.:3.000	1st Qu.:1.000
## Median :4.000	Median :4.000	Median :4.000	Median :1.000
## Mean :3.415	Mean :3.579	Mean :3.987	Mean :1.913
## 3rd Qu.:5.000	3rd Qu.:5.000	3rd Qu.:5.000	3rd Qu.:3.000
## Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000
## NA's :3	NA's :5	NA's :4	NA's :4
## Empathy	Eating.to.survive	Giving	Compassion.to.animals
## Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
## 1st Qu.:3.000	1st Qu.:1.000	1st Qu.:2.000	1st Qu.:3.000
## Median :4.000	Median :2.000	Median :3.000	Median :4.000
## Mean :3.859	Mean :2.229	Mean :2.976	Mean :3.971
## 3rd Qu.:5.000	3rd Qu.:3.000	3rd Qu.:4.000	3rd Qu.:5.000
## Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000
## NA's :5		NA's :6	NA's :7
## Borrowed.stuff	Loneliness	Cheating.in.school	Health
## Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
## 1st Qu.:3.000	1st Qu.:2.000	1st Qu.:3.000	1st Qu.:3.000
## Median :4.000	Median :3.000	Median :4.000	Median :3.000
## Mean :4.018	Mean :2.887	Mean :3.745	Mean :3.251
## 3rd Qu.:5.000	3rd Qu.:4.000	3rd Qu.:5.000	3rd Qu.:4.000
## Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000
## NA's :2	NA's :1	NA's :4	NA's :1
## Changing.the.past	God	Dreams	Charity
## Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
## 1st Qu.:2.000	1st Qu.:2.000	1st Qu.:3.000	1st Qu.:1.000
## Median :3.000	Median :3.000	Median :3.000	Median :2.000
## Mean :2.952	Mean :3.303	Mean :3.297	Mean :2.104
## 3rd Qu.:4.000	3rd Qu.:5.000	3rd Qu.:4.000	3rd Qu.:3.000

##	Max.	:5.000	Max.	:5.000	Max.	:5.000	Max.	:5.000
##	NA's	:2	NA's	:2			NA's	:3
##	Number.of.friends	Punctuality		Lying		Waiting		
##	Min.	:1.000	Length:1010		Length:1010		Min.	:1.000
##	1st Qu.:	3.000	Class :character		Class :character		1st Qu.:	2.000
##	Median	:3.000	Mode :character		Mode :character		Median	:3.000
##	Mean	:3.344					Mean	:2.672
##	3rd Qu.:	4.000					3rd Qu.:	3.000
##	Max.	:5.000					Max.	:5.000
##							NA's	:3
##	New.environment	Mood.swings		Appearance.and.gestures		Socializing		
##	Min.	:1.000	Min.	:1.000	Min.	:1.000	Min.	:1.000
##	1st Qu.:	3.000	1st Qu.:	3.000	1st Qu.:	3.000	1st Qu.:	2.000
##	Median	:4.000	Median	:3.000	Median	:4.000	Median	:3.000
##	Mean	:3.475	Mean	:3.258	Mean	:3.598	Mean	:3.158
##	3rd Qu.:	4.000	3rd Qu.:	4.000	3rd Qu.:	4.000	3rd Qu.:	4.000
##	Max.	:5.000	Max.	:5.000	Max.	:5.000	Max.	:5.000
##	NA's	:2	NA's	:4	NA's	:3	NA's	:5
##	Achievements	Responding.to.a.serious.letter		Children		Assertiveness		
##	Min.	:1.000	Min.	:1.000	Min.	:1.000	Min.	:1.000
##	1st Qu.:	2.000	1st Qu.:	2.000	1st Qu.:	3.000	1st Qu.:	3.000
##	Median	:3.000	Median	:3.000	Median	:4.000	Median	:4.000
##	Mean	:2.963	Mean	:3.071	Mean	:3.621	Mean	:3.519
##	3rd Qu.:	4.000	3rd Qu.:	4.000	3rd Qu.:	5.000	3rd Qu.:	4.000
##	Max.	:5.000	Max.	:5.000	Max.	:5.000	Max.	:5.000
##	NA's	:2	NA's	:6	NA's	:4	NA's	:2
##	Getting.angry	Knowing.the.right.people		Public.speaking		Unpopularity		
##	Min.	:1.000	Min.	:1.000	Min.	:1.000	Min.	:1.000
##	1st Qu.:	2.000	1st Qu.:	3.000	1st Qu.:	3.000	1st Qu.:	3.000
##	Median	:3.000	Median	:4.000	Median	:4.000	Median	:3.000
##	Mean	:3.015	Mean	:3.486	Mean	:3.522	Mean	:3.462
##	3rd Qu.:	4.000	3rd Qu.:	4.000	3rd Qu.:	5.000	3rd Qu.:	4.000
##	Max.	:5.000	Max.	:5.000	Max.	:5.000	Max.	:5.000
##	NA's	:4	NA's	:2	NA's	:2	NA's	:3
##	Life.struggles	Happiness.in.life		Energy.levels		Small...big.dogs		
##	Min.	:1.000	Min.	:1.000	Min.	:1.000	Min.	:1.000
##	1st Qu.:	2.000	1st Qu.:	3.000	1st Qu.:	3.000	1st Qu.:	2.000
##	Median	:3.000	Median	:4.000	Median	:4.000	Median	:3.000
##	Mean	:3.032	Mean	:3.706	Mean	:3.634	Mean	:2.973
##	3rd Qu.:	4.000	3rd Qu.:	4.000	3rd Qu.:	4.000	3rd Qu.:	4.000
##	Max.	:5.000	Max.	:5.000	Max.	:5.000	Max.	:5.000
##	NA's	:3	NA's	:4	NA's	:5	NA's	:4
##	Personality	Finding.lost.valuables		Getting.up		Interests.or.hobbies		
##	Min.	:1.000	Min.	:1.000	Min.	:1.000	Min.	:1.000
##	1st Qu.:	3.000	1st Qu.:	2.000	1st Qu.:	3.000	1st Qu.:	3.000
##	Median	:3.000	Median	:3.000	Median	:4.000	Median	:4.000
##	Mean	:3.292	Mean	:2.872	Mean	:3.592	Mean	:3.551
##	3rd Qu.:	4.000	3rd Qu.:	4.000	3rd Qu.:	5.000	3rd Qu.:	5.000
##	Max.	:5.000	Max.	:5.000	Max.	:5.000	Max.	:5.000
##	NA's	:4	NA's	:4	NA's	:5	NA's	:3
##	Parents..advice	Questionnaires.or.polls		Internet.usage		Finances		
##	Min.	:1.000	Min.	:1.000	Length:1010		Min.	:1.000
##	1st Qu.:	3.000	1st Qu.:	2.000	Class :character		1st Qu.:	2.000
##	Median	:3.000	Median	:3.000	Mode :character		Median	:3.000

```

## Mean :3.266 Mean :2.749 Mean :3.024
## 3rd Qu.:4.000 3rd Qu.:3.000 3rd Qu.:4.000
## Max. :5.000 Max. :5.000 Max. :5.000
## NA's :2 NA's :4 NA's :3
## Shopping.centres Branded.clothing Entertainment.spending Spending.on.looks
## Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000
## 1st Qu.:2.000 1st Qu.:2.000 1st Qu.:2.000 1st Qu.:2.000
## Median :3.000 Median :3.000 Median :3.000 Median :3.000
## Mean :3.234 Mean :3.051 Mean :3.202 Mean :3.106
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.000
## Max. :5.000 Max. :5.000 Max. :5.000 Max. :5.000
## NA's :2 NA's :2 NA's :3 NA's :3
## Spending.on.gadgets Spending.on.healthy.eating Age Height
## Min. :1.00 Min. :1.000 Min. :15.00 Min. : 62.0
## 1st Qu.:2.00 1st Qu.:3.000 1st Qu.:19.00 1st Qu.:167.0
## Median :3.00 Median :4.000 Median :20.00 Median :173.0
## Mean :2.87 Mean :3.558 Mean :20.43 Mean :173.5
## 3rd Qu.:4.00 3rd Qu.:4.000 3rd Qu.:22.00 3rd Qu.:180.0
## Max. :5.00 Max. :5.000 Max. :30.00 Max. :203.0
## NA's :2 NA's :7 NA's :20
## Weight Number.of.siblings Gender Left...right.handed
## Min. : 41.00 Min. : 0.000 Length:1010 Length:1010
## 1st Qu.: 55.00 1st Qu.: 1.000 Class :character Class :character
## Median : 64.00 Median : 1.000 Mode :character Mode :character
## Mean : 66.41 Mean : 1.298
## 3rd Qu.: 75.00 3rd Qu.: 2.000
## Max. :165.00 Max. :10.000
## NA's :20 NA's :6
## Education Only.child Village...town
## Length:1010 Length:1010 Length:1010
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
##
##
##
## House...block.of.flats
## Length:1010
## Class :character
## Mode :character
##
##
##
##

```

Prije nego počnemo s uporabom deskriptivne statistike i manipulacijom podataka, neka nam misao vodilja budu predložena istraživačka pitanja:

Istraživačko pitanje 1: Razlikuju li se izraženi strahovi ispitanih žena i muškaraca?

Pocinjemo odvajanjem skupa podataka na skup podataka gdje su ispitanici žene i drugog gdje su ispitanici muškarci.


```
zene = pitanja[pitanja$Gender == "female", ]
muskarci = pitanja[pitanja$Gender == "male", ]
```

Pogledajmo sada o kojim se to strahovima radi, odnosno o kojim strahovima imamo prikupljene podatke.

```
for(column_name in names(pitanja[64:73]))
  print(column_name)
```

```
## [1] "Flying"
## [1] "Storm"
## [1] "Darkness"
## [1] "Heights"
## [1] "Spiders"
## [1] "Snakes"
## [1] "Rats"
## [1] "Ageing"
## [1] "Dangerous.dogs"
## [1] "Fear.of.public.speaking"
```

Kako bismo vidjeli kako su ispitanici rangirali razinu straha (ocjena 1-5) moramo proći kroz svaki pojedini strah za naše dvije skupine, vizualizirati podatke i i mjere centralne tendencije za stjecanje boljeg uvida u podatke; žene i muškarce.

```
zene = zene[complete.cases(zene['Rats']),]
muskarci = muskarci[complete.cases(muskarci['Rats']),]

cat('Srednja vrijednost iskazanog straha od štakora kod žena iznosi ', mean(zene$Rats), '\n')

## Srednja vrijednost iskazanog straha od štakora kod žena iznosi 2.721284
cat('Srednja vrijednost iskazanog straha od štakora muškaraca iznosi ', mean(muskarci$Rats), '\n')

## Srednja vrijednost iskazanog straha od štakora muškaraca iznosi 1.963325
cat('Podrezana srednja vrijednost iskazanog straha od štakora kod žena iznosi ', mean(zene$Rats, trim =

## Podrezana srednja vrijednost iskazanog straha od štakora kod žena iznosi 2.651899
cat('Podrezana srednja vrijednost iskazanog straha od štakora muškaraca iznosi ', mean(muskarci$Rats, t

## Podrezana srednja vrijednost iskazanog straha od štakora muškaraca iznosi 1.787234
cat('Medijan iskazanog straha od štakora kod žena iznosi ', median(zene$Rats), '\n')

## Medijan iskazanog straha od štakora kod žena iznosi 3
cat('Medijan iskazanog straha od štakora muškaraca iznosi ', median(muskarci$Rats), '\n')

## Medijan iskazanog straha od štakora muškaraca iznosi 2
cat('Standardna devijacija iskazanog straha od štakora kod žena iznosi ', sd(zene$Rats), '\n')

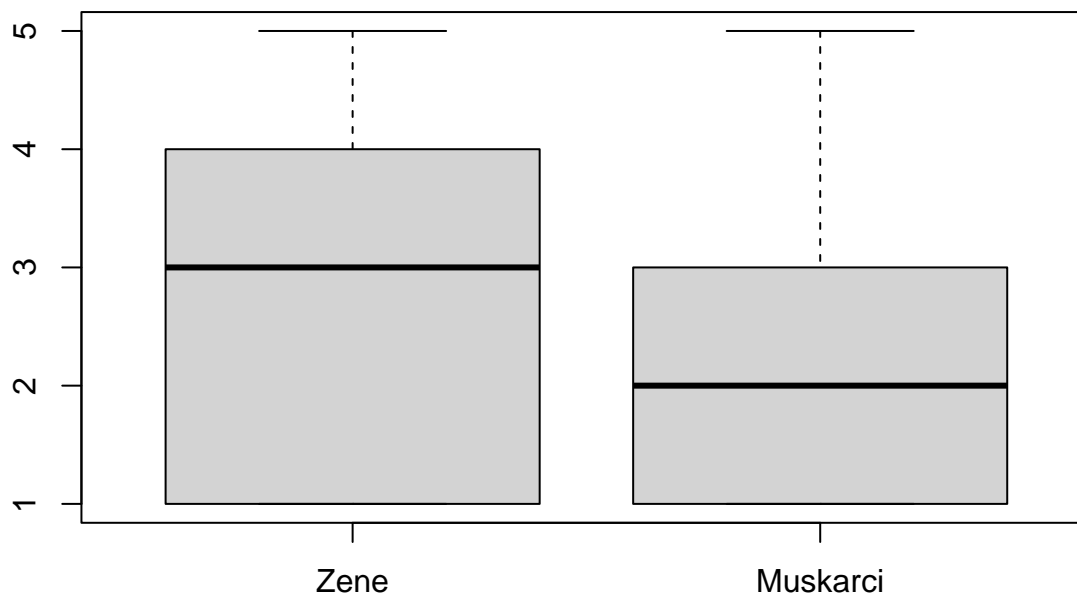
## Standardna devijacija iskazanog straha od štakora kod žena iznosi 1.468802
cat('Standardna devijacija iskazanog straha od štakora muškaraca iznosi ', sd(muskarci$Rats), '\n')

## Standardna devijacija iskazanog straha od štakora muškaraca iznosi 1.161526
```

Nakon što smo vidjeli koliko iznose mjere centralne tendencije, zanima nas kako su te vrijednosti raspoređene, imamo li vrijednosti koje odskakuju i sl., a to najbolje možemo uvidjeti vizualizacijom podataka. Nadalje, histograma ćemo koristiti kako bismo doznali oblik naše distribucije i gustoće podataka.

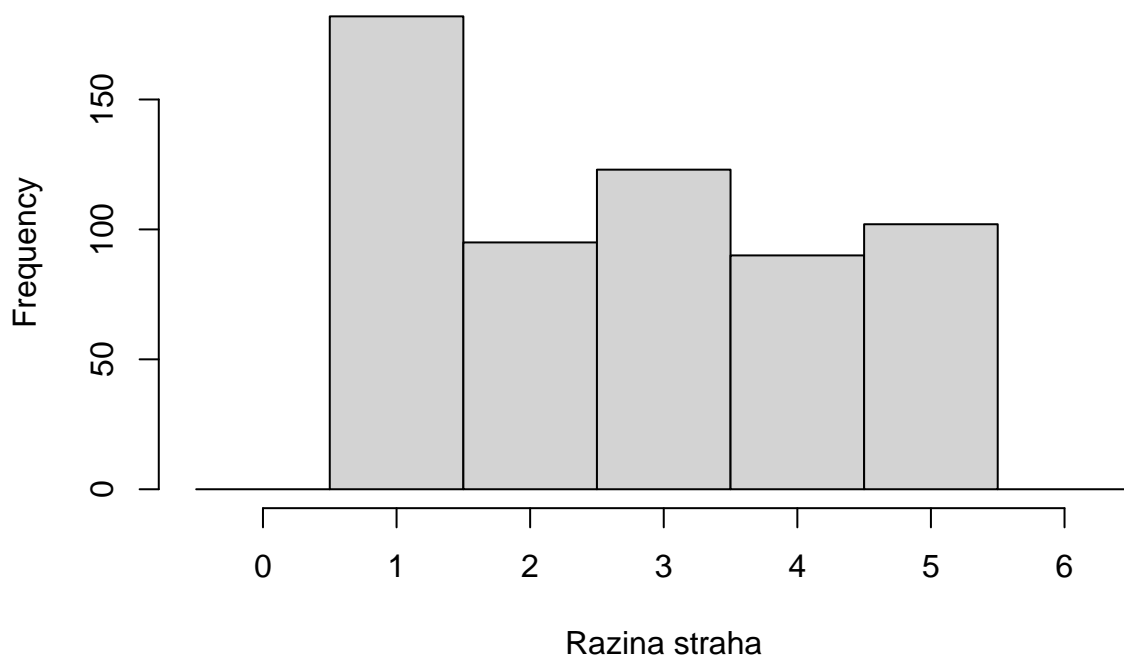
```
# Pravokutni dijagrami dviju skupina za strah od stakora:
boxplot(zene$Rats, muskarci$Rats,
        names = c('Zene', 'Muskarci'),
        main='Pravokutni dijagram razine straha od stakora u zena i muskaraca')
```

Pravokutni dijagram razine straha od stakora u zena i muskaraca



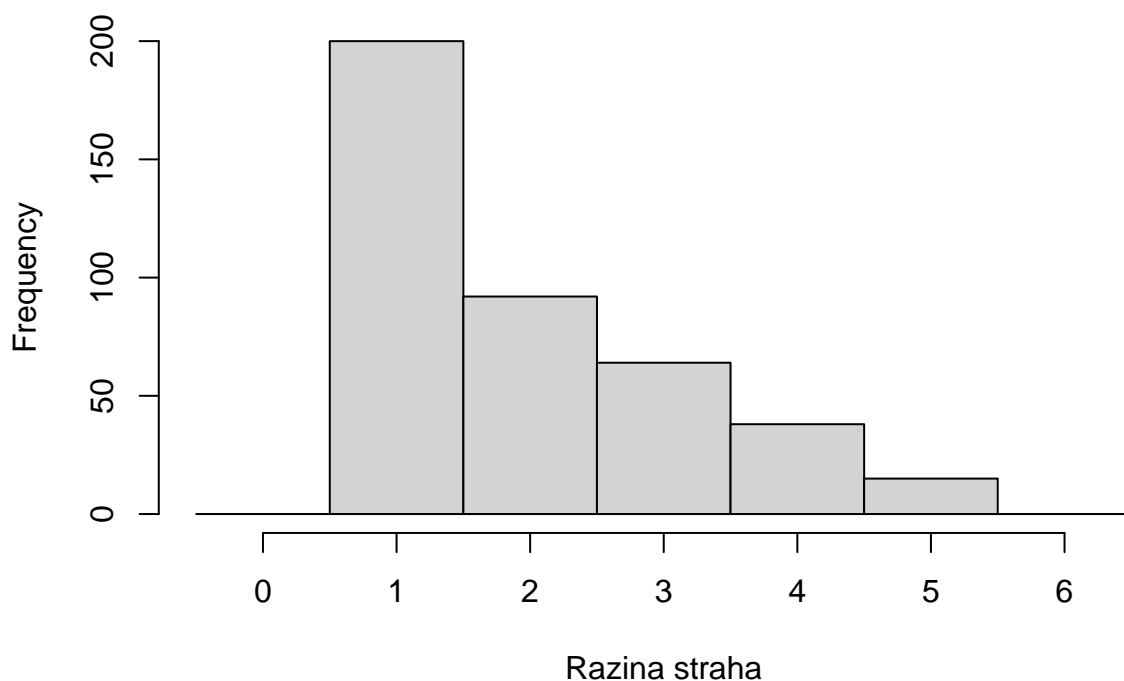
```
hist(zene$Rats,
     breaks=seq(min(zene$Rats)-1.5,max(zene$Rats)+1.5,1),
     main='Histogram razine straha od stakora kod zena',
     xlab='Razina straha')
```

Histogram razine straha od stakora kod žena



```
hist(muskarci$Rats,  
     breaks=seq(min(muskarci$Rats)-1.5,max(muskarci$Rats)+1.5,1),  
     main='Histogram razine straha od stakora kod muskaraca',  
     xlab='Razina straha')
```

Histogram razine straha od stakora kod muskaraca



možemo pretpostaviti da zaista postoji razlika među iskazanim strahom od štakora u žena i muškaraca, Vizualno

ali kako bismo to zaista i dokazali potrebno je provesti statistički test koji će testirati jednakost srednjih vrijednosti dviju populacija.

Ovakvo ispitivanje možemo provesti t-testom.

Testiranje jednakosti srednjih vrijednosti dvije populacije

Neka su $X_1^1, X_1^2, \dots, X_1^{n_1}$ i $X_2^1, X_2^2, \dots, X_2^{n_2}$ dva nezavisna slučajna uzorka koji dolaze iz normalnih distribucija s očekivanjima μ_1 i μ_2 te s nepoznatim, ali jednakim varijancama σ . Zajednička disperzija uzorka se računa kao težinska sredina disperzija S_{X_1} i S_{X_2} :

$$S_X^2 = \frac{1}{n_1 + n_2 - 2} [(n_1 - 1)S_{X_1}^2 + (n_2 - 1)S_{X_2}^2].$$

Slučajna varijabla

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

ima jediničnu normalnu distribuciju. Slučajna varijabla

$$W^2 = \frac{(n_1 - 1)S_{X_1}^2 + (n_2 - 1)S_{X_2}^2}{\sigma^2}$$

ima χ^2 razdiobu s $n_1 + n_2 - 2$ stupnja slobode. Zato slučajna varijabla

$$T = \frac{Z \sqrt{n_1 + n_2 - 2}}{W} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_X \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

ima egzaktnu t distribuciju s $n_1 + n_2 - 2$ stupnja slobode.

Ukoliko imamo 2 nezavisno normalno distribuirana uzorka, ali ovoga puta sa različitim varijancama, tada koristimo testnu statistiku

$$T' = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_{X_1}^2}{n_1} + \frac{s_{X_2}^2}{n_2}}}$$

koja ima aproksimativnu t -distribuciju sa stupnjevima slobode

$$v = \frac{(s_{X_1}^2/n_1 + s_{X_2}^2/n_2)^2}{(s_{X_1}^2/n_1)^2/(n_1 - 1) + (s_{X_2}^2/n_2)^2/(n_2 - 1)}$$

gdje je

$$s_{X_i}^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (X_i^j - \bar{X}_i)^2$$

za $i = 1, 2$.

Hipoteze tada glase:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

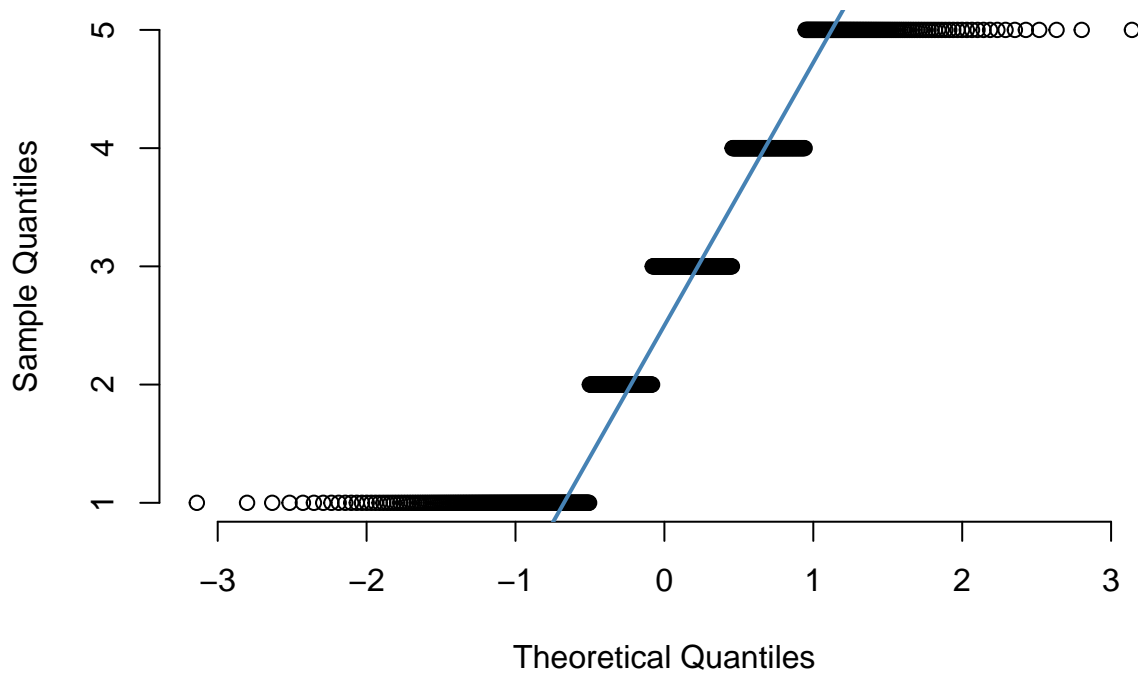
Kako bismo mogli provesti test, moramo najprije provjeriti pretpostavke normalnosti i nezavisnosti uzorka. Već iz histograma možemo vidjeti da bismo mogli imati problem s normalnošću naših podataka, ali ono što nam ide u prilog da će naša statistika ipak biti robusna jest veličina skupa podataka.

Obzirom da razmatramo dva uzoraka dvaju različitih spolova, možemo pretpostaviti njihovu nezavisnost.

Sada, dakle, trebamo provjeriti normalnost podataka koju najčešće provjeravamo: histgoramom (kojeg smo prethodno već iscrtali), qq-plotom te KS-testom (kojim provjeravamo pripadnost podataka distribuciji).

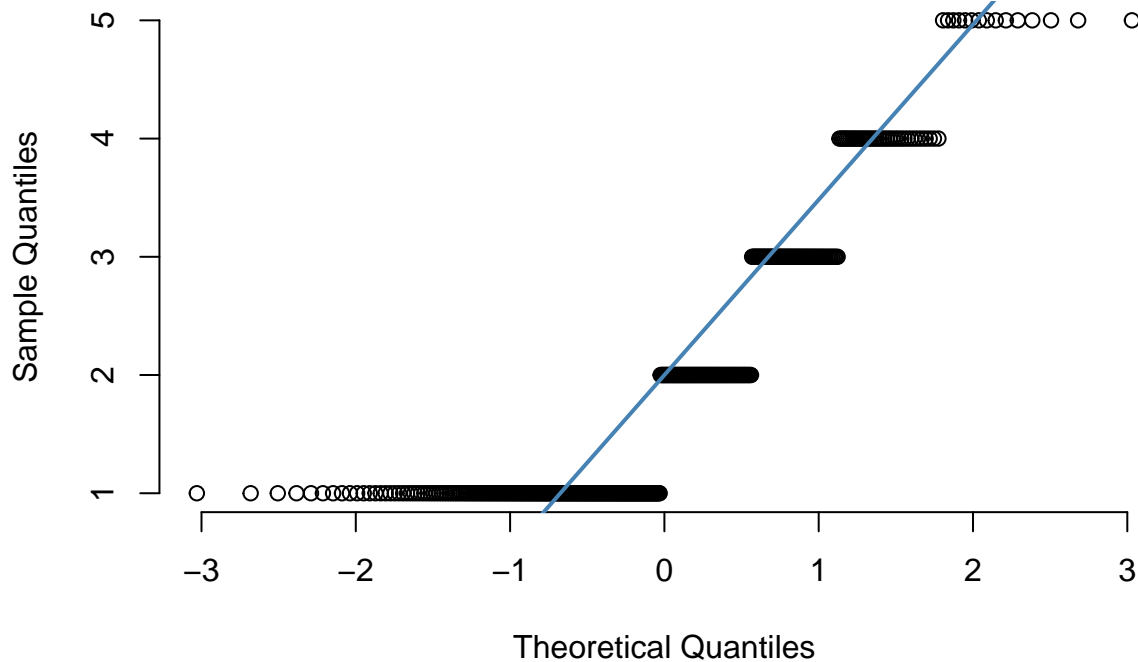
```
qqnorm(zene$Rats, pch = 1, frame = FALSE, main='Strah od stakora kod zena')
qqline(zene$Rats, col = "steelblue", lwd = 2)
```

Strah od stakora kod zena



```
qqnorm(muskarci$Rats, pch = 1, frame = FALSE, main='Strah od stakora kod muskaraca')
qqline(muskarci$Rats, col = "steelblue", lwd = 2)
```

Strah od stakora kod muskaraca



Na temelju qq-plota možemo vidjeti da su naše sumnje u normalnost podataka zaista opravdane. Ali, kao što smo već prije spomenuli, statistika bi idalje mogla biti robusna obzirom na veličinu našeg skupa podataka. Provjerimo zadovoljavamo li ostale preduvjete za provođenje t-testa, uzevši u obzir već izračunate varijance naših podataka. Testirajmo prvo jesu li naše varijance značajno različite.

Test o jednakosti varijanci

Ako imamo dva nezavisna slučajna uzorka $X_1^1, X_1^2, \dots, X_1^{n_1}$ i $X_2^1, X_2^2, \dots, X_2^{n_2}$ koji dolaze iz normalnih distribucija s varijancama σ_1^2 i σ_2^2 , tada slučajna varijabla

$$F = \frac{S_{X_1}^2 / \sigma_1^2}{S_{X_2}^2 / \sigma_2^2}$$

ima Fisherovu distribuciju s $(n_1 - 1, n_2 - 1)$ stupnjeva slobode, pri čemu vrijedi:

$$S_{X_1}^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_1^i - \bar{X}_1)^2, \quad S_{X_2}^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (X_2^i - \bar{X}_2)^2.$$

Hipoteze testa jednakosti varijanci glase:

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

U programskom paketu R test o jednakosti varijanci je implementiran u funkciji `var.test()`, koja prima uzorke iz dvije populacije čije varijance uspoređujemo.

Dakle, ispitajmo jednakost varijanci naših danih uzoraka.

```
var.test(zene$Rats, muskarci$Rats)
```

```
##
```

```
## F test to compare two variances
```

```
##
## data: zene$Rats and muskarci$Rats
## F = 1.5991, num df = 591, denom df = 408, p-value = 4.314e-07
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 1.335516 1.908986
## sample estimates:
## ratio of variances
## 1.599076
```

p-vrijednost od 4.314e-07 nam govori da nećemo odbaciti hipotezu H_0 da su varijance naša dva uzorka jednaka.

Provedimo sada t-test uz pretpostavku jednakosti varijanci.

```
# Uvijek se držimo istog poretka
t.test(zene$Rats, muskarci$Rats, alt = "two.sided", var.equal = TRUE)
```

```
##
## Two Sample t-test
##
## data: zene$Rats and muskarci$Rats
## t = 8.7206, df = 999, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 0.5874004 0.9285168
## sample estimates:
## mean of x mean of y
## 2.721284 1.963325
```

Zbog jako male p-vrijednost možemo odbaciti H_0 hipotezu o jednakosti izrazenih strahov u korist H_1 , odnosno možemo reći da se izraženi strahovi kod muškaraca i žena razlikuju.

Pokušajmo sada nešto robusnije od t-testa kako bismo dokazali svoju pretpostavku, vidjeli smo da naši podaci ne odgovaraju normalnosti, što ćemo i dodatno potvrditi Shapiro testom. Provedimo dakle neparametarski test kako bismo ustanovili razlikuju se izraženi strahovi žena i muškaraca, dakle koristeći iste hipoteze. Koristit ćemo Mann-Whitney i Wilcox signed-rank test.

```
shapiro.test(muskarci$Rats) #podaci nisu normalno distribirani - studentov t-test nije prikladan
```

```
##
## Shapiro-Wilk normality test
##
## data: muskarci$Rats
## W = 0.78849, p-value < 2.2e-16
```

```
wilcox.test(muskarci$Rats, zene$Rats, paired = FALSE)
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: muskarci$Rats and zene$Rats
## W = 86003, p-value = 5.698e-16
## alternative hypothesis: true location shift is not equal to 0
wilcox.test(zene$Rats, muskarci$Rats, paired = FALSE)
```

```
##
## Wilcoxon rank sum test with continuity correction
```

```
##
## data:  zene$Rats and muskarci$Rats
## W = 156125, p-value = 5.698e-16
## alternative hypothesis: true location shift is not equal to 0
```

Zaključno, nakon što smo Shapiro testom potvrdili da se podaci ne ravnaaju po normalnoj distribuciji, uz 95% interval povjerenja, možemo odbaciti H_0 hipotezu o jednakosti izrazenih strahova u korist H_1 , odnosno možemo reći da se izraženi strahovi kod muškaraca i žena razlikuju.

Možemo primjetiti i da se različitim redoslijedom argumenata W vrijednost mijenja u Mann-Whitney u-testu. Uzevši u obzir uobičajne prakse, zadržavamo vrijednost manje statistile i bilježimo da je $W = 86003$ iako se p-vrijednosti pritom, naravno, ne mijenjaju.

Provedimo još sada Wilcox signed-rank test s našim uparenim uzorcima. Potrebno će dakle biti uskladiti veličine uzoraka te ćemo se zadržati pri tome da veći uzorak smanjimo na veličinu manjeg uzorka, iako postoje sofisticiranije metode, ovdje ćemo se ipak zadovoljiti ovakvim pristupom.

```
uzorak = pitanja[complete.cases(pitanja['Rats']),]
zene = uzorak[uzorak$Gender == "female", ]
muskarci = uzorak[uzorak$Gender == "male", ]

dim(zene)
```

```
## [1] 592 150
```

```
dim(muskarci)
```

```
## [1] 409 150
```

```
zene2 = zene[1:409,] #broj muškaraca u skupu podataka
```

```
wilcox.test(muskarci$Rats, zene2$Rats, paired = TRUE)
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data:  muskarci$Rats and zene2$Rats
## V = 13676, p-value = 2.198e-14
## alternative hypothesis: true location shift is not equal to 0

wilcox.test(zene2$Rats, muskarci$Rats, paired = TRUE)
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data:  zene2$Rats and muskarci$Rats
## V = 39299, p-value = 2.198e-14
## alternative hypothesis: true location shift is not equal to 0
```

Provođenjem Wilcox signed-rank test, uz 95% interval povjerenja, također, možemo odbaciti H_0 hipotezu o jednakosti izrazenih strahova u korist H_1 , odnosno možemo reći da se izraženi strahovi kod muškaraca i žena razlikuju. Ovdje također bilježimo vrijednost manje statistike, tj, $V = 13676$.

Istovjetne postupke, vizualizacije i testiranja proveli smo i na svim drugim strahovima te su rezultati više-manje istovjetni i možemo ustvrditi da se strahovi zaista razlikuju između žena i muškaraca, ali se zbog opsežnosti ove bilježnice nećemo upuštati u provođenje testova. Valja naglasiti da se i kod ostalih skupina podaci ne ravnaaju po normalnoj distribuciji te je učinkovitije bilo provoditi neparametarske testove.