

Sređivanje podataka

Nedelja 2 - Vežbe

Dr Nikola N. Grubor

**Grupisanje je proces razvrstavanja
(organizovanja, klasifikovanja) podataka.**

- Apsolutna učestalost = Prebrojavanje podataka po kategorijama / intervalu

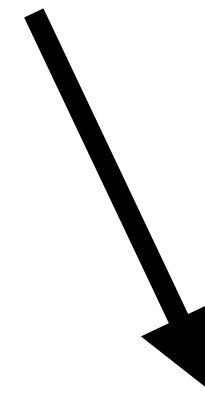
- Relativna učestalost = $\frac{\text{Apsolutna učestalost}}{\text{Ukupan broj podataka}}$ (%)

Tabelarno prikazivanje podataka

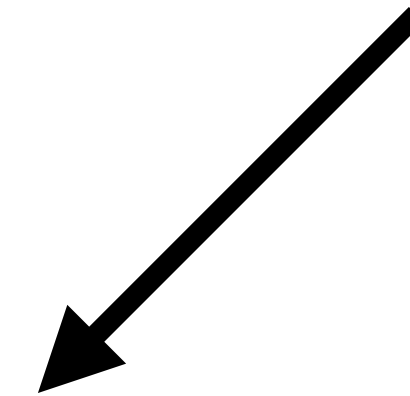
Navika pušenja	n	%
Pušači	23	46
Nepušači	27	54
Ukupno	50	100

Tabelarno prikazivanje podataka

Apsolutne učestalosti



Relativne učestalosti



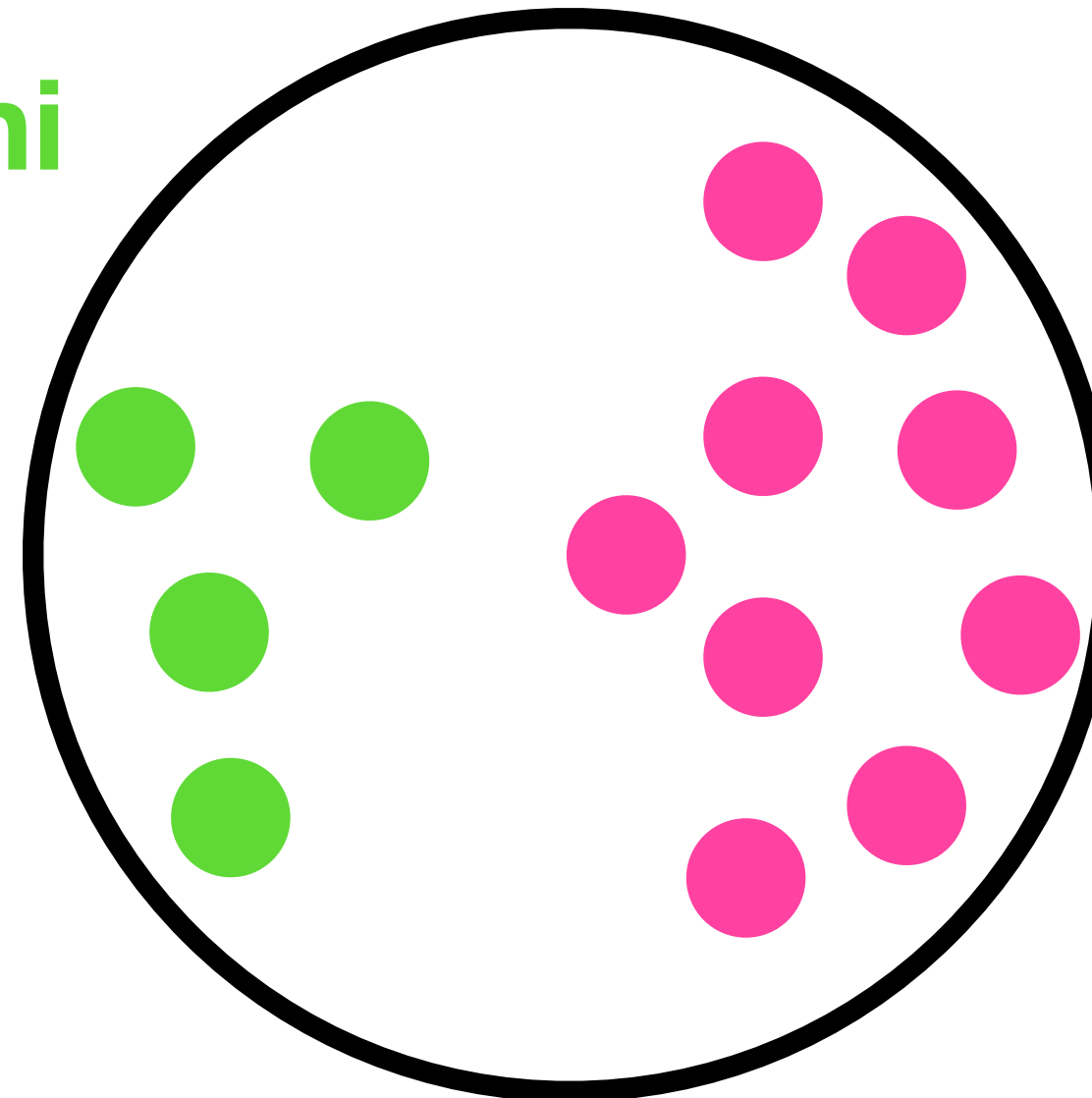
Navika pušenja	n	%
Pušači	23	46
Nepušači	27	54
Ukupno	50	100

Hospitalizovani sa COVID-19



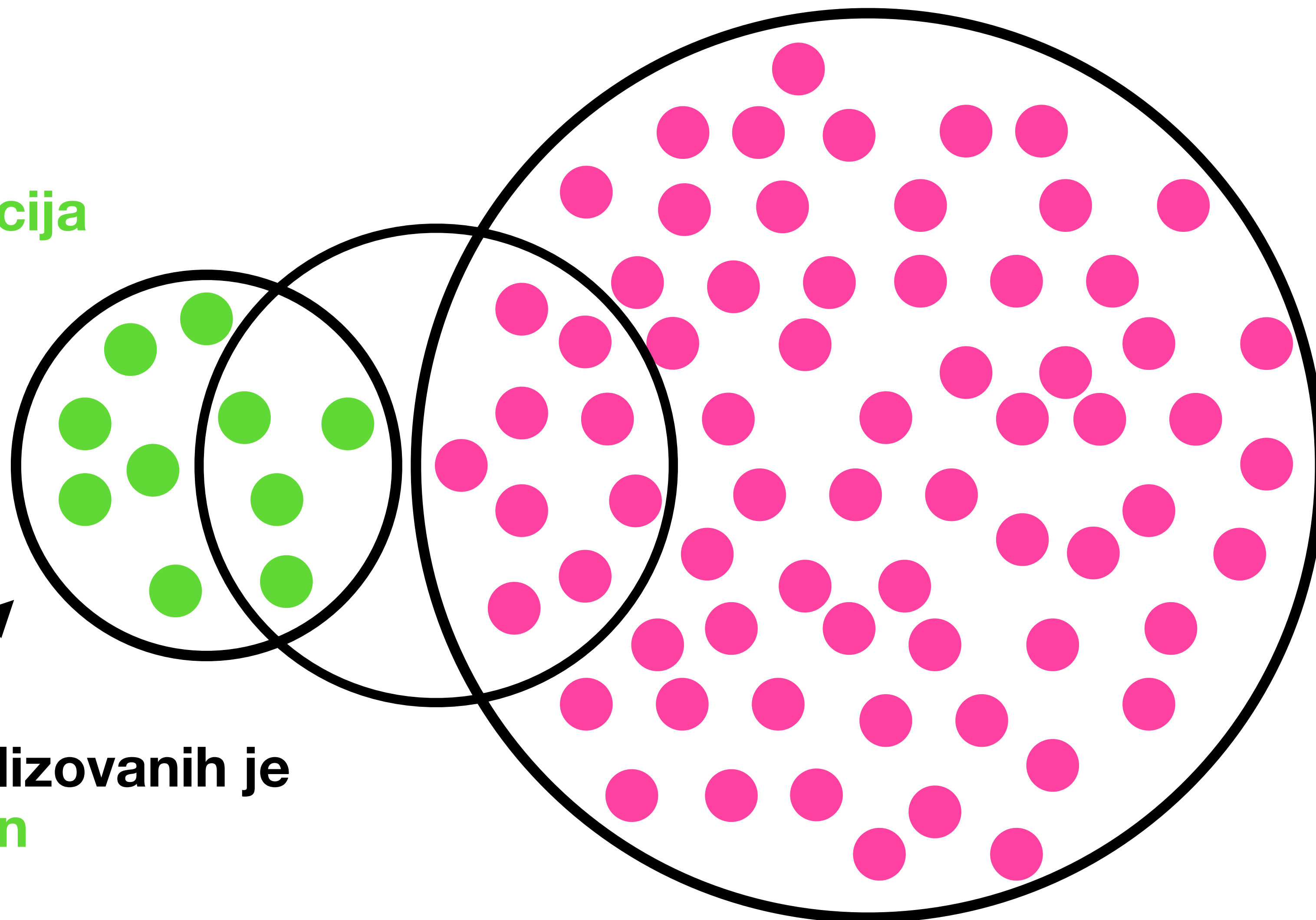
Nevakcinisani

Vakcinisani



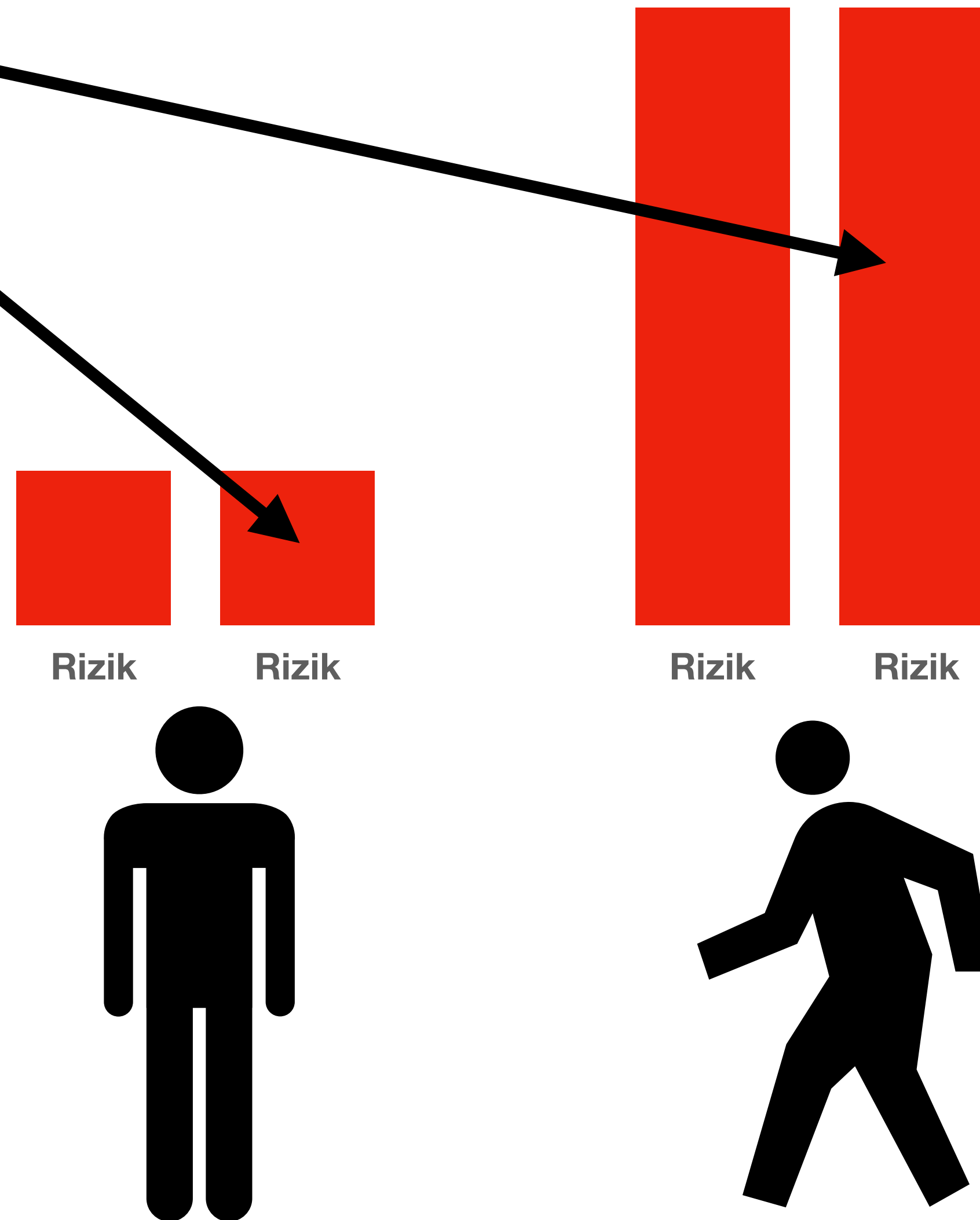
Vakcinisana populacija

Nevakcinisana populacija



Veći procenat hospitalizovanih je
nevakcinisan

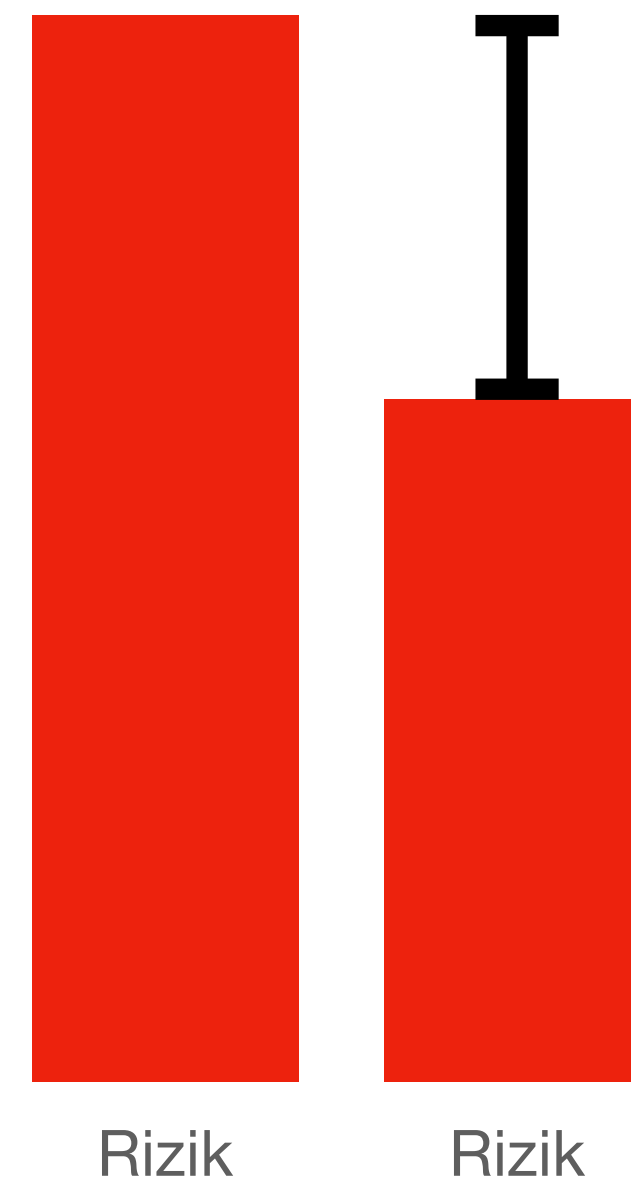
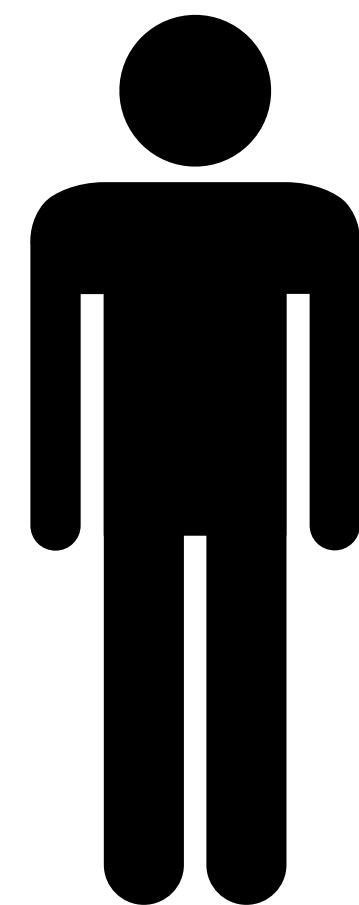
Šta će se desiti ako smanjimo rizik?



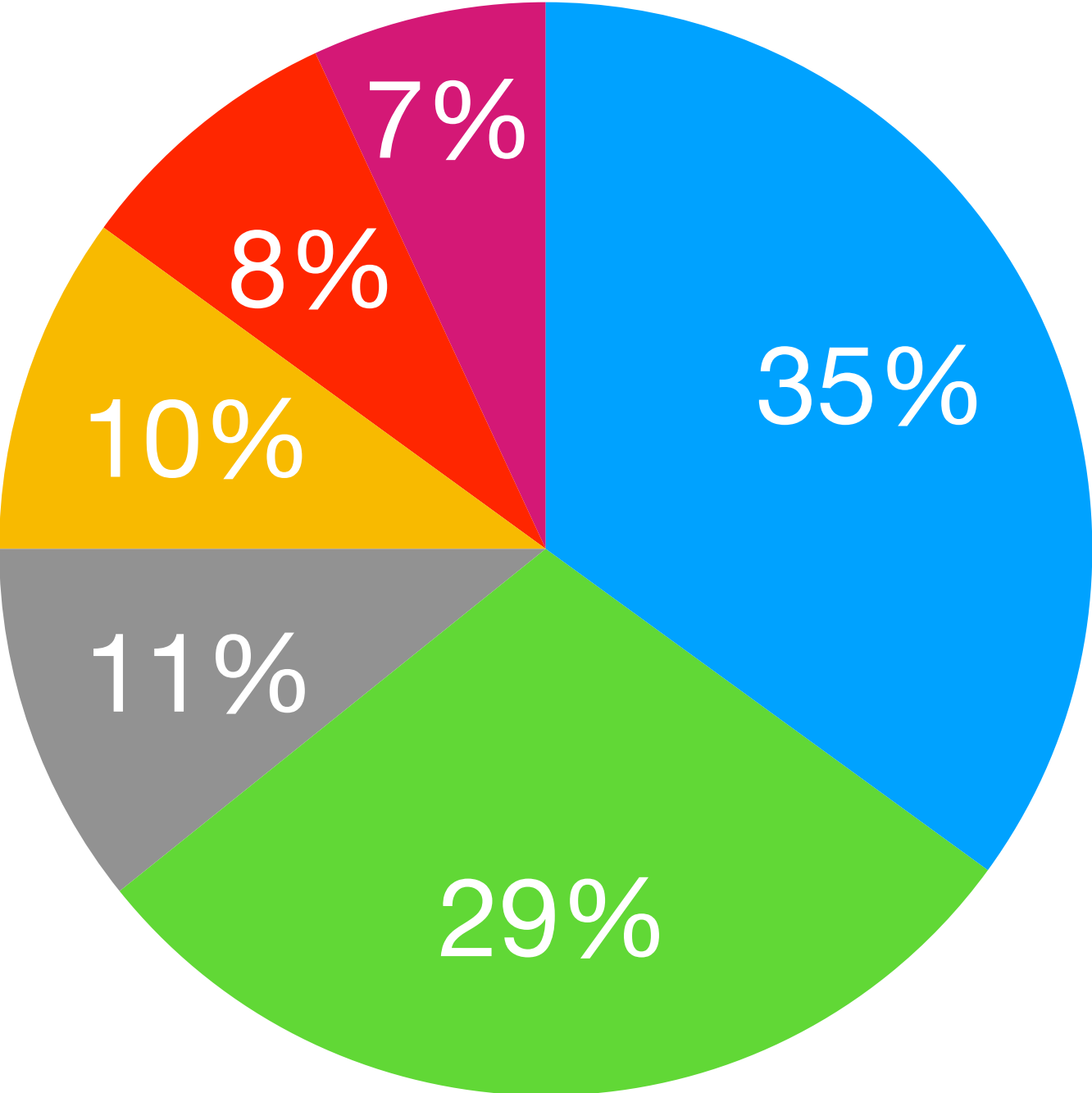
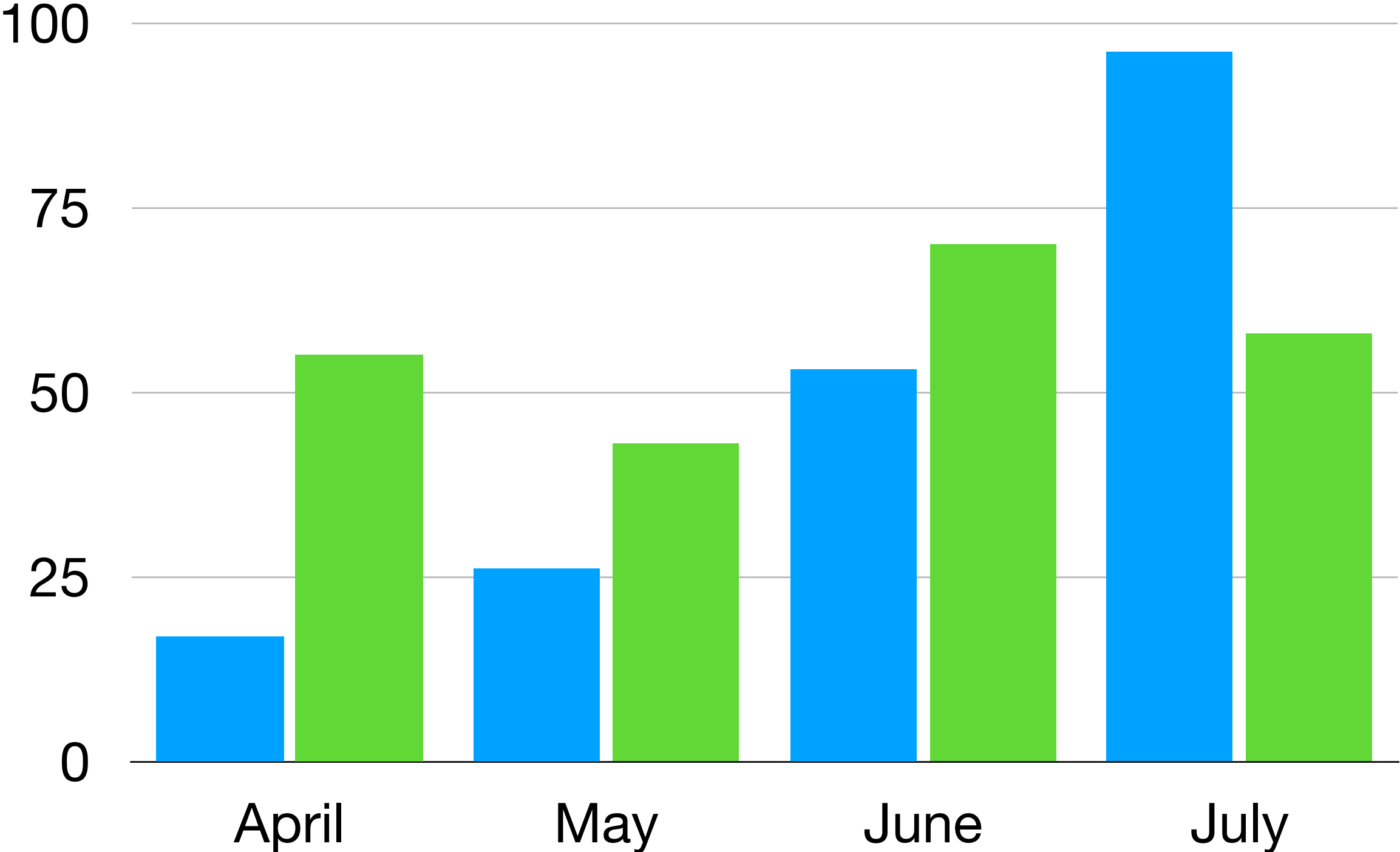
10/5 mmHg

=

20%

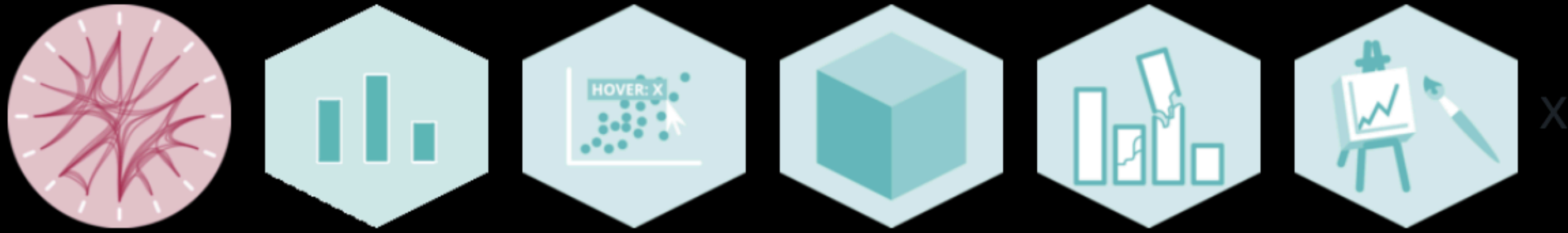


Grafičko prikazivanje podataka



Grafičko prikazivanje podataka

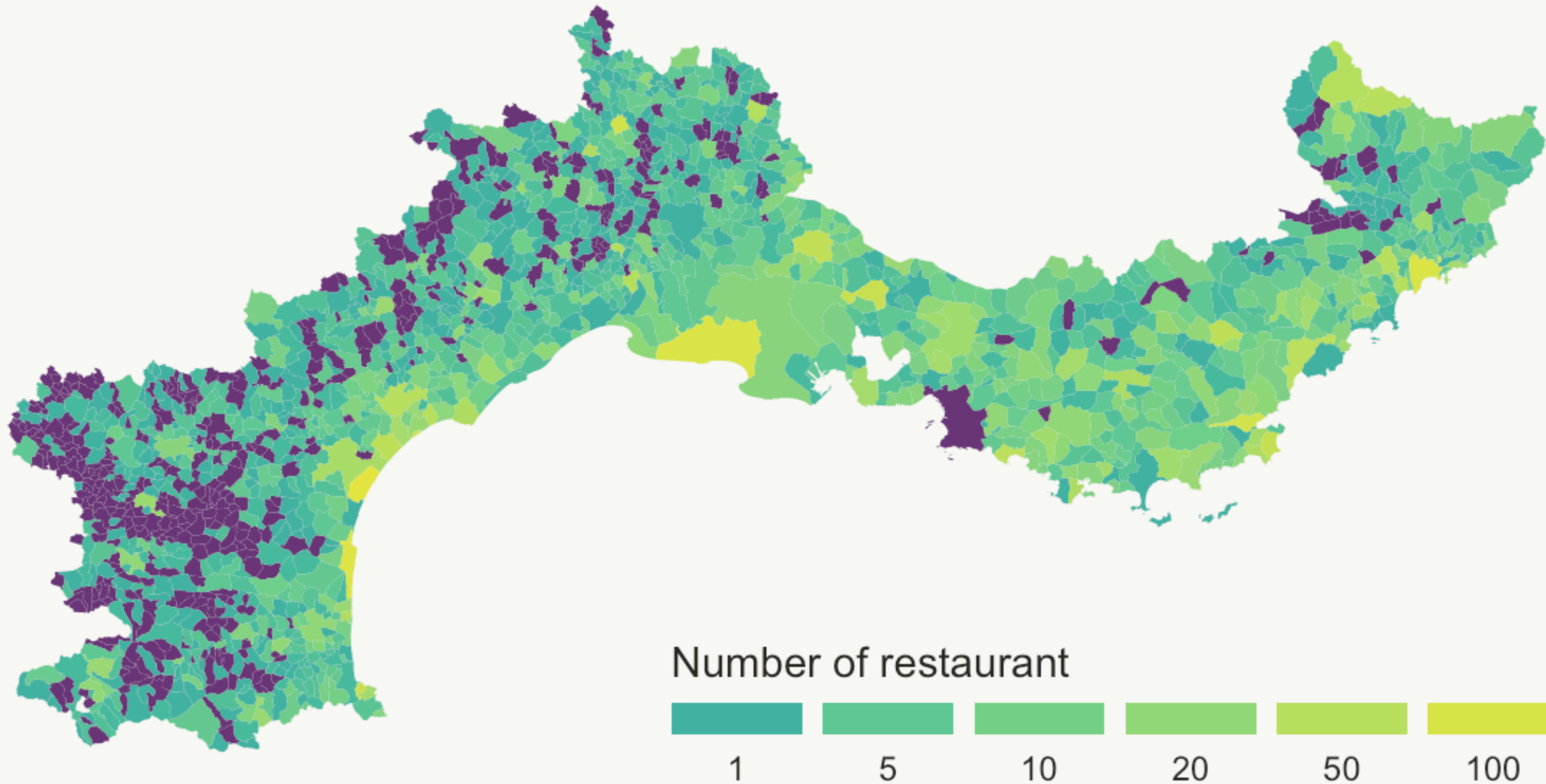




x

South of France Restaurant concentration

Number of restaurant per city district



Nominalni podaci

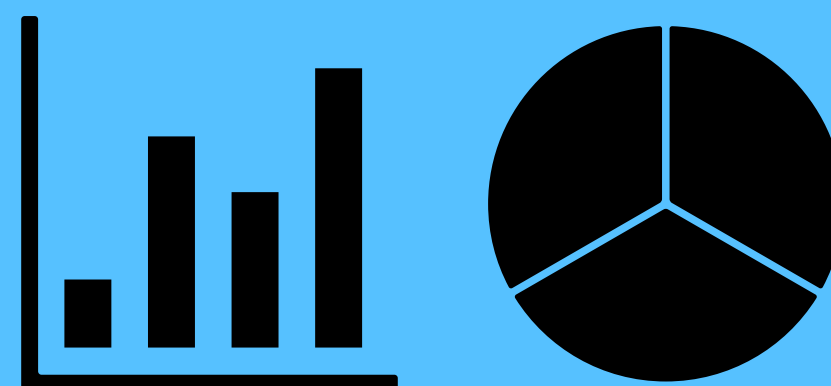
Grupišu se po **kategorijama**

Učitati bazu: “zadovoljstvo zdravstvenom zaštitom”

Apsolutne i relativne učestalosti za **pol** i **pušenje**

Grafički prikaz

- stubičasti dijagram (bar chart)
- kružni dijagram (pie chart)
- sačuvati grafikon



R Frequency Distributions ✕

Click pressing Ctrl key to select multiple variables

Variables (pick one or more)

hr_bronhitis
ID
pol
pusenje
starost
zadovoljstvo_zz

Chi-square goodness-of-fit test (for one variable only) ☐

Show missing data ☐

Show percent ☒

Show graph ☒



Help



OK



Cancel

Output



Submit

```
> #####Frequency Distributions#####
```

```
> .Table <- NULL
```

```
> (.Table <- table(Dataset$pol)) # counts for pol
```

muski	zenski
21	14

apsolutne učestalosti

```
> round(100*.Table/sum(.Table), 2) # percentages for pol
```

muski	zenski
60	40

relativne učestalosti

Graphs and tables Tools Help Original me

Bar graph(Frequencies)
Pie chart(Frequencies)
Stem-and-leaf display
Histogram
Quantile-comparison plot
Bar graph(Means)
Line graph(Means)
Line graph(Repeated measures)
Boxplot
Dot chart
Ordered chart

Scatterplot
Scatterplot matrix

Adjusted survival curve
Adjusted cumulative incidence curve
Stacked cumulative incidences

Graph settings
Graph colors
Graph detailed colors

Summary table of sample characteristics
Summary table of results

R Pie chart(Frequencies) X

Variable (pick one)

hr_bronhitis
ID
pol
pusenje
starost
zadovoljstvo_zz

☒ Draw in color

Include in Segment Labels

☒ Percentages
☐ Frequency counts
☐ Neither

Condition to limit samples for analysis. Ex1. age>50 & Sex==0 Ex2. age<50 | Sex==1

<all valid cases>

< >

Help Reset OK Cancel Apply

R Graphics: Device 2 (ACTIVE)

File History Resize

Save as >

Copy to the clipboard >

Print... Ctrl+P

close Device

Metafile...

Postscript...

PDF...

Png...

Bmp...

TIFF...

Jpeg >

50% quality...

75% quality...

100% quality...

quency

15

10



Ordinalni podaci

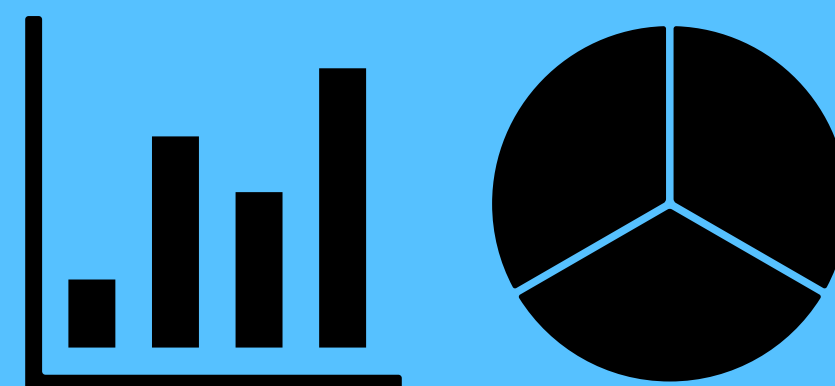
Grupišu se po **kategorijama**

Učitati bazu: "zadovoljstvo zdravstvenom zaštitom"


Apsolutne i relativne učestalosti za **zadovoljstvo_zz**

Grafički prikaz

- stubičasti dijagram (bar chart)
- kružni dijagram (pie chart)
- sačuvati grafikon



Output

 Generate report

```
> #####Frequency Distributions#####
```

```
> .Table <- NULL
```

```
> (.Table <- table(Dataset$zadovoljstvo_zz)) # counts for zadovoljstvo_zz
```

1	2	3	4	5
1	4	15	12	3

apsolutne učestalosti

```
> round(100*.Table/sum(.Table), 2) # percentages for zadovoljstvo_zz
```

1	2	3	4	5
2.86	11.43	42.86	34.29	8.57

relativne učestalosti

Numerički diskontinuirani podaci

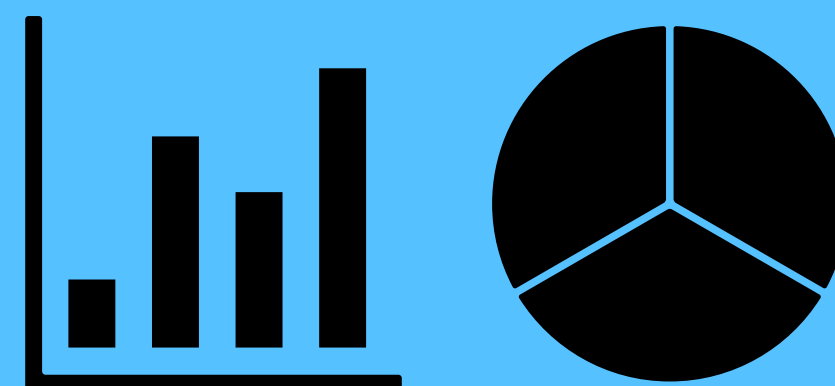
Grupišu se po **diskretnim vrednostima** ili **klasnim intervalima**

Učitati bazu: "zadovoljstvo zdravstvenom zaštitom"

Apsolutne i relativne učestalosti za **hr_bronhitis**

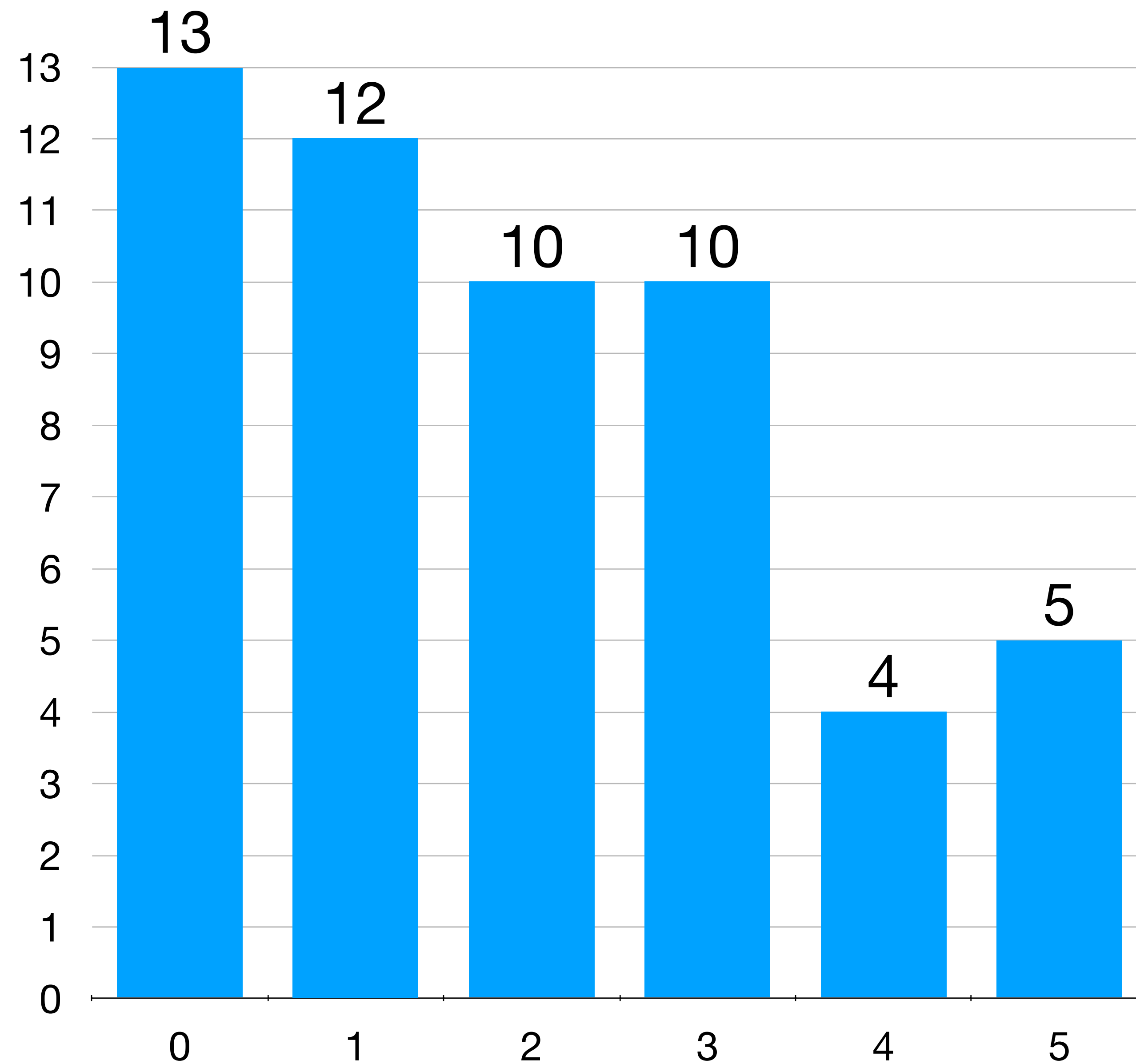
Grafički prikaz

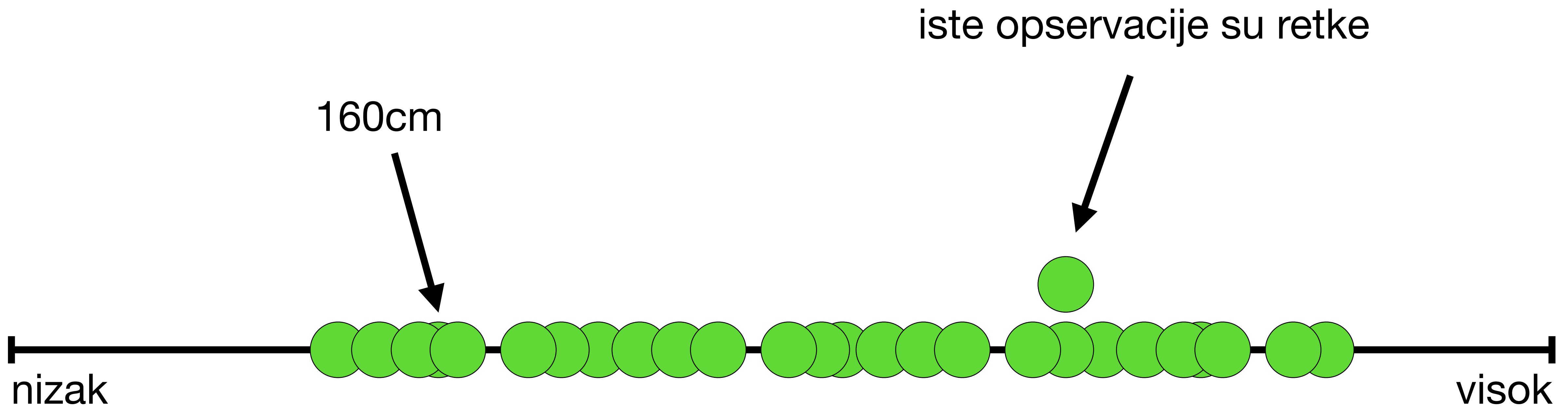
- stubičasti dijagram (bar chart)
- kružni dijagram (pie chart)
- sačuvati grafikon



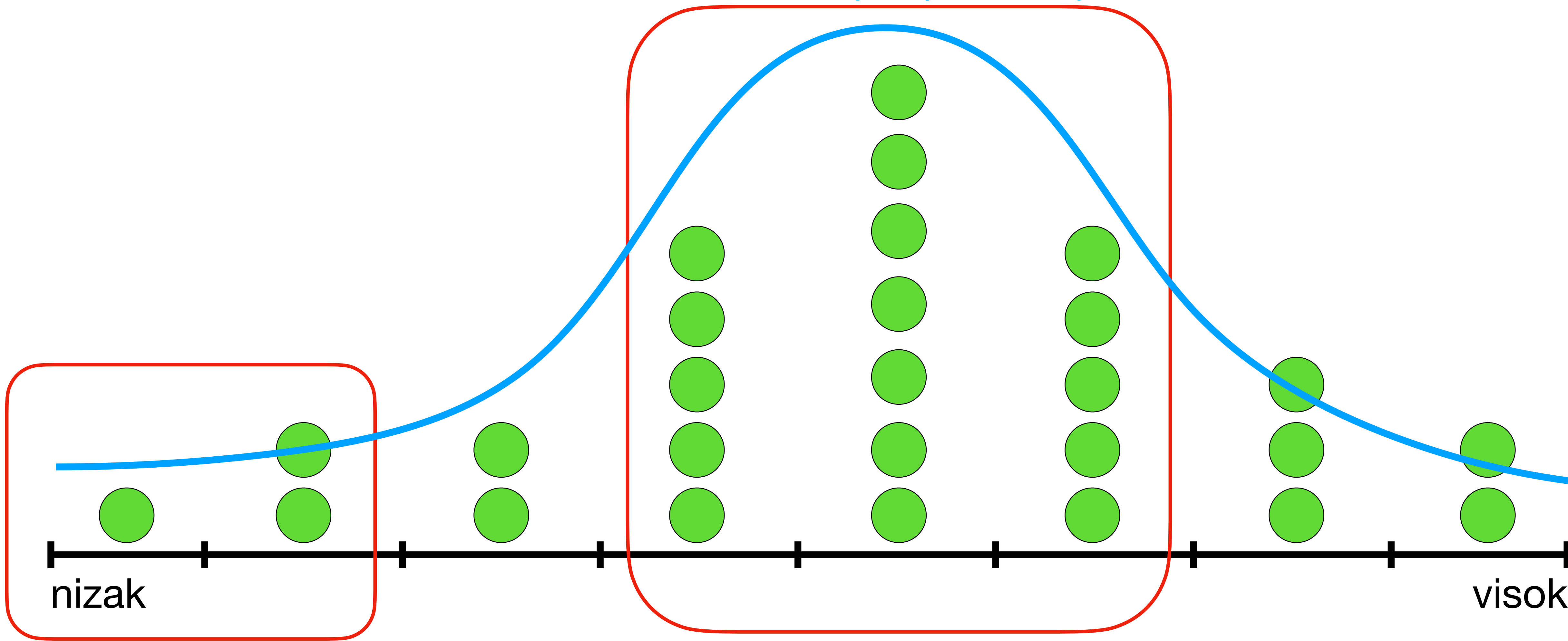
Broj respiratornih infekcija	n	%
0	13	26
1	12	24
2	10	20
3	10	20
4	3	6
5	2	4
Ukupno	50	100

Broj respiratornih infekcija



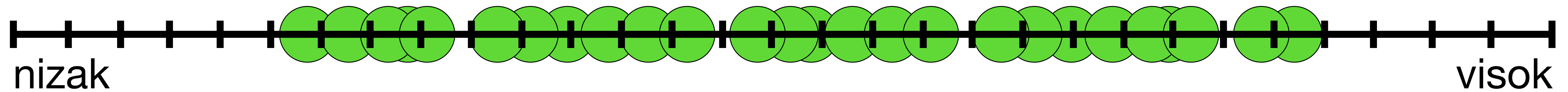


distribucija opservacija

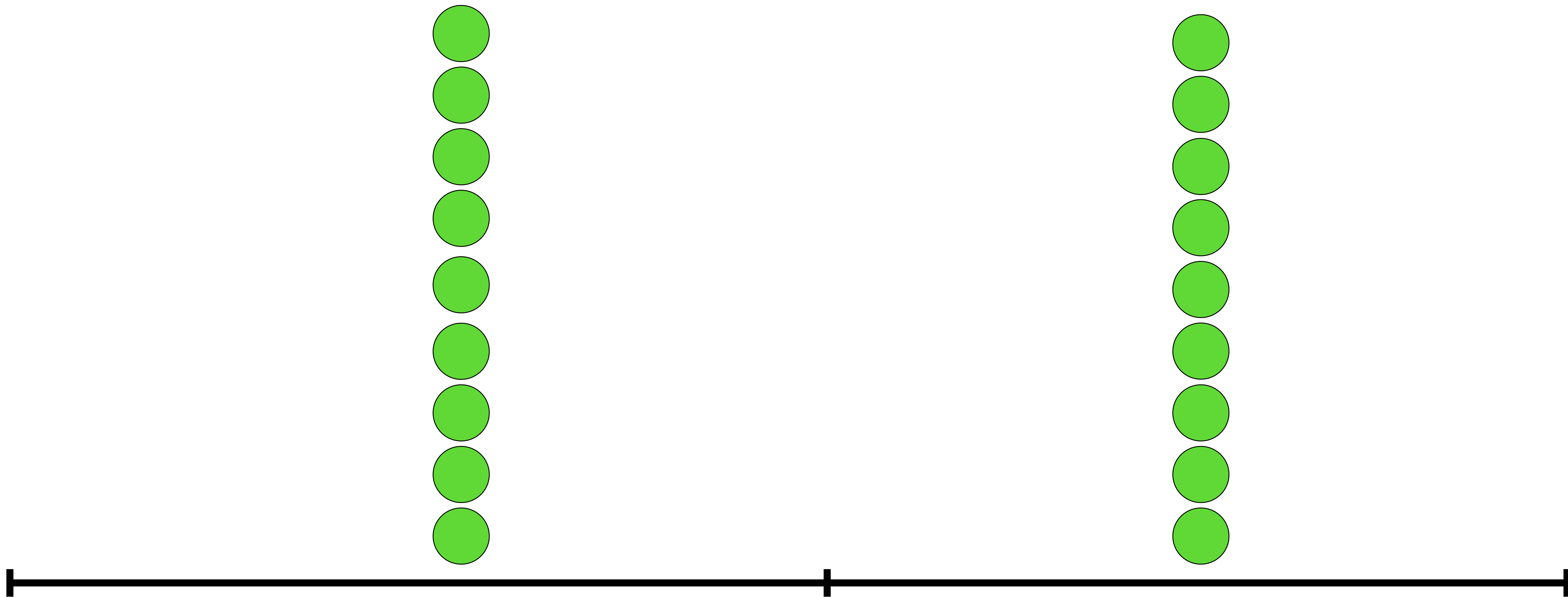


manje verovatne opservacije

verovatnije opservacije



uski intervali nisu korisni



široki intervali nisu korisni

Numerički kontinuirani podaci

Grupišu se po **klasnim intervalima**

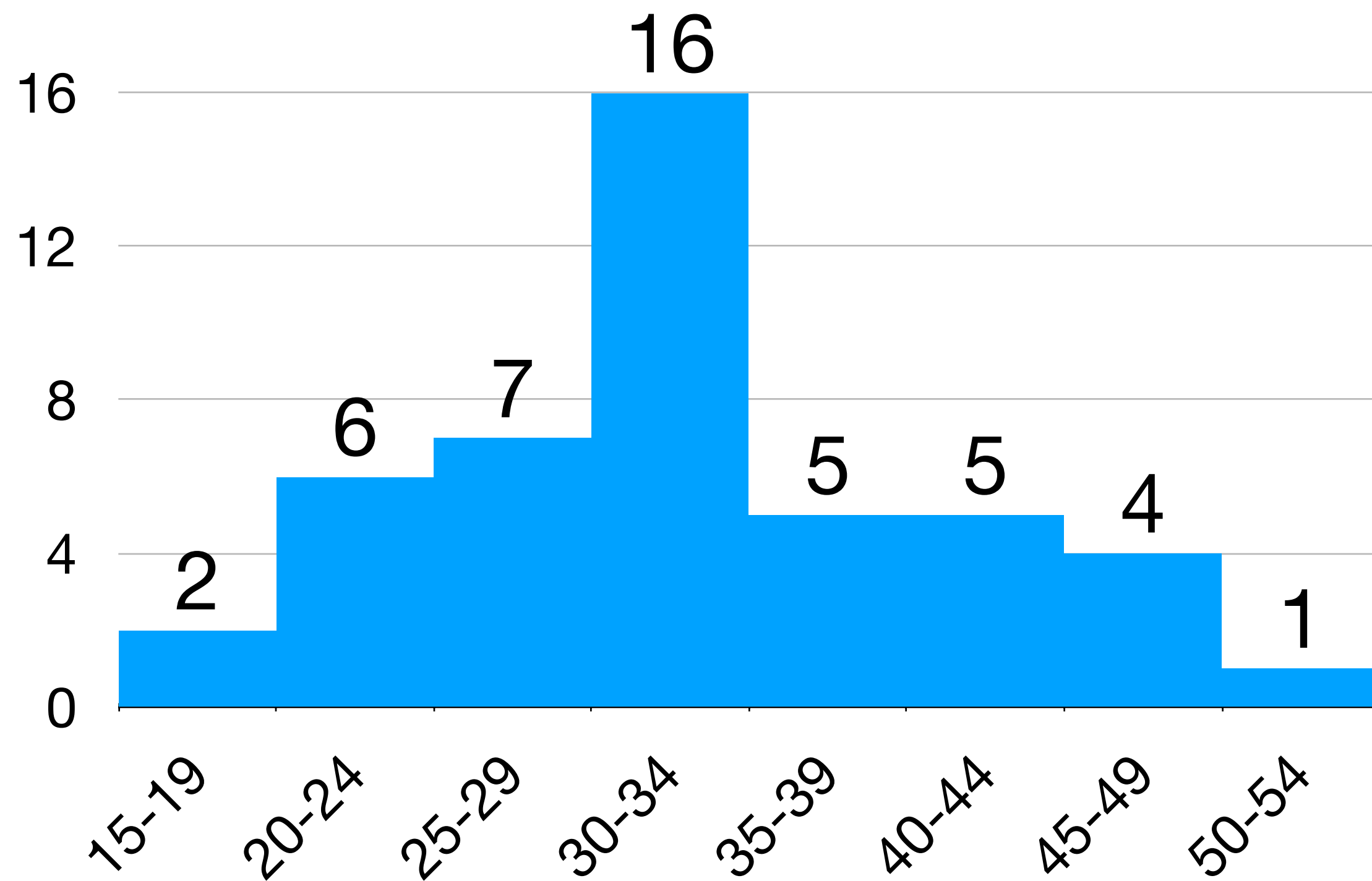
Učitati bazu: "zadovoljstvo zdravstvenom zaštitom"

Apsolutne i relativne učestalosti za **starost**

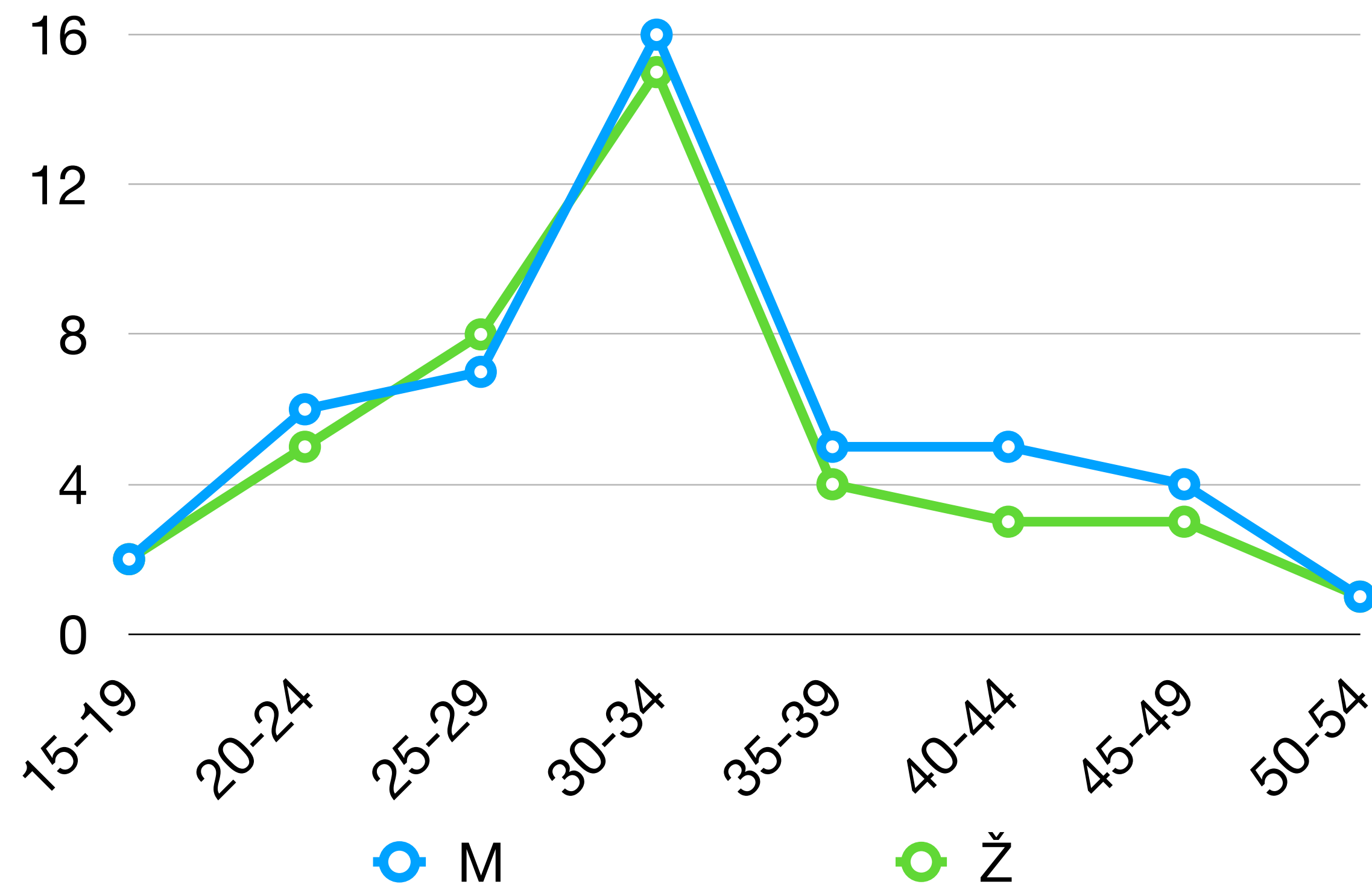
Grafički prikaz

- histogram
- poligon frekvencija
- sačuvati grafikon


Histogram



Poligon frekvencija



Graphs and tables → Histogram

 Histogram ✕

Variable (pick one)

- hr_bronhitis
- ID
- starost**
- zadovoljstvo_zz

Grouping variable(pick 0 or 1)

- hr_bronhitis
- ID
- pol
- pusenje
- starost
- zadovoljstvo_zz

☐ Draw in color (when grouped)

Number of sections (when not grouped)






Y axis (when not grouped)

☒ Frequency counts

☐ Percentages

☐ Densities

Condition to limit samples for analysis. Ex1. age > 50 & Sex == 0 Ex2. age < 50 | Sex == 1

 Help  Reset  OK  Cancel  Apply

V1. Napraviti Tabelu

Koja lokalizacija infarkta je najčešća, a koja najređa?

DZ, P, DZ, DZ, AD, P, DZ, DZ, P, P, DZ, P, P, DZ, DZ, DZ, P, P, DZ, DZ

P (Prednji),
AD (Anteroseptalni/dijafragmalni),
DZ (Donji/zadnji)

Variable	Variable	Variable
Opervacije	Vrednost	Vrednost
Opservacije	Vrednost	Vrednost
Opervacije	Vrednost	Vrednost
Ukupno	Ukupno	Ukupno

V1. Rešenje

Koja lokalizacija infarkta je najčešća, a koja najređa?

Lokalizacija	Broj	%
P	8	40
DZ	11	55
AD	1	5
Ukupno	20	100

Istraživački zadatak

1. Skinuti bazu 'DZ' sa kursa
2. Učitati je u EZR
3. Izračunati apsolutne i relativne učestalosti za:
 - Hipertenziju
 - Uhranjenost
 - Starost
4. Grafički ih prikazati odgovarajućom metodom