GRAFIČKI PRIKAZ PODATAKA

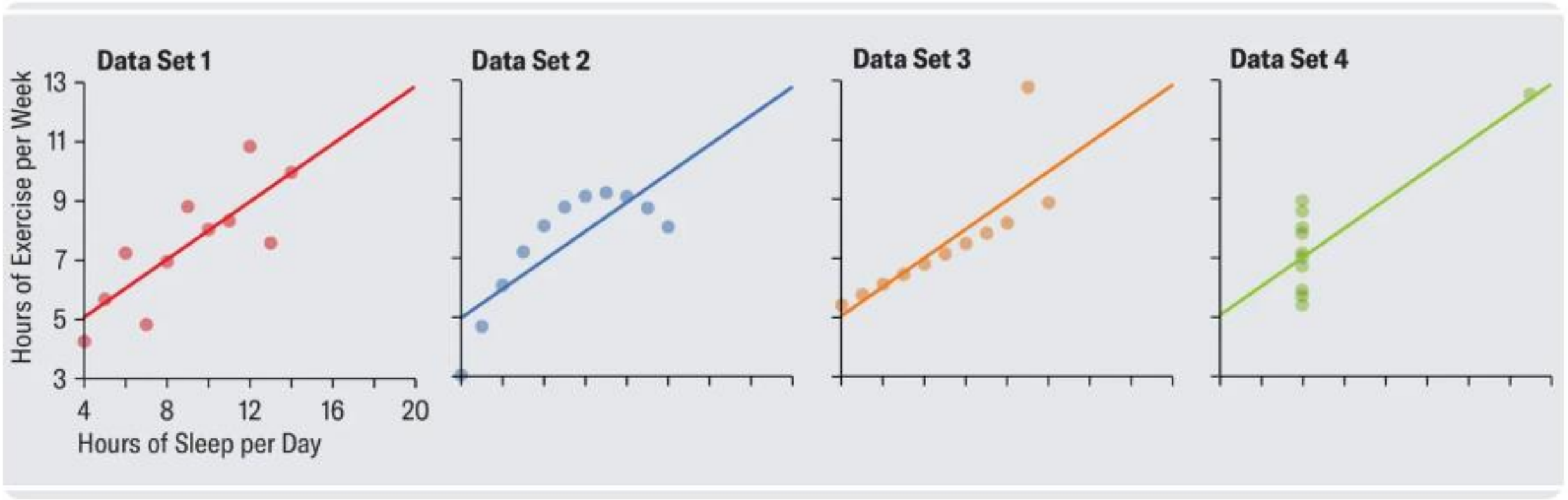


“There are three kinds of lies: lies, damned lies, and statistics.”

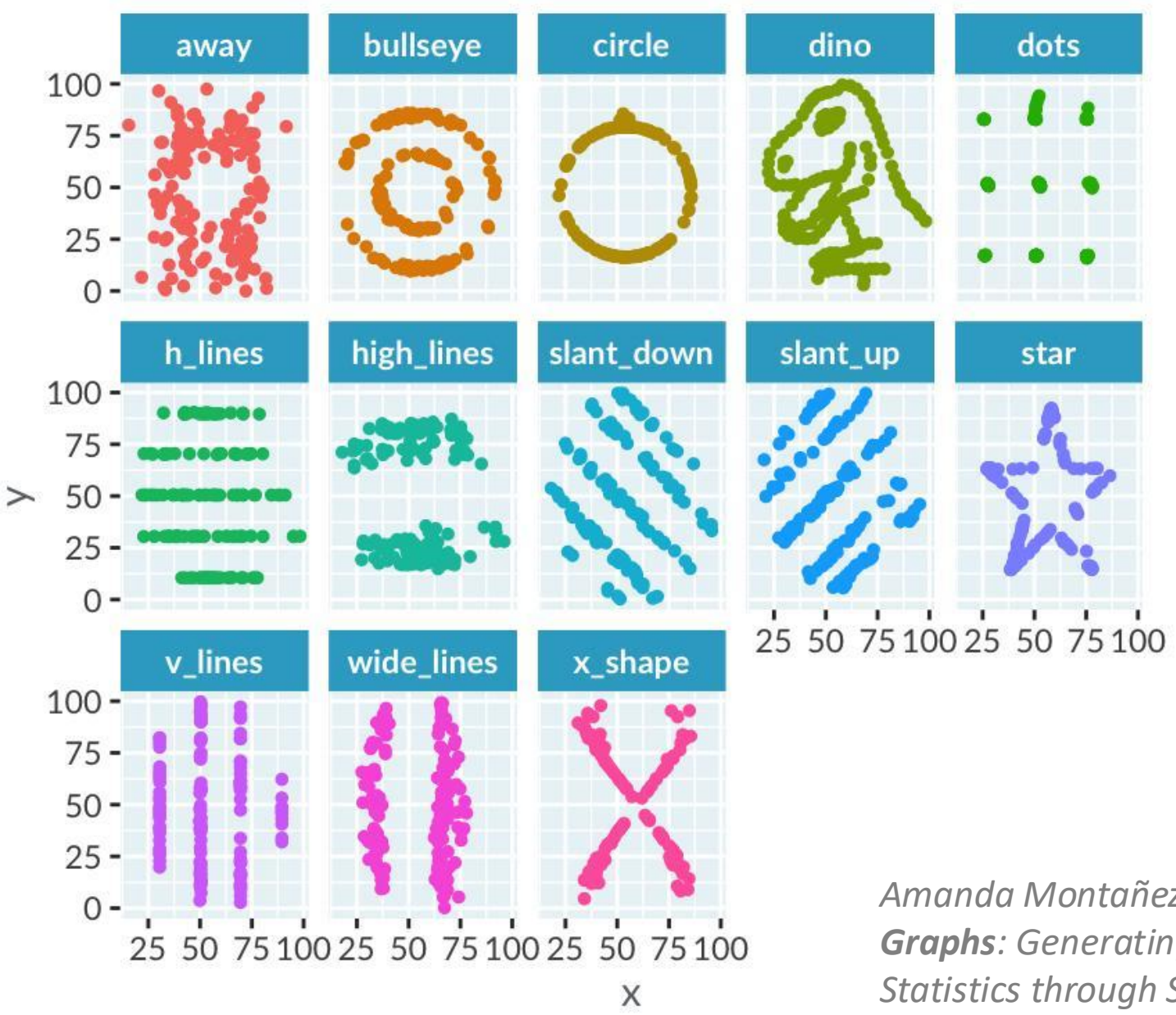
Mark Twain

Statističar Francis Anscombe o važnosti vizualizacije podataka(1973)

Hours of exercise per week: average, 7.5; standard deviation, 2.03
Hours of sleep per day: average, 9; standard deviation, 3.32
Correlation between the two: 0.816

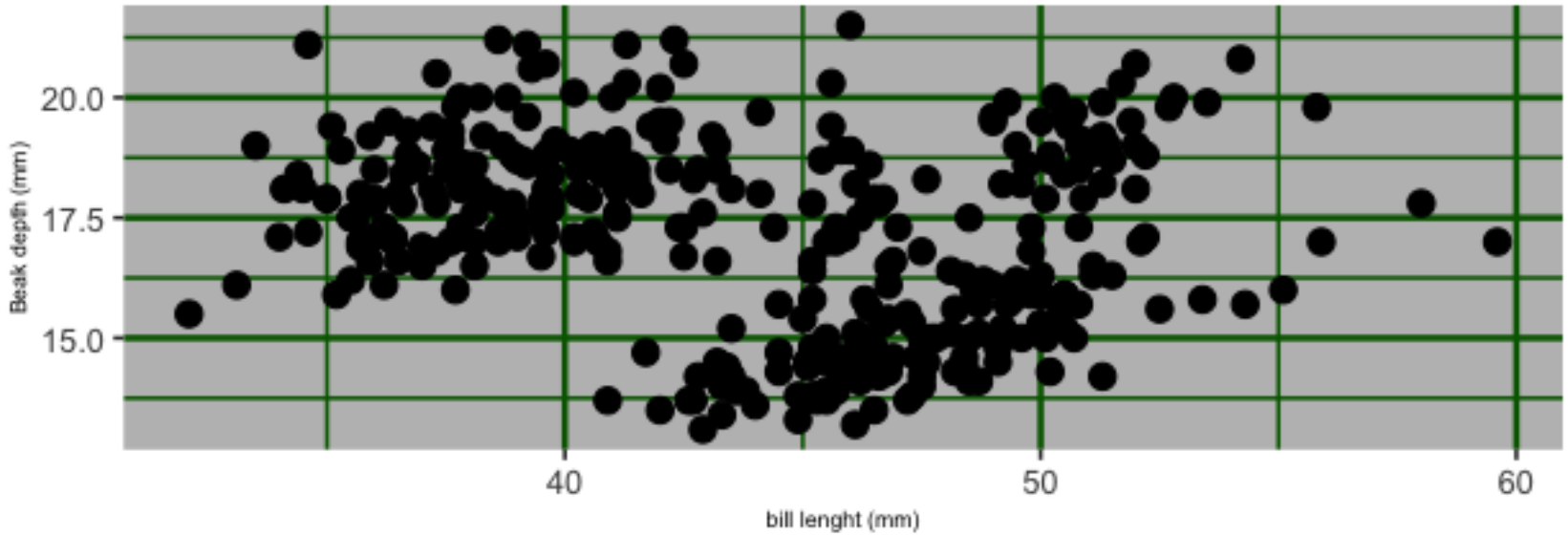


Amanda Montañez; Sources: R: A Language and Environment for Statistical Computing. R Core Team. R Foundation for Statistical Computing, 2021; "Graphs in Statistical Analysis," by F. J. Anscombe, in *American Statistician*, Vol. 27, No. 1; February 1973

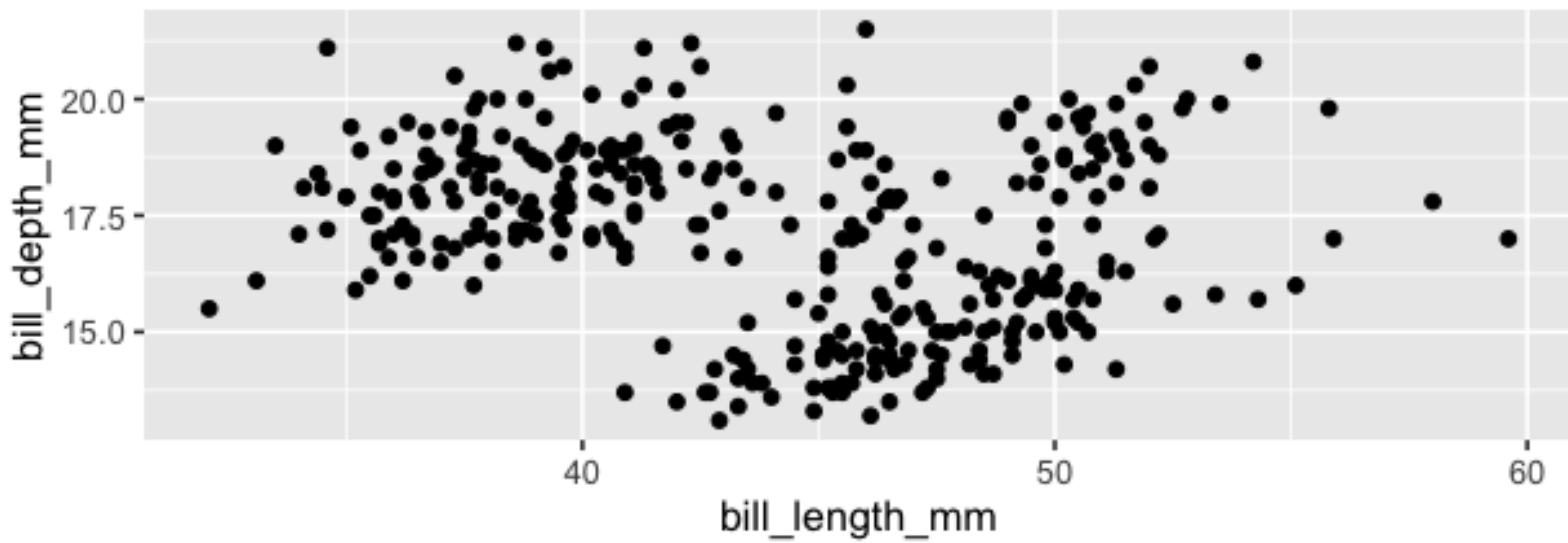


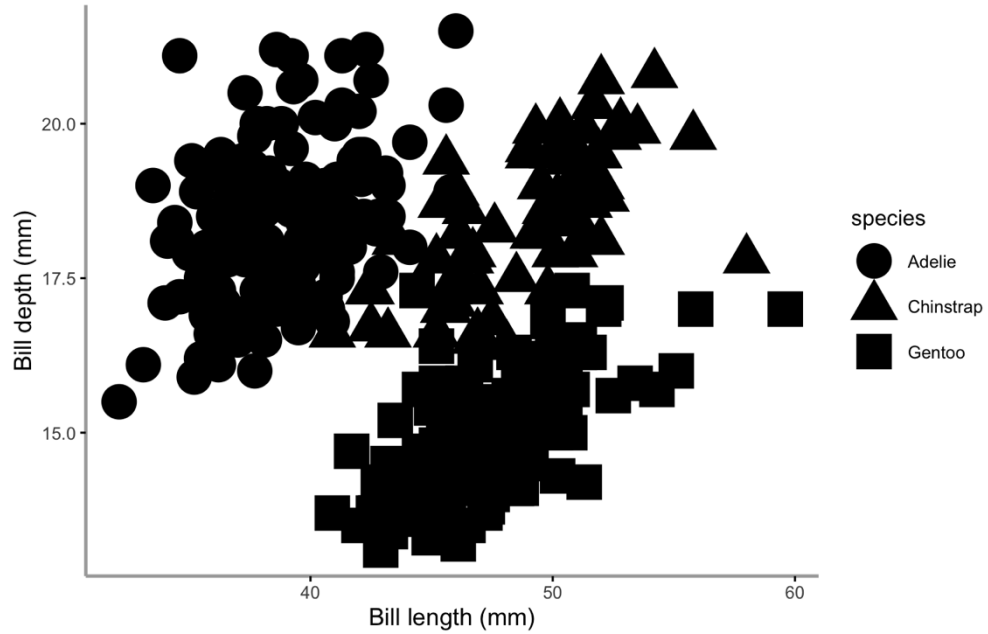
- **Deskriptivna statistika**
- **Grafički prikazi otkrivaju skrivene obrasce**
- **Vizualizacija je dio analize podataka!**

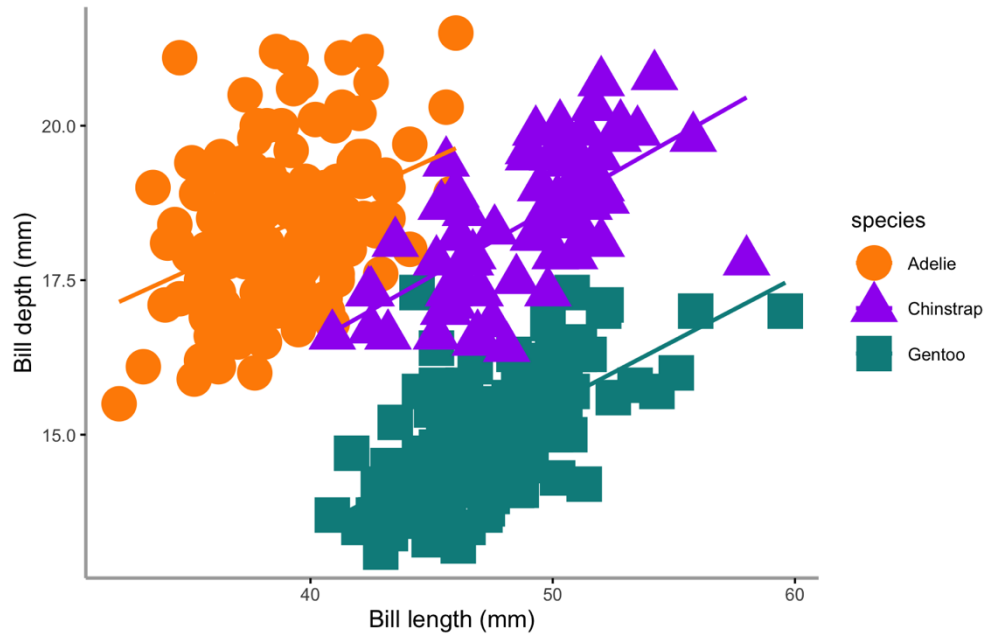
*Amanda Montañez; Sources: Jumping Rivers; "**Same Stats, Different Graphs**: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing,": Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*

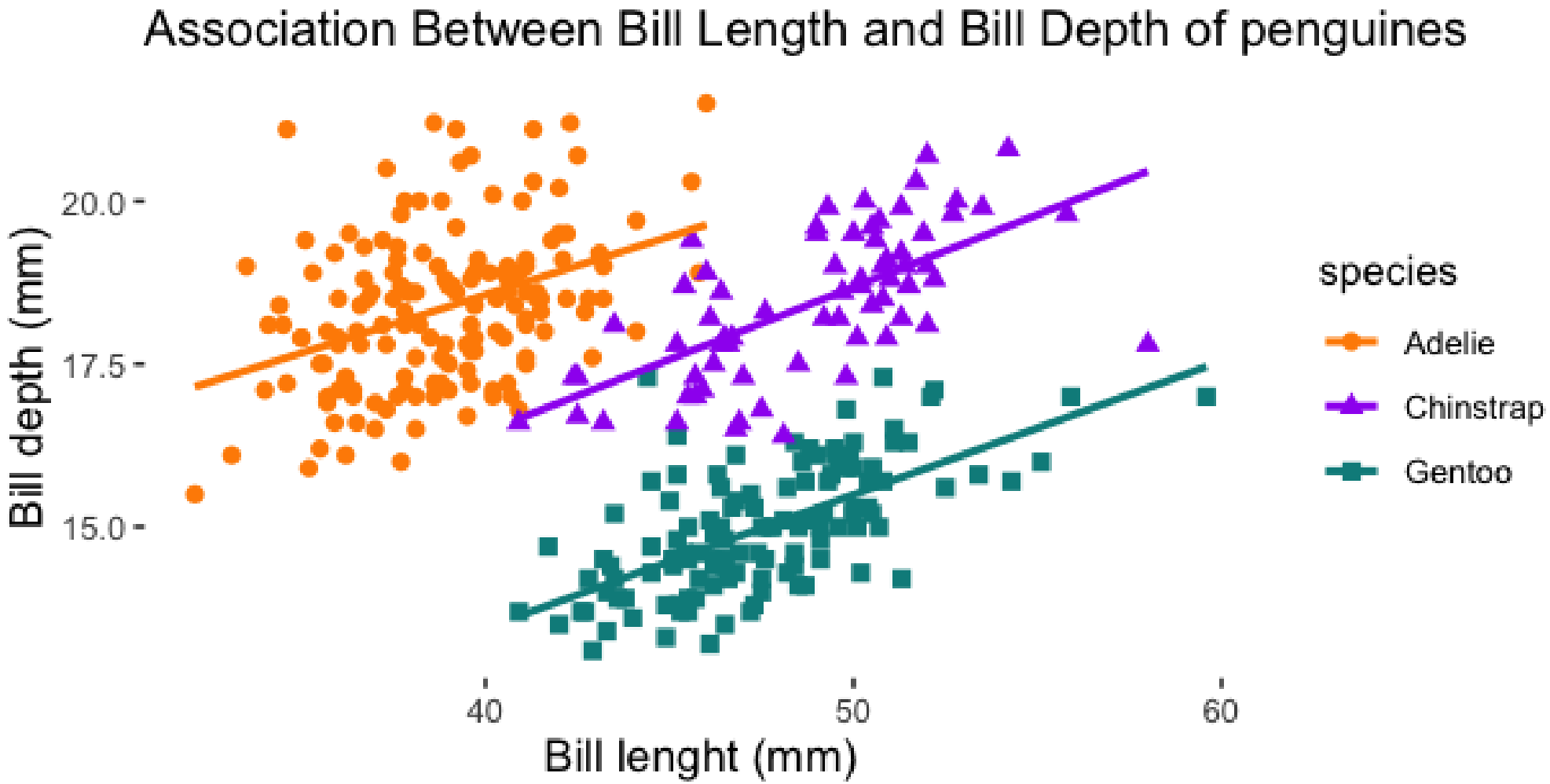

Beak depth (mm)
20.0
17.5
15.0
40
50
60
bill lenght (mm)




bill_depth_mm
20.0
17.5
15.0
40
50
60
bill_length_mm


Bill depth (mm)
20.0
17.5
15.0
40
50
60
Bill length (mm)
species
Adelie
Chinstrap
Gentoo


Bill depth (mm)
20.0
17.5
15.0
40
50
60
Bill length (mm)
species
Adelie
Chinstrap
Gentoo


Association Between Bill Length and Bill Depth of penguines
Bill depth (mm)
20.0
17.5
15.0
40
50
60
Bill lenght (mm)
species
Adelie
Chinstrap
Gentoo

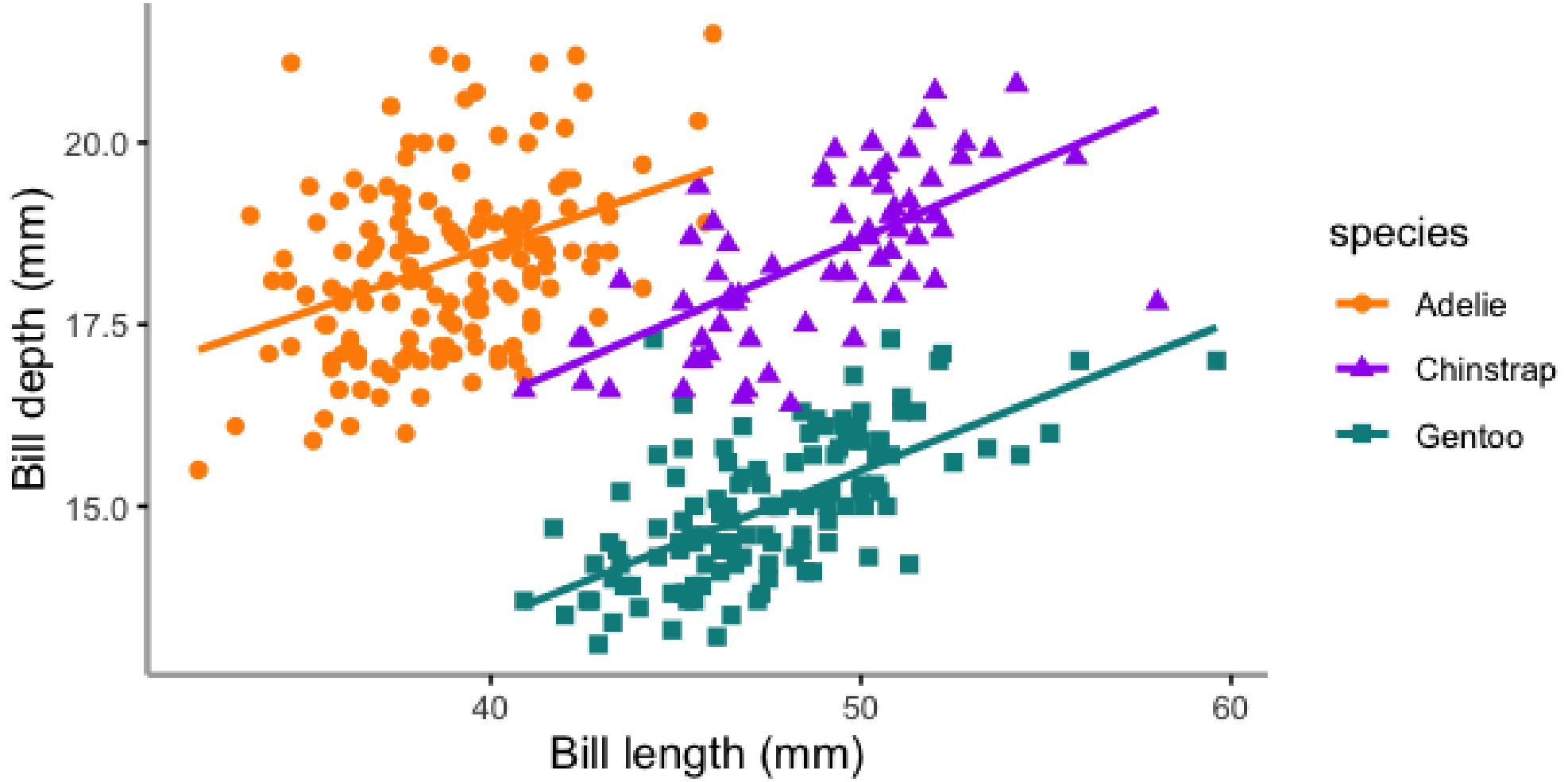


Fig. 1. Association of bill depth and bill length in three penguin species: *Pygoscelis adeliae* (Adelie, N=151), *Pygoscelis papua* (Gentoo, N=123) *and Pygoscelis antarctica* (Chinstrap, N=68)

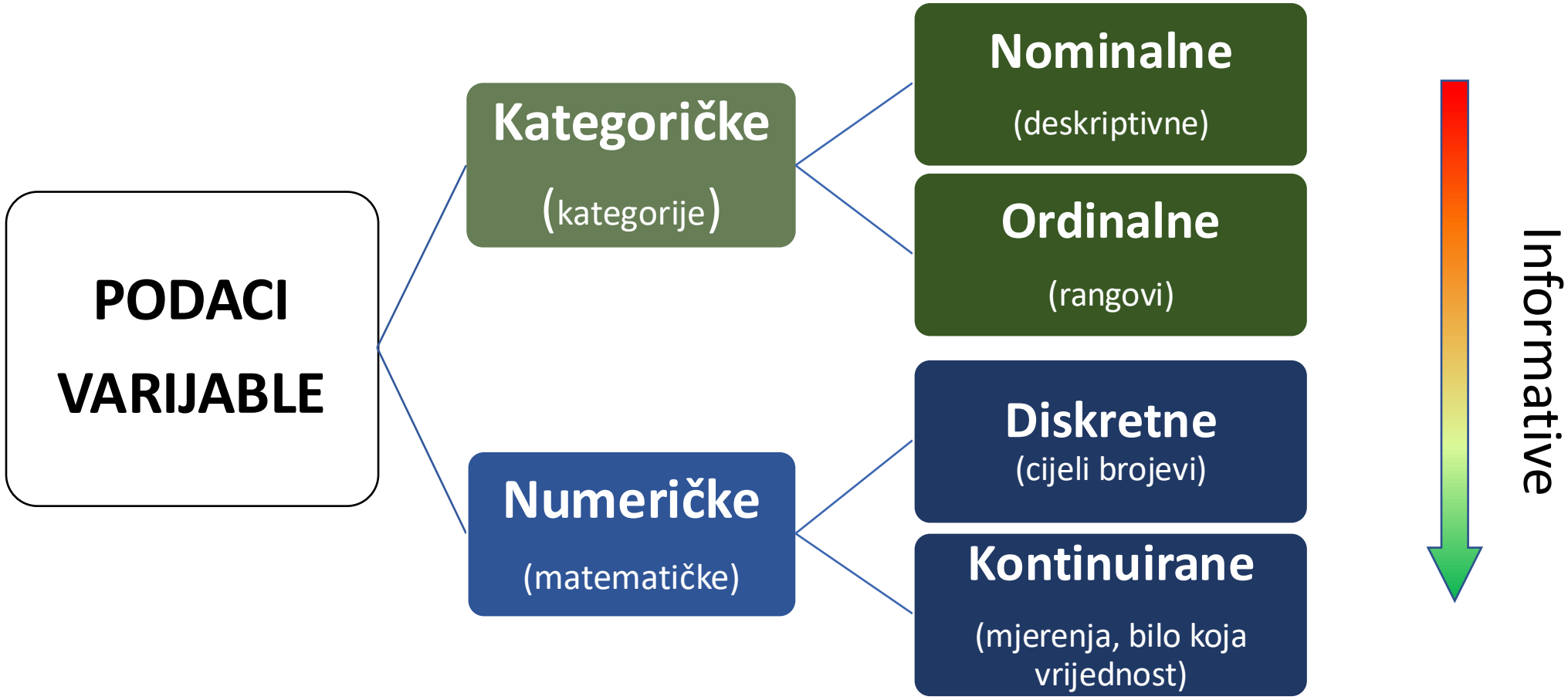
GRAFIČKI PRIKAZ PODATAKA

Služi vizualnoj analizi podataka ali i prezentaciji rezultata!

- Grafički prikaz podataka mora biti **jasan i informativan**
- Grafički izričaj treba biti **što čišći i jednostavniji!**
- **Font** informacija na grafu (nazivi varijabli, jedinice) mora biti **lako čitljiv i uočljiv**

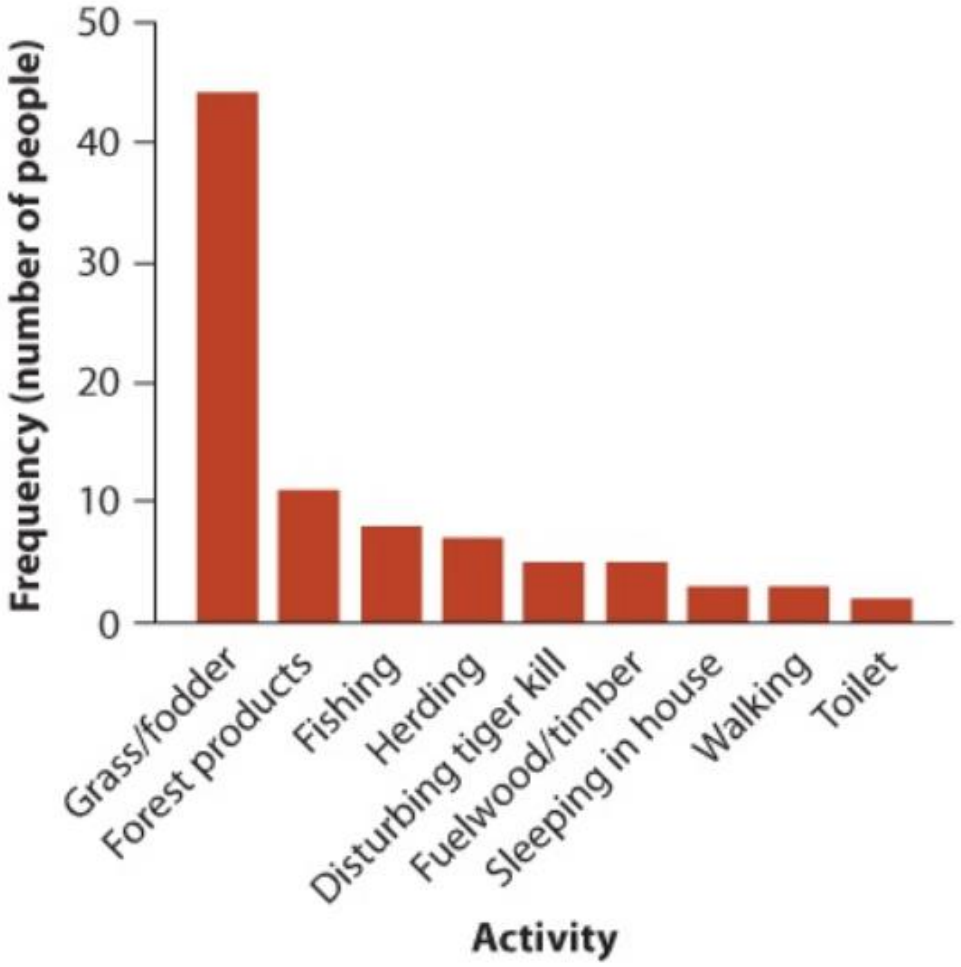
- **Legenda i opis moraju biti dostatni za razumijevanje rezultata**

- **Za kvantitativne podatke** po mogućnosti treba sadržavati **relevantna mjerila srednje vrijednosti i varijacije**
- **Magnitude** trebaju biti prikazane realno


PODACI
VARIJABLE
Kategoričke
(kategorije)
Numeričke
(matematičke)
Nominalne
(deskriptivne)
Ordinalne
(rangovi)
Diskretne
(cijeli brojevi)
Kontinuirane
(mjerenja, bilo koja vrijednost)
Informative

Jedna kategorička varijabla - stupčasti dijagrami *(bar plots)*

- pokazuju apsolutne ili relativne frekvencije





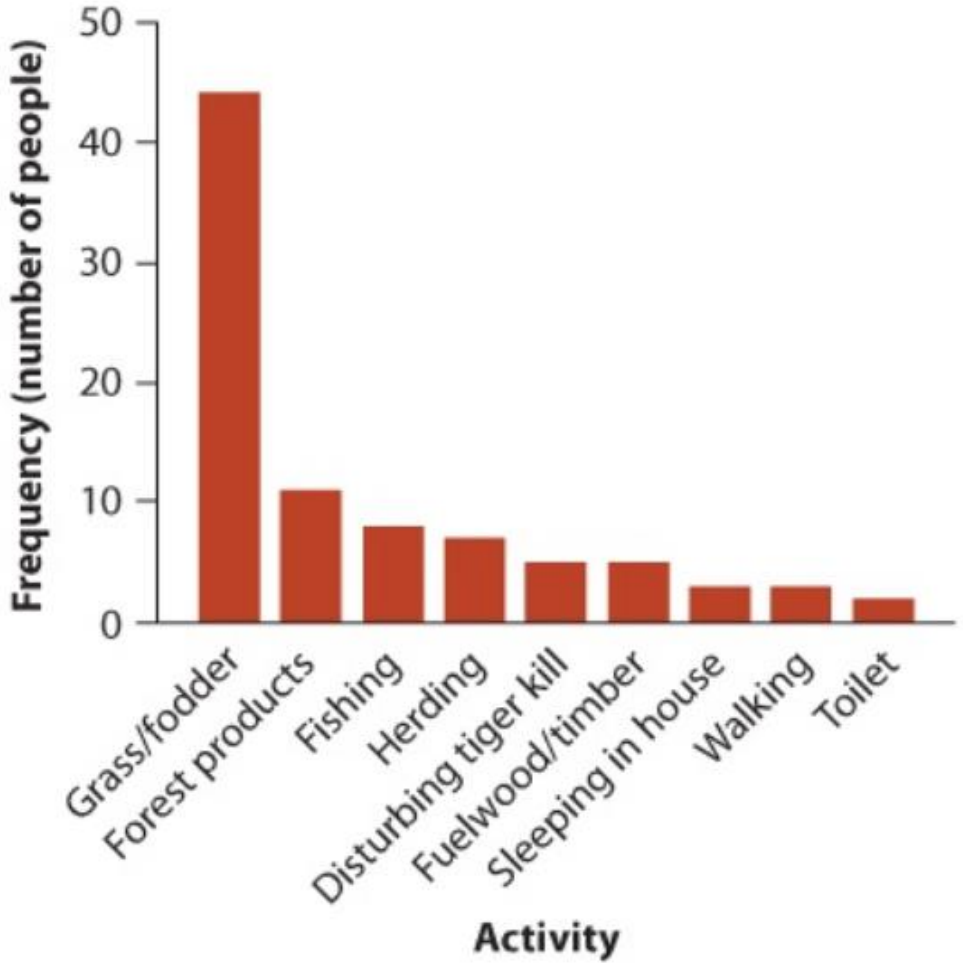
- y os treba započeti s nulom zbog realne magnitude
- apsolutna frekvencija
- kategorije poredane po učestalosti

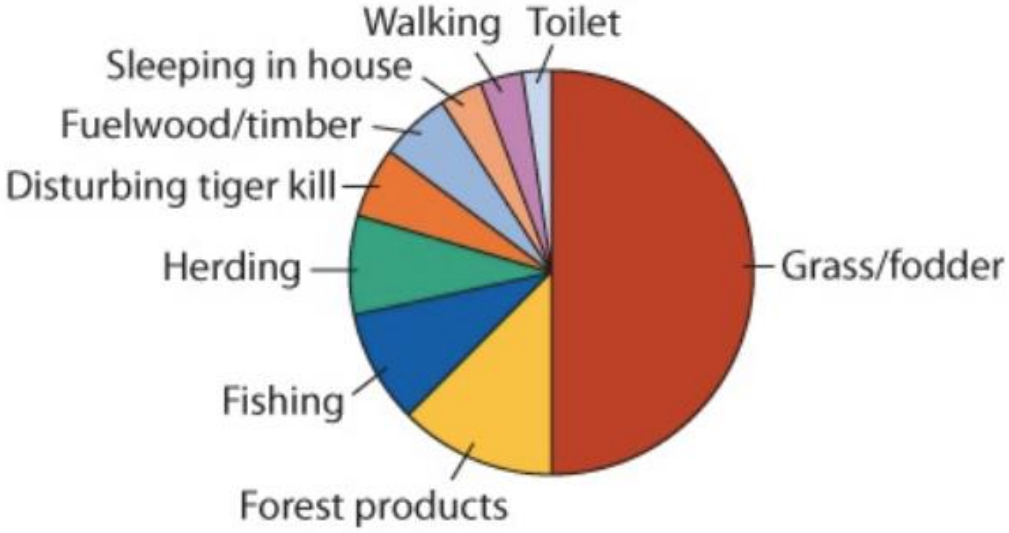
Stupčasti graf prikazuje aktivnosti ljudi u trenutku smrtnog napada tigrova na području Chitwan National Park, Nepal, u razdoblju 1979 do 2006. Ukupni broj žrtava N=88.

Figure 2.2-1
Whitlock et al., *The Analysis of Biological Data*, 2e, © 2015 W. H. Freeman and Company

Jedna kategorička varijabla - stupčasti dijagrami *(bar plots)*

- za kategoričke varijable se smatraju boljima od tortnih grafova (pie chart)





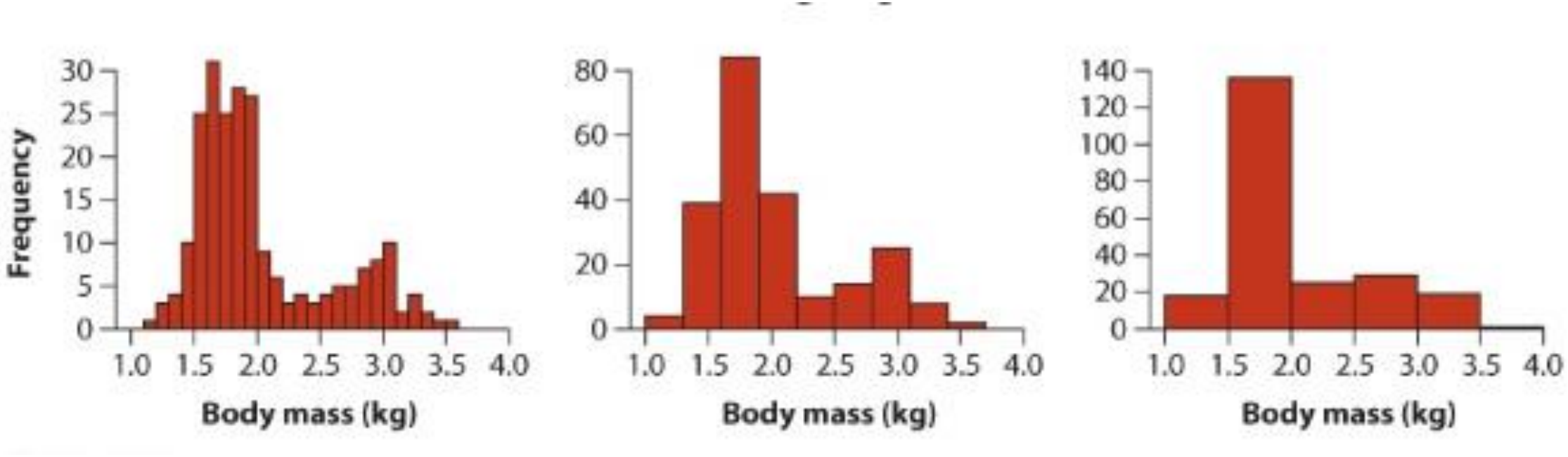
Aktivnosti ljudi u trenutku smrtnog napada tigrova na području Chitwan National Park, Nepal, u razdoblju 1979 do 2006. Ukupan broj žrtava N=88.

Figure 2.2-1
Whitlock et al., *The Analysis of Biological Data*, 2e, © 2015
W. H. Freeman and Company

Jedna numerička varijabla – histogram

- histograme koristimo za vizualizaciju frekvencija (ujedno i distribucija)
- stupičasti grafovi u kojima su numerički podaci podijeljeni u intervale





interval 0.1 kg *interval 0.3 kg* *interval 0.5 kg*

Masa lososa iz 228 jezera Pick Creek na Alaski

Figure 2.2-5
Whitlock et al., *The Analysis of Biological Data*, 2e, © 2015 W. H. Freeman and Company

Frequency of bird species in Organ Pipe Cactus National Monument. Total number of species, n = 43.

TABLE 2.2-3 Frequency distribution of bird species abundance at Organ Pipe Cactus National Monument.

Abundance	Frequency (Number of species)
0–50	28
50–100	4
100–150	3
150–200	3
200–250	1
250–300	2
300–350	1
350–400	0
400–450	0
450–500	0
500–550	0
550–600	0
600–650	1
Total	43

Jedna numerička varijabla – histogram

- u histogramu nema razmaka između stupaca

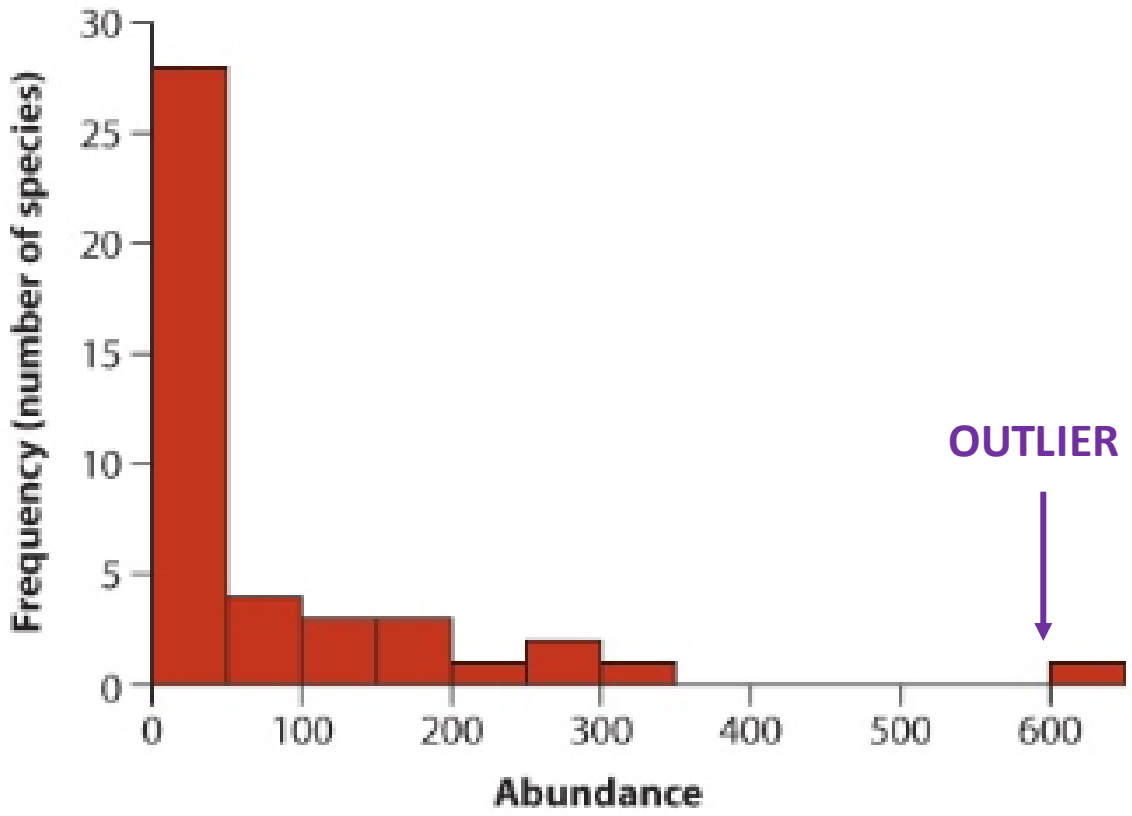


TABLE 2.2-3 Frequency distribution of bird species abundance at Organ Pipe Cactus National Monument.

Abundance	Frequency (Number of species)
0–50	28
50–100	4
100–150	3
150–200	3
200–250	1
250–300	2
300–350	1
350–400	0
400–450	0
450–500	0
500–550	0
550–600	0
600–650	1
Total	43

Frekvencija učestosti vrsta ptica u Nacionalnom Parku Organ Pipe Cactus National Monument. Ukupan broj vrsta, *n* = 43.

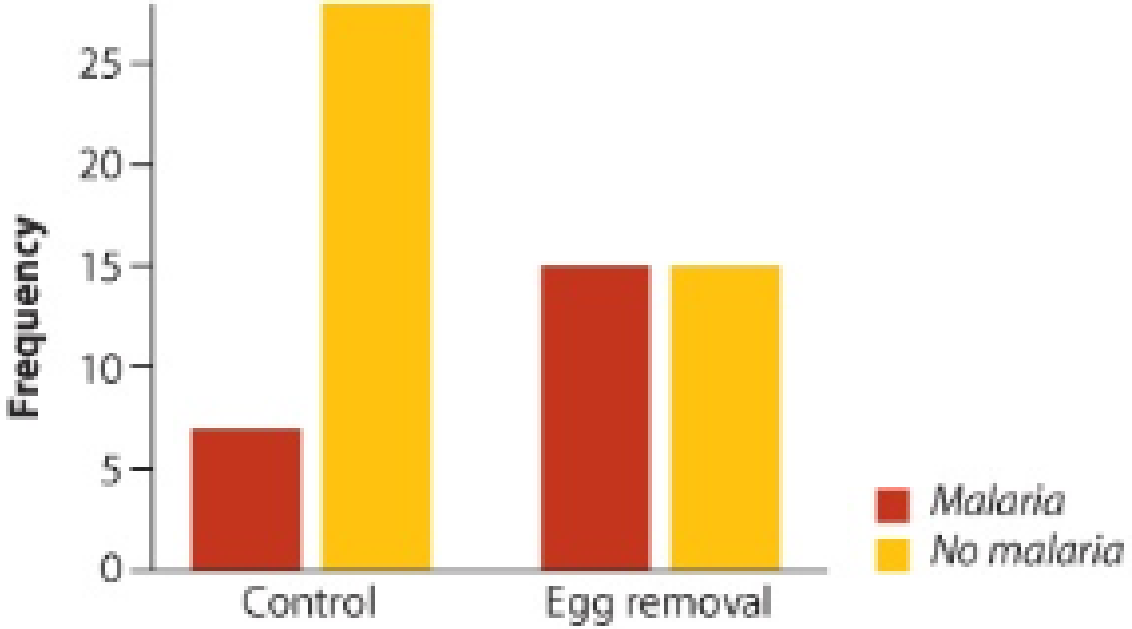
Povezanost kategoričkih varijabli – stupičasti grafovi

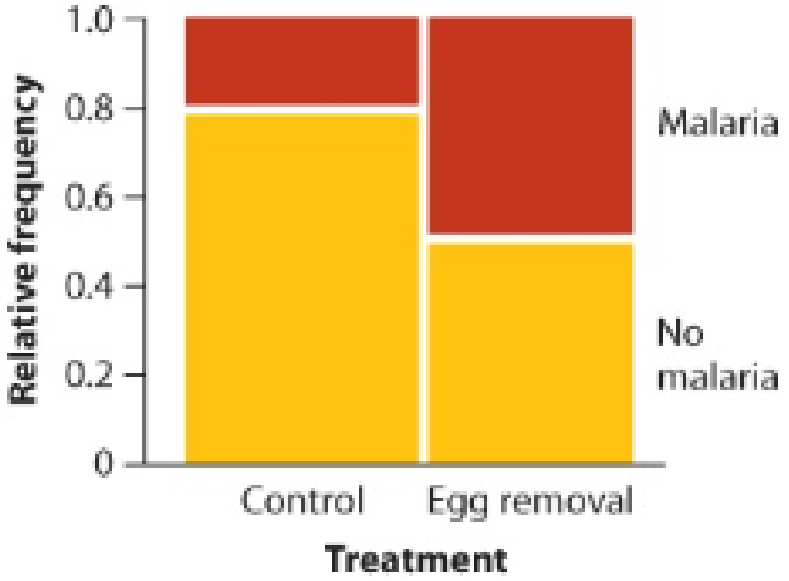
- **grupirani stupičasti grafovi ili mozaički grafovi (relativne ili apsolutne frekvencije)**



	Control group	Egg-removal group	Row total
Malaria	7	15	22
No malaria	28	15	43
Column total	35	30	65

apsolutne frekvencije





Opplinger et al 1996 Nature

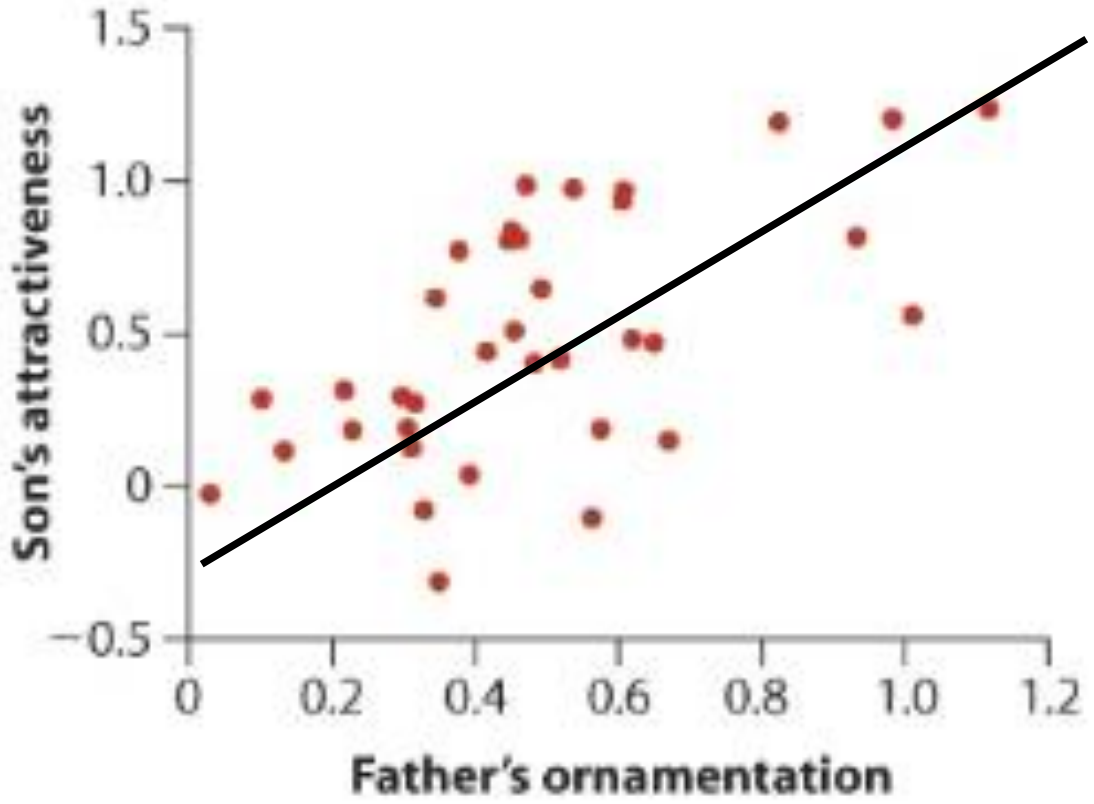
Utjecaj reproduktivnog stresa (odstranivanje jaja iz gnijezda) na osjetljivost velike sjenice na patogene

Figure 2.3-1
Whitlock et al., *The Analysis of Biological Data*, 2e, © 2015 W. H. Freeman and Company





Povezanost numeričkih varijabli – grafovi raspršenosti *(scatterplot)*





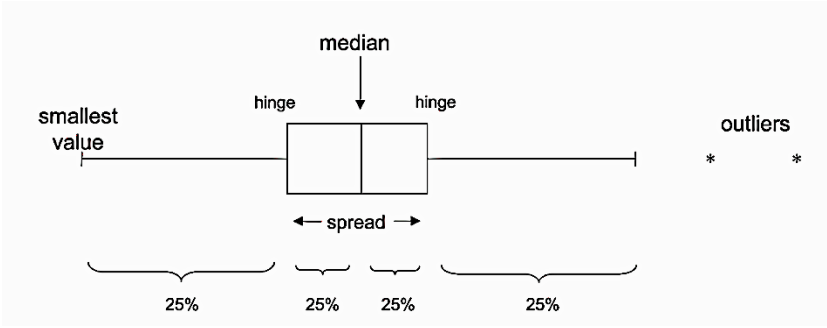
x os – nezavisna (*predictor*)
y os – zavisna (*response*)

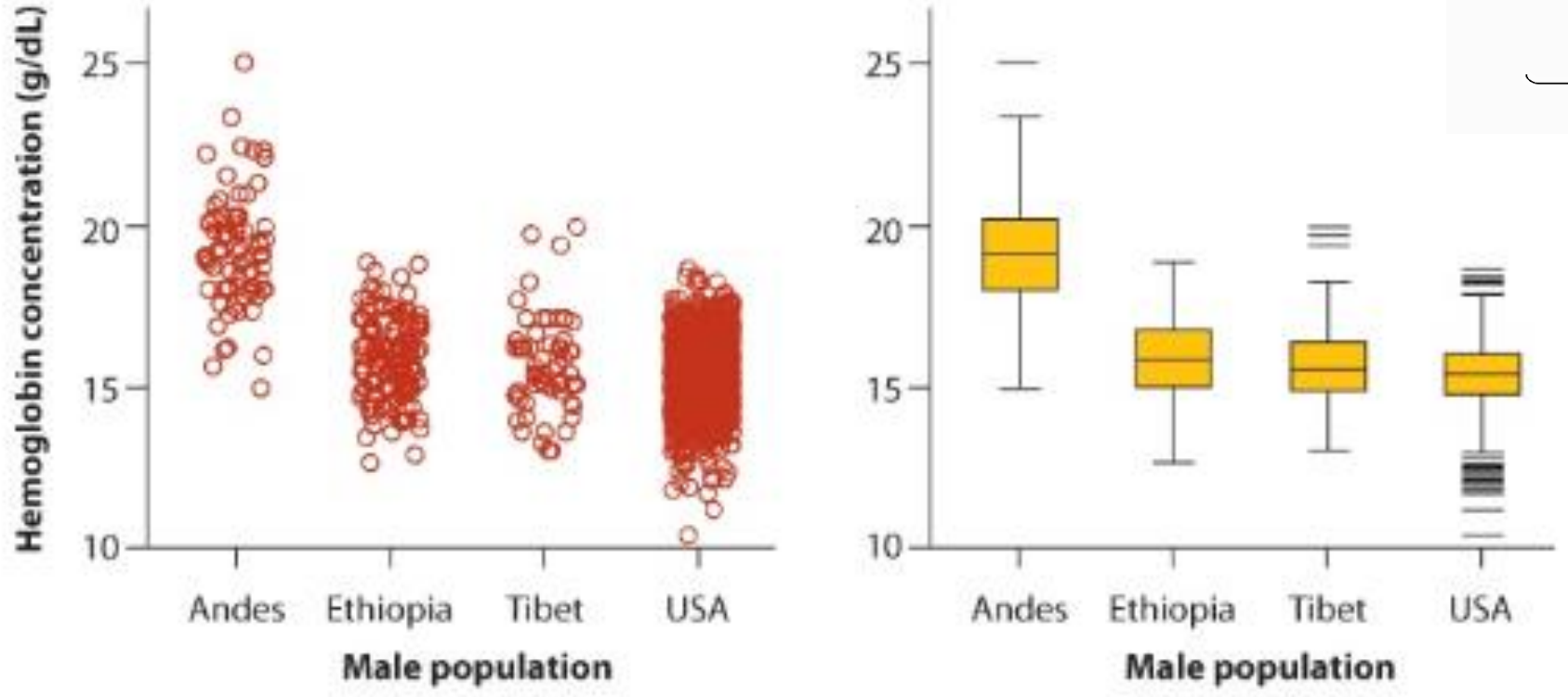
Točkasti graf prikazuje odnos između ornamentacije mužjaka gupija i prosječne privlačnosti njihovih sinova ženkama. Ukupan broj obitelji: *n* = 36.

Brooks 2000, Nature

Povezanost numeričke i kategoričke varijable

- **pravokutni** *(boxplot) i "***stripe***" grafovi*



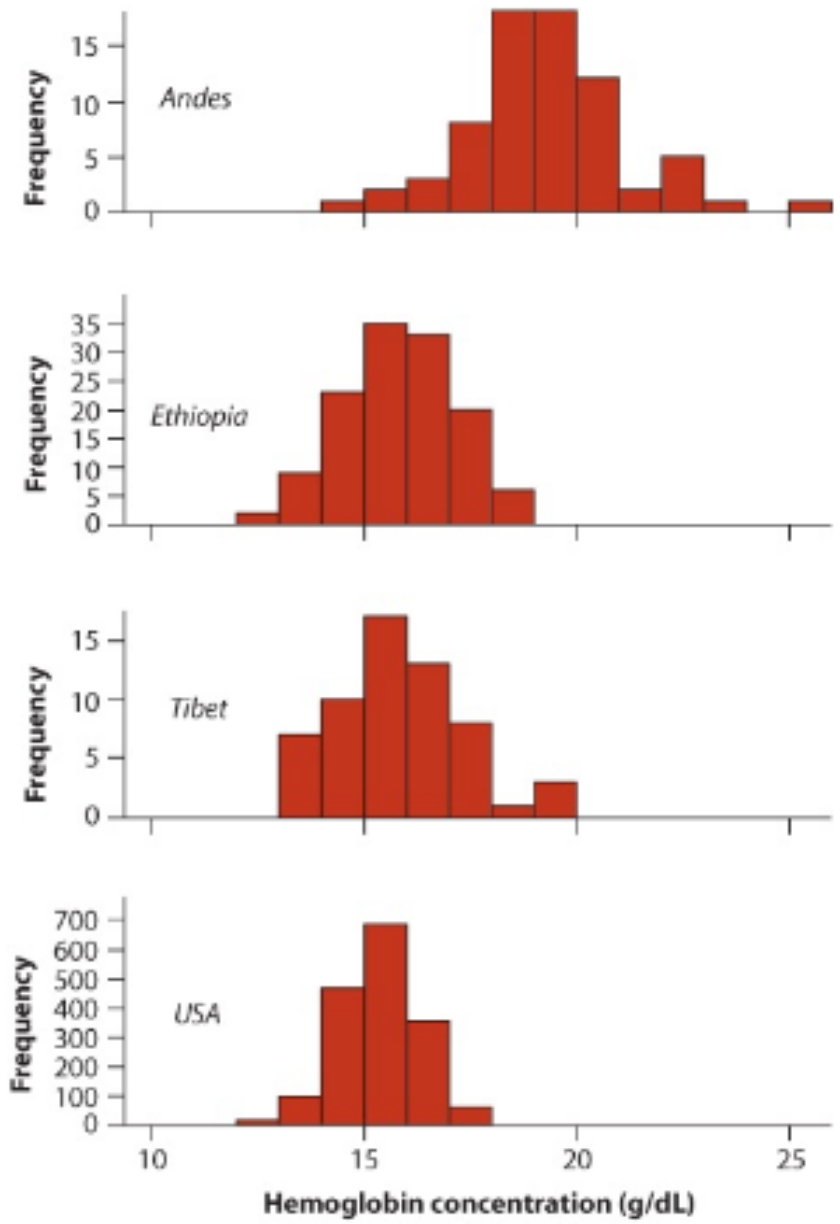


Koncentracija hemoglobina u muškaraca tri područja viših n.v.: Ande (71), Ethiopia (128), Tibet (59). Populacija od 1704 muškaraca koji žive na razini mora (USA) je kontrola.

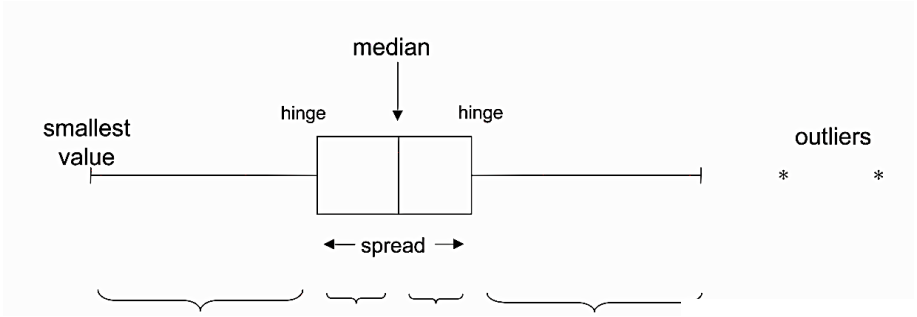
Bell et al. PNAS 2002

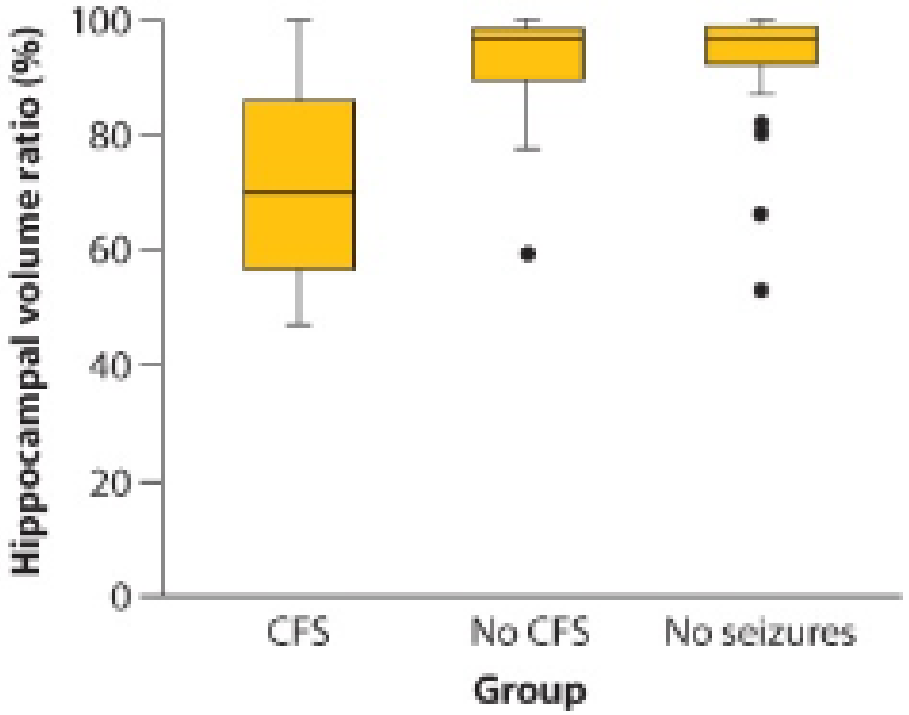
Figure 2.3-4
Whitlock et al., *The Analysis of Biological Data*, 2e, © 2015 W. H. Freeman and Company

Povezanost numeričke i kategoričke varijable - histogrami

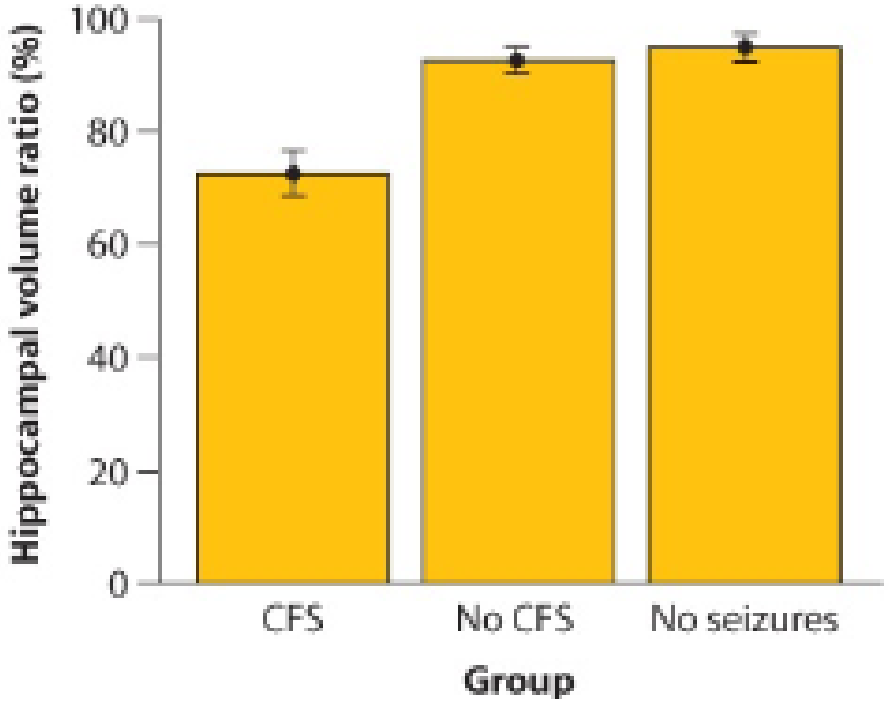


Povezanost numeričke i kategoričke varijable – stupičasti grafovi



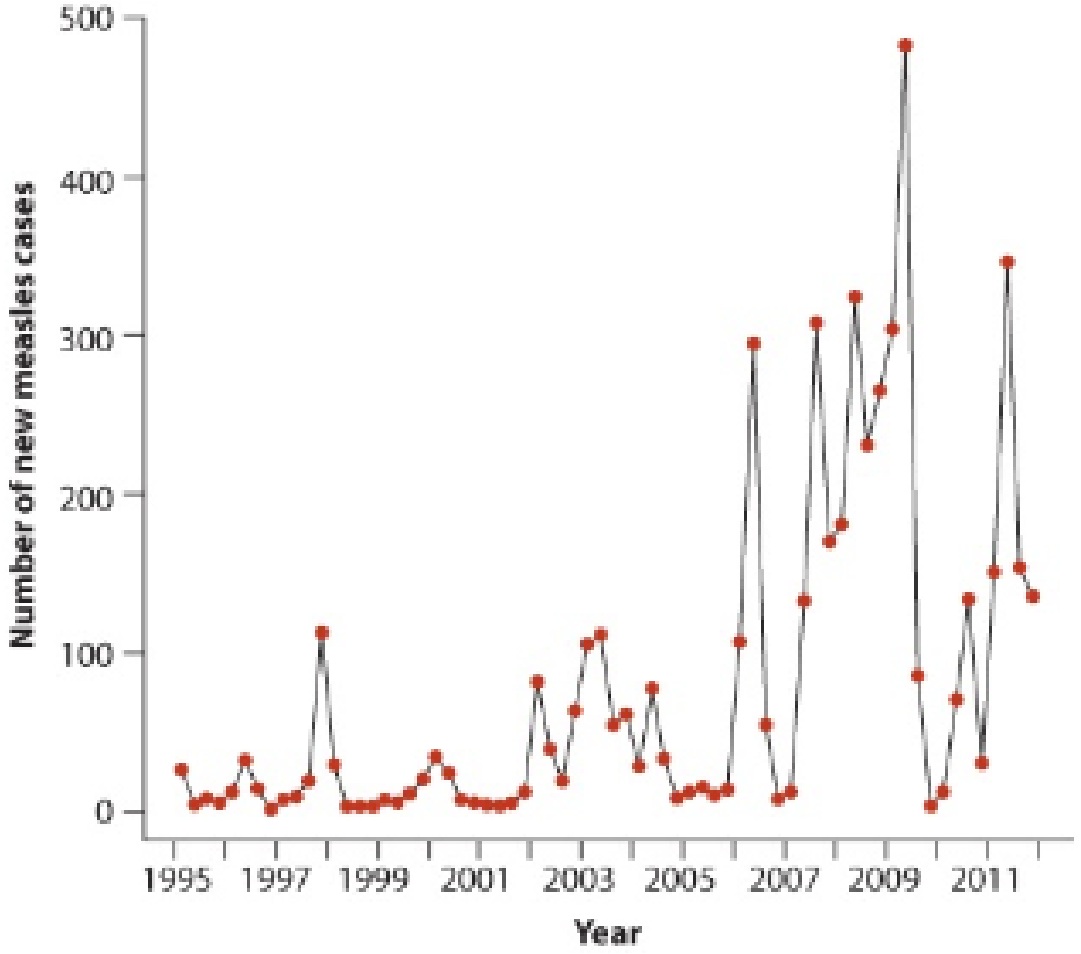


srednja vrijednost +- SE



Gubitak volumena hipokampusa u 107 epileptičnih pacijenata: CFS – pacijenti s dječjim febrilnim konvulzijama, NO CFS – pacijenti s dječjim nefebrilnim konvulzijama, No seizures – kontrola, bez epilepičnih napadaja ili konvulzija

Vremenski trendovi – linijski graf



nagib odražava brzinu promjene

Slučajevi ospica u Engleskoj i Walesu 1995 – 2011 (svako tromjesečje je jedan podatak)

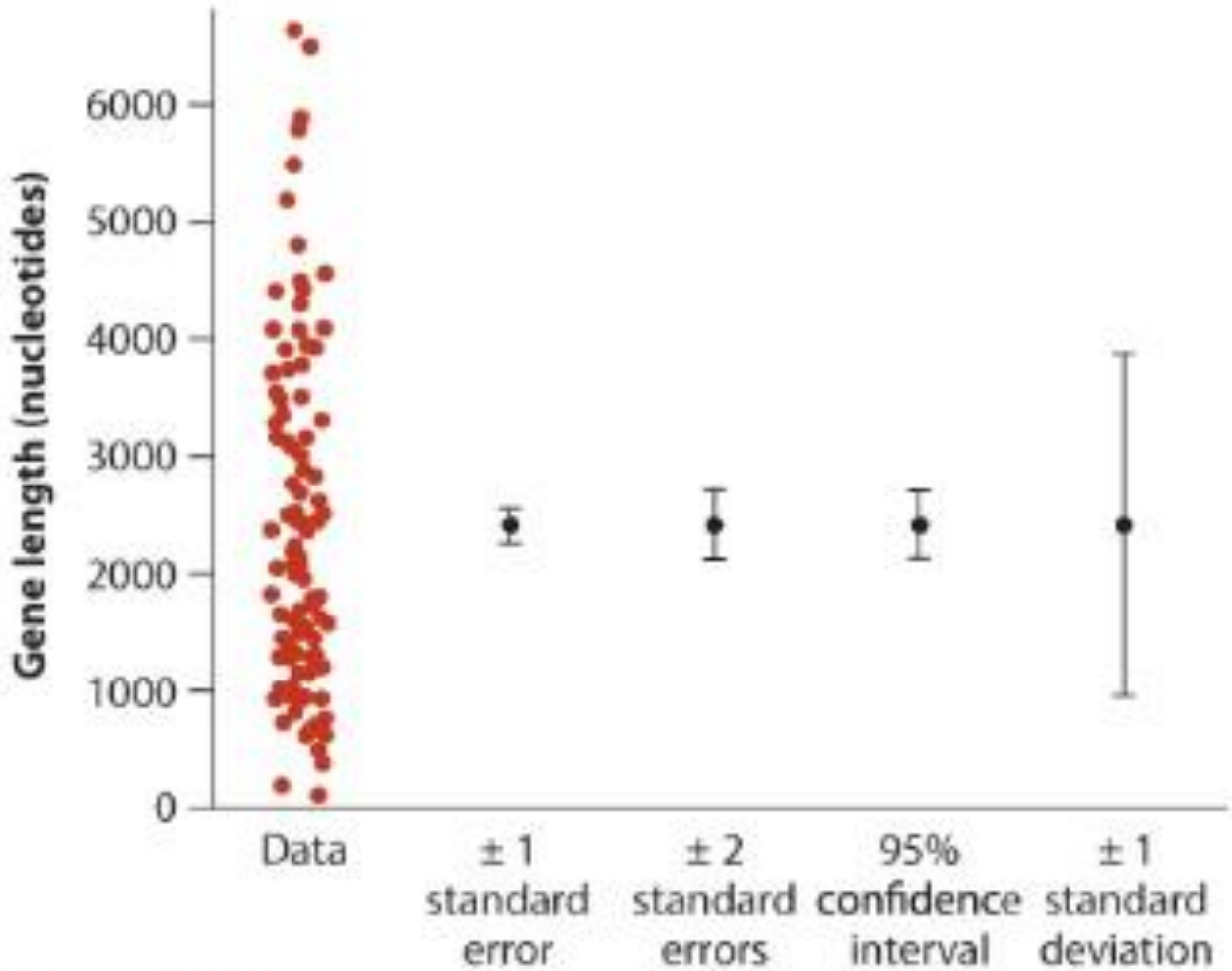


Figure 4.4-2
Whitlock et al., *The Analysis of Biological Data*, 2e, © 2015 W. H. Freeman and Company


EVERY SINGLE PERSON WHO CONFUSES CORRELATION AND CAUSATION ENDS UP DYING
timoelliott.com