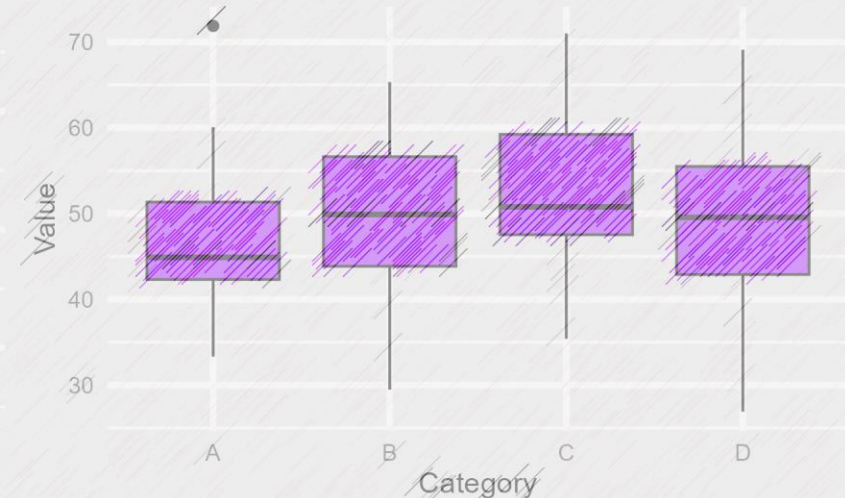
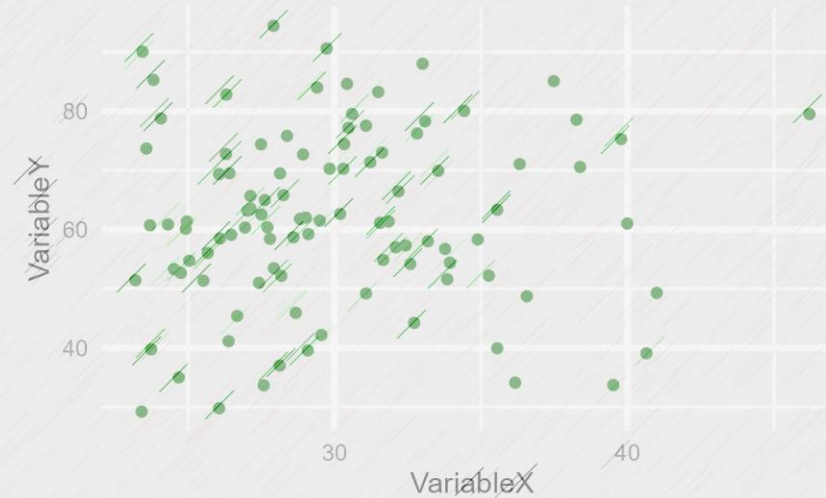
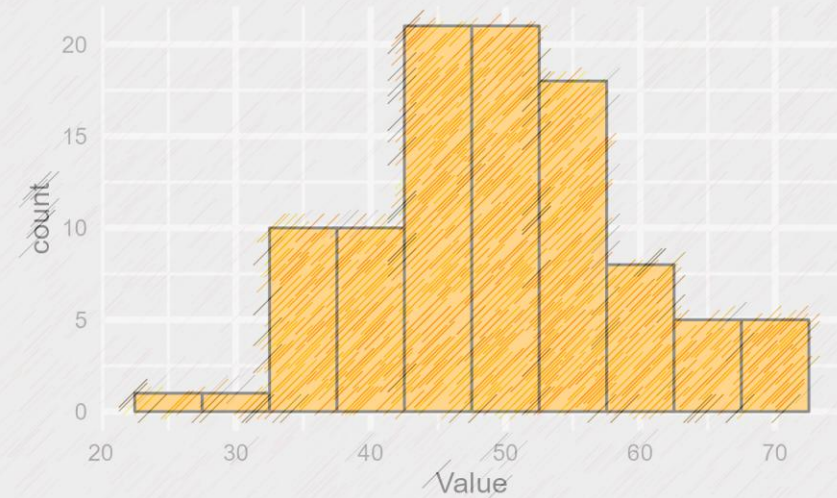
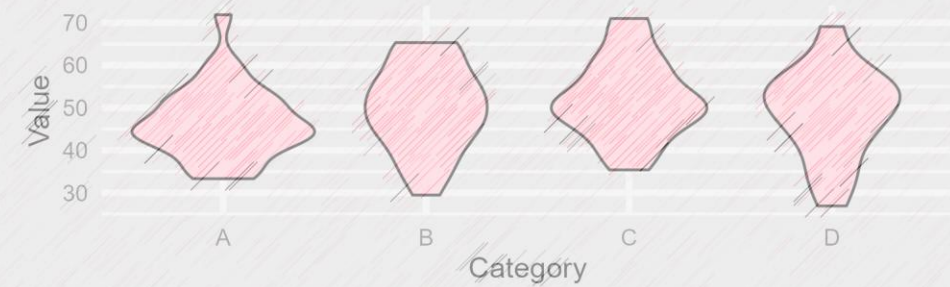
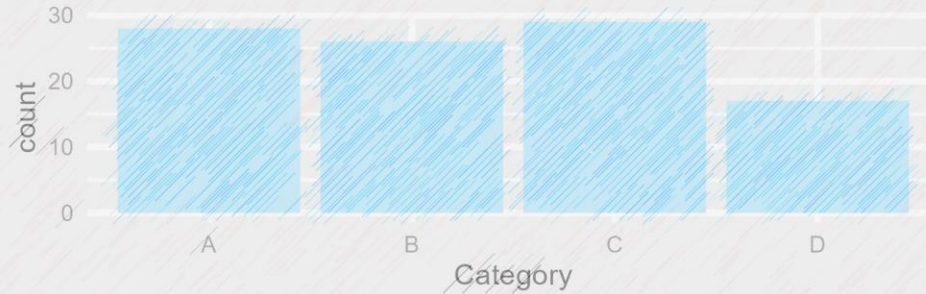


min	max	median	iqr	mean	sd
4.2	21.5	9.8	5.0	10.6	4.5
13.6	27.3	19.2	7.1	19.7	4.4
18.5	33.9	25.9	4.3	26.1	3.8



Seminar 5: Tipovi varijabli, grafički i numerički sažeci

Tipovi varijabli

Numeričke varijable

Diskretne ili kontinuirane

Kategoričke varijable

Nominalne ili ordinalne

Intervalne

Intervalna varijabla je ona kod koje postoji poredak i razlika između dvije vrijednosti je smisljena
Omjer dviju vrijednosti nema smislenu interpretaciju
Npr. Temperatura (F, °C),

Omjerne (*ratio*)

Omjerna varijabla ima sva svojstva intervalne varijable i jasnu definiciju 0,0. Kada je varijabla jednaka 0,0, nema te varijable.
Npr. Aktivnost enzima, koncentracija tvari

Tipovi varijabli I

- Jesu li sljedeće varijable kvantitativne (diskretne ili kontinuirane, s kojim mjernim jedinicama?), ili kvalitativne (nominalne ili ordinalne, s kojim razinama?)?
 - Sistolički krvni tlak.
 - Smjer diplomskog studija.
 - Akademski razina (univ.bacc.; mag.; dr.sc;).
 - Koliko je puta osoba posjetila liječnika prošle godine.

Tipovi varijabli II

- Studija o ljudima koji su umrli dok su snimali selfije zabilježila je lokaciju. Što su tablici varijable (i navedite tip varijable)?

- Lokacija.
- Broj ljudi koji su umrli na svakoj lokaciji.
- Postotak ljudi koji su umrli na svakoj lokaciji.

	Number	Percentage
Nature, associated environments	48	43.2
Train, railway, associated structures	22	19.9
Buildings, associated structures	17	15.3
Road, bridge, associated structures	12	10.8
Dams, associated structures	7	6.3
Fields, farms, associated structures	4	3.6
Others	1	0.9

Tipovi varijabli III

- Koji su tipovi varijabli prisutni u ovom setu podataka (podaci o pacijenticama koje boluju od raka jajnika)?
- Uočavate li neke probleme s organizacijom podataka?
- Koji tip numeričkih sažetaka je primjeren za svaku od varijabli? (srednja vrijednost, medijan, omjer?)

Patient_ID	Tumor Stage	HER2 Expression	CRP	WBC	CA_125
P001	1	12.27	12.17	-	1.06
P002	1	9.07	18.74	8.74	6.7
P003	3	9.34	9.16	6.82	18.71
P004	2	12.97	2.67	7.14	NA
P005	1	7.82	4.01	3.64	56.85
P006	4	8.94	19.73	9.23	2.13

Naputci za dobre grafove

- Ne zatrpavajte graf nepotrebnim dodacima (uzorci, 3D efekti, nepotrebne legende) – što jednostavnije to bolje
- Uključite sve potrebne informacije (označite obje osi). Simboli, boje i uzorci trebaju biti definirani u legendi (na slici ili ispod slike). Rezultati na grafu moraju se moći razumjeti bez čitanja glavnog teksta.
- Budite oprezni s korištenjem boja u grafovima (trebaju biti podobne za crno-bijelo printanje i za daltoniste) .
- Nemojte pokušavati prikriti ili krivo interpretirati rezultate.

Grafičke metode I – loši grafovi

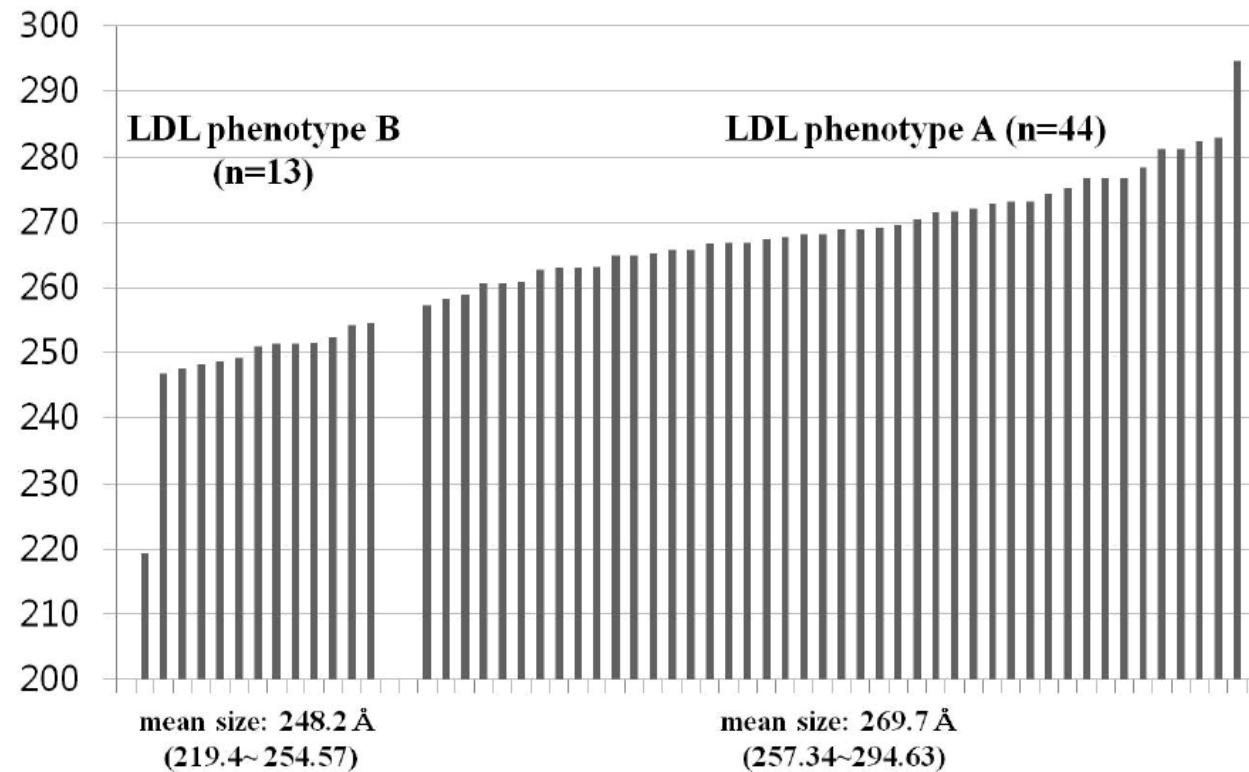


Fig. 1. Distribution of low-density lipoprotein (LDL) particle size in all study subjects (LDL phenotypes A and B). *LDL phenotype A group* (mean size: 269.7 Å, n = 44), subjects with buoyant-mode profiles [peak LDL particle diameter ≥ 264 Å] including intermediate LDL subclass pattern [$256 \text{ Å} \leq \text{peak LDL particle diameter} \leq 263 \text{ Å}$]; *LDL phenotype B group* (mean size: 248.2 Å, n = 13), subjects with dense-mode profiles [peak LDL particle diameter $\leq 255 \text{ Å}$]

Grafičke metode II – loši grafovi

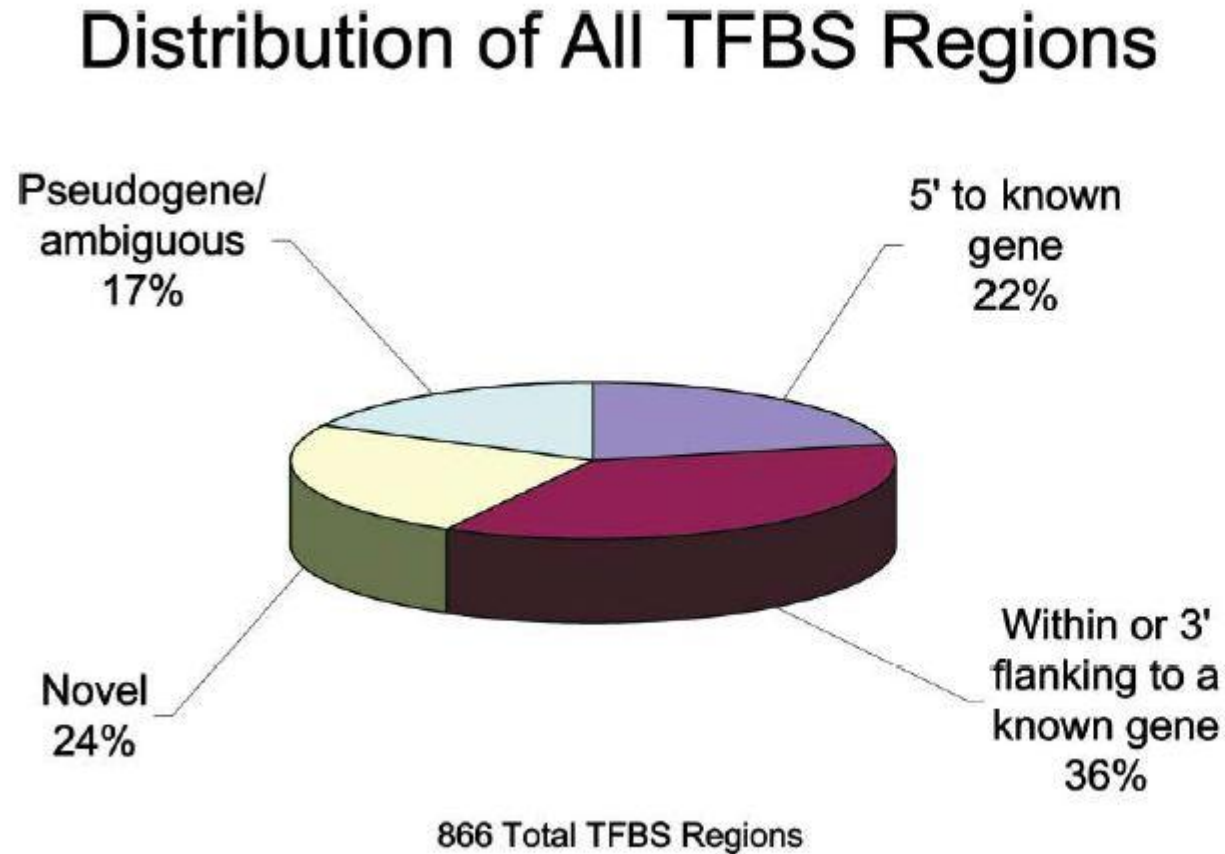


Figure 1. Classification of TFBS Regions
TFBS regions for Sp1, cMyc, and p53 were classified based upon proximity to annotations (RefSeq, Sanger hand-curated annotations, GenBank full-length mRNAs, and Ensembl predicted genes). The proximity was calculated from the center of each TFBS region. TFBS regions were classified as follows: within 5 kb of the 5' most exon of a gene, within 5 kb of the 3' terminal exon, or within a gene, novel or outside of any annotation, and pseudogene/ambiguous (TFBS overlapping or flanking pseudogene annotations, limited to chromosome 22, or TFBS regions falling into more than one of the above categories).

Grafičke metode III – loši grafovi

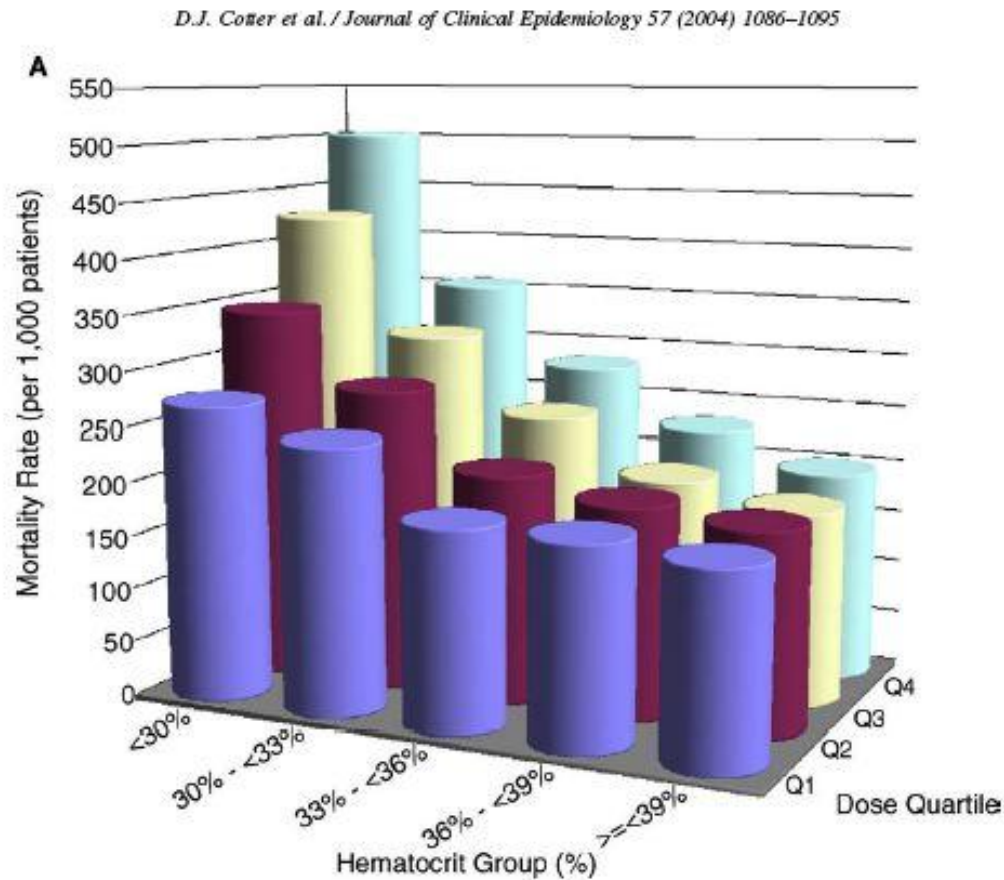
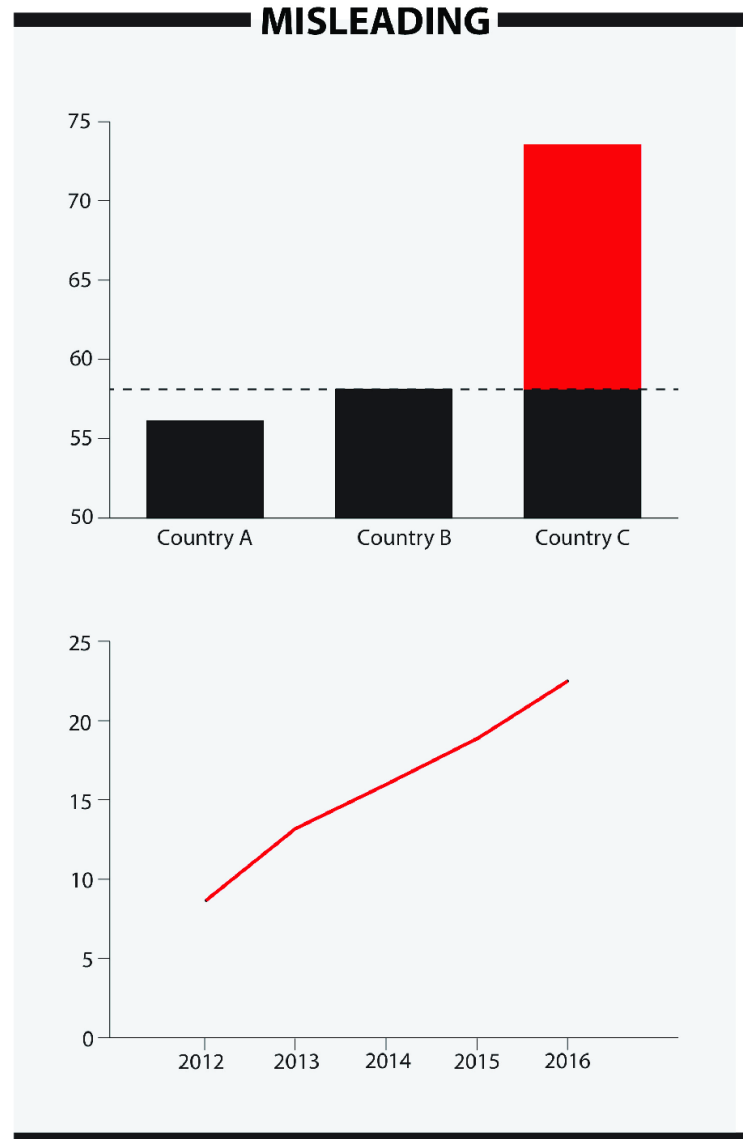
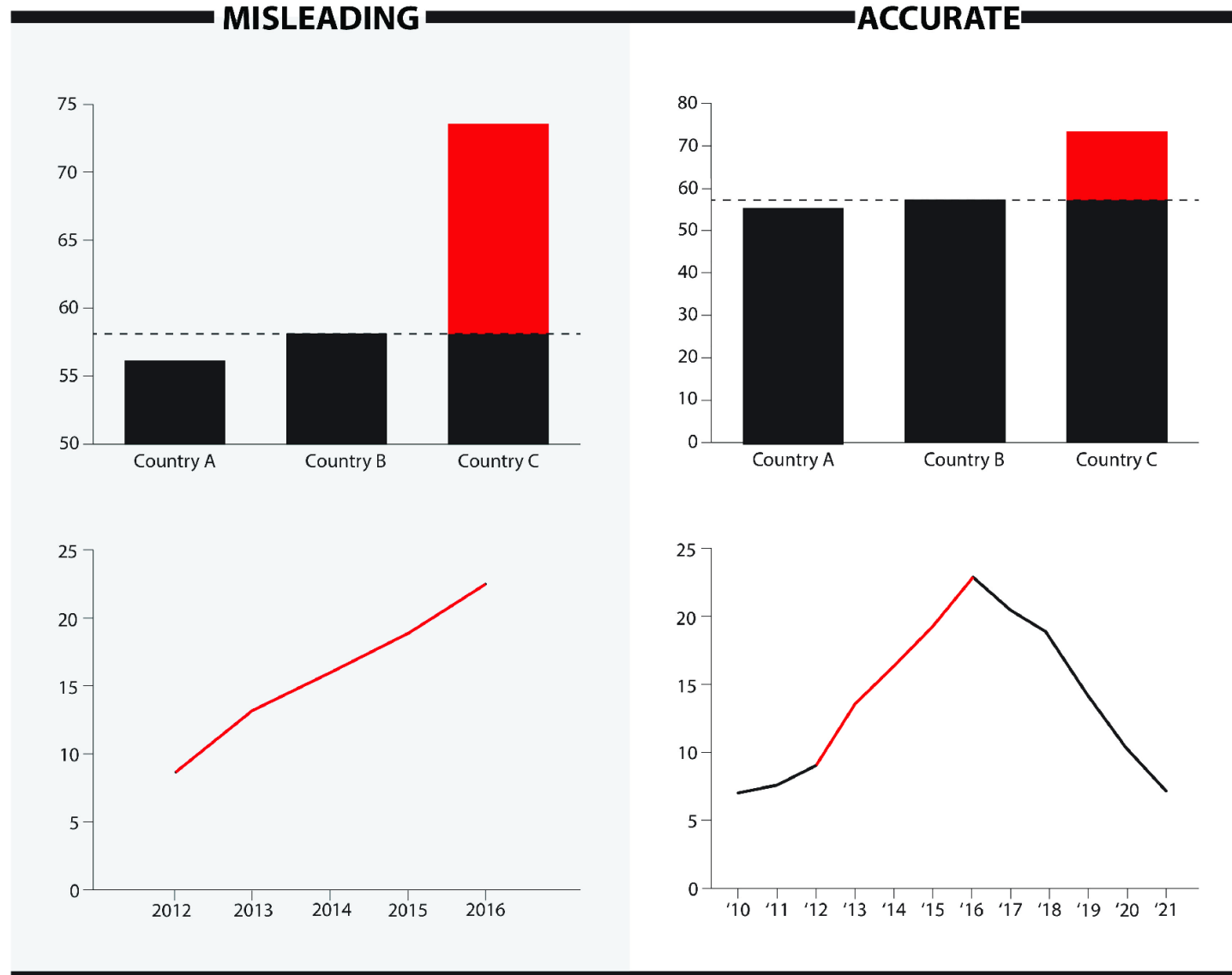


Fig. 2. (A) Unadjusted 1-year mortality rates by hematocrit group disaggregated by epoetin dose quartile. Within each epoetin dose quartile, there is a trend toward increasing mortality as the observed study hematocrit decreases, most notably in the fourth quartile ($>21,692$ units/wk). Similarly, there is a trend toward increasing mortality as the epoetin dose increases within each observed study hematocrit range, most notably in the lowest ($<30\%$) hematocrit range. (B) Relative risk of death by hematocrit group

Grafičke metode IV – loši grafovi



Grafičke metode IV – loši grafovi



Grafičke metode V – odaberi primjereni graf

- Za istraživanje odnosa između pH vode i temperature vode, u raznim potocima, koji bi grafikon bio najkorisniji?
 - Kutijasti dijagram (boxplot)
 - Složeni stupčasti grafikon (stacked barplot)
 - Histogram
 - Dijagram raspršenosti (scatterplot)

Grafičke metode VI – odaberi primjereni graf

- Istraživačka studija ispitala je način na koji su studenti obično dolazili na sveučilište (autobusom; osobnim automobilom; automobil; itd.) i njihov poštanski broj. Koji bi grafikon bio najbolji za prikaz ovih podataka?
 - Kutijasti dijagram (*boxplot*)
 - Dijagram raspršenosti (*scatterplot*)
 - Stupčasti dijagram (jedan uz drugi) (*side-by-side barplot*)

Grafičke metode VII – što je krivo

- Istraživanje 173 ženke potkovnjaka zabilježilo je, između ostalog, boju oklopa (jedna od "Svijetla srednja", "Srednja", "Tamna srednja" ili "Tamna") i stanje oklopa (jedno od "oboje u redu", "jedan u redu", "nijedan u redu"). Kritizirajte dijagram raspršenosti korišten za opis podataka.

