
Grafikkens grammatikk

MEDISIN OG TALL

ARE HUGO PRIPP

apripp@ous-hf.no

Are Hugo Pripp er seniorforsker og biostatistiker ved Oslo senter for biostatistikk og epidemiologi, Forskningsstøtteavdelingen, Oslo universitetssykehus, og tilknyttet K.G. Jebsen-senter for hjernevæskeforskning. Han er professor II ved Fakultet for helsevitenskap ved OsloMet – storbyuniversitetet.

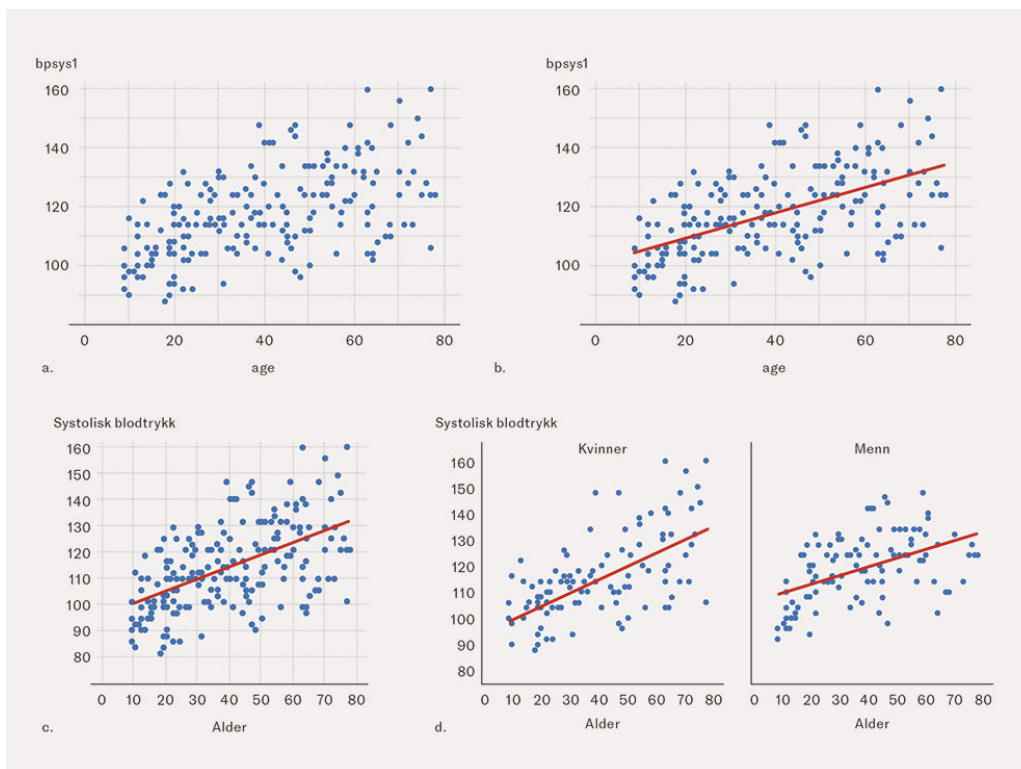
Forfatteren har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

Grafisk presentasjon av data og statistiske resultater er ofte lettere tilgjengelig enn numeriske oppsummeringer og tabeller. Grammatikk er verktøyet som strukturerer formidlingen, både språklig og grafisk.

For å få til noe meningsfullt er det fornuftig å sette sammen elementer etter en bestemt struktur. Det gjelder så vel for språk, som for bygningskonstruksjoner og pasientbehandling. Dessverre blir ofte grafisk fremstilling av data og statistiske resultater til underveis, uten en bestemt plan for datautvalg, grafikktype og utseende. Leland Wilkinson (1944–2021) har utviklet et grammatisk system for grafer basert på syv grunnleggende elementer [\(1\)](#). Grafikkens grammatikk gir både struktur og fleksibilitet og er et nyttig verktøy for å bygge opp ulike grafikktyper.

De syv elementene

Grafikkens grammatikk består av syv elementer. Som for enhver statistisk analyse begynner det hele med dataene, altså variablene og observasjonene. Deretter følger hvordan vi kobler sammen dataene – ofte omtalt som estetikken eller estetiske koblinger i en graf. Det tredje elementet er hvordan dataene skal fremstilles geometrisk, for eksempel med punkter, linjer, stolper eller kakediagrammer. Disse første tre elementene er helt nødvendige, som vist i figur 1a [\(2\)](#). Der er alder plottet mot blodtrykk uten videre bearbeiding verken statistisk eller visuelt.



Figur 1 Alder og blodtrykk er grafisk vist ved hjelp av a) de første tre elementene i grafisk grammatikk, som består av dataene, koblingen mellom disse og geometriske former. b) Deretter er det lagt inn en regresjonslinje som det fjerde elementet, c) tilpasset skalaene og koordinatene som femte og sjette element og til slutt d) delt opp i to delgrafer (fasettering) ved å legge inn det syvende elementet, i tillegg til en visuell tilpasning. Dataene er et tilfeldig utvalg fra NHANES: Data from the US National Health and Nutrition Examination Study, R-pakke versjon 2.1.0 (2).

Det fjerde elementet i grafikkens grammatikk er statistiske transformasjoner og beregninger, noe som ofte er nyttig, men ikke alltid nødvendig. Et eksempel på dette kan være å sette inn en enkel regresjonslinje (figur 1b). Deretter er det femte elementet å tilpasse skalaene til grafen med angitte grenser, intervaller og beskrivelser. I figur 1c er x- og y-aksen avgrenset til henholdsvis 0 til 80 og 80 til 160, der begge har intervall på 10. Beskrivelsene er endret fra de engelske variabelnavnene i databasen til egnede norske navn, slik at det er enkelt å forstå at alder og blodtrykk er plottet mot hverandre. Valg av koordinatsystem er regnet som det sjette elementet, og er også illustrert i figur 1c ved å angi lik avstand og lengde på aksene. Såkalt fasettering, der data blir delt opp og vist i flere delgrafer med samme struktur, er det siste og syvende elementet. I figur 1d er grafen vist separat for kvinner og menn.

I tillegg til disse syv grammatikalske elementene i en graf kommer eventuelle visuelle tilpasninger, som for eksempel skrifttype og -størrelse, bakgrunnsfarger, rutenett, hjelpelinjer og utseendet til aksene – ofte betegnet som *themes*. Utseendet til figur 1d er derfor endret med en annen bakgrunnsfarge, visuell tilpasning av aksene og fjerning av rutenettet. Ifølge grafisk grammatikk blir justering av utseendet en siste tilpasning etter at grafens struktur er på plass, som definert av de syv foregående elementene.

ggplot2

Grafikkens grammatikk gir oss et system og en struktur for å fremstille data og statistiske resultater som er nyttig, enten man lager figurer for hånd eller ved hjelp av enkle grafiske programmer. Ved å tilpasse dette systemet til statistiske programmer oppnår man mye mer. Det finnes flere programmer basert på grafikkens grammatikk, men kanskje det mest brukte av dem er ggplot2 (3) til programpakken R (The R Project for Statistical Computing) og RStudio (Posit Software, PBC). ggplot2 er også en del av den større tidyverse-pakken (4), som gir en utvidet og modernisert bruk av R og RStudio. Programmene kan lastes ned gratis fra internett og er også tilgjengelige via de fleste høyskoler og universiteter, samt sykehusenes egne datatjenester. Lærebøker og presentasjoner på YouTube er til stor hjelp, men etter noe grunnleggende innføring vil ofte kunstig intelligens, som ChatGPT (OpenAI Foundation), bli en uvurderlig assistent, og den bidro også til å lage figur 1.

Å lage figurer basert på grafikkens grammatikk med avanserte programpakker kan føles som ekstraarbeid og noe man helst vil overlate til de spesielt interesserte. Etter egen erfaring har verktøy for kunstig intelligens gjort alt mye enklere. Hjelp og råd, enten fra kunstig eller menneskelig intelligens, fungerer imidlertid ofte best når man vurderer dem kritisk selv.

LITTERATUR

1. Wilkinson L. The Grammar of Graphics. New York, NY: Springer, 2005.
2. NHANES. Data from the US National Health and Nutrition Examination Study, R-pakke versjon 2.1.0. <https://cran.r-project.org/web/packages/NHANES/index.html> Lest 25.1.2026.
3. Wickham H. red. ggplot2: Elegant Graphics for Data Analysis. Cham: Springer International Publishing, 2016.
4. Wickham H, Averick M, Bryan J et al. Welcome to the Tidyverse. J Open Source Softw 2019; 4: 1686. [CrossRef]

Publisert: 19. mars 2026. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.26.0031

Opphavsrett: © Tidsskriftet 2026 Lastet ned fra tidsskriftet.no 11. juli 2026.