

Jan Erik Kristiansen

Visualisering av statistikk
Fra tabell til diagram

Notater

Forord

“Tallenes tale” er ikke alltid like tydelig. Ofte taler tallene til oss i form av diagrammer og vår fortolkning av disse avhenger i stor grad av hvordan de er utformet. Å lyve med statistikk er langt lettere med diagrammer enn med tabeller. Etter hvert som statistikk og diagrammer blir en stadig viktigere del av arbeidsdagen (og hverdagen) for fler og fler, blir det også viktig kjenne til de muligheter og begrensninger som diagrammer byr på.

Det følgende er en elementær innføring og veiledning i bruk og utforming av diagrammer. Målgruppen er alle som presenterer statistikk i form av diagrammer – som avsendere eller mottakere – enten det nå er i semesteroppgaver, rapporter, artikler, lærebøker, aviser, markedsundersøkelser, utredninger/notater, undervisning eller foredrag.

Hovedvekten er lagt på en fyldig behandling de vanligste typene diagrammer og ulike varianter av disse. Én type grafiske framstillinger som er mye brukt, men som ikke behandles her, er ulike former for tematiske kart. Heller ikke skjematiske diagrammer, som organisasjonskart, flytdiagrammer o.l. omtales. Hovedvekten er lagt på diagrammer som har koordinatsystemet som utgangspunkt, dvs. numerisk grafikk.

Utvalget av diagrammer er nok tematisk preget av at min arbeidsplass er Statistisk sentralbyrå, selv om jeg har forsøkt å inkludere eksempler også fra andre områder.

Med tanke på en senere revisjon og utvidelse av notatet, tar jeg svært gjerne i mot kommentarer og forslag fra leserne.

Oslo, juli 2002

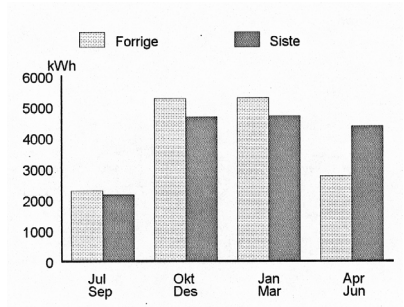
Innhold

1. Innledning: Hvorfor bruke diagrammer?	3
2. Diagrammets arkitektur: Noen generelle prinsipper og regler	7
2.1. Noen tommelfingerregler	7
2.2. Grafisk misvisning	12
2.3. Hva slags diagram skal jeg lage?	14
3. Søylediagrammer: Å sammenlikne grupper	16
3.1. Enkle vertikale søylediagrammer	16
3.2. Grupperte søylediagrammer:	17
3.3. S sammensatte søylediagrammer	18
3.4. Horisontale søylediagrammer	19
4. Linjediagram: Tid og mengde	21
4.1. Enkle linjediagrammer	21
4.2. Multiple linjediagrammer.....	22
4.3. Gruppert linjediagram	23
4.4. Arealdiagram	23
4.5. To verdiakser	24
4.6. Indeksdiagram	25
4.7. Glatting: Glidende gjennomsnitt og trendlinjer.....	25
4.8. Delårsdata.....	26
4.9. Kumulative linjediagrammer.....	26
5. Sektordiagram: Prosentfordeling	27
5.1. Én enkelt fordeling.....	27
5.2. Sammenlikne fordelinger?	28
5.3. Varianter.....	28
6. Andre typer diagrammer	30
6.1. Kombinasjonsdiagrammer	30
6.2. Punktdiagram	31
6.3. Min - maks diagram (variasjonsdiagram)	33
6.4. Flate- og volumdiagram	33
6.5. Rangdiagram	34
6.6. Z-diagram	34
6.7. Radar-diagram.....	35
6.8. Diagram-matrise.....	35
7. Kort om bruk av farger	36
7.1. Hvilke farger?.....	36
7.2. Fargerikt fellesskap?	37
7.3. Farger hvor?	37
Litteratur	38
De sist utgitte publikasjonene i serien Notater	39

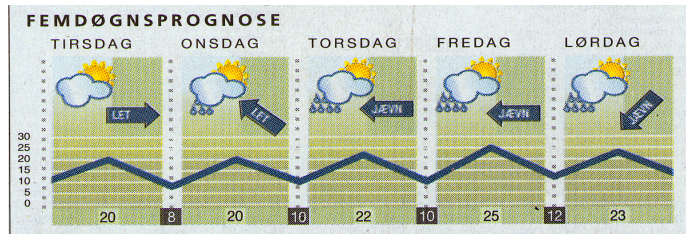
1. Innledning: Hvorfor bruke diagrammer?

Diagrammer og andre former for grafiske framstillinger eller illustrasjoner brukes i dag i økende grad både i lærebøker, fagbøker, brosjyrer, tidsskrifter og årsmeldinger, men også i aviser, på TV og Internett. Særlig er avisenes økonomi- og næringslivssider fulle av diagrammer, hvor renteendringer, oljepris, aksjekurser og børsutvikling framstilles grafisk. Også i utredninger, undervisning, presentasjoner og foredrag blir tall og statistikk i økende grad presentert i form av diagrammer.

Men diagrammene har også fått nye bruksområder: På strømregningen forsøker elektrisitetsleverandøren på en pedagogisk måte å tydeliggjøre strømforbruket mitt ved hjelp av et søylediagram (dette eksemplet er fra Viken Energinetts faktura, men de fleste strømleverandørene bruker liknende diagrammer).



Også i avisenes værvarsler kan diagrammer benyttes for å presentere temperaturer og nedbør. Eksemplet her er fra danske *Jyllands-Posten*:

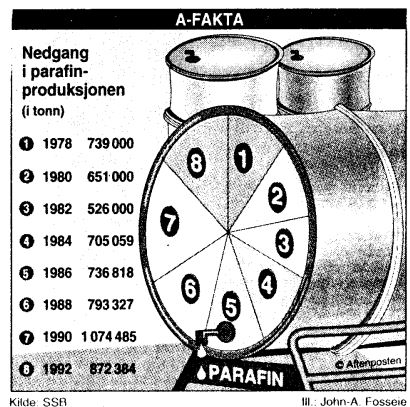
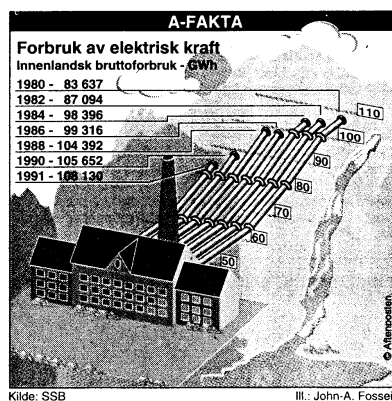


I tillegg finner vi også stadig oftere diagrammer i annonser (for eksempel for aksjefond o.l.), mens eksemplet til høyre viser hvordan et stilisert diagram kan brukes som logo.

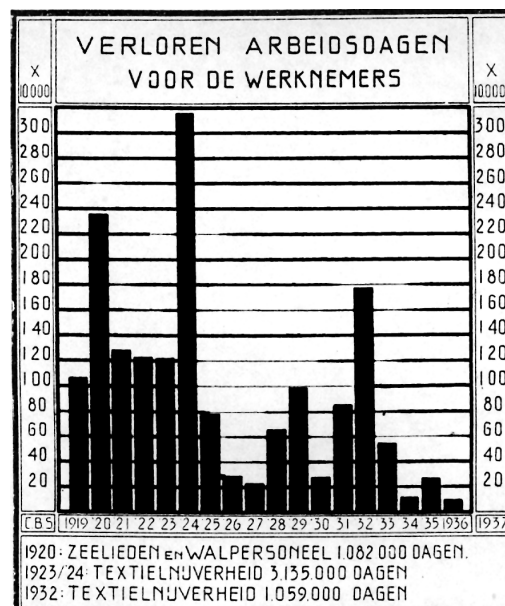
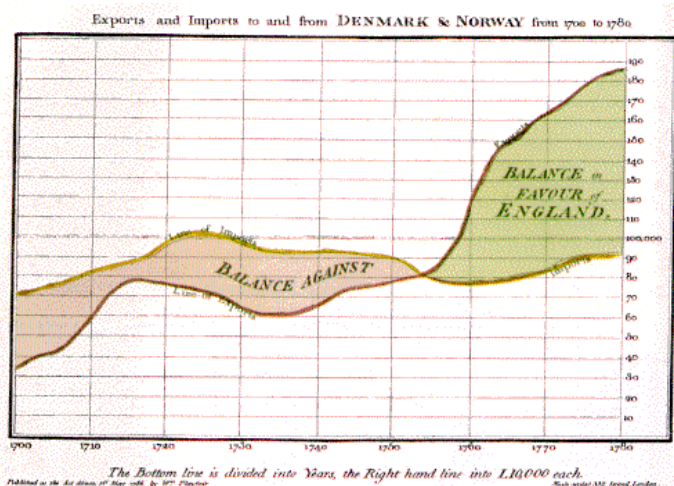


Denne økende bruken av diagrammer skyldes nok delvis en økt interesse for og bruk av ulike typer statistikk. Men utviklingen har nok også for en stor del sammenheng med tilgjengeligheten av grafiske dataprogrammer¹, som gjør det langt enklere å framstille ulike diagrammer. Og som de fleste andre tekniske nyvinninger kan også disse programmene misbrukes. Resultatet blir gjerne "fikse" diagrammer overlesset med effekter, hvor oljefat eller kronestykker stables på en måte som gjør at informasjonen drukner i grafisk støy.

Eksemplene til høyre er fra *Aftenposten* for noen år tilbake (de lager heldigvis ikke så mange diagrammer av denne typen lenger). Diagrammene må vel nærmest betraktes som illustrerte tabeller, og de er mer forvirrende enn informative. Dette er eksempler på det Tufte (1983) kaller "chart junk".



Her er det derfor fristende å gå noen år tilbake i tiden og vise et par eksempler fra den gang diagrammer ble tegnet for hånd. Det første (nedenfor til venstre) er fra 1786 og er signert William Playfair. Diagrammet til høyre er fra Nederland (ca. 1937). Poenget med disse historiske eksemplene er å vise at moderne datateknologi verken er en nødvendig *eller* tilstrekkelig forutsetning for å lage gode, lesbare, elegante og informative diagrammer.



Men hvorfor i det hele tatt bruke diagrammer, når en tabell i de fleste tilfeller gir langt mer informasjon? Svaret er: Nettopp derfor! Tabeller gir ikke bare mer informasjon, men også ofte for mye. Særlig i store og omfattende tabeller kan det være vanskelig å se skogen for bare tall: Informasjonen drukner lett i numerisk støy.

Et diagram derimot, gir et visuelt, helhetlig inntrykk som gjør det lettere raskt å oppfatte for eksempel hovedtendenser i et utviklingsforløp, forskjeller mellom grupper eller sammenhenger mellom ulike fenomener. Diagrammet demonstrerer et poeng!

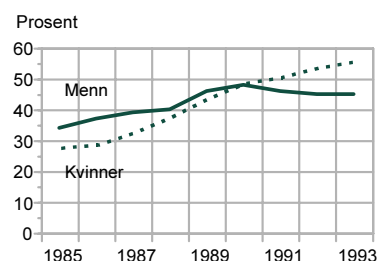
Ta for eksempel følgende (fiktive) tabell:

	1985	1986	1987	1988	1989	1990	1991	1992	1993
Menn:	34	37	39	40	46	48	46	45	45
Kvinner:	28	29	33	38	44	49	51	54	56

Denne tabellen viser en rekke ting, bl.a.:

- at andelen har økt både for menn og kvinner i perioden, men mest for kvinner (henholdsvis 11 og 28 prosentpoeng)
- at andelen menn i årene 1985-87 var klart høyere enn andelen kvinner
- at andelen menn og kvinner i 1988-90 var omtrent det samme
- at andelen kvinner passerte andelen menn i 1990 og siden har vært høyere, og endelig
- at denne forskjellen er økende.

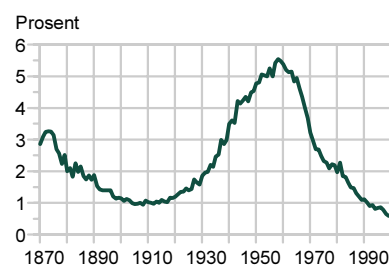
Alt dette framgår selvfølgelig av tabellen, men det kreves relativt inngående studier av tallene for å få med seg alt dette (og dette er en liten tabell!). Diagrammet derimot, gir på et øyeblikk et umiddelbart visuelt inntrykk av og overblikk over hovedtendensen i utviklingen, og er derfor spesielt godt egnet til bruk i mer populære framstillinger og presentasjoner rettet mot et publikum som ikke er trent i å lese og bruke tabeller.



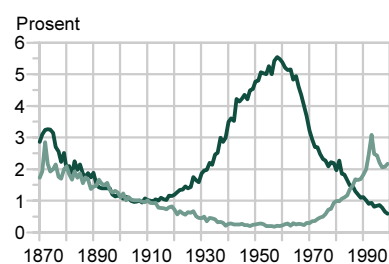
Ett annet eksempel på diagrammets komprimerende funksjon: Tabellen nedenfor viser prosentandelen jenter som hvert år har fått fornavnet *Anne* i perioden 1870-1999.

År:	Pst.:	År:	Pst.:	År:	Pst.:	År:	Pst.:	År:	Pst.:	År:	Pst.:
1870	2,83	1892	1,39	1914	0,97	1936	2,49	1958	5,51	1980	1,94
1871	3,06	1893	1,36	1915	1,06	1937	2,96	1959	5,45	1981	2,24
1872	3,21	1894	1,36	1916	1,01	1938	2,83	1960	5,35	1982	1,82
1873	3,23	1895	1,36	1917	0,99	1939	2,96	1961	5,17	1983	1,79
1874	3,22	1896	1,36	1918	1,13	1940	3,46	1962	5,10	1984	1,61
1875	3,11	1897	1,17	1919	1,12	1941	3,57	1963	5,12	1985	1,45
1876	2,67	1898	1,10	1920	1,16	1942	3,49	1964	4,80	1986	1,43
1877	2,53	1899	1,13	1921	1,24	1943	4,19	1965	4,91	1987	1,27
1878	2,20	1900	1,11	1922	1,31	1944	4,11	1966	4,58	1988	1,16
1879	2,48	1901	1,03	1923	1,32	1945	4,20	1967	4,32	1989	1,06
1880	1,97	1902	1,09	1924	1,42	1946	4,32	1968	3,97	1990	1,08
1881	2,07	1903	1,05	1925	1,36	1947	4,18	1969	3,64	1991	0,98
1882	1,80	1904	0,96	1926	1,40	1948	4,46	1970	3,19	1992	0,87
1883	2,22	1905	0,93	1927	1,71	1949	4,50	1971	2,92	1993	0,90
1884	1,95	1906	0,94	1928	1,61	1950	4,74	1972	2,67	1994	0,78
1885	2,12	1907	0,97	1929	1,55	1951	4,77	1973	2,66	1995	0,81
1886	1,82	1908	0,91	1930	1,83	1952	5,03	1974	2,45	1996	0,83
1887	1,71	1909	1,04	1931	1,93	1953	5,00	1975	2,29	1997	0,75
1888	1,84	1910	0,99	1932	1,95	1954	4,97	1976	2,24	1998	0,62
1889	1,70	1911	0,97	1933	2,17	1955	5,22	1977	2,06	1999	0,56
1890	1,85	1912	0,94	1934	2,11	1956	4,97	1978	2,19		
1891	1,50	1913	1,01	1935	2,43	1957	5,39	1979	2,16		

Tabellen inneholder 130 datapunkter eller observasjoner og det er svært vanskelig å danne seg noe klart bilde av *Annes* "vekst og fall" bare ved å studere tallrekkene. I figuren derimot, framstår bølgebevegelsen tydelig. Figuren er intuitivt forståelig og krever ingen kunnskaper om statistikk eller prosentregning. I tillegg er diagrammet langt mindre plasskrevende (det har større datatetthet) enn tabellen og denne komprimeringsfordelen blir selvfølgelig større jo flere tallrekker som skal presenteres eller sammenliknes. Dette ser vi tydelig hvis vi legger til ytterligere en dataserie.



Dette eksemplet illustrerer også en annen viktig funksjon; nemlig *den pedagogiske*: Et diagram sier ofte mer enn tusen tall. Og visuelt framstilte sammenhenger eller forløp er ikke bare lettere å oppfatte, de er også langt lettere å huske enn variasjoner i én eller flere tallrekker. (Jeg kan her nevne at jeg sammen med en kollega i 1998 utga en bok som nesten i sin helhet besto av diagrammer (*Fornavn i Norge - Navnemoter og motenavn*. Ad Notam Gyldendal). Boka var vel ingen bestselger, men solgte rimelig bra. Poenget her er at en slik bok neppe kunne vært laget med tabeller i stedet for diagrammer; både fordi den da ville blitt 5-6 ganger så stor og fordi bare et lite fåtall spesielt interessert da ville funnet den interessant).



Diagrammet gir altså et raskt helhetsinntrykk og overblikk. Men samtidig gir det nødvendigvis et *forenklet* bilde hvor detaljene går tapt, idet man sjelden kan avlese nøyaktige verdier: Man kan f.eks. ikke i diagrammet ovenfor se om andelen som fikk navnet *Anne* i 1980 egentlig er 1,9, 2,0 eller 2,1 prosent (og slett ikke at det er 1,94). Derfor: I den grad det er viktig at leseren kan avlese nøyaktige verdier (f.eks. med tanke på videre bearbeiding), bør man bruke tabeller i stedet. Diagrammer er altså først og fremst egnet til en mer "populær" presentasjon av statistikk, hvor detaljer og desimaler er mindre viktig. I formidling av statistikk og forskning som primært er rettet mot kolleger og forsk-

ningsmiljøer (som selv ønsker å bearbeide dataene) vil derfor tabeller ofte være en bedre måte å presentere tallene på (Eventuelt kan tallene gjengis i et eget tabellvedlegg).

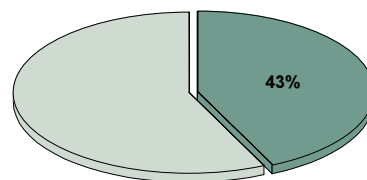
En annen grunn til å bruke grafiske framstillinger, er at enkelte typer informasjon er vanskelig å framstille i tabeller eller tekst. Ett eksempel er tematiske kart med geografisk informasjon, f.eks. bo-settingskart.

Endelig kan figurer/diagrammer – på samme måte som bilder eller andre illustrasjoner – fungere som interessevekkende visuelle blikkfang og/eller grafiske elementer som myker opp teksten og gjør den mer innbydende og lettere å lese.

Men først og fremst er altså grafiske framstillinger *et effektivt hjelpemiddel for informasjonsformidling* og eventuelle dekorative eller illustrerende funksjoner bør være underordnet den informative funksjonen. Diagrammer bør ikke brukes som fyllstoff, men skal først og fremst bidra til å understreke eller fremheve et poeng: en økning/nedgang, en forskjell, en sammenheng eller f.eks. en tendens til mindre eller større forskjeller.

Én definisjon av et diagram er: “... en skjematisk framstilling som blir brukt i naturvitenskap, statistikk o.a. for å kunne gi et bilde av forholdet mellom to eller flere størrelser” (Store norske leksikon, 1996).

Det sentrale her er "to eller flere størrelser". Å sammenlikne tall er kjernen i all deskriptiv statistikk – enten den presenteres tabellarisk eller grafisk – og det gir liten mening å framstille ett enkelt tall grafisk. Av og til ser vi imidlertid sektordiagrammer som skal illustrere ett tall, for eksempel at skilsmissehyppigheten er 43 prosent. Men også et slik diagram er jo egentlig basert på to tall; nemlig prosentandelen som forventes å bli skilt og andelen som ikke blir det. Et diagram basert på bare ett tall – det være seg antallet fødsler eller antallet trafikkulykker i løpet av ett år – er nokså meningsløst.



Men selv med to tall er det bare sjelden grunn til å gjøre bruk av et diagram. Først med tre tall – og dermed tre mulige sammenlikninger – begynner det å bli et behov for å framstille forholdet grafisk. Generelt kan man si at med et økende antall "størrelser" øker også antallet mulige sammenlikninger dramatisk, og dermed også behovet for den forenkling og oversikt som diagrammet tilbyr.

1. En del av de løsningene som diskuteres/anbefales i det følgende kan være avhengig av hvilket verktøy eller program som brukes. Det vanligste programmet er i dag sannsynligvis Microsofts *Excel*. For mange formål er dette tilfredsstillende, men programmet har sine klare begrensinger. For det første gir ikke *Excel* brukeren full kontroll over redigeringen, blant annet fordi programmet ikke inneholder en linjal eller andre hjelpemidler for å sette størrelsen på diagrammet. For det andre kan ikke *Excel* generere Postscript-filer og den eneste måten å få diagrammene inn i *Word* (eller et annet redigeringsverktøy) blir dermed ved å klippe og lime. Dette fører imidlertid ofte til forvrengninger av enkelte elementer i diagrammet, særlig raster/mønster og fonter.

Et godt alternativ til *Excel* er *DeltaGraph*, fra SPSS (de fleste diagrammene i denne publikasjonen er laget med *DeltaGraph*). I tidligere versjoner har det bl.a. vært problemer med eksportfunksjonen når programmet kjøres under Windows NT. I den siste versjonen (5.0) synes dette å være løst, i alle fall under Windows 2000.

Et tredje alternativ er selvfølgelig Adobes *Illustrator*. Dette generelle design-verktøyet inneholder imidlertid ikke noen regneark-funksjoner. Derimot gir det full kontroll over design og layout av diagrammet.

2. Diagrammets arkitektur: Noen generelle prinsipper og regler

Det er selvfølgelig vanskelig å gi noen helt generelle regler for hvordan grafiske framstillinger/ diagrammer skal utformes. Dette vil avhenge av flere forhold, blant annet:

- hvem som er målgruppen (er den vant til å lese diagrammer?)
- om man lager et arbeidsnotat eller en årsmelding
- om man hovedsakelig henvender seg til kolleger (spesialister) eller til et større publikum
- hvilken sammenheng diagrammene står i (publikasjonens generelle layout, om de står alene eller med forklarende tekst)
- og ikke minst: Hvilket dataprogram som benyttes og hvilke valgmuligheter som finnes.

Det finnes likevel noen generelle regler som bør følges. På samme måte som språket har sin grammatikk, finnes det en slags dia-grammatikk for grafisk framstilling.

2.1. Noen tommelfingerregler

Typografiens første bud er: *Konsekvens*. Når man har valgt en layout med bestemte marger, skrifttyper/størrelser, plassering av overskrifter, etc. i en publikasjon, skal denne malen følges nøye. Noe annet skaper et tilfeldig, rotete og slurvete inntrykk. Det samme gjelder også utformingen av diagrammer: Lag en mal eller modell – og følg den!

Størrelse

Diagrammenes bredde følger ofte av publikasjonens format og layout (1, 2 eller 3 spalter). Men også høyden bør være fast; velg 2 eller 3 ulike høydeformater og bruk disse. Ikke "tøy" diagrammet for å fylle siden eller spalten.

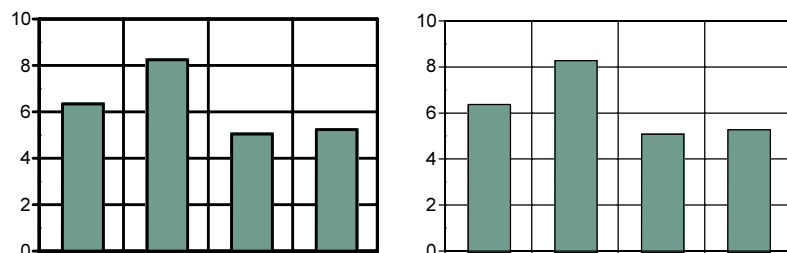
En av diagrammets hovedfunksjoner er å forenkle og visuelt presentere hovedtendensen i et tallmateriale. Dette betyr at detaljer stort sett er mindre viktig. Satt på spissen kan man si at et diagram som er utformet i samsvar med regelen om enkelhet, bør kunne reduseres til frimerkestørrelse. Som eksempel vises her (i full størrelse) ett av de to frimerkene som ble laget til Statistisk sentralbyrås 100 års jubileum i 1976. Selv om dette formatet nok for mange formål er i minste laget, er det bare sjelden behov for å la diagrammer oppta en halv (eller hel) side. Det formatet som gjennomgående er brukt i denne publikasjonen (50 x 35 mm) er, for de fleste formål, stort nok. Unntak finnes selvfølgelig for kompliserte, analytiske diagrammer med mange elementer, men hovedregelen er: Smått er godt! (Se for eksempel diagrammene i det engelske tidsskriftet *The Economist*).



Frimerke utgitt til SSBs 100 års jubileum

Strektykkelse

Tynne streker gir diagrammet et "lettere" preg, mens tykke streker gir et "tyngre" preg. Valg av strektykkelse må tilpasses utskriftsmedium (laser-skriver eller fotosetter), men vanligvis er 1/2 pkt passende.



Strektykkelsen betyr mye for diagrammet utseende/uttrykk

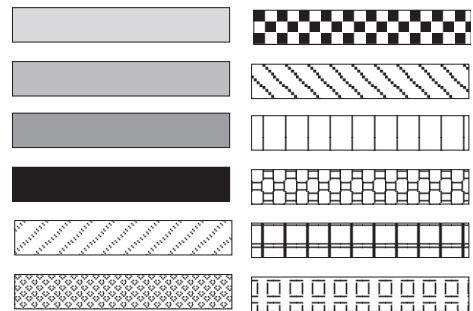
¼ pkt _____
 ½ pkt _____
 ¾ pkt _____
 1 pkt _____

Med tanke på bruk av fotosetter og offset-trykk bør man ikke bruke svært tynne streker, da de lett kan forsvinne i trykkeprosessen. Kurver/linjer bør være klart tykkere enn diagrammets øvrige elementer, f.eks. 1,5-2 pkt.

Skravering/gråtoner

Utvalget av gråtoner og ulike former for skraveringer eller fyll vil variere fra program til program. Ved bruk av gråtoner er en generell regel ikke å bruke mer enn 3-4 ulike prosenter; f.eks. 1, 15, 50 og 100 prosent. Selv om programmet har flere muligheter og de klart lar seg skille fra hverandre på en laseroriginal, blir resultatet ofte uleselig etter trykking. Rasterene bør plasseres i "naturlig" rekkefølge, fra mørkt til lyst.

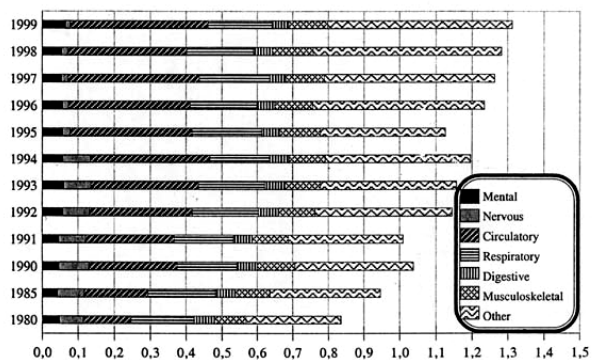
Fire ulike gråtoner er imidlertid ofte ikke nok og de fleste programmer har da også en rekke andre skraveringer eller mønster å by på. Ikke alle mønster er imidlertid like egnet. Mønstrene lengst til høyre bør unngås, bl.a. fordi forklaringsboksene i diagrammene ofte er små og mønstrene blir derfor vanskelig å tyde. Velg heller en type linjemønster. Men også her er det grunn til forsiktighet: Unngå horisontale og vertikale linjer, bruk heller skråstilte.



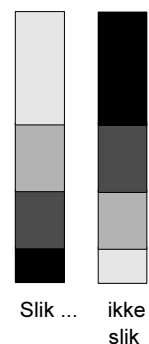
Mønstrene til høyre anbefales ikke

Eksemplet til høyre viser hvordan resultatet kan bli når man bruker denne typen mønster (eksemplet er hentet fra Statistisk årbok for Kypros). Særlig dårlig blir lesbarheten her, fordi søylene er så smale.

FIGURE 10: OUTPATIENTS ATTENDANCES BY SELECTED DISEASE CATEGORY, 1980-1999 (ALL MEDICAL INSTITUTIONS)

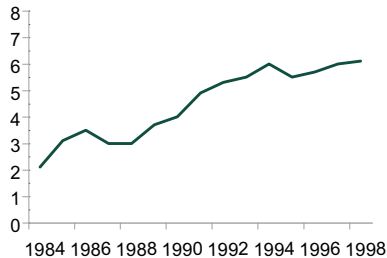


Ved bruk av gråtoneraster, er det vanlig å starte med det mørkeste rasteret i bunn, med den lyseste gråtonen på toppen (ellers blir søylene lett "topptunge"). Med liggende søyler er det vanlig med de mørkeste feltene til venstre, nærmest aksene.

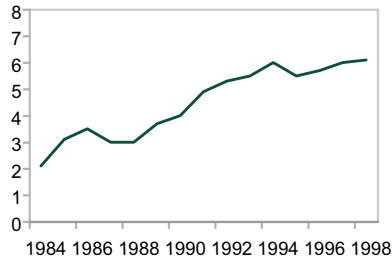


Hjelpelinjer

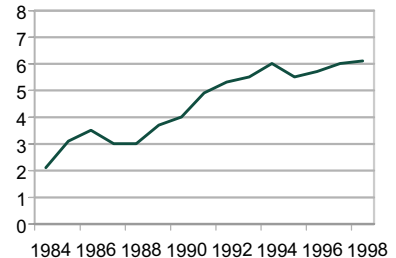
Hjelpelinjer (engelsk: grids) skal hjelpe leseren til å avlese verdiene på aksene. I søylediagrammer er det stort sett bare behov for hjelpelinjer på Y-aksen. I linjediagrammer derimot, finnes det svært varierende praksis. Enkelte bruker horisontale linjer for å hjelpe avlesningen av Y-aksen, men andre bruker vertikale linjer for lettere å identifisere tidspunktet eller perioden. Alternativt kan man bruke både horisontale og vertikale linjer.



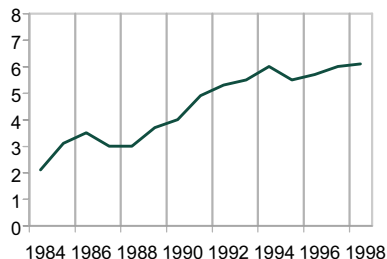
Dette er en løsning uten hjelpelinjer og uten ramme



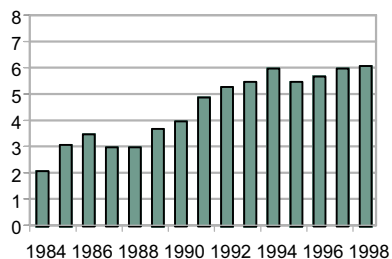
Med ramme, men uten hjelpelinjer blir diagrammet virkende litt "tomt"



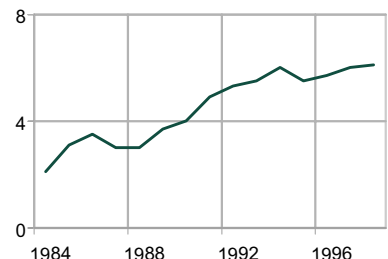
Her er det bare horisontale hjelpelinjer, som gjør det lettere å lese av verdiene på y-aksen



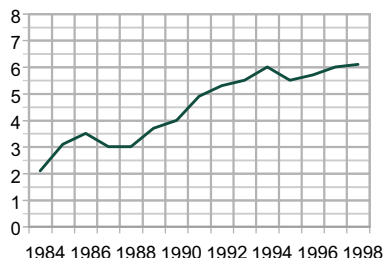
Vertikale hjelpelinjer gjør det lettere å identifisere tidspunktene/-periodene



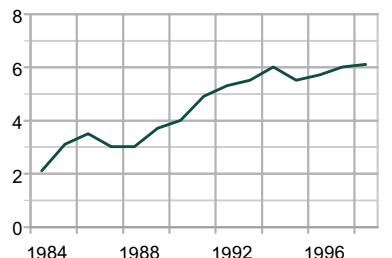
I søylediagrammer er det strengt tatt ikke behov for vertikale hjelpelinjer



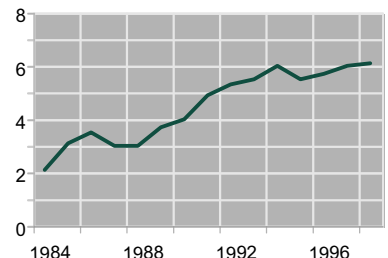
Her er det kanskje litt for få hjelpelinjer, minimum antall bør være 4



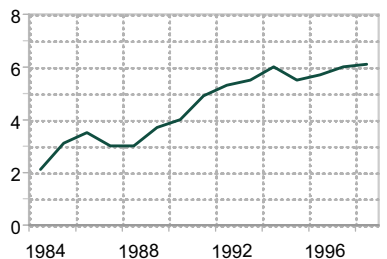
..., mens her er det litt for mange (maksimum = 10)



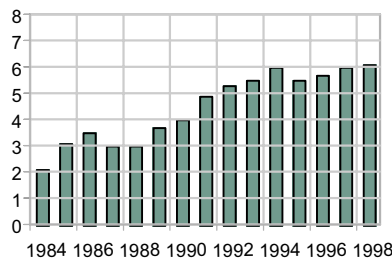
Dette er kanskje passe?



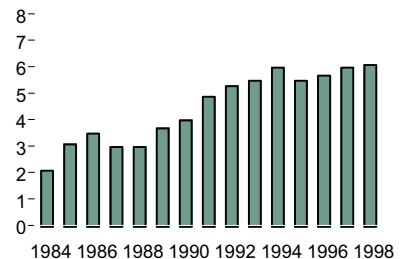
Dette er en annen løsning: Grå bakgrunn med hvite linjer



Denne løsningen benytter prikkete linjer, som skiller seg klart fra kurven



Her er hjelpelinjene lagt over søylene

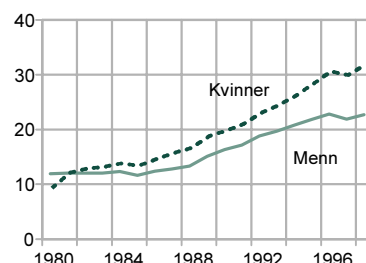
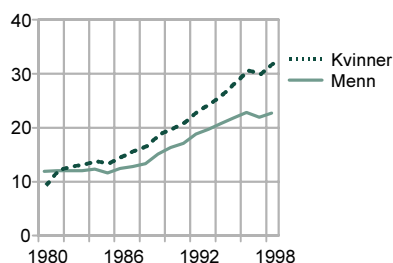
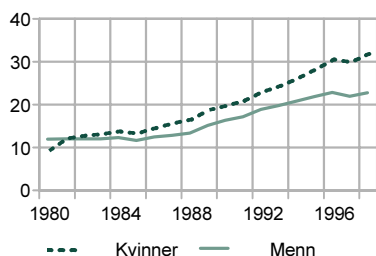


En "minimalistisk" versjon, hvor også aksene er fjernet

Uansett hvilken løsning man velger, bør antallet linjer verken være for stort eller for lite (en regel kan være minimum fire, maksimum ti). Hvilken type linjer man skal benytte (f.eks. heltrukne eller prikkete), er ofte et spørsmål om smak og behag. Men hjelpelinjene bør alltid være tynnere enn selve data-linjen/kurven.

Forklaring

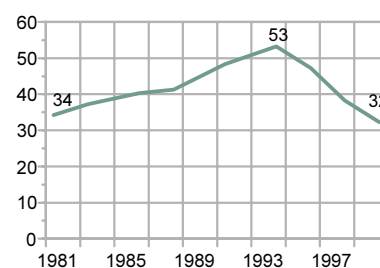
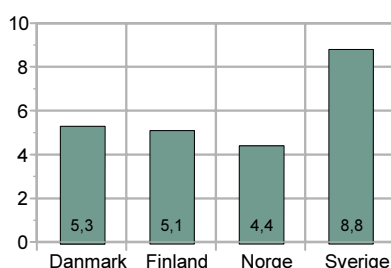
De fleste dataprogrammene tilbyr forskjellige plasseringer av tegnforklaringene, f.eks. under figuren eller til høyre. Dette blir ofte rotete og lite lesbart og i svært mange tilfeller er det derfor en bedre løsning å ta teksten inn i diagrammet (Forutsetningen er selvfølgelig at teksten (tegnforklaringen) er relativt kort og får plass inne i diagrammet). Dette er litt mer arbeidskrevende enn bare å akseptere programmets plassering, men det øker lesbarheten betraktelig. I tillegg får data-arealet på denne måten den dominerende plassen det fortjener.



Tall i diagrammet?

Skal diagrammer også inneholde tallene? Og i så fall; hvor mange og hvor? Generelt blir jo en grafisk fremstilling brukt for å gi overblikk og hovedtendenser, ikke detaljer. Derfor er hovedregelen ikke å gjengi tallene. Diagrammer med mange tall blir en slags Ole Brumm-løsning: "Ja takk, begge deler".

Hvis du likevel ønsker å gjøre dette, så anbefales det å sette tallene inn i søylene. For linjediagrammer kan det noen ganger være nyttig å oppgi den første og siste verdien i en tidsserie (eventuelt også minimum/maksimum).

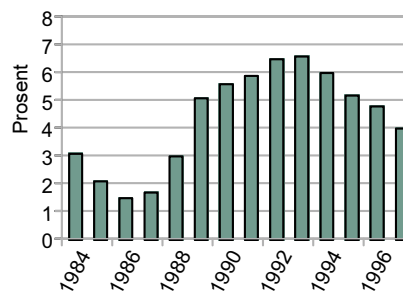


Tekst i diagrammet

Skrifttyper/-størrelse: Velg én skrifttype (gjerne en grotesk, f.eks. Arial/Helvetica, Frutiger, Trebuchet eller en annen skrifttype som er tydelig forskjellig fra brødteksten) for hele diagrammet, eventuelt med halvfet eller kursiv som variasjoner. Hvis skrifttypen ikke er gitt av sammenhengen, bruk gjerne en "narrow", "thin", "light" eller "condensed", dermed får du plass til mer tekst på samme linje. Plasseringen av ulike tekstelementer bør være den samme i de ulike diagrammene. Fontstørrelsen må tilpasses diagrammets/publikasjonens format/størrelse, men for de fleste formål er 7 pkt tilstrekkelig. Kildehenvisninger og noter kan være litt mindre (6 pkt). Tittel/overskrift bør være større, for eksempel 9 pkt.

Tittel: Tittelen bør være kort, men presis. I den grad det kan være noen tvil om hva som "telle-enheten", må dette presiseres (hus/leilighet, familie/husholdning, person/ektepar). Også måleenheten angis ofte i tittelen (antall/prosent/ rater/indeks/kroner). Tittelen bør plasseres over diagrammet, venstrejustert eller midtstilt. Når dataene gjelder ett bestemt år, skal dette oppgis. Når det gjelder tidsserier, vil årangivelsen som oftest framgå av tidsaksen.

Akse-etiketter/verdier: Som all annen tekst bør også denne plasseres horisontalt. Man skal ikke måtte drive med nakke-gymnastikk for å lese skråstilt eller vertikal tekst, som i dette eksemplet. Hvis det blir trangt om plassen, så kan man sløyfe noen årstall. Alternativt kunne man i dette tilfellet skrive årstallene slik: '84 '85 etc. Det dårligste alternativet er å bruke en mindre skrifttype.



Noter: Noter bør, så langt det er mulig, unngås, men er av og til likevel nødvendige for å presisere eller utdype forhold i diagrammet.

Kilde: Hvis alle diagrammene i en rapport/publikasjon er basert på ett og samme datamateriale, er det ikke nødvendig å oppgi kilde i det enkelte diagram. Det er når data hentes fra mange ulike kilder, at det er viktig at disse oppgis. Kildene er vanligvis institusjoner eller publikasjoner (bøker/rapporter/tidsskrifter) eller begge deler. Eksempler:

Kilde: Statistisk årbok 2001, Statistisk sentralbyrå
Norges Bank/Oslo Børs

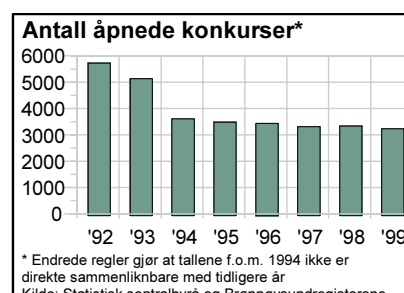
En datakilde som blir mer og mer vanlig, er Internett. Kildehenvisninger har da vanligvis formen:

Kilde: <http://www.ssb.no/konkurs/>

Det er vanlig å plassere kildehenvisningen til slutt, etter eventuelle noter.

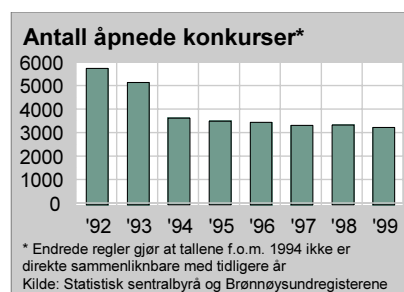
Bruk av "rammer"

Av og til ser man at figuren rammes inn av en ytre ramme eller boks. En ramme brukes ofte for å samle alle elementene i figuren (overskrift, forklaring, noter, referanser) og å avgrense dette mot teksten. Vanligvis er imidlertid dette unødvendig, særlig når figuren har andre skrifttyper enn teksten. En ytre ramme blir ofte "dobbel opp", fordi de fleste figurer også har en indre ramme, dannet av x- og y-aksen.



Her blir det for trangt mellom selve diagrammet/teksten og rammen

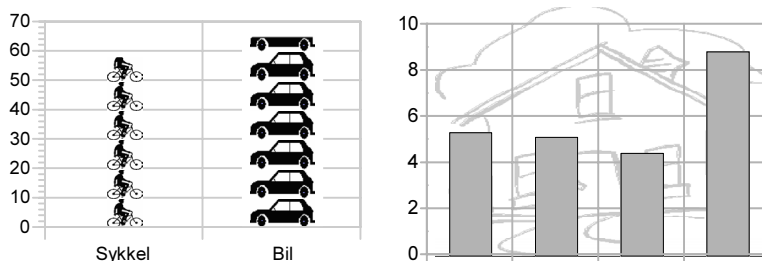
Men hvis du bruker en ytre ramme (eller et bakgrunnsraster), sørg da for at det blir tilstrekkelig luft mellom selve diagrammet/teksten og rammen.



Enkelhet

En annen grunnleggende regel er at diagrammene bør utformes så enkelt som mulig: Dette gjelder både omfanget av data som framstilles og bruken av effekter. Ikke forsøk å si alt i ett diagram, bruk heller flere enkle diagrammer enn ett komplisert. Elementer som ikke er nødvendig, bør fjernes eller bare brukes med stor forsiktighet (f.eks. rammer/skygger).

Enkelhet innebærer også at man ikke leses på med unødige effekter i form av clip-art, bakgrunnsillustrasjoner, logoer o.l. Bruken av såkalte pikto-grammer anbefales heller *ikke*, da de sannsynligvis forvirrer langt flere enn de imponerer.



Eksempler på bruk av visuelle effekter som sjelden øker diagrammets informasjonsverdi: snarere tvert imot

Enkelheten må imidlertid ikke gå på bekostning av *fullstendighet*: Diagrammet skal stå på egne ben og må inneholde alle opplysninger som er nødvendige for forståelsen og tolkningen av data. Man skal ikke være avhengig av f.eks. å lese teksten for å få med seg forbehold eller nødvendige forklaringer.

Men ingen regler uten unntak: Mange av de rådene som er gitt ovenfor er gitt med tanke på diagrammer som skal trykkes i notater, rapporter, bøker eller tidsskrifter. Et annet viktig bruksområde er for eksempel transparenter til bruk i undervisning og foredrag. Reglene om konsekvens og enkelhet gjelder selvfølgelig også her. Når det gjelder størrelsen, kan man imidlertid her med fordel lage diagrammene så store som mulig. Linjetykkelse og skriftstørrelse bør da økes tilsvarende.

2.2. Grafisk misvisning

Reglene ovenfor angår hovedsakelig diagrammets utseende. Dette er selvfølgelig viktig for lesbarheten, men det er andre aspekter ved utformingen som har langt større betydning for informasjoninnholdet og tolkningen. Særlig er aksenes lengde og forholdet mellom dem viktig.

Y-aksens lengde

En ikke uvanlig måte å "dramatisere" data på, er å utelate verdier på y-aksen og starte med minimumsverdien. Slike "kutt" av aksene kan noen ganger være berettiget, men bør vanligvis unngås. Eksempel til høyre er hentet fra *Aftenposten* og illustrerer antallet barn under 15 år som er pågrepet av politiet. Diagrammet gir inntrykk av en nærmest eksplosiv økning og en mangedobling av antallet, mens i realiteten er økningen "bare" 70 prosent.

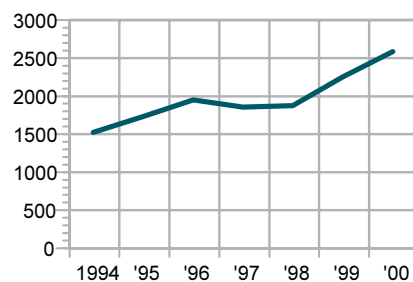
Også pilen i enden av kurven har en klart dramatiserende effekt, fordi den på en måte indikerer at økningen vil fortsette, noe man jo selvfølgelig ikke kan anta.



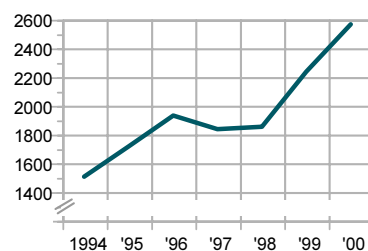
(Fra *Aftenposten* 18/11 2001)

Figuren til høyre – derimot – gir et langt mindre dramatisk (og mer korrekt) bilde av utviklingen. Grunnen til dette er at verdiaksen her ikke er kuttet, men begynner på 0.

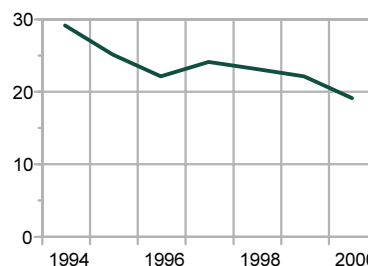
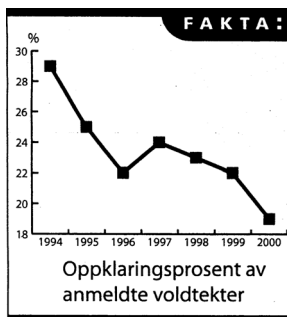
Nå finnes det selvfølgelig situasjoner hvor det er legitimt å foreta slike kutt på aksene. For eksempel ville avisenes daglige rente- og børsdiagrammer nærmest bli meningsløse uten slike kutt. Når det er detaljene i utviklingen som er i fokus, er dette både nødvendig



og tillatt. Men det bør ikke gjøres for å dramatisere et ellers "kjedelig" utviklingsforløp. Og hvis y-aksen ikke starter på 0, bør dette "bruddet" tydelig markeres på aksen.



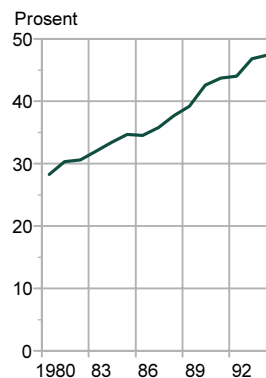
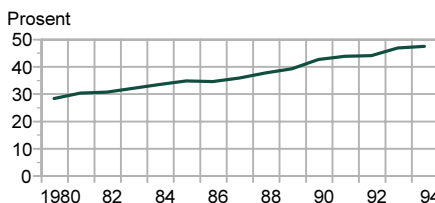
Et annet eksempel: I *Klassekampen* illustrerer dette diagrammet den markert nedgang i andelen oppklarte voldtektssaker. Også her er verdiaksen kuttet, noe som kan gi inntrykk av at oppklaringsprosenten nå nærmer seg null. Når vi inkluderer 0-punktet på y-aksen, blir nedgangen langt mindre dramatisk.



Diagrammet til venstre er fra *Klassekampen* 8/3 2002

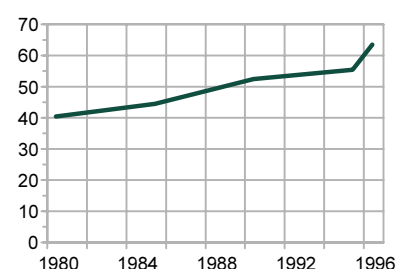
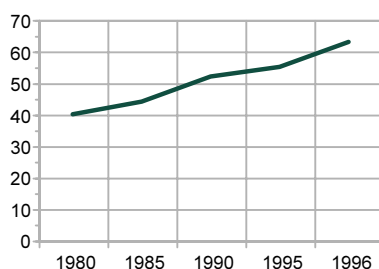
Diagrammets proporsjoner

Selv uten kutt på y-aksen, kan innholdet og fortolkningen av et diagram påvirkes av lengdene på aksene og forholdet mellom dem. Eksempelet til høyre (utviklingen i skilsmissehyppigheten) viser dette: Her har Y-aksene samme skala, men diagrammenes ulike proporsjoner formidler ulike inntrykk av økningen. Det "stående" diagrammet til høyre gir – på grunn av formatet – inntrykk av langt mer markert økning enn det "liggende" (kfr. diagrammet fra *Aftenposten* på forrige side). Det vanligste formatet er et liggende format, omtrent som standardformatet i denne publikasjonen; bredde : høyde = ca. 1 : 1,5.



Avstand mellom tidspunkter

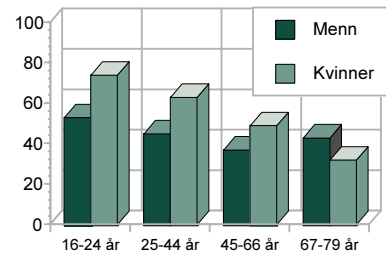
Når man presenterer tidsserier, så er det viktig at det er proporsjonal avstand mellom tidspunktene. Ofte har man i begynnelsen av en tidsserie bare tall for enkelte år, mens man på slutten har tall for hvert år, f.eks.: 1980, 1985, 1990, 1995 og 1996. Et diagram med lik avstand mellom disse fem datapunktene gir inntrykk av en jevn økning i perioden. Men når vi endrer avstanden mellom de første fire observasjonene, ser vi imidlertid at det har vært en markert økning fra 1995 til 1996.



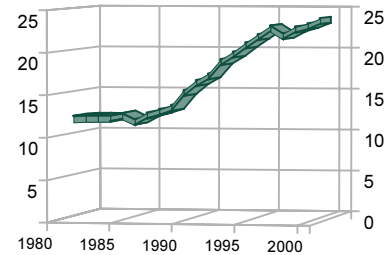
Den markerte økningen fra 1995 til 1996 blir helt borte i diagrammet til venstre, fordi avstanden mellom de to årene er den samme som mellom f.eks. 1990 og 1995

"Dybde"

Dagens dataprogrammer har etter hvert fått mulighet for å lage diagrammer med dybdevirkning. Dette betegnes ofte som 3-dimensjonalitet, men er det ikke. Effekten er en grafisk illusjon og er oftest bare egnet til å forvanske informasjonsinnholdet i grafikken: Den minsker lesbarheten fordi den gjør det vanskeligere å avlese søylenes verdi på y-aksen.



Og heller ikke linjediagrammer får noen økt informasjonsverdi ved å bruke dybde-effekter; snarere tvert imot.

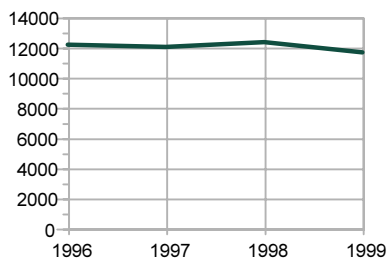


2.3. Hva slags diagram skal jeg lage?

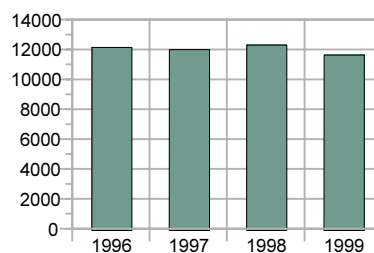
Nedenfor gjengis en relativt liten, enkel tabell som viser antallet skadde og drepte menn og kvinner i perioden 1996-1999, fordelt på fire ulike grupper av trafikanter.

Antallet skadde og drepte menn og kvinner, fordelt på trafikantgrupper								
	Bil		Motorsykkkel		Sykkel		Fotgjenger	
	M	K	M	K	M	K	M	K
1996	4 842	4 115	937	219	641	364	494	549
1997	4 984	3 860	990	211	636	334	500	507
1998	5 266	3 914	987	210	587	326	494	542
1999	5 007	3 727	1 002	185	549	272	446	474

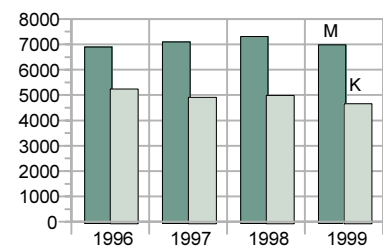
Med utgangspunkt i denne tabellen kan det nå lages en lang rekke ulike diagrammer. Her er et utvalg:



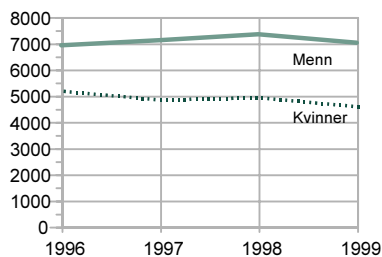
Enkelt linjediagram som viser det totale antallet skadde og drepte i perioden



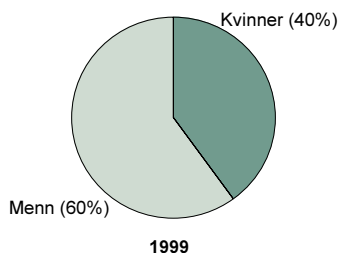
Enkelt søylediagram som viser det samme som linjediagrammet til venstre



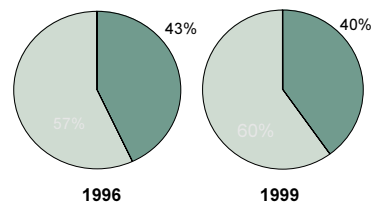
Gruppert søylediagram som viser antallet skadde og drepte menn/kvinner i perioden



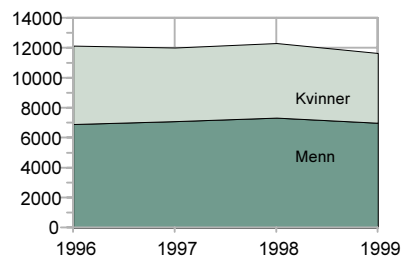
Linjediagram med tidsserier for menn og kvinner; viser det samme som det grupper-te søylediagrammet ovenfor



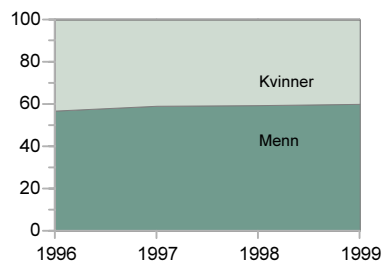
Sektordiagram som viser fordelingen på menn og kvinner et bestemt år



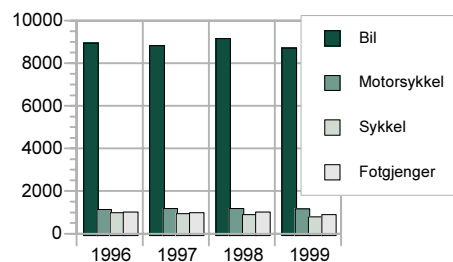
Sektordiagram som viser andelen kvinner i 1996 og 1999



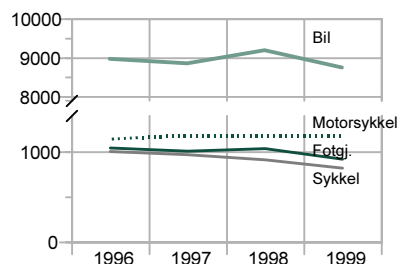
Arealdiagram som viser utviklingen for menn og kvinner, og totalt



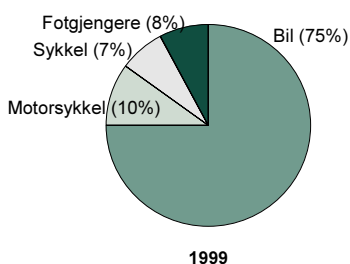
Arealdiagram som viser utviklingen i den relative fordelingen på menn og kvinner



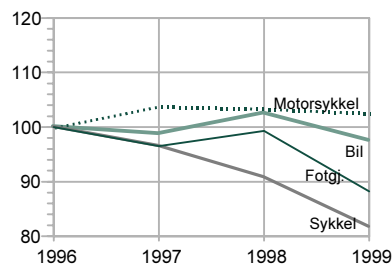
Gruppert søylediagram som viser antallet skadde og drepte i ulike trafikantgrupper



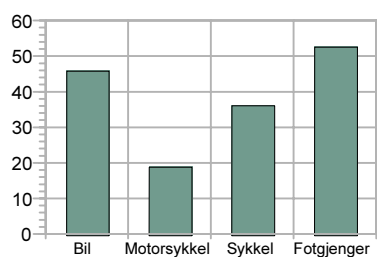
Linjediagram som viser utviklingen for ulike trafikantgrupper



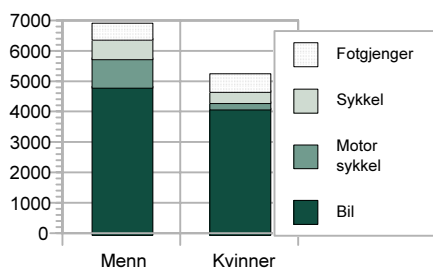
Sektordiagram som viser fordelingen på trafikantgrupper et bestemt år



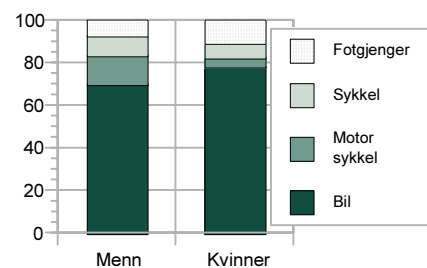
Linjediagram som viser den relative utviklingen for ulike trafikantgrupper. Indeks 1996 = 100



Søylediagram som viser andelen drepte og skadde kvinner i ulike trafikantgrupper



Stablede søyler; viser hvordan menn og kvinner er fordelt på ulike trafikantgrupper



Stablede søyler som viser det samme som diagrammet til venstre, men relativt

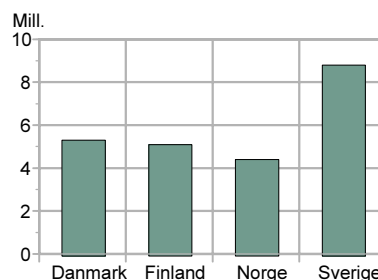
Som de mange eksemplene viser, er valget av diagramtype helt avhengig av hva man ønsker å vise. Bare sjelden er det gitt i selve datamaterialet hva slags diagram som er det "beste". Diagrammet bør ha et klart budskap og utformingen avhenger av hva dette budskapet er; hva man vil poengtere eller legge vekt på.

3. Søylediagrammer: Å sammenlikne grupper

Det vertikale søylediagrammet (også kalt stolpediagrammet) er på mange måter basisdiagrammet og utgangspunktet for de fleste andre typer diagrammer.

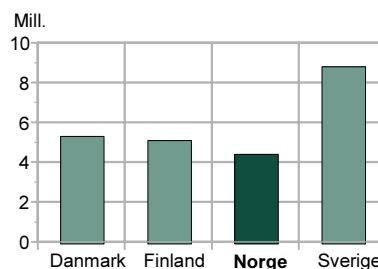
3.1. Enkle vertikale søylediagrammer

Enkle søylediagrammer inneholder alltid to variable og brukes for å sammenlikne f.eks. antall, andeler, frekvenser eller gjennomsnitt i ulike grupper; f.eks. andelen arbeidsledige i ulike aldersgrupper eller – som her – folkemengden i de nordiske land. Andelen arbeidsledige og folkemengde er her avhengige variabler, alder og land er uavhengige.

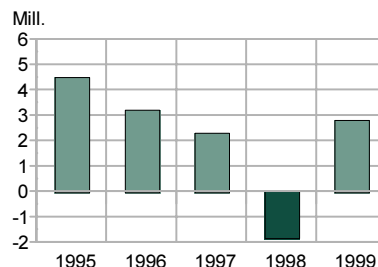


Vertikale søylediagrammer kan også brukes til å vise utviklingen over tid, f.eks. ved relativt korte tids-serier (mellom 2 og 10 tidspunkter) og særlig når tidsserien viser absolutte tall og ikke rater eller hyppigheter: F.eks. er det vanlig å vise *antallet* drepte i trafikken eller *antallet* fødte som søylediagram, mens rater (døde pr. 100 000 innbyggere eller fødselsrater) bør vises som linjediagram (forutsatt at man har mange nok perioder eller tidspunkter, dvs. mer enn 4-5). Tid (= år) er her å betrakte som en uavhengig variabel.

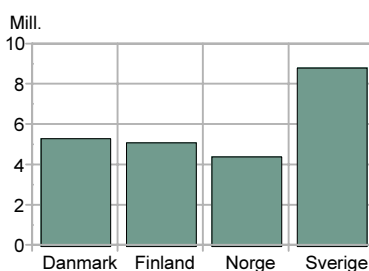
Når man som i dette eksemplet bare har én uavhengig variabel (land), er det vanlig å la søylene ha samme mønster eller raster (evt. farge). Om man ønsker å fremheve ett av landene (f.eks. Norge), kan dette gjøres ved å bruke et annet raster og/eller ved å bruke uthevet skrift.



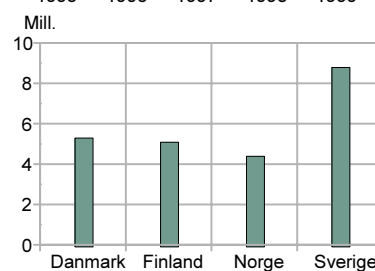
Hvis diagrammet inneholder både positive og negative verdier (for eksempel overskudd/underskudd, vekst/nedgang), bør man bruke forskjellige raster/mønster /farger for å skille mellom dem.



Når det gjelder den grafiske utformingen, bør søylene i søylediagrammer være omtrent like brede som mellomrommet mellom dem; de bør altså verken være for brede eller for smale.

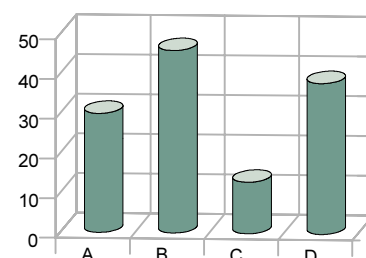
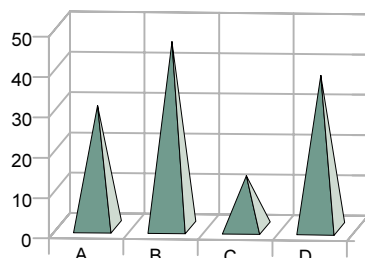


Her er søylene for brede...



...og her er de for smale

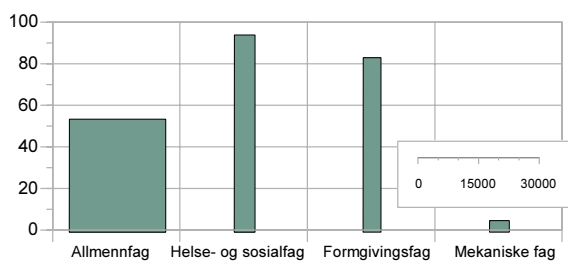
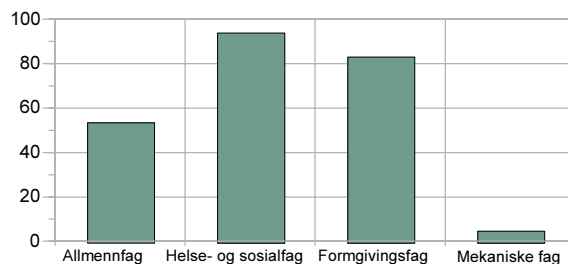
De ulike grafikkprogrammene tilbyr ofte alternative utforminger av søylene. Til høyre vises to muligheter: Pyramider og runde søyler. Disse – og liknende utforminger – bør bare brukes med stor forsiktighet (om i det hele tatt). Fordi søylene



inneholder en tredimensjonal effekt, blir det vanskelig å lese av verdiene på aksene. Gruppe A i de to diagrammene viser egentlig verdien 30, men ser mer ut som 27-28 (på grunn av perspektivet). Og pyramidene til venstre bør etter min mening bare brukes hvis man skal illustrerer høyden på fjelltopper i Rondane.

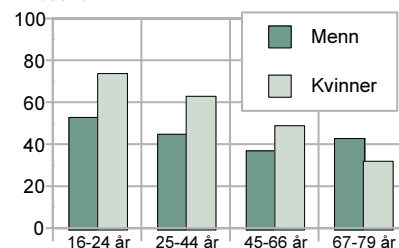
En spesiell type søylediagrammer er diagrammer hvor også bredden på søylene har betydning. For eksempel: Hvis vi sammenlikner andelen jenter på ulike studieretninger i videregående skole, vil et "vanlig" søylediagram se ut omtrent som det til høyre.

Men siden antallet elever er svært forskjellig på de ulike studieretningene, ønsker vi kanskje også å vise dette. Dette gjøres ved å la søylenes bredde reflektere antallet elever (allmennfag har omtrent fem ganger så mange elever som de andre studieretningene). En egen forklaring angir hvor mye bredden på søylene tilsvarer (allmennfag = ca. 24 000 elever).

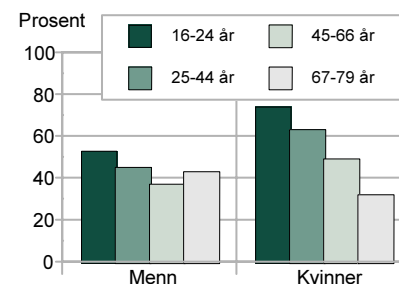


3.2. Grupperte søylediagrammer:

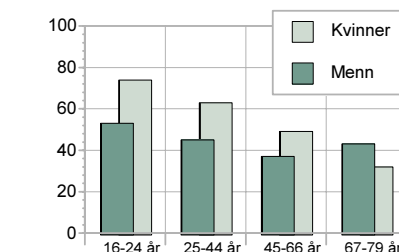
Mange søylediagrammer inneholder bare to variable; én uavhengig og én avhengig (som for eksempel land og folkemengde eller alder og arbeidstid). Men svært ofte ønsker vi å sammenlikne betydningen av og samspillet mellom to (eller flere) faktorer, f.eks. både kjønn og alder. Eksempelet til høyre viser andelen biblioteksbrukere siste år blant menn og kvinner i ulike aldersgrupper.



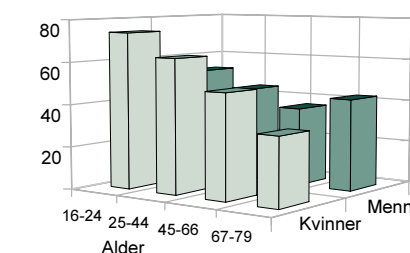
Med to uavhengige variable, kan disse grupperes på to måter, avhengig av hva vi fokuserer på; hvilken sammenlikning som er viktigst å få fram: Vi sammenlikner først og fremst de søylene som er gruppert sammen (uten mellomrom). Eksemplet øverst inviterer til en sammenlikning av menn og kvinner i ulike aldersgrupper: "Forskjellen mellom menn og kvinner er størst blant de yngste og avtar med alderen".



Det nederste grupperingen – derimot – inviterer til å sammenlikne aldersgruppene separat for menn og kvinner: "Biblioteksbruken reduseres med økende alder, men mer for kvinner enn for menn".



En annen måte å plassere/gruppere søylene på, er å la dem "overlappe". I enkelte grafikkprogrammer kan søylene plasseres halvveis bak hverandre, i stedet for ved siden av hverandre. Det vanlige er da å sette søylene med lavest verdier foran, slik at "høye" søyler ikke skygger for de mindre (men dette er selvfølgelig ikke alltid mulig).



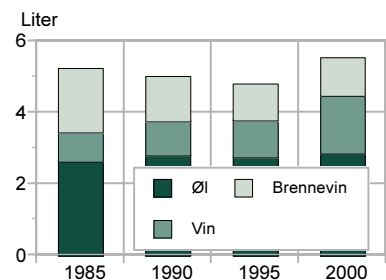
Søylediagrammer med to uavhengige variable kan også presenteres tredimensjonalt; dvs. med en egen akse (z-akse) for den tredje variabelen. De fleste grafikkprogrammer har nå etterhvert en slik mulighet. Denne muligheten anbefales ikke og bør i alle fall bare brukes med stor forsiktighet. Legg f.eks. merke til at høye søyler i forgrunnen lett kan skygge for lavere søyler bak.

3.3. Sammensatte søylediagrammer

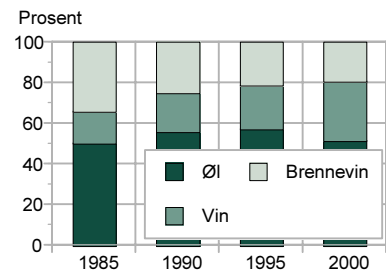
Sammensatte søylediagrammer brukes for å vise hvordan én størrelse er sammensatt av eller fordelt på ulike enkeltkomponenter. For eksempel ønsker vi å sammenlikne ikke bare hvordan det totale antallet asylsøkere har utviklet seg over tid, men om og hvordan sammensetningen av gruppen eventuelt har endret seg. Eller vi ser på alkoholkonsumets utvikling over tid, både totalt og fordelt på øl, vin og brennevin.

Rekkefølgen på kategoriene kan bestemmes på forskjellige måter: Noen ganger finnes det kanskje en "naturlig" rekkefølge (f.eks. utdanningsnivå, alder). Andre ganger kan rekkefølgen bestemmes på grunnlag av størrelse eller "viktighet" (de største eller viktigste nederst). For å øke lesbarheten, legges ofte den komponenten som varierer minst i bunnen av søylene. Og som ellers, gjelder også her regelen om at mange verdier/ komponenter reduserer lesbarheten.

I eksemplet til høyre vises alkoholkonsumet per innbygger i absolutte tall (alkohol-liter). Siden både totalen og fordelingen kan endre seg over tid, blir det mange ganger vanskelig å se for eksempel hvilke kategorier som endrer seg mest. For å lette sammenlikningen, kan man se på den relative fordelingen; dvs. la totalen være 100 prosent.

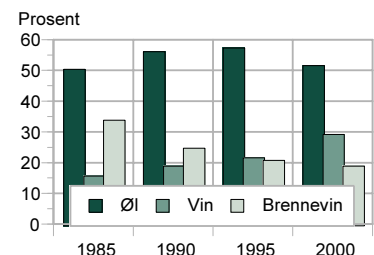


De enkelte søylene i et sammensatt 100% søylediagram viser altså en relativ fordeling og kan således sammenliknes med et sektordiagram. Men fordi søylediagrammet har en Y-akse, er det langt lettere å sammenlikne to eller flere søyler enn det er å sammenlikne fordelingene i sektordiagrammer. Dette er grunnen til at man sjelden bruker sektordiagrammer for å sammenlikne fordelinger i flere grupper eller på flere tidspunkter.

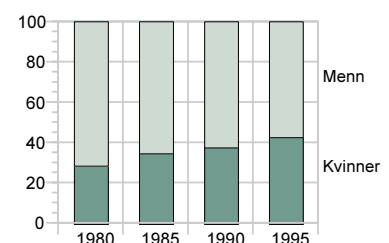


Som ellers er det viktig å tilstrebe enkelhet i presentasjonen; i denne sammenhengen innebærer det at man ikke bør benytte fordelinger med mer enn 4-5 verdier eller grupper.

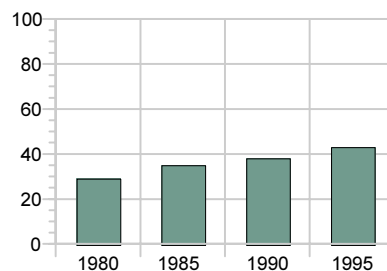
Noen ganger ser man fordelinger presentert side om side som om det var andeler, slik som til høyre. Dette gjøres for lettere å kunne sammenlikne gruppene, ved at alle har et felles nullpunkt. Men når det er snakk om hele fordelinger som summerer til 100 prosent, er stablede søyler å foretrekke.



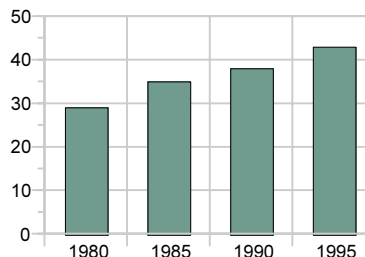
En spesiell situasjon er når et sammensatt 100-prosent søylediagram viser en fordeling med bare to verdier/grupper, for eksempel utviklingen i andelen mannlige og kvinner studenter i en bestemt utdanning. Dette kan da presenteres slik:



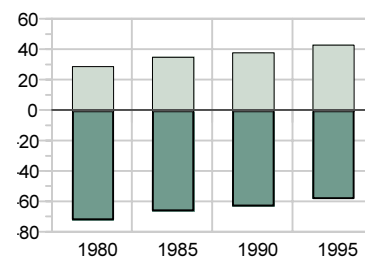
Men siden en slik fordeling alltid vil summere til 100 prosent, er det her vanlig å sløyfe den ene verdien eller gruppen og for eksempel bare presentere andelen kvinner:



I slike "andelsdiagrammer" er det imidlertid ikke nødvendig å la verdiaksen gå til 100 prosent. Når maksimumsverdien – som her – er 43 prosent, er det naturlig å la verdiaksen gå til 50 prosent.



En tredje måte å presentere det samme på, er ved hjelp av et såkalt "flytende" søylediagram. (For at verdiaksene skal få en "symmetriske" inndeling, lar vi her kvinnenens akse gå til 60 prosent, siden mennenes går til 80).



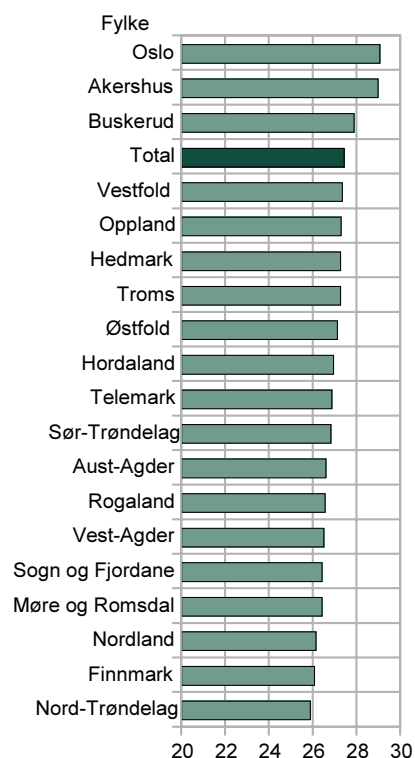
3.4. Horisontale søylediagrammer

Horisontale søylediagrammer har mange av de samme funksjonene som de vertikale; disse viser også antall, andeler, frekvenser eller gjennomsnitt i ulike grupper. Til forskjell fra vertikale søyle-diagrammer, brukes imidlertid de horisontale *sjelden* til foreta sammenlikninger over tid (unntatt endringer fra ett tidspunkt til et annet).

Det er særlig to situasjoner hvor horisontale søylediagrammer egner seg godt: Den ene er hvis variabelverdiene på den uavhengige variabelen har lange betegnelser. Det er da mye lettere å få plass til (og ikke minst å lese) betegnelsene når teksten står horisontalt. For å øke lesbarheten, bør teksten her høyrejusteres.

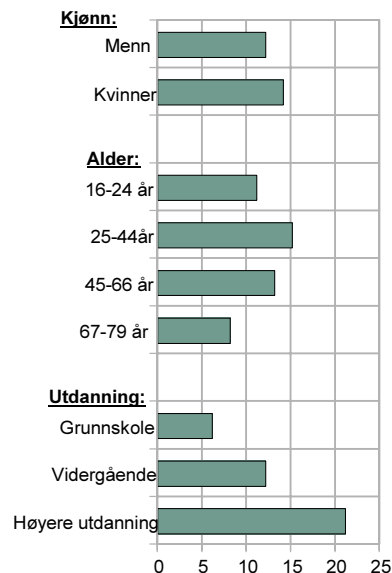
En annen situasjon er når vi har mange variabelverdier, f.eks. fylker. Slike "lange" diagrammer lar seg vanskelig plassere den andre veien ("i bredden").

Når enhetene ikke har noen naturlig rekkefølge (f.eks. alder eller utdanningslengde) kan man sortere (rangere) dem etter verdien på den avhengige variabelen. Her ser man da lettere hvilke fylker som har høyest/lavest verdi.

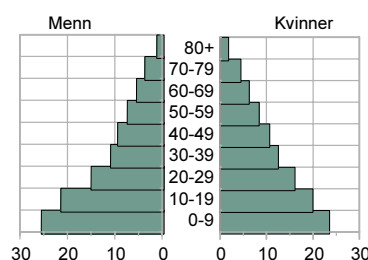


Legg også merke til at det her er vanlig å ta med landsgjennomsnittet, noe som er mer sjelden i stående søylediagrammer. Og på samme måte som vertikale søylediagrammer, kan også horisontale være gruppert eller sammensatt.

En tredje situasjon hvor det horisontale søylediagrammet egner seg bedre enn det vertikale, er når vi ønsker å vise hvordan én avhengig variabel varierer i mange, ulike grupper. For eksempel hvordan andelen som reiser på påskeferie varierer *både* med kjønn, alder og utdanning.



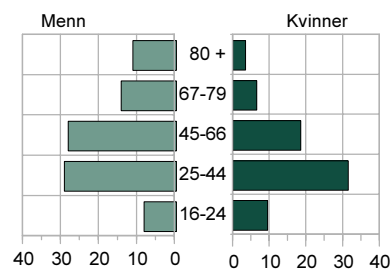
En spesiell type horisontalt søylediagram er *befolkningspyramiden*. Dette er i utgangspunktet to horisontale søylediagrammer som er satt sammen, "rygg mot rygg". Legg også merke til at – til forskjell fra "vanlige" søylediagrammer – har befolkningspyramider vanligvis ikke noe mellomrom mellom søylene, som ligger "vegg i vegg". Befolkningspyramider kan være "enkle", dvs. at de inneholder bare to variable, vanligvis kjønn og alder. Men de kan også inneholde en tredje variabel, f.eks. ekteskapelig status.



Hver av de to seriene her er egentlig en fordeling av en kontinuerlig variabel (alder), og diagrammet er derfor et slags histogram, derfor brukes det vanligvis ikke mellomrom mellom søylene i befolkningspyramider.

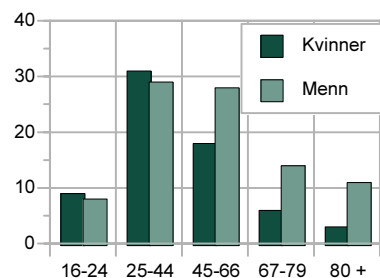
Ingen av de vanligste grafikkprogrammene har imidlertid befolkningspyramider som en egen diagramtype. Dette betyr litt "håndarbeid": Man må lage to separate diagrammer (for menn og kvinner); deretter må det ene speilvendes og de to settes sammen.

Denne typen "rygg mot rygg" -diagrammer kan selvfølgelig også brukes for andre formål, særlig finner vi dem ofte brukt i sammenlikning av menn og kvinner. Et eksempel kan være andelen menn og kvinner i ulike aldersgrupper med utdanning på universitets- og høyskolenivå. (Her må alltid høyre og venstre verdiakse ha samme skala, dvs. samme maksimum).



Men ofte er et (vertikalt) gruppert søylediagram likevel å foretrekke i denne situasjonene, fordi det er langt lettere å sammenlikne f.eks. menn og kvinner i hver aldersgruppe, når søylene står ved siden av hverandre enn når de står "rygg mot rygg".

"Rygg-mot-rygg"- diagrammer kan også brukes hvis man vil illustrere ulike aspekter (måleenheter) ved et fenomen, som mengde og verdi, f.eks. fangstmengde og fangstverdi av ulike fiskeslag (målt i henholdsvis tonn og kroner).



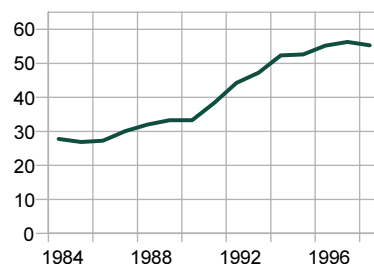
4. Linjediagram: Tid og mengde

Linjediagram (også kalt kurvediagram) brukes først og fremst for å vise utvikling over tid og dette er en svært vanlig diagramtype, ikke minst på avisenes økonomisider. Nå kan selvfølgelig også tidsserier vises som søylediagrammer, særlig hvis antallet tidspunkter er lite: En tommelfingerregel kan være at linjediagrammer bare skal benyttes når man har fire eller flere observasjoner (datapunkter); har man data bare for to eller tre tidspunkter/perioder, bør det benyttes (stående) søylediagrammer. En annen situasjon hvor søylediagrammer kan være å foretrekke, er når avstanden mellom datapunktene er svært uregelmessig; for eksempel 1980, 1987, 1991, 1996 og 1998.

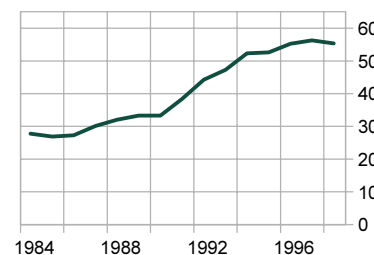
Det finnes to hovedtyper tidsseriedata: Data som gjelder *perioder* (for eksempel år, kvartal, måneder, dager eller også timer/minutter) og data som gjelder bestemte *tidspunkter* (for eksempel 1/1 hvert år).

4.1. Enkle linjediagrammer

Mange linjediagrammer inneholder bare én linje/tidsserie, som f.eks. arbeidsløsheten i en periode, skilsmissehyppigheten, fødselsrater, renteutviklingen eller barnehagedekningen. Linjen tegnes da vanligvis som en enkel linje (men likevel klart tykkere enn eventuelle hjelpelinjer).

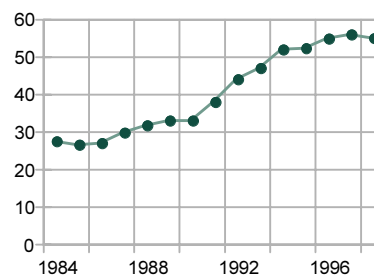


Tradisjonelt blir verdiaksen plassert på venstre side (i de fleste dataprogrammer er dette standard eller "default"). Men i en tidsserie foregår ofte den mest "interessante" utviklingen på høyre side, derfor kan verdiaksen ofte med fordel flyttes dit.

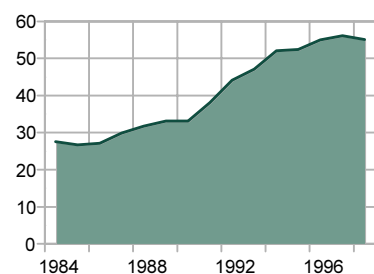


I linjediagrammer plasseres alltid tiden på den horisontale aksen og alltid fra venstre mot høyre.

Datapunktene i en tidsserie markeres ofte med symboler eller såkalte "markører" (f.eks.: □ ○ ● ▲ ◆). Men med bare én dataserie, er det ikke noe behov for å skille mellom flere linjer og bruken av symboler er derfor unødvendig her.



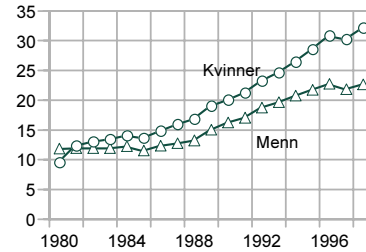
Alternativt kan man fylle/skravere arealet under linjen: Denne typen kalles ofte silhuett-diagrammer og er en form for arealdiagram (mer om disse senere).



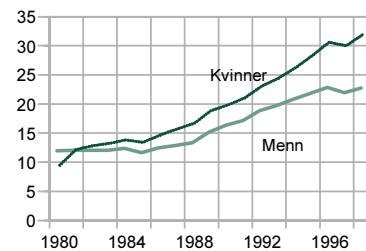
4.2. Multiple linjediagrammer

Svært ofte ønsker man å sammenlikne utviklingen over tid for ulike grupper (For eksempel menn og kvinner, ulike aldersgrupper eller flere land). En tommelfingerregel er at et linjediagram bare sjelden bør inneholde mer enn 5-6 tidsserier. Dette antallet er selvfølgelig avhengig både av størrelsen på diagrammet, om linjene ligger svært nær hverandre og om de krysser hverandre.

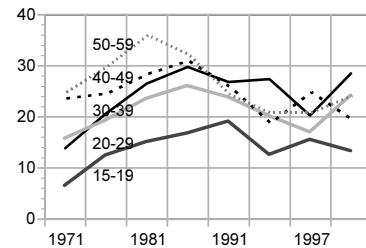
Diagrammene til høyre viser prosentandelen menn og kvinner i alderen 19-24 år som er i høyere utdanning. I det øverste eksemplet er det brukt ulike symboler for å skille mellom de to linjene; menn og kvinner. Men med mange linjer og/eller (som her) mange datapunkter, bidrar bruken av symboler til å overlesse diagrammet. Og heller ikke med bare to linjer er det noe behov for å bruke markører. Snarere bidrar markørene til markert mindre lesbarhet.



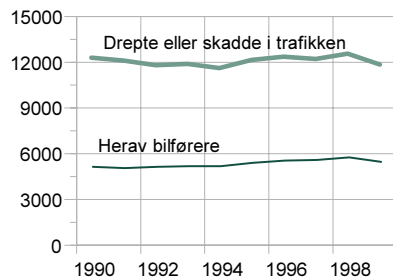
I stedet kan man bruke ulike linjetyper, som her. Alternativt kan man bruke gråtoner (eller farger) i tillegg til sort.



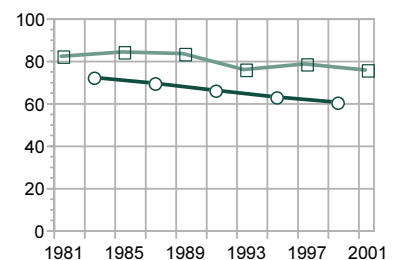
Linjediagrammer bør ikke inneholde for mange dataserier eller linjer; 4-5 serier regnes ofte for å være maksimalt. Avhengig, selvfølgelig, av avstanden mellom linjene og om de er parallelle eller om de krysser hverandre. Eksemplet til høyre (selvmordshyppighet i utvalgte aldersgrupper) inneholder bare fem serier, men er likevel på grensen til det uleselige.



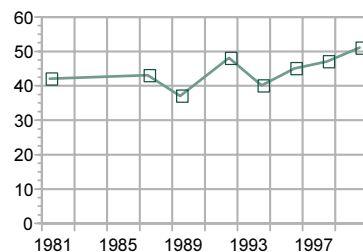
Vanligvis bør linjene ha samme tykkelse, noe som signaliserer at gruppene (f.eks. menn og kvinner) er likeverdige. Ulike linjetykkelser bør reserveres for spesielle situasjoner, for eksempel når en av tidsseriene viser en total og en annen viser antallet i en undergruppe. Eksemplet til høyre viser det totale antallet drepte eller skadde i trafikken hvert år og antallet bilførere. Også når linjer ligger svært nær/krysser hverandre, kan man med fordel bruke ulik linjetykkelse. Et tredje tilfelle er når en av linjene viser gjennomsnittet for en rekke undergrupper.



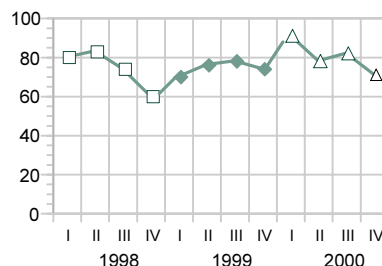
I noen situasjoner bør likevel symboler brukes. For eksempel: I situasjoner hvor man ikke har data for hvert enkelt år, bidrar symbolene til å markere hvilke år man har data for. I et diagram som viser valgdeltakelsen både ved kommune- og Stortingsvalgene er dette viktig, siden de to seriene jo har data som refererer til forskjellige år.



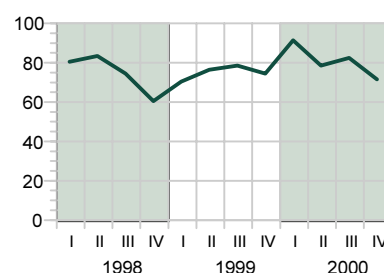
En annen situasjon der markører har en funksjon, er når det i en tidsserie er ulik avstand mellom tidspunktene eller periodene. I diagrammet til høyre markerer symbolene klart hvilke år man har data for. Dermed blir det også tydelig at den tilsynelatende lange, stabile perioden på 1980-tallet kan henge sammen med manglende data, og ikke nødvendigvis er uttrykk for at utviklingen var så stabil.



En tredje situasjon hvor en med fordel kan bruke symboler, er for eksempel ved presentasjon av kvartalsstatistikk. Her kan man bruke ulike symboler for de forskjellige årgangene, slik:

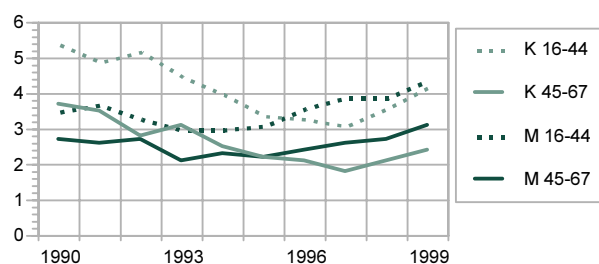


Alternativt kunne de ulike årgangene/periodene her markeres med en egen bakgrunn, for eksempel ved at annenhver årgang gis en grå bakgrunn, slik:

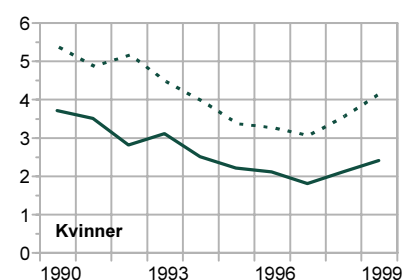
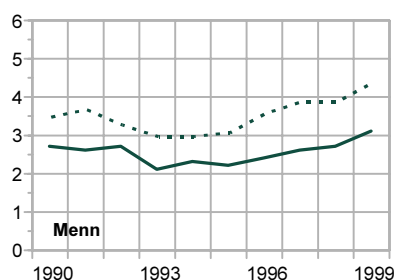


4.3. Gruppert linjediagram

Noen ganger inneholder et linjediagram tre såkalte uavhengige variable; i tillegg til tidsdimensjonen kan dette f.eks. være kjønn og alder. For å lette sammenlikningen er det da viktig å "gruppere" linjene (kfr. grupperte søylediagrammer). Dette kan skje slik som til høyre, ved at mennenes kurver er i sort og kvinnenes i grått, samtidig som de yngste skiller fra de eldste ved å bruke henholdsvis prikkede og heltrukne linjer.



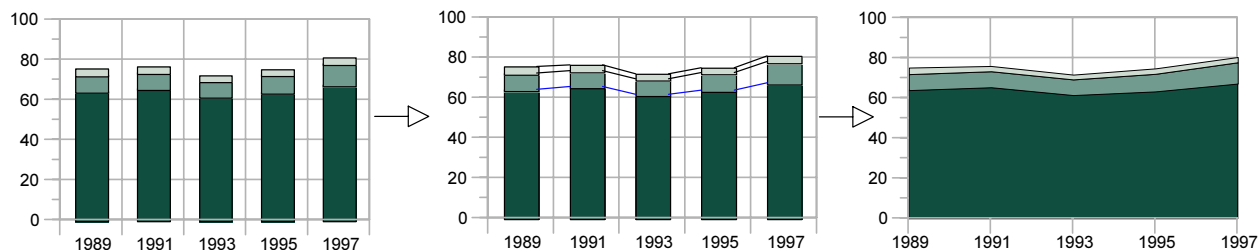
Men det å måtte tolke to slike sett med "koder" er ikke direkte brukervennlig. En bedre løsning er ofte å lage separate diagrammer; som her for menn og kvinner. Disse bør da plasseres ved siden av hverandre, snarere enn over/under hverandre, da blir det lettere å sammenlikne nivåene



4.4. Arealdiagram

Arealdiagram tenkes vanligvis på som en form for linjediagram, hvor linjene er lagt "oppå hverandre" og arealet mellom dem er skravert (det kalles derfor også av og til "akkumulert linjediagram"). Men vi kan like gjerne se på arealdiagrammet som en variant av sammensatte søylediagram. I eksemplet

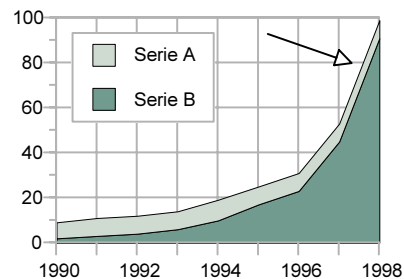
nedenfor blir likheten tydelig: På samme måte som et sammensatt søylediagram, viser arealdiagrammet *både* det totale antallet/mengden *og* hvordan denne totalen er sammensatt.



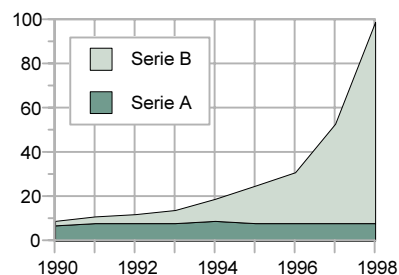
Når arealdiagrammet ofte betraktes som en form for linjediagram, har nok dette sammenheng med at det (på samme måte som linjediagrammet) hovedsakelig brukes for å illustrere utviklingen over tid.

På samme måte som stablede søylediagrammer, finnes arealdiagrammet i to utgaver, som viser henholdsvis absolutt og relativ fordeling (= 100 prosent).

Når man utformer arealdiagrammer, er rekkefølgen av tidsseriene ofte av stor betydning for lesbarheten og fortolkningen. I det øverste diagrammet til høyre får man lett det inntrykket at den øverste serien (serie A) er synkende.



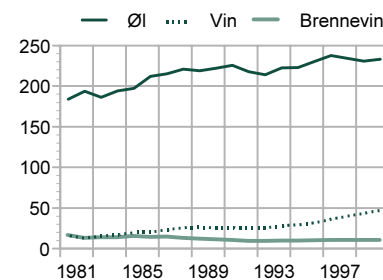
Men diagrammet nedenfor viser at dette ikke er tilfellet. En generell regel kan derfor være at den tidsserien som varierer minst, bør ligge i "bunn" av diagrammet.



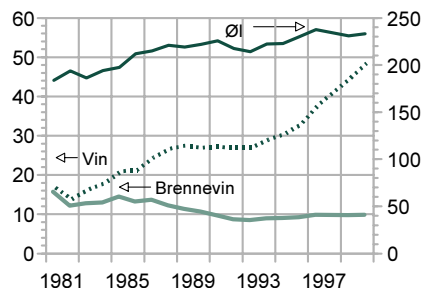
Et annet tilfelle hvor rekkefølgen er viktig, er når en eller flere av seriene varierer mye. Serien med minst variasjon bør da plasseres i bunn.

4.5. To verdiakser

Enkelte ganger ønsker man å sammenlikne utviklingen i fenomener som befinner seg på svært ulike nivåer på måleskalaen. La oss anta at vi ønsker å sammenlike forbruket av øl, vin og brennevin. Et diagram som viser utviklingen (1 000 vareliter) vil da se omtrent slik ut:



Men siden ølvolumet er langt høyere enn vin og brennevin, blir det vanskelig å se tydelig detaljene i utviklingen av vin- og brennevinforbruket. En løsning på dette er å operere med to skalaer; én for øl og én for vin og brennevin, som vist i diagrammet nedenfor. Økningen i vinkonsumet og nedgangen for brennevin blir nå langt tydeligere. Men samtidig oppstår faren for misforståelser, ved at leseren kan komme til å lese av verdien på feil akse. Det kan derfor være grunn til å bruke f.eks. piler for å gjøre det helt tydelig hvilken akse som måler hva.

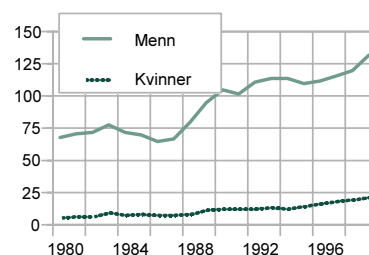


To akser kan også være formålstjenlig når man ønsker vise den samme utviklingen på to forskjellige måter/i forskjellige enheter; f.eks. både i kroner og dollar eller kilometer og miles.

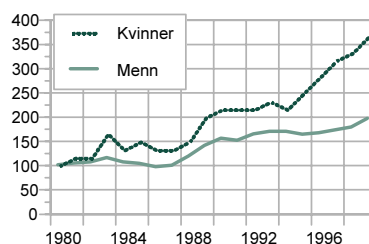
En annen situasjon hvor to skalaer er helt nødvendig, er hvis man sammenlikner "epler" og "pærer"; dvs. variabler med ulike måleenheter. For eksempel om vi sammenlikner utviklingen i vareimporten både når det gjelder verdi og mengde, f.eks. kroner og tonn. En annen måte å løse slike problemer på, er å bruke et såkalt indeksdiagram.

4.6. Indeksdiagram

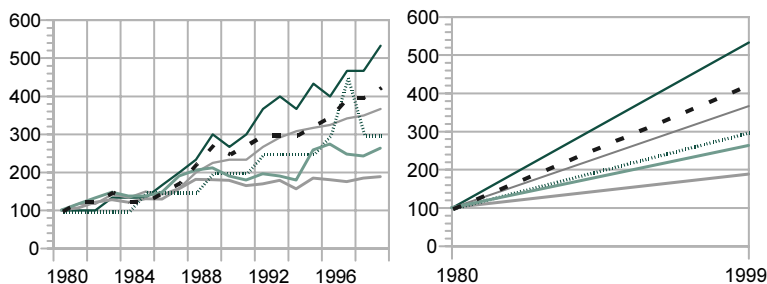
Et indeksdiagram viser relative verdier (vanligvis i prosent) i forhold til en basisverdi et bestemt år. Denne basisverdien settes vanligvis til 100. Formålet er å lettere kunne sammenlikne utviklingen i to eller flere tidsserier, hvor enten nivået er svært forskjellig eller som måles på ulike skalaer. Ta for eksempel to tidsserier som viser antall siktede for forbrytelser per 1 000 innbyggere for menn og kvinner. Vi ser da at begge ratene har økt siden 1980, men det er – fordi nivåene er så forskjellige – svært vanskelig for eksempel å si noe om økningen er størst for menn eller kvinner.



Alternativet er da å lage et indeksdiagram, hvor begge serienes verdier i 1980 settes lik 100. Det blir nå helt tydelig at antallet kvinnelige siktede har økt langt mer (relativt sett) enn antallet mannlige. (Legg også merke til at rekkefølgen av *Menn* og *Kvinner* i forklaringen nå er byttet om, og svarer til rekkefølgen av kurvene). Problemet med indeksdiagrammer er at leseren lett kan forveksle endringstakt og nivå: I diagrammet til høyre kan man komme til å tro at kvinner begår flere forbrytelser enn menn.



En spesiell variant av indeksdiagrammet kalles viftediagram (på grunn av sin form). Dette er en forenklet versjon som bare består av to tidspunkter. Dermed kan man langt lettere vise utviklingen for flere tidsserier. Til høyre vises utviklingen i etterforskede forbrytelser, fordelt på seks ulike typer kriminalitet. Med hele tidsserien (19 tidspunkter), blir resultatet vanskelig å lese, bl.a. fordi mange av linjene krysser hverandre. Med bare to tidspunkter, blir det langt mer oversiktlig.

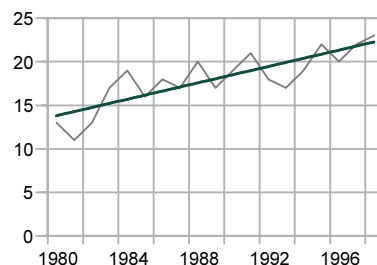
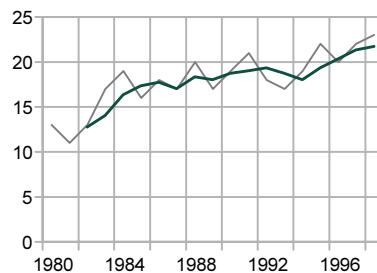


4.7. Glattung: Glidende gjennomsnitt og trendlinjer

Mange tidsserier likner en "berg- og dalbane" fordi de ofte inneholder tilfeldige variasjoner. Dette kan bl.a. være fordi de er basert på (små) utvalg. Tidsserien blir da gjerne "hakkete" og det kan være vanskelig å se om det egentlig er noen tendens i utviklingen (kfr. den tykke linjen i diagrammet til høyre neste side). En vanlig løsning på dette problemet er å justere tallene, f.eks. med glidende gjennomsnitt (ved f.eks. tre-års glidende gjennomsnitt brukes gjennomsnittet for de tre siste år). Dermed fjernes noe av den tilfeldige variasjonen og tendensen blir tydeligere.

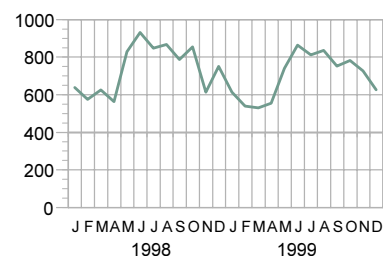
En annen måte å tydeliggjøre tendensen i et tidsforløp på, er å legge inn en såkalt "trendlinje". Dette er – grovt sagt – en måte å indikere den generelle tendensen i data på. Til høyre vises en linjær trendlinje (linjær regresjon). Også glidende gjennomsnitt og andre typer justeringer benevnes ofte trendlinjer.

Trendlinjer benyttes ofte til å projisere eller lage en prognose for fremtidig utvikling. Dette må imidlertid ikke gjøres ukritisk; de fleste utviklingstrender snur før eller senere ("What goes up, must come down").

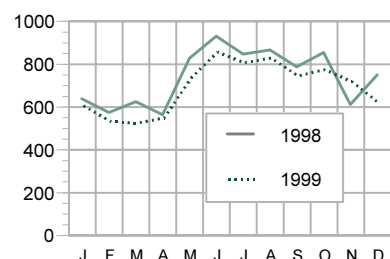


4.8. Delårsdata

Diagrammet ovenfor (med kvartalsdata) er et eksempel på en spesiell type tidsserier, nemlig såkalte delårsdata eller data med et syklisk mønster. Dette kan være daglig, ukentlig, månedlig eller halvårlig statistikk. Et typisk eksempel er den månedlige trafikkulykkesstatistikken. På samme måte som kvartalstallene, kan også disse presenteres som én fortløpende linje, slik som i diagrammet til høyre.



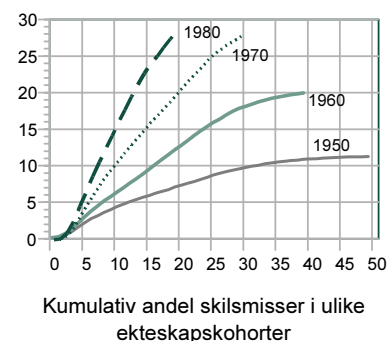
Men slike data ofte inneholder store sesongmessige variasjoner, og den relevante sammenlikningsperioden er da ofte ikke forrige måned, men samme måned året før. En slik sammenlikning er ikke lett å gjøre her. Alternativet er da å presentere seriene slik:



4.9. Kumulative linjediagrammer

For mange tidsserier vil det være interessant å ikke bare se på forekomsten av et fenomen hver måned eller hvert år, men også å se på det kumulative resultatet. Ett eksempel: De årlige skilsmissetallene kan summeres (kumuleres) for de enkelte ekteskapskohorter (årganger). Deretter kan man beregne den totale andelen skilsmisser i de ulike kohortene.

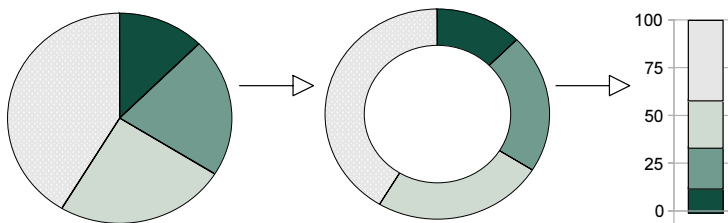
Tilsvarende kan man for eksempel vise andelen kvinner som har fått sitt første barn ved ulike aldre eller andelen unge som har flyttet hjemmefra ved ulike aldre.



5. Sektordiagram: Prosentfordeling

5.1. Én enkelt fordeling

Sektordiagram (også kalt sirkel- eller kakediagram) brukes for å vise relative fordelinger for kvalitative variable. Til forskjell fra andre diagrammer, har sektordiagrammet ingen akser. Men om vi sammenlikner det med et sammensatt søylediagram, ser vi at dette bare er to ulike måter å framstille en fordeling på: Når vi "ruller ut" sirkelen, blir omkretsen (360 grader) gjort om til en 100 prosent akse.



Fra sektor- til søylediagram

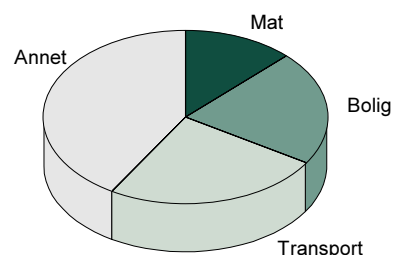
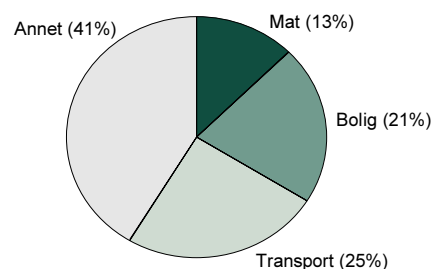
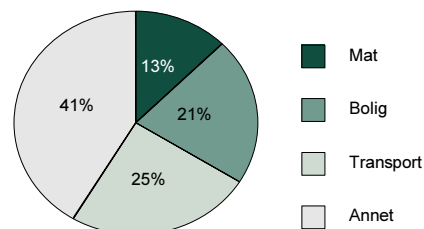
Sektordiagrammet viser altså en *relative* fordeling; hvordan en helhet (som alltid er 100 prosent) fordeles på undergrupper.

Eksempler kan være:

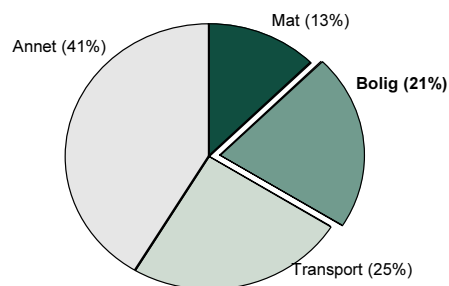
- hvordan arbeidsreiser fordeler seg på ulike transportmidler
- hvordan husholdningenes utgifter fordeler seg på ulike typer forbruk
- hvordan feriereiser fordeler seg på ferietyper, eller
- hvordan alkoholkonsumet fordeler seg på øl, vin og brennevin

Som ellers er det vanlig å la mønsterene gå fra mørkt til lyst. Siden denne diagramtypen mangler verdiakse, er den best egnet til å gi et grovt bilde av fordelingen. Men det er også mulig å oppgi verdiene for de ulike andelene. I stedet for å bruke en egen boks med forklaringer, anbefales det å knytte betegnelsene direkte til de ulike sektorene, enten utenfor sirkelen eller inne i de enkelte sektorene.

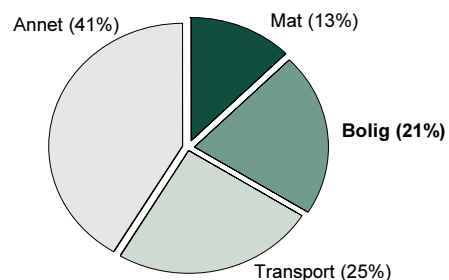
Også sektordiagrammer kan gis "dybde" eller 3-dimensjonalitet. Men fordi det segmentet som ligger foran dermed blir tykkere, kan det lett også få en økt betydning. Dette anbefales derfor ikke.



Som for de fleste andre diagrammer gjelder også her regelen om ikke å bruke for mange kategorier eller grupper (maksimum 5-6). Hvis vi ønske å fremheve eller fokusere på en av gruppene, kan dette gjøres ved å trekke ut sektoren, evt. å utheve betegnelsen.

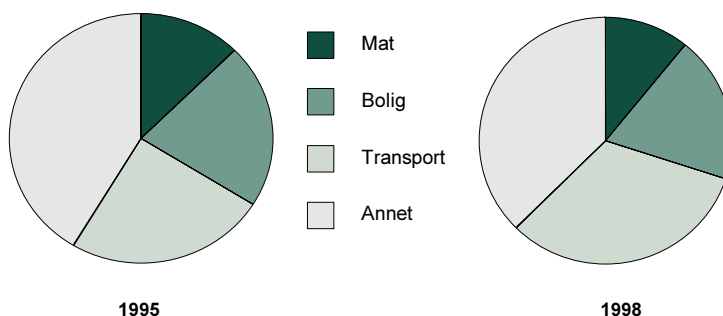


Å "eksplodere" hele diagrammet virker mot sin hensikt og er mer forvirrende enn fokuserende. Virkningen er omtrent den samme som å skrive et helt notat med uthevet skrift.

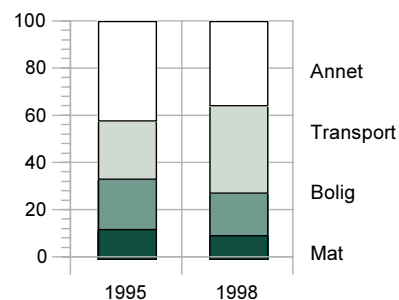


5.2. Sammenlikne fordelinger?

Noen ganger ønsker vi å sammenlikne to eller flere fordelinger; for ulike grupper eller tidspunkter. Dette lar seg selvfølgelig gjøre ved å plassere to sektordiagrammer ved siden av hverandre, slik:

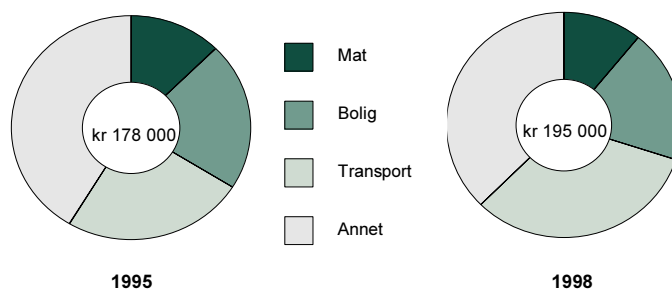


Men vanligvis er det da mer hensiktsmessig å lage et sammensatt søylediagram. Dette har en verdiakse, noe som gjør det langt lettere å sammenlikne gruppene; særlig hvis det er snakk om mer enn to. Og dette gjelder spesielt når forskjellene – som her – er relativt små.

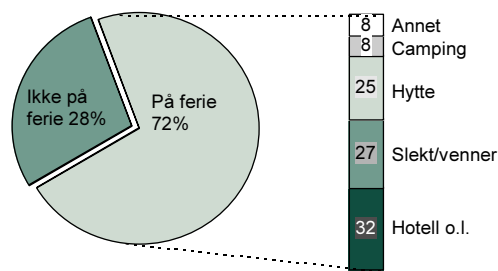


5.3. Varianter

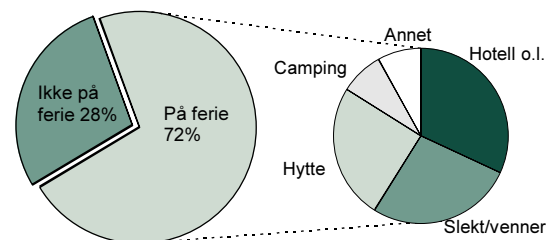
En variant av sektordiagrammet kalles "smult-ring", fordi den har et hull i midten. Dette kan brukes til å gi tilleggsopplysninger, for eksempel totaltall (her: total forbruksutgift).



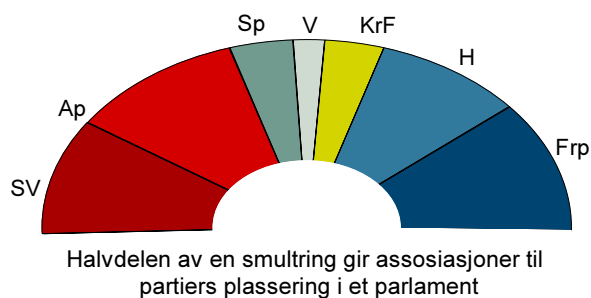
Noen ganger ønsker vi å fokusere spesielt på én av gruppene i et sektordiagram; og mer enn bare ved å trekke ut gruppen. Vi vil kanskje også vise hvordan denne gruppen igjen fordeler seg på ulike undergrupper; for eksempel hvordan de som har vært på ferie, har overnattet. Dette kan gjøres ved å vise fordelingen for undergruppen som en egen søyle, på denne måten:



En annen løsning er å også vise denne fordelingen som et sektordiagram. Men det andre sektordiagrammet bør da være tilsvarende mindre, for å illustrere at dette er fordelingen for en undergruppe.



En spesiell variant av smultringen brukes ofte når man skal illustrere fordelingen av stemmer på – eller mandatfordelingen for – ulike politiske partier. (Dette er forøvrig et eksempel på at bruk av flere farger kan ha en funksjon, siden disse fargene faktisk også har en viss forankring i "virkeligheten" (jf. diskusjonen om bruk av farger i kap. 7)).



6. Andre typer diagrammer

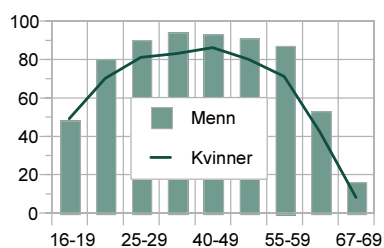
6.1. Kombinasjonsdiagrammer

I forrige kapittel så vi hvordan et sektordiagram kunne kombineres med en søyle, for å vise fordelingen for en undergruppe. Det å kombinere diagramtyper er svært vanlig og vi skal her se på noen av de mest typiske sammensetningene.

Søyle - linje

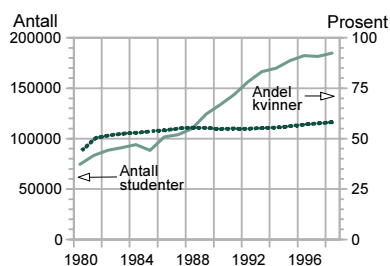
Den aller vanligste kombinasjonen er kanskje bruken av *både* linjer og søyler i samme diagram. Det kan være flere grunner til at man ikke alltid bruker bare linjer eller bare søyler. For eksempel ønsker man av og til å sammenlikne grupper eller kategorier som ikke vanligvis ikke sammenliknes ("epler og pærer", kfr. diskusjonen om to verdiakser i kap. 4).

En annen situasjon er når én gruppe er en slags "referanse" for en annen. Et eksempel kan være kvinners yrkesdeltakelse i ulike aldersgrupper, presentert som et linjediagram. Svært ofte ønsker vi å sammenlikne dette med de tilsvarende yrkesprosentene for menn. Vi kan da legge inn disse som søyler bak kvinnene, som en slags referansegruppe.

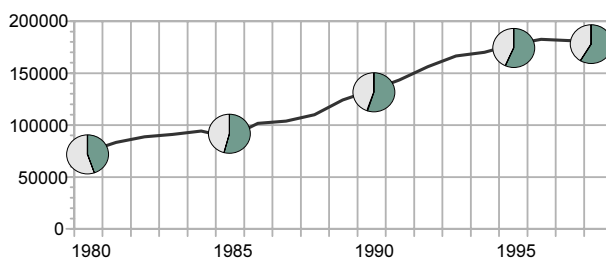


Linje - sektor

I en periode med sterk økning i antallet studenter, har også andelen kvinner blant studentene økt og kvinnene er nå i flertall. Dette kan fremstilles i et linjediagram med to linjer og separate verdiakser for de to avhengige variablene (antall studenter og andel kvinner). Problemet med to akser er (som nevnt tidligere) at det må gjøres helt tydelig hvilken akse som gjelder hvilken variabel (hvis ikke, kunne man her lett få inntrykk av at andelen kvinner blant studentene snart er 100 prosent).

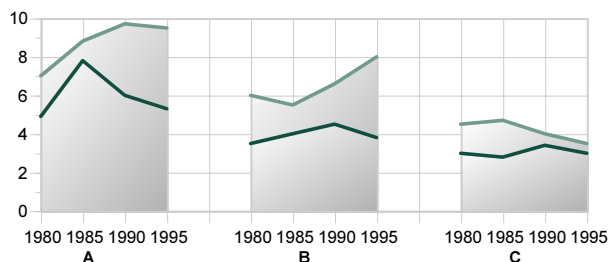


En alternativ løsning er legge inn sektordiagrammer med andelen kvinner på linjen på utvalgte tidspunkter, som en slags tilleggsopplysning (for ikke å overlesse diagrammet, er kjønnsfordelingen gjengitt bare for enkelte år).



"Linje-som-søyle"

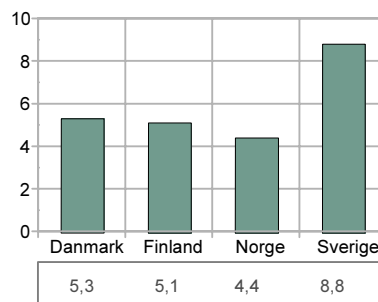
La oss anta at vi har data om tre ulike fritidsaktiviteter (A, B og C) for menn og kvinner for fire forskjellige tidspunkter. Dette blir lett for mange variable for ett linjediagram. Alternativet er å lage tre separate diagrammer. Men i stedet for å presentere dem som tre frittstående diagrammer kan de settes sammen, side ved side. Ved å skraverer arealet under linjene får det



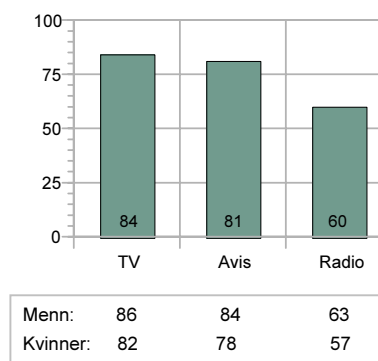
enkelte diagram (den enkelte aktivitet) karakter av en søyle, og vi kan nå lett sammenlikne både nivået for de tre aktivitetene og studere utviklingen over tid for aktivitetene for menn og kvinner.

Diagram m/tabell

Dette består rett og slett i at man i en egen tabell oppgir verdiene som diagrammet bygger på. Denne "Ole Brumm-løsningen" anbefales ikke generelt, men kan være ønskelig eller nødvendig noen ganger. Dataprogrammer som *Excel* og *DeltaGraph* inneholder begge slike muligheter, og den vanligste løsningen er å plassere dataene/verdiene *under* diagrammet, slik som i eksemplet til høyre. Denne løsningen forutsetter imidlertid at det er relativ få kategorier på x-aksen, hvis ikke blir en slik tabell uleselig. Lange tidsserier egner seg derfor ikke for denne typen kombinasjon.



En annen variant er å presentere en tabell som supplerer og "utdyper" diagrammet. Mens diagrammet viser totaltall (her: andelen som har brukt ulike massemedier en dag), kan tabellen vise fordelingen på f.eks. menn og kvinner eller ulike aldersgrupper.

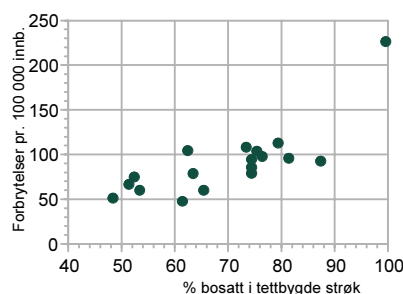


Ved siden av slike kombinasjonsdiagrammer, finnes det også en lang rekke andre diagramtyper. Noen av dem er relativt spesialiserte og brukes bare for spesielle formål. Her skal vi presentere noen av de mer generelt anvendelige.

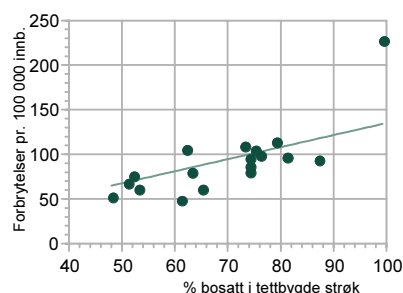
6.2. Punktdiagram

Denne diagramtypen er svært vanlig i forskning og statistisk analyse, men kan også med fordel brukes for mer populære presentasjoner. Punktdiagrammet har to kvantitative skalaer og viser en toveis fordeling. Det brukes til å illustrere sammenhengen mellom to kvantitative variable (det kalles derfor også ofte korrelasjonsdiagram). Diagramtypen brukes særlig for regionale data, f.eks. kommuner, fylker og land, men kan også brukes for individ-data, når antallet ikke er for stort.

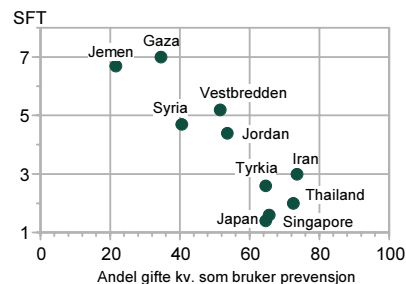
Eksemplet til høyre illustrerer forholdet mellom befolkningstetthet og antallet forbrytelser pr. 100 000 innbyggere i fylkene (det er vanlig å plote den uavhengige variabelen på den horisontale aksene og den avhengige på den vertikale): Hvert punkt (symbol) representerer altså ett fylke og vi ser at de har en tendens til å gruppere seg langs diagonalen, noe som indikerer en positiv sammenheng: Jo høyere befolkningstetthet, jo flere forbrytelser. Siden denne typen diagram først og fremst skal belyse sammenhenger snarere enn nivå, er det vanlig å ikke la aksene gå til 0.



Sammenhengen i et punktdiagram er ikke alltid like lett å få øye på (og noen ganger er det heller ikke noen sammenheng). For å tydeliggjøre tendensen i materialet er det derfor vanlig å legge inn en såkalt "trendlinje" (de fleste dataprogrammer har denne muligheten). Denne linjen er uttrykk for f.eks. den linjære (positive eller negative) sammenhengen mellom de to variablene.



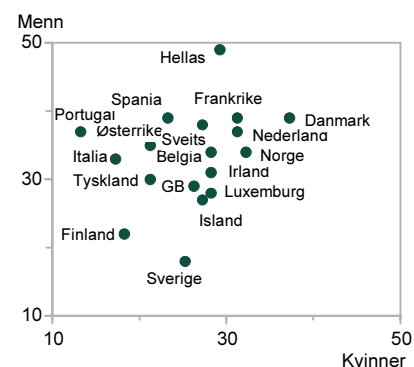
Når punkt-diagrammet ikke inneholder alt for mange enheter, kan man identifisere de enkelte enhetene. Til høyre vises sammenheng- en mellom prevensjonsbruk og samlet fruktbarhetstall (SFT) i utvalgte land. Identifisering av enhetene (landene) gir tilleggsopp- lysninger som øker diagrammets informasjonsverdi.



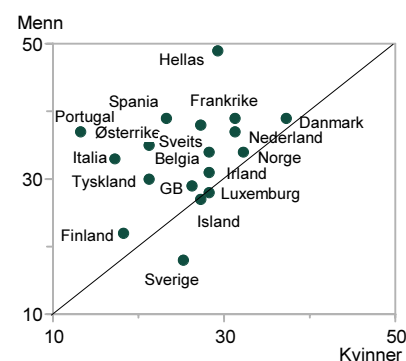
Hvis man for eksempel opererer med to eller flere (under-) grupper av enheter, kan disse identifiseres med ulike symboler, for eksempel nordiske og ikke-nordiske land:



I eksemplene ovenfor har alle diagrammene en uavhengig variabel som antas å påvirke en annen, avhengig variabel. Men denne dia- gramtypen kan generelt brukes til å studere mønstre/ grupperinger i data. Ett eksempel: Diagrammet til høyre viser andelen menn og kvinner som er dagligrøykere i en del europeiske land. Uten at noen av variablene her kan sies å være uavhengig eller avhengig av den andre, er det likevel interessant å se på sammenhengen mellom dem. Ved første øyekast er bildet litt kaotisk.

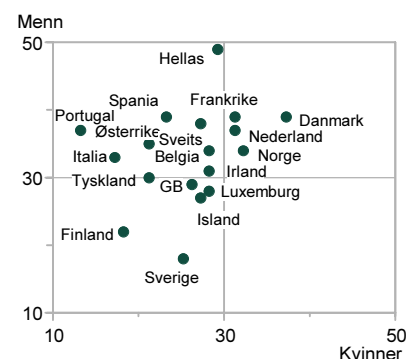


Men om vi legger inn en diagonallinje, ser vi f.eks. tydeligere at noen av landene (Irland, Luxemburg, Storbritannia, Island, Norge og Danmark) befinner seg på eller nær diagonalen; dvs. at andelen mannlige og kvinnelige dagligrøykere her er omtrent like store. Sveriges spesielle situasjon, med flere kvinnelige enn mannlige dagligrøykere, blir også nå tydeligere.

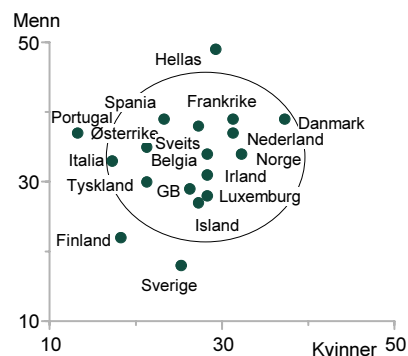


En annen måte å "gruppere" landene på er å se på andelen menn og kvinner over og under f.eks. 30 prosent. Diagrammet visualiserer dermed en inndeling i fire grupper:

- Land hvor mennenes andel er over 30 prosent og kvinnenes andel er under. Det framstår som det "typiske (sør-europiske) mønsteret".
- Land hvor både menns og kvinners andel er over 30 prosent
- Land hvor både andelen menn og kvinner lavere enn 30 prosent
- Til slutt ser vi at ikke i noe land er kvinnenes andel over 30 prosent og mennenes under.

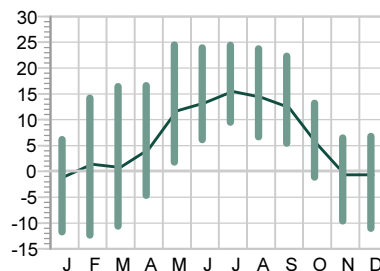


En tredje måte å visualiser et mønster i data på, er å se på eventuelle atypiske enheter. I diagrammet til høyre er de "typiske" landene gruppert sammen ved hjelp av en sirkel, mens fire land – hvert på sin måte – framstår som atypiske. Avhengig av hva man vil framheve, kan altså utformingen av et punktdiagram være med på å visualisere og understreke budskapet.



6.3. Min - maks diagram (variasjonsdiagram)

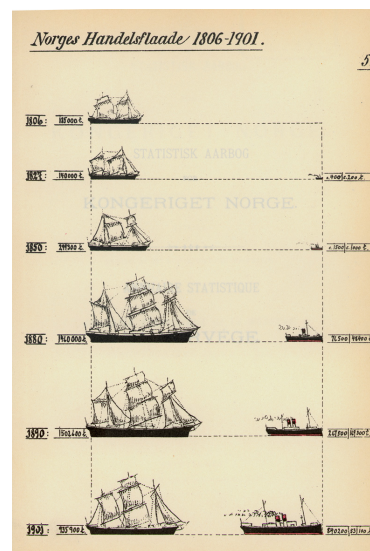
Også kalt "flytende" søylediagram (siden verdiaksen ikke har noe "naturlig 0-punkt). Dette er en diagramtype som angir variasjonsbredden i et datamateriale, f.eks. høyeste og laveste verdi, eller startverdi/sluttverdi. Ett slik eksempel kan være maksimums og minimumstemperaturer per måned. I tillegg kan man legge inn f.eks. gjennomsnittstemperatur som egen linje.



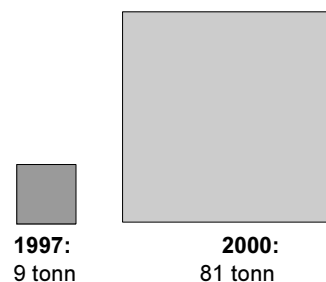
6.4. Flate- og volumdiagram

I kapitlet om linjediagrammer nevnte vi arealdiagrammer som en form for linjediagram, hvor linjene er stablet i høyden og arealet mellom dem er skravert. Dette er den betegnelsen som vanligvis brukes om denne typen diagrammer (f.eks. i *Excel*). Men historisk benyttes ofte betegnelsen "flatediagram".

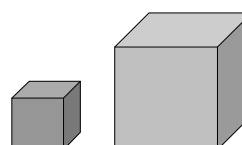
Denne typen diagrammer var svært populære i begynnelsen av forrige århundre. Når Statistisk årbok i 1902-utgaven brakte dette diagrammet, som viste utviklingen i Norges handelsflåte, så var dette helt i tråd med tidens dominerende praksis



I sin enkleste form består et flatediagram av to eller flere kvadrater; hvor sidene i kvadratet er lik kvadratroten av verdiene. Et eksempel: Eksporten av en vare økte fra 9 tonn i 1997 til 81 tonn i 2000. Siden denne diagramtypen ikke har verdiakser, har ikke størrelsen noen betydning; det er forholdet mellom arealene som er viktig. Det at verdiaksene mangler, gjør at verdiene må oppgis.



Volumdiagrammer bygger på samme prinsippet, men er litt mer komplisert å beregne, siden man her får en dimensjon til.



Flate- og volumdiagrammer benyttes i dag svært sjelden, og med god grunn: For det første er dette diagrammer som ikke finnes i de vanligste dataverktøyene; slike diagrammer må lages "for hånd". For det andre: Selv om diagrammene ofte er elegant utformet, er de vanskelige å tolke. Blant annet fordi de er to- eller tredimensjonale; dvs. at både bredden og høyden (og ev. dybden) har betydning. I stedet for å sammenlikne f.eks. høydene på to søyler, må vi i et flatediagram sammenliknet arealet av to flater; noe som i praksis er umulig, utover å fastslå at "den ene er større enn den andre". Denne typen diagrammer har derfor sjelden annet enn en rent illustrerende funksjon.

6.5. Rangdiagram

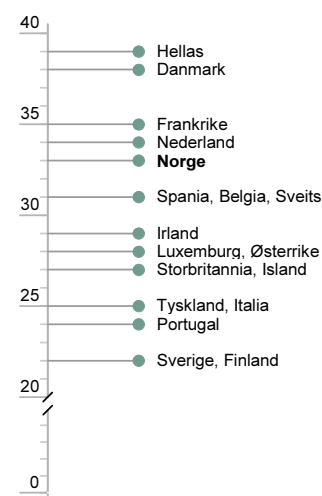
Et eksempel på en rangert liste kan være Statistisk sentralbyrås statistikk over de mest populære fornavnene i året som gikk. En slik, rangert liste kan rett og slett se slik ut:

1. Emilie
2. Ida
3. Thea
4. Julie
5. Ingrid
6. Sara
7. Malin
8. Marte
9. Nora
10. Karoline

En slik rangering sammenliknes ofte med tilsvarende liste for året før. En alternativ variant er da - for hvert navn - å angi økning/nedgang i forhold til forrige år ved hjelp av symboler. Symbolene kan være f.eks. ++ + = ÷÷ ÷ eller piler som peker opp og ned og symboliserer sterk eller svak økning/nedgang og stabilitet. Et slik "diagram" kan se omtrent slik ut:

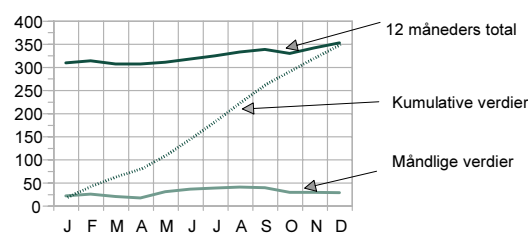
- Emilie →
- Ida ↗
- Thea ↘
- Julie ↗
- Ingrid ↘

I diagrammet til høyre rangeres europeiske land på grunnlag av andelen dagligrøykere i landene omkring 1995. Sammenliknet med en tabell eller liste som bare viser andelen i de ulike landene, får vi her i tillegg også et visuelt inntrykk av avstanden mellom dem. På grunn av formen, kalles dette av og til et termometer-diagram.

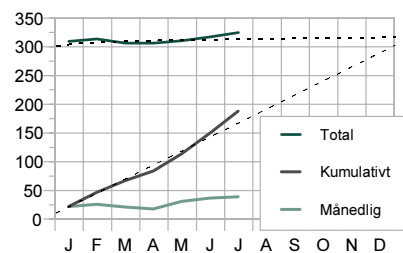


6.6. Z-diagram

Dette er egentlig ikke en egen diagramtype, men en spesiell anvendelse av et kumulativt linjediagram for å lage en prognose (eksemplet til høyre viser tydelig hvorfor det har fått navnet). Det egner seg spesielt for tidsserier med månedlige data, for eksempel antall døde i trafikkulykker. Diagrammet viser tre ulike aspekter ved en variabel, her: Månedlige verdier, kumulative månedlige verdier og 12 måneders rullerende total. Eksempel: For mai viser det antall døde i mai, antallet døde i perioden januar-mai og totalen siste 12 måneder (juni-mai).

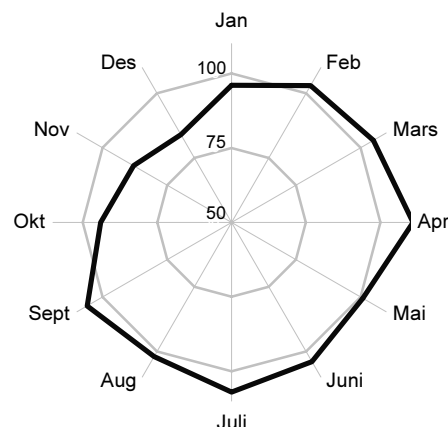


Etter hvert som man får tall for årets første måneder, kan man bruke to av tidsseriene til å "predikere" det endelige, årlige antallet. Dette gjøres ved å legge inn såkalte trendlinjer i både den kumulative og totale serien.



6.7. Radar-diagram

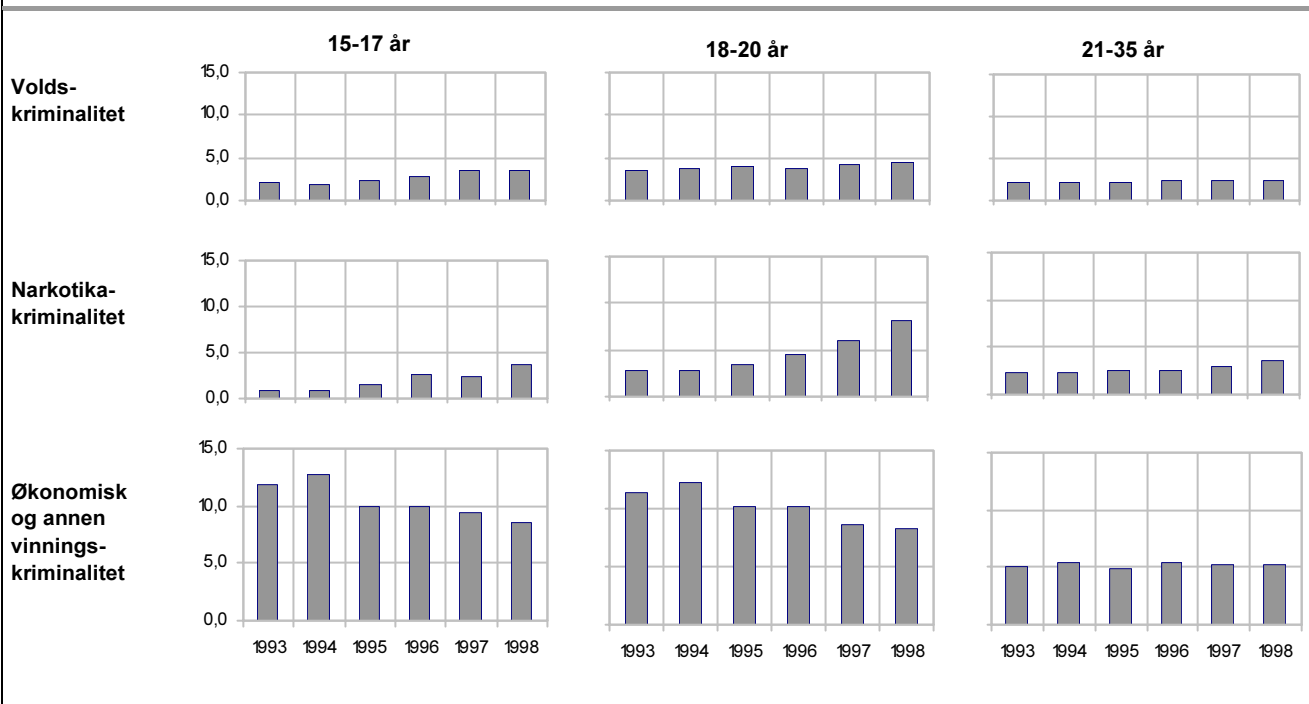
En spesiell form for linjediagram er "radar"-diagrammet. Dette er et sirkulært linjediagram og egner seg godt til å visualisere forskjellige typer månedlig statistikk (man tenker seg da året som et hjul eller sirkel). Til høyre vises fødslenes fordeling over året; dvs. antall levende fødte hver måned som et avvik fra gjennomsnittet (= 100).



6.8. Diagram-matrise

Når man har behov for å framstille tre eller flere variable grafisk, kan en diagram-matrise ofte være løsningen: Her kan leseren sammenlikne effekten av flere variabler samtidig. I eksemplet nedenfor kan vi studere lovbrudd pr. 1 000 innbyggere, fordelt på type lovbrudd, siktedes alder og år.

Siktede for forbrytelser. Etter utvalgte lovbruddsgrupper og alder. Pr. 1000 innbygger. 1993-1998



7. Kort om bruk av farger

I det foregående er alle diagrammer laget i sort eller sort + en farge og forskjeller mellom søyler/linjer er markert med ulike gråtoner og/eller fargetoner. Og for mange (kanskje de fleste?) formål er bruk av gråtoner og raster tilstrekkelig. Det er også det billigste: Å trykke i farger er fortsatt dyrere enn sort/hvitt. Men noen ganger er farger hensiktsmessig, og heller ikke dyrere. Dette gjelder f.eks. ved bruk av diagrammer i presentasjoner, enten det nå er via transparenter eller skjerm.

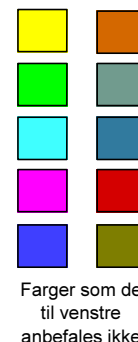
Bruk av farger kan sies å ha to hovedfunksjoner: Farger er et effektivt virkemiddel som kan bidra til å forbedre lesbarheten (og dermed forståelsen) av et diagram. I tillegg kan farger bidra til å gjøre diagrammet eller presentasjonen mer attraktiv eller "spennende". Når det er sagt, bør det også sies at bruk av farger er vanskelig og det bør skje i samråd med noen som har kunnskap og erfaring. Særlig er dette viktig når man lager noe som skal sendes til et trykkeri.

7.1. Hvilke farger?

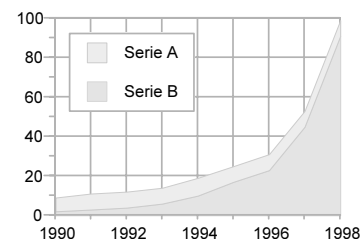
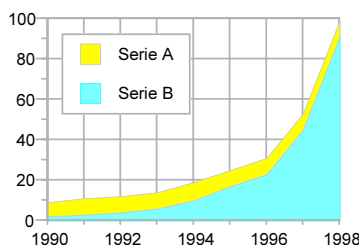
Hvilke farger man velger er selvfølgelig i stor grad et spørsmål om "smak og behag" og individuelle preferanser. Men det finnes likevel én enkel regel:

Bruk ikke rene, lyse eller svært sterke farger. Benytt heller såkalte "skitne" farger; dvs. farger som inneholder sort.

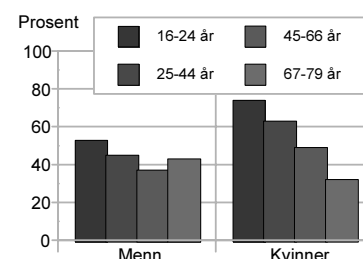
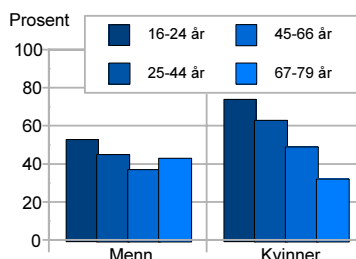
- For det første vil de rene fargene ofte lett bli (for) intense eller "skingrende".
- I tillegg er det problematisk å benytte prosenter av rene farger (på samme måte som vi bruker gråtoner i sort/hvitt).
- For det tredje egner disse fargene seg også dårlig til tekst.



Ved valg av farger må man også tenke på hvordan de oppfører seg i kopieringsmaskinen: Følgende diagram egner seg f.eks. *ikke* til kopiering. Med farger fungerer det rimelig bra, men i sort/hvitt-varianten til høyre blir det vanskeligere å skille mellom seriene.

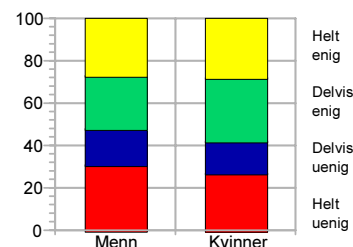


Et annet eksempel: I diagrammet til høyre lar aldersgruppene seg så vidt skille ved bruk av ulike nyanser i blått. Men etter en tur gjennom kopi-maskinen blir forskjellene nesten helt borte.

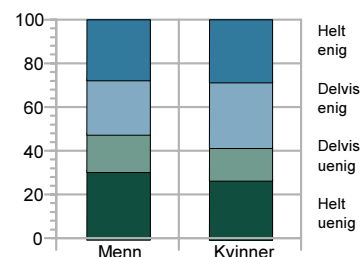


7.2. Fargerikt fellesskap?

Antallet mulige eller tilgjengelige farger er nærmest uendelig. Men som ellers gjelder også her regelen om *enkelhet*: Det er ikke nødvendigvis slik at *flere* farger øker informasjonsverdien av et diagram. I eksempelet til høyre er det brukt fire ulike farger for å skille mellom svarene på et spørsmål om enighet/uenighet i en påstand. Fargebruken her virker tilfeldig og lite planlagt.

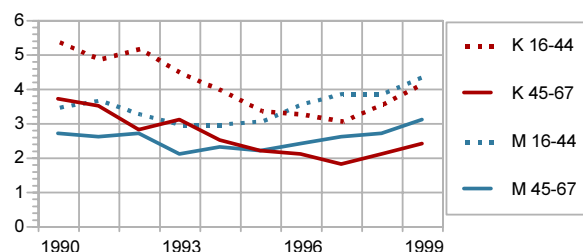


Siden svaralternativene her er gradert (helt/delvis enig/uenig) er det naturlig å la fargebruken gjenspeile dette; for eksempel ved at vi bruker bare to farger og graderer disse, slik at mørke farger tilsvarer "helt (u)enig" og lysere farger: "delvis (u)enig". Dvs. at "styrken" på fargen tilsvarer graden av enighet.



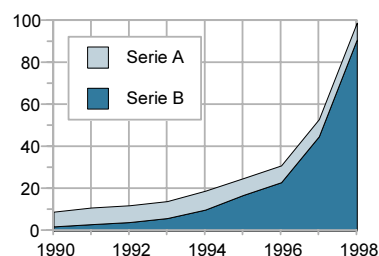
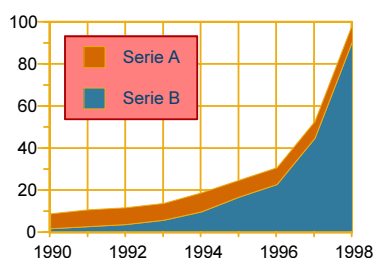
Tilsvarende i linjediagrammene til høyre. Her vises utviklingen for menn og kvinner i to aldersgrupper. I stedet for å bruke fire ulike farger, bør fargene her brukes til å skille mellom menn og kvinner, mens aldersgruppene skiller ved ulike strektyper.

Hadde linjene representert fire ulike land – derimot – hadde kanskje fire farger vært mer naturlig.



7.3. Farger hvor?

Her er hovedregelen å *fargelegge data*; dvs. å bruke fargene først og fremst for å gjøre det lettere å sammenlikne grupper eller serier, ikke for å pynte opp et ellers "kjedelig" diagram. Bruk av farger på andre elementer enn dataseriene bidrar i



stor grad til å ta oppmerksomheten bort fra det som er viktig. Selv om en publikasjon trykkes i flere farger, er dette ingen grunn til å bruke dem alle! To farger (f.eks. sort + 1) kan ofte være vel så effektivt som mange farger.

Litteratur

Bounford, Trevor (2000): *Digital Diagrams - Effective design and presentation of statistical information*. London: Cassell & Co.

Harris, Robert L. (1999): *Information Graphics - A Comprehensive Illustrated Reference*. London and New York: Oxford University Press.

Kristiansen, Jan Erik og Jørgen Ouren (1998): *Fornavn i Norge - Navnemoter og motenavn*. Oslo: Ad Notam Gyldendal.

Kosslyn, Stephen M. (1994): *Elements of graph design*. New York: W. H. Freeman and Company.

Plain Figures (1996). Second edition by Myra Chapman and Cathy Wykes. London: The Stationary Office.

Tufte, Edward R. (1983): *The Visual Display of Quantitative Information*. Cheshire: Graphics Press.

Wallgren, A. et. al. (1994): *Statistikens bilder - att skapa diagram*. Stockholm: Statistiska centralbyråen. (Finnes også i engelsk utgave)

White, Jan V. (1984): *Using charts and graphs: 1000 ideas for visual persuasion*. New York and London: R. R. Bowker Company.

De sist utgitte publikasjonene i serien Notater

- 2002/19 F. Gundersen og L. Solheim: Regionalisering av FoU-statistikken. 43.s
- 2002/20 L. Vågane: Omnibusundersøkelsen november/desember 2001. Dokumentasjonsrapport. 56s.
- 2002/21 G. Claus, O. Haugen, P.M. Holt og E. Knutsen: Regnskapsstatistikk. Næringsoppgaver for ikke-finansielle aksjeselskaper, 1999. Dokumentasjon. 34s.
- 2002/22 M. Takle: Befolkningsstatistikk på rutene. Dokumentasjon. 35s
- 2002/23 D. Roll-Hansen, S. Ferstad, M. Stålnacke, P. Tuhus og R. Nøtnæs: En spørreskjemametodisk gjennomgang av datainn-samling gjennom Grunnskolens informasjonssystem (GSI). 109s.
- 2002/24 T.P. Bøe og I. Håland: Dokumentasjon av arbeidskraftundersøkelsen (AKU). 85s.
- 2002/25 A. Akselsen og T. Sandnes: FD - Trygd: Dokumentasjonsrapport. Stønader til enslig forsørger. 1992-2000. 46s.
- 2002/26 E. Rønning: Statistisk sentralbyrås tidsbruksundersøkelse 2000/01. Dokumentasjon og resultater fra intervjuet. 125s.
- 2002/27 S. Myro og C. Torp: Stedsfesting av bedrifter i Bedrifts- og foretaksregisteret. Hovedprosjekt. 37s.
- 2002/28 C. Nordseth og T. Sandnes: FD - Trygd: Dokumentasjonsrapport. Foreløpig uførestønad. 1992-2000. 37s.
- 2002/29 S. Derakhshanfar og T. Sandnes: FD - Trygd: Dokumentasjonsrapport. Økonomisk sosialhjelp. 1992-2000. 36s.
- 2002/30 I. Johansen: Undersøking om foreldrebetaling i barnehagar, januar 2002. 42s.
- 2002/31 T.M. Køber, H. Moafi, E. Rønning og Ø. Sivertstøl: Bruk av forløpsdatabaser i Statistisk sentralbyrå. 60s.
- 2002/32 T.M. Normann: Omnibusundersøkelsen februar/mars 2002. Dokumentasjonsrapport. 37s.
- 2002/33 S. Reid: Bosettingskriteriene i inntekts-systemet til kommunene. Erfaringer med overgang til ny beregningsmåte og nye bosettingskriterier, 2002. 43s.
- 2002/34 K.E. Engebretsen, P.E. Gjedtjernet, S. Kristoffersen, P.G. Larssen og J.H. Wang: Mottak og tilrettelegging av SLN-data. 49s.
- 2002/35 D. Rafat: Analyse av sammenheng mellom ektefellers sysselsetting i en familie. 27s.
- 2002/36 A. Bruvoll og T. Bye: En vurdering av avfallspolitikkens bidrag til løsning av miljø- og ressursproblemer. 31s.
- 2002/37 K.I. Bøe: B.R. Joneid: KOSTRA revisjonssystem. Malverk for generelt revisjonssystem - KOSTRA-data. Revidert utgave. 66s.
- 2002/38 N. Arnesen, G. Daugstad, O.E. Hallingstad, E. Skretting Lunde og B.Vold: Kvalitetssikring i KOSTRA. Forslag til dokumentasjonsrutiner med erfaring fra FylkesKOSTRA-helsetjenester, somatikk. 54s.
- 2002/39 H. Moafi: Omlegging av folkehøgskolestatistikk. Overgang til elektronisk rapportering. 31s.
- 2002/40 Ø. Kleven: Mediebrukundersøkelsen 2001. Dokumentasjonsrapport. 43s.
- 2002/41 Ø. Kleven: Samordnet levekårsundersøkelse 2000 -panelundersøkelsen. Dokumentasjonsrapport. 129s.
- 2002/42 L. Solheim: Foreløpige tall i FoB2001 Utvalg,vekter, estimering og usikkerhet. 64s.
- 2002/43 A. Andersen, E. Birkeland, J. Epland og M. I. Kirkeberg: Økonomi og levekår for ulike grupper trygdemottakere 2001. Foreløpig rapport. 214s.