

Datavisualisering

Peter M. Dahlgren, JMG

2016-08-05

Att presentera sitt forskningsresultat visuellt är viktigt för att dels kommunicera forskning och dels skapa bättre förståelse för resultaten. Här får du praktiska tips på hur du kan göra tabeller och diagram som är någorlunda snygga, korrekta och pedagogiska.

Se detta som tips på hur du kan utforma dina tabeller och diagram, inte som regler som slaviskt måste följas. Tipsen riktar sig först och främst till vetenskapliga tabeller och diagram, men kan också användas till journalistiska tabeller och diagram.

Mjukvara

- Gör diagram i Microsoft Excel.
- Tabeller kan göras i Microsoft Word eller Excel.
- Använd inte Numbers för Mac. Det är för dåligt.
- Kopiera aldrig tabeller eller diagram från SPSS. De är horribla. Gör egna i stället.

Grunderna i datavisualisering

Ibland brukar man prata om **explorativ datavisualisering**. Det betyder att man skapar diagram och andra visualiseringar för att se vad som döljer sig bakom datan, till exempel när du använder statistikprogram. Explorativ datavisualisering är med andra ord endast till för att forskaren ska kunna bilda sig en uppfattning om datan.

Denna text handlar inte om explorativ datavisualisering, utan om visualisering som du vill publicera eller visa upp för andra.

Tips

1. **Diagram och tabeller bör kunna läsas fristående från texten.**
Tanken är att någon som skummar texten också kan ta till sig det viktiga bara genom att titta på diagrammet eller tabellen.
2. **Gå från det enkla till det komplicerade.** När man presenterar diagram och tabeller är det viktigt att inte ge allt för mycket information på en gång. Undvik att besvara *alla* frågeställningar med en enda tabell eller diagram. Börja i stället enkelt och presentera beskrivningar av datan (deskriptiv analys). Därefter kan du visa på intressanta samband (bivariat analys) för att därefter presentera mer komplicerade modeller (multivariat analys). Börja med andra ord med det enkla och gå successivt mot det mer komplicerade. Vad man ska välja att presentera? Du måste veta vad din forskningsfråga är, och därmed vad du bör välja bort. Därför underlättar det betydligt om man har ett genomtänkt syfte och åtminstone en idé om hur det färdiga resultatet kan se ut.

Dataformat

Det mest tidskrävande med datavisualisering är ofta att få datan i rätt format.

När man förbereder data för Excel kan datan ibland behöva vara i rätt format. Jag rekommenderar att du använder långt dataformat.

	A	B
1	Årtal	Antal
2	2010	51
3	2011	67
4	2012	83
5	2013	72
6	2014	51
7	2015	62

Långt dataformat:

	A	B	C	D	E	F	G
1	Årtal	2010	2011	2012	2013	2014	2015
2	Antal	51	67	83	72	51	62

Brett dataformat:

Tabeller

Alla diagram börjar med en tabell.

Tabellens delar

En tabell i en vetenskaplig publikation består ofta av tre delar: tabellhuvud, tabellresultat och tabellkommentar.

Tabell 1 *Andel som uppger att de någon gång blivit tillfrågad om att betala någon form av muta (procent)*

		Ja ¹	Nej	Summa procent	Antal svar
<i>Offentlig myndighet/tjänsteman</i>	2009	1,2	98,8	100	1 562
	2013	1,1	98,9	100	1 573
	2015	0,8	99,2	100	1 545
<i>Privat företag/anställd</i>	2009	1,3	98,7	100	1 553
	2013	2,4	97,6	100	1 564
	2015	2,5	97,5	100	1 537

Kommentar: Frågan lyder: 'Har du i Sverige under de senaste 12 månaderna blivit tillfrågad om att betala någon form av muta i kontakt med: Offentlig myndighet/tjänsteman' och 'Privat företag/anställd'. Svarsalternativen är 'Ja, flera gånger', 'Ja, någon gång' samt 'Nej'. ¹Kategorin 'Ja' bygger på en sammanslagning av alternativen 'Ja, flera gånger' samt 'Ja, någon gång'. Procentbasen utgörs av dem som besvarat frågan.

Källa: De nationella SOM-undersökningarna 2009, 2013 och 2015.

Figure 1: Exempel från boken Ekvilibrium (red. Ohlsson, Ekengren Oscarsson & Solevid) från 2016, s. 94.

I **tabellhuvudet** skrivs först och främst tabellens nummer för att kunna hänvisa till tabellen i den löpande texten. Det gör det enkelt för läsaren att hitta rätt tabell.

Efter tabellnumret kommer en kort rubrik. Den bör vara slagkraftig och koncis men också redovisa tidsperiod, analysenheter och enheten (exempelvis procent). Om du redovisar resultatet av en frågeundersökning kan själva frågan också skrivas i rubriken. Skriv enkelt, krångla inte till språket i onödan.

I **tabellresultatet** är den viktigaste informationen. Du bör inte använda decimaler, utan avrunda helst om möjligt. Om du redovisar procent så avrunda också summan till 100 procent. Då ser man också om procenten summeras på varje rad (som på bilden ovan) eller i varje kolumn.

- Om du jämför två tidsperioder kan det vara bra att framhäva skillnaderna i en egen kolumn.
- Om du jämför två grupper med varandra (män respektive kvinnor) kan det också vara bra att använda en referenscolumn för samtliga tillsammans (totalen).
- Om tabellen blir stor och svår att överblicka kan du lösa det på två sätt. Genom att markera det viktigaste genom fetmarkering eller inringning.

Du kan också försöka slå ihop flera kategorier så att de blir färre.

I **tabellkommentaren** skriver du mer utförlig information om exempelvis hur variabler har slagits ihop eller skapats, källa till datamaterialet eller hur tabellen ska tolkas.

- Använd lite mindre textstorlek i tabellkommentaren.
- Beskriv vad kategorin "Övrigt" innehåller.
- Skriva hur många analysenheter är, såsom N=1 585, eller mer pedagogiskt "antal svar".
- Skriv signifikansnivåerna om du använder asterisk (*) i tabellresultatet. Vanligtvis skrivs de:

Stjärnor	p-värde	Signifikansnivå
*	p < 0,05	95 %
	p < 0,01	99 %
*	p < 0,001	99,9 %

Effektparametertabell

En effektparametertabell är ett bra sätt att ta en väldigt stor korstabell med tre variabler och minska ned den till de viktigaste beståndsdelarna.

På SPSS-Akuten finns en bra guide om hur du gör en effektparametertabell från en korstabell.

Tabell 3. Andel (procent) som tror på evolutionen bland de som röstade på presidentkandidaterna Barack Obama och John McCain respektive bland religiösa och ej religiösa.

	Ej religiös	Religiös	Effekt av religiös:
Obama	84,2	49,7	-34,5
McCain	64,2	27,1	-37,1
Effekt av McCain:	-20,0	-22,6	

Figure 2: Exempel på effektparametertabell.

Diagram

Att tänka på när du gör diagram

Innehåll

- Försök inte trycka in allt för många saker i ett enda diagram. Behöver du visa många saker, gör många diagram.
- Låt någon utomstående titta på ditt diagram. Förstår han eller hon? Om inte, gör om!

Axlar

- Lägg den oberoende variabeln på den horisontella axeln (X).
- Lägg den beroende variabeln på den vertikala axeln (Y).
- Låt axeln börja på 0 (exempelvis 0 till 100 procent), om du inte har goda skäl för något annat.

Linjer

- Använd olika typer av linjer (streckade, heldragna m.fl.) för att skilja dem åt.

Perspektiv och färger

- Undvik 3D-diagram.
- Använd hög kontrast.
- Följ ett färgschema, se exempelvis Practical Rules for Using Color in Charts.
- Diagram bör kunna skrivas ut i svartvitt. Förlita dig därför inte enbart på olika färger för att skilja linjer åt.

Många diagram

Ska du visa många diagram jämte varandra? Tänk då på att:

- Använd samma skalstorlek (exempelvis 0-100 procent) på alla diagrams axlar. Om du har 0-20 på ett diagram och 0-70 på ett annat blir de svårare att jämföra.
- Gör ett snyggt diagram en gång. Kopiera och klistra sedan in det flera gånger för att spara tid.

Pajdiagram

Pajdiagram är runda och ser ut som pajer. Dessa bör du inte använda av två skäl:

- Det är svårt att uppskatta den relativa storleken på respektive andel.
- Det går inte att jämföra två pajdiagram med varandra på ett enkelt sätt.

Om du ändå vill använda pajdiagram bör du tänka på att:

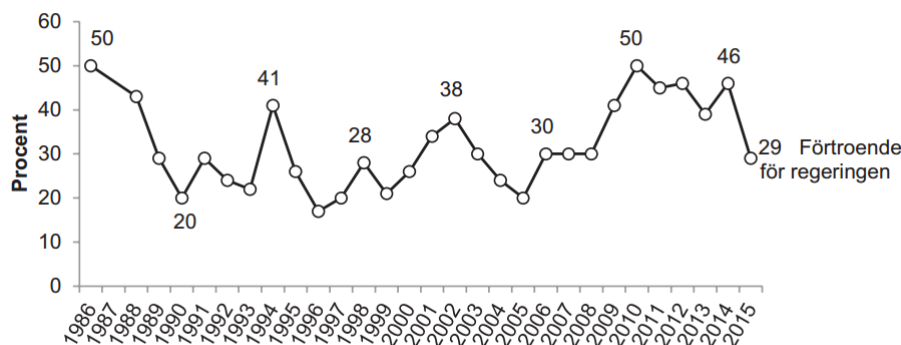
- Använd inte mer än 7 kategorier. Annars blir det svårt att överblicka. Sikta på 2-3 kategorier (till exempel män/kvinnor).
- Kategorierna presenteras från störst till minst. Om du exempelvis sorterar de fem största städerna så bör Stockholm vara först, följt av Göteborg och så vidare.
- Den största kategorin placeras med början klockan 12 och rör sig sedan medsols mot exempelvis klockan 4 (Stockholm). Därefter följer kategorierna i fallande ordning, från klockan 4 till klockan 6 (Göteborg), och så vidare.

Linjediagram

Linjediagram är särskilt användbara när du ska visa trender eller tidsserier.

Lägg gärna etiketterna (i det här fallet “Förtroende för regeringen”) i slutet av linjen. Det gör det lättare att se vad linjerna representerar, utan att man behöver hoppa fram och tillbaka med blicken mellan linjen och en förklaringsruta.

Figur 2 Andel stort förtroende för regeringen 1986–2015 (procent)



Kommentar: Avser andelen mycket/ ganska stort förtroende bland samtliga som svarat.

Källa: Den nationella SOM-undersökningen 2015.

Figure 3: Exempel från boken Ekvilibrium (red. Ohlsson, Ekengren Oscarsson & Solevid) från 2016, s. 58.

Stapeldiagram

Stapeldiagram är användbara när du vill jämföra två eller fler kategorier.

Använd liggande stapeldiagram när du har många kategorier och man ska kunna läsa dem med enkelhet. Liggande stapeldiagram är också användbart när

kategorinamnen (i det här fallet Aftonbladet, Expressen etc.) är långa och blir svårlästa på X-axeln.

Sortera listan på ett relevant sätt, exempelvis från högsta till lägsta värdet. Lägg värdena i slutet av staplarna så att dessa framgår tydligt.

Diagram 6. Unika besökare på webbplats under vecka 6 (tusental besökare).

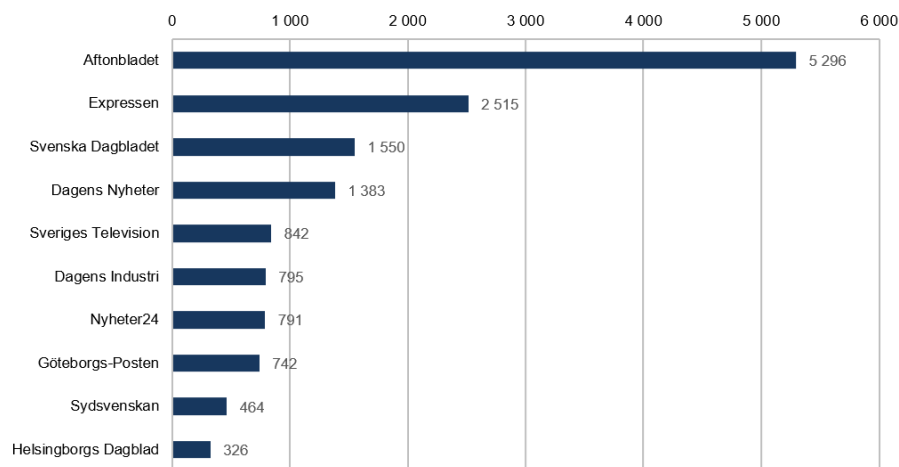


Figure 4: Exempel på liggande stapeldiagram.

Grupperat stapeldiagram

Data: 2 kategoriska variabler, korstabell

Grupperade stapeldiagram är användbara om du behöver jämföra grupper (korstabell).

Tänk på att förenkla så att det inte blir för många staplar. Du kan exempelvis slå ihop kategorier. Lägg gärna värdet ovanför eller i stapeln.

Sortera gärna staplarna på ett logiskt sätt, som i det här fallet utifrån ökande ålder.

Staplad fraktionsstapel

Data: Kategoriska variabler, Likert-skala, brett dataformat

En staplad fraktionsstapel visar andelen (fraktionen) av helheten. När du gör en fraktionsstapel behöver du inte göra om din data till procent, Excel gör det automatiskt.

Diagram 8. Jämförelser med ett grupperat stapeldiagram (tusental kronor).

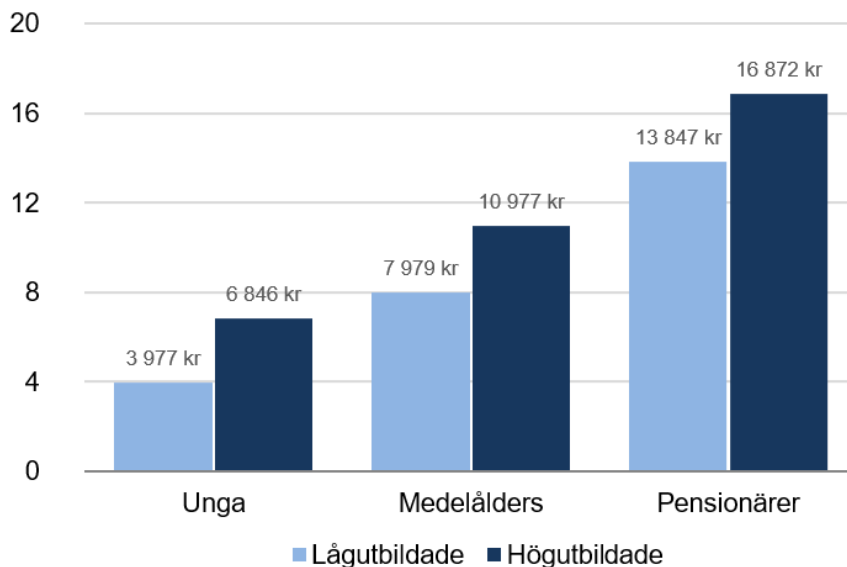


Figure 5: Exempel på grupperat stapeldiagram.

En staplad fraktionsstapel fungerar bra för att visa Likert-skalar. De kan enkelt jämföras genom att man placerar flera staplar jämte varandra. Siffrorna i respektive stapel uppgår till hundra procent.

Fraktionsstapeln kan också placeras liggande om man har många kategorier:

Glöm inte den löpande texten

Du behöver inte presentera allt som diagram. Glöm inte bort att utnyttja den löpande texten på ett effektivt sätt.

Har du gjort en surveyundersökning kan du börja med beskrivande statistik av deltagarna, vilket ofta tar upp mindre plats än tabeller eller diagram:

Respondenterna var 18–45 år ($M=31$; $SD=7$) och majoriteten (56 %) hade en högskoleexamen.

Denna korta mening ger information om åldersspannet (18-45), vad medelåldern är (M) samt standardavvikelsen för åldern (SD). Med denna information går det någorlunda enkelt att få en bild av åldersdistributionen.

Därefter kan du gå vidare med att beskriva intressanta samband, även det i den löpande texten:

Diagram 5. Vill införa obligatoriskt lördagsgodis (procent).

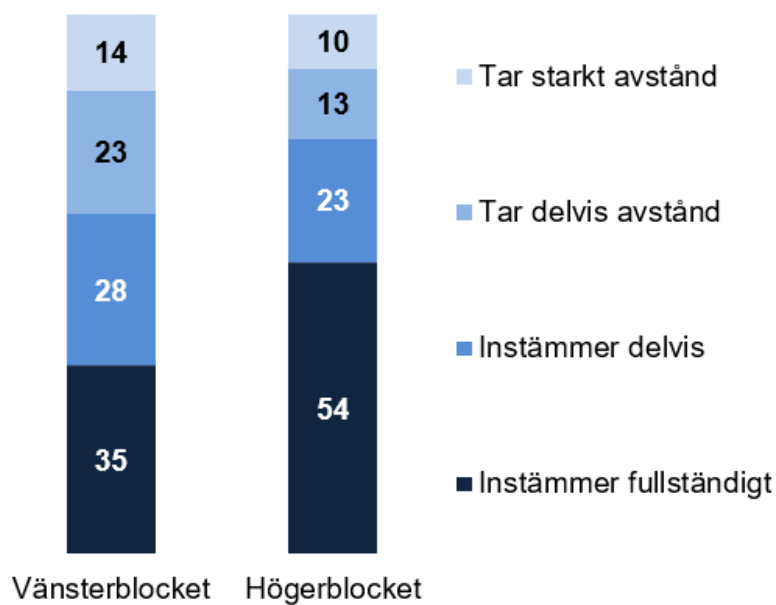


Figure 6: Exempel på hur en Likert-skala kan omvandlas till ett diagram.

Diagram 5. Liggande fraktionsstaplar är lämpliga när det finns många alternativ (procent).

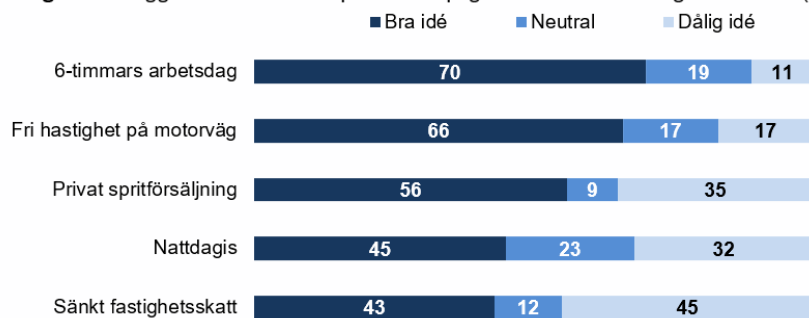


Figure 7: Exempel på liggande fraktionsstapel.

Det fanns ett starkt positivt samband mellan nyhetskonsumtion och politisk kunskap ($r=0,61$; $p<0,001$), men inte mellan nyhetskonsumtion och attityd ($r=0,02$; $p=0,528$).

I det här fallet står **r** för Pearsons korrelationskoefficient och **p** står för sannolikheten från signifikanstestet. Ibland kan det också vara bra att skriva ut vilken typ av signifikanstest det är (enkelsidigt eller dubbelsidigt), om det inte framgår av sammanhanget.

Skriv alltid ut det fullständiga p-värdet med tre decimaler, till exempel $p=0,039$. Då undviker du avrundningsfel, så att $p=0,051$ inte blir $p<0,05$ av misstag, och du undviker också att p-värdet behandlas dikotomt som signifikant ($p<0,05$) respektive icke-signifikant ($p>0,05$).

Kritiska frågor

Har du gjort en tabell eller diagram, eller ska du bedöma någon annans? Använd dessa frågor för att hitta svagheter och uppskatta kvaliteten.

Diagram och tabeller:

1. Finns det en rubrik där det tydligt framgår vad det visar?
2. Kan den läsas fristående från texten?

Diagram:

1. Är den oberoende variabeln placerad på X-axeln?
2. Är den beroende variabeln placerad på Y-axeln?
3. Är någon utav axlarna beskurna så att eventuella värden inte börjar på 0?

Video

Introduction to data visualization

An introduction to the practice of data visualization, with theory, examples, and good humor. Studio rerecording of a presentation at McGill University graduate students.

Data Visualization Best Practices

Learn how to get the most out of your business intelligence platform including how to choose the right chart type, and how to optimise your charts to make them easier to understand and more attractive.

Lästips

- Cairo, A. (2016). *The Truthful Art: Data, Charts, and Maps for Communication*. New Riders.
- Few, S. (2012). *Show Me the Numbers: Designing Tables and Graphs to Enlighten (2nd edition)*. Burlingame, Calif.: Analytics Press.
- Tufte, E. (2001). *The Visual Display of Quantitative Information (2nd edition)*. Cheshire, Conn: Graphics Pr.
- Wallgren, A., & Wallgren, B. (2008). *Diagram - statistikens bilder*. Stockholm: Norstedts juridik.