

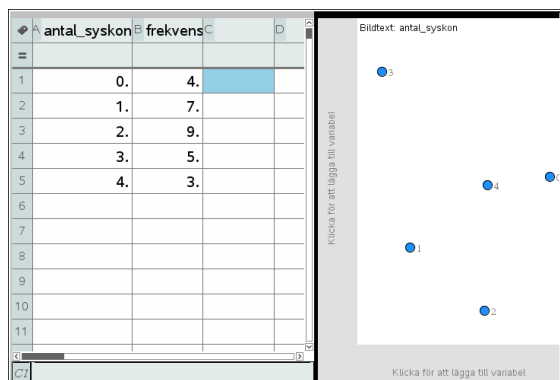
Grundläggande statistik kurs 1

Problem 1

Arbeta med frekvenstabeller

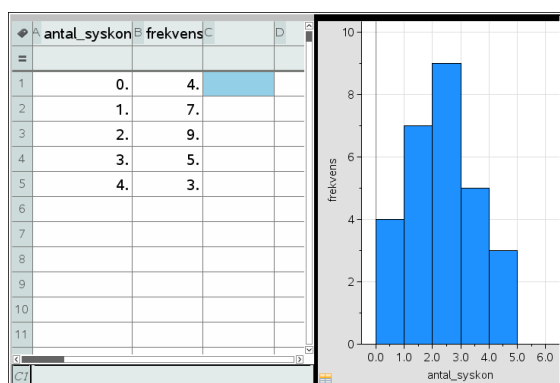
Sid 2:

Så här ser sidan 2 ut. Vi har alltså en delad sida med kalkylbladet till vänster och en Data&Statistik-sida till höger. I den sidan har vi ännu inte gjort någonting.

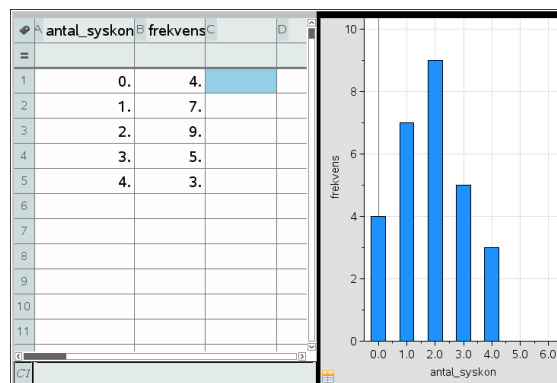
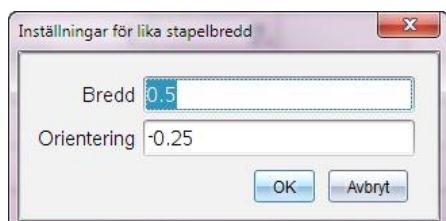


Vi ska nu rita ett diagram för de data i vi har i kalkylarket. Klicka först under den vågräta axeln och välj *antal_syskon*.

På den lodräta axeln ska du ha frekvensen men eftersom vi redan har en frekvenstabell så högerklickar du och väljer *Lägg till Y-sammanfattningslista*. Välj sedan *frekvens*. Då kommer diagrammet upp på den delade sidan.



Så här vill vi inte ha diagrammet. Nu är det ett histogram och det ser ut som vi t.ex. har data i intervallet 0 till 1 med frekvensen 4. Vi har ju 4 data med värdet 0. Högerklicka i diagrammet och välj alternativet *Inställningar av stapelvärden/Lika stapelbredd*. Väljer du enligt dialogrutan nedan får du ett snyggt diagram.

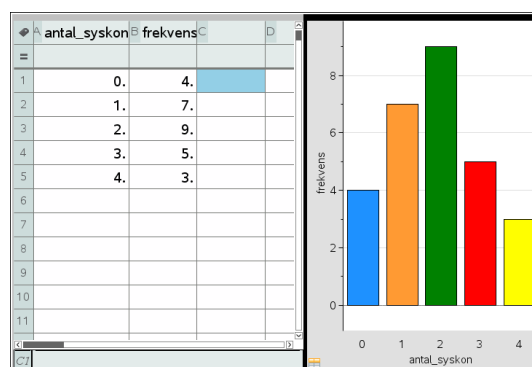


Det finns ett annat sätt att rita diagrammet. Högerklicka på den vågräta axeln och välj *Lägg till X-variabel med översiktslista*. Då kommer följande dialogruta upp. Eftersom vi här bara har två kolumner (variabler) så är det redan ifyllt. Har du fler kolumner med data som har ett namn så får du klicka på pilarna och välja.

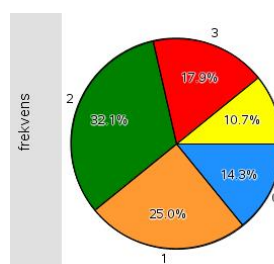


När du har klickat OK så visas diagrammet. Ändra sedan stapelbredden som tidigare.

Det finns ett tredje sätt: När du har ditt stolpdigram högerklickar du under den vågräta axeln och väljer *Forcera X i kategoriform*. Då får vi ett stapeldiagram utan numerisk skala.

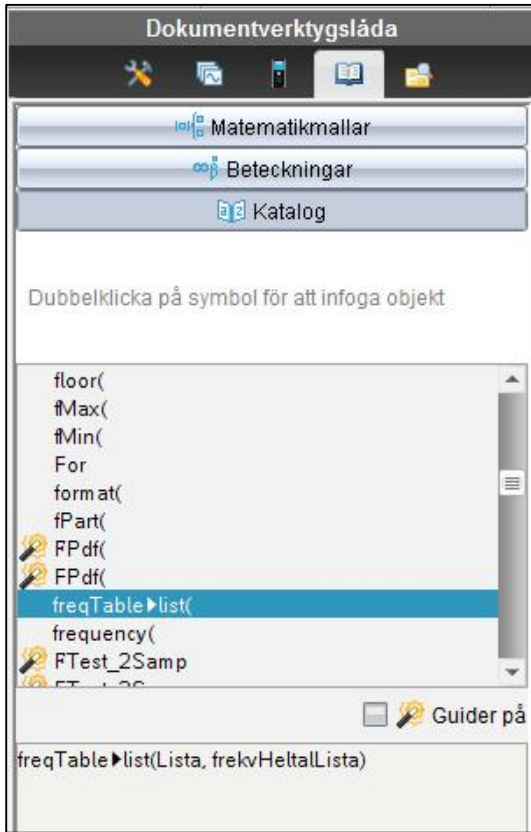


Högerklickar du i diagrammet kan du också välja att rita ett cirkeldiagram. Man kan välja att rita med Etiketter och då ser vi andelarna i procent.

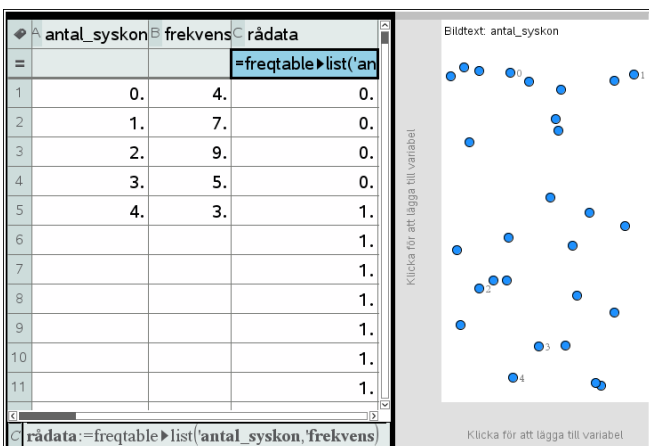


Sid 3:

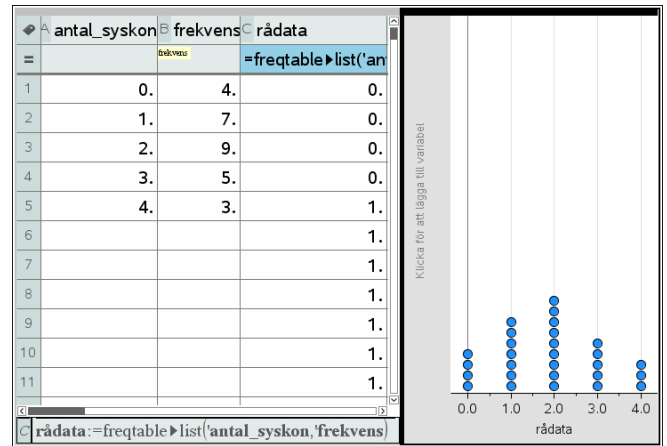
Att vi har en frekvenstabell betyder att någon har skapat den utifrån rapporterade värden från varje elev. Vi ska nu först skapa en tredje variabel – *Rådata* i kalkylarket. I formelfältet ska du sedan använda ett speciellt kommando för detta. I katalogen hittar du detta Se nedan.



Om du dubbelklickar infogas kommandot på formelraden och du klickar sedan i variablerna *antal_syskon* och *frekvens*. Vi får nu vår rådatalista sorterad. Elegant 😊.

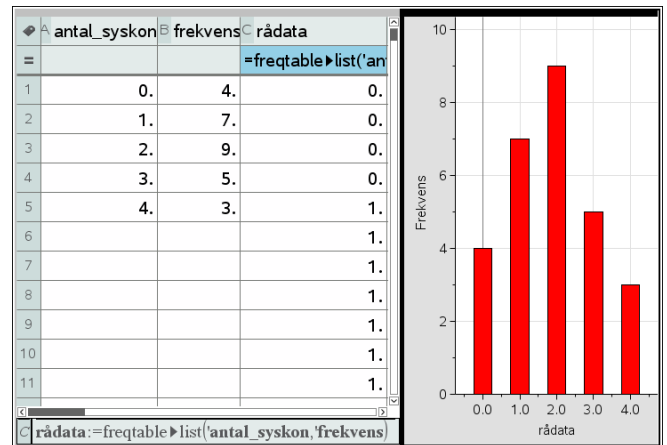


Nu kan vi rita ett diagram igen och vi behöver bara arbeta med en variabeln *rådata*. Vi väljer alltså variabeln *Rådata* under den vågräta axeln och diagrammet är klart.



Nu ritas ett *Punktdiagram*. Man ser alltså alla värden från rådatalistan som punkter.

Vi kan nu ändra till Histogram och ställa om stapelbredd enligt tidigare. Om du högerklickar och väljer *Skala* kan man genom att välja Procent få den relativa frekvensen i diagrammet.



Sid 4:

Här har vi gjort beräkningar av medelvärdet för antal syskon på några olika sätt. Det finns ju inbyggda kommandon för detta men det är viktigt att förstå hur beräkningen går till.

Vi kan nu t.ex. beräkna medelvärdet för antalet syskon på två sätt:

```
mean(antal_syskon, frekvens) → 1.85714285714  
mean(rådata) → 1.85714285714
```

Ett krångligare sätt som visar hur man gör om man inte använder de inbyggda funktionerna ovan är:

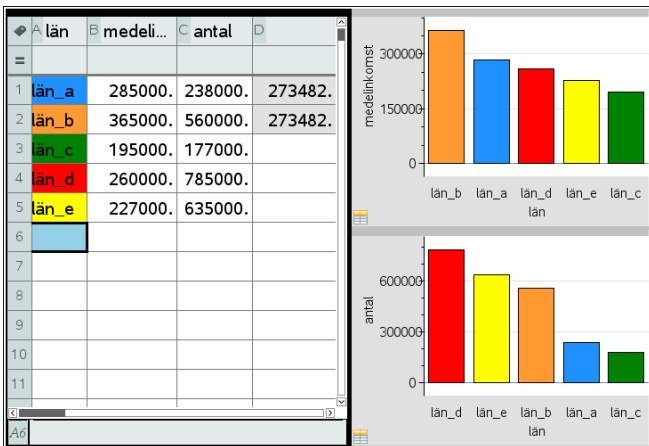
```
sum(antal_syskon * frekvens) / sum(frekvens) → 1.85714285714
```

Det är dock viktigt att man förstår hur beräkningen går till om man har en frekvenstabell. Uttrycket ovan visar ju detta.

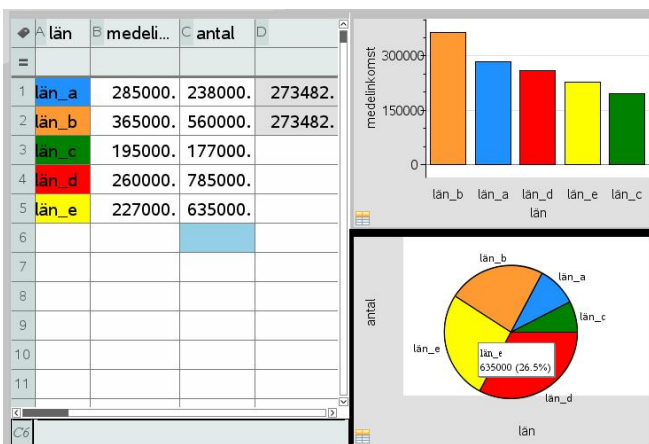
Om du nu går in och ändrar värden i frekvenstabellen på sid 2 så uppdateras rådatalistan och beräkningarna av medelvärdet på denna sida. Prova detta och ändra sedan tillbaka. Hade vi haft med några diagram så hade dessa också uppdaterats.

Problem 2

I detta problem ritas vi nu diagram där vi har en icke numerisk variabel. Det handlar om medelinkomst och antalet invånare i fem olika län. I kolumn D har vi beräknat den totala medelinkomsten på två olika sätt.



Variabeln antal passar att åskådliggöra i ett cirkeldiagram eftersom variabelns värden är andelar av en total. Om man flyttar markören över sektorerna så ser man värde och andel i procent!



Problem 3:

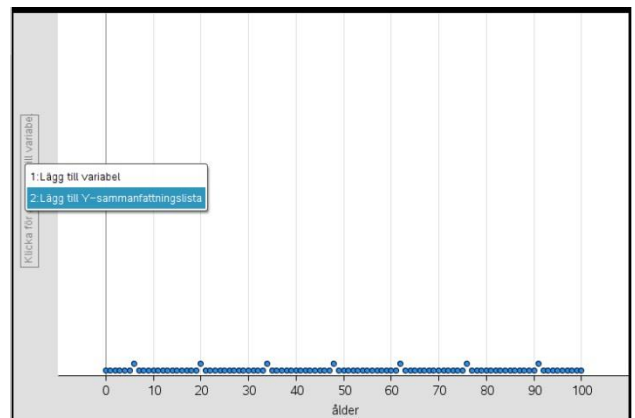
På nästa sida har vi i kalkylarket i kolumn B och C data från SCB om antalet personer i Sverige i olika ettårs-åldersklasser uppdelat på kön för år 2014. Personer 100 år och äldre har placerats i ålderskategorin "100".

I kolumnerna F och G har vi istället summerat data från kolumnerna B och C så att vi får 10-årsklasser. I cell F1 t.ex. så har vi med formeln $=\text{sum}(b1:b10)$ fått med alla kvinnor i åldern 0-9 år. I cell F2 har vi använt formeln $=\text{sum}(b11:b20)$ osv.

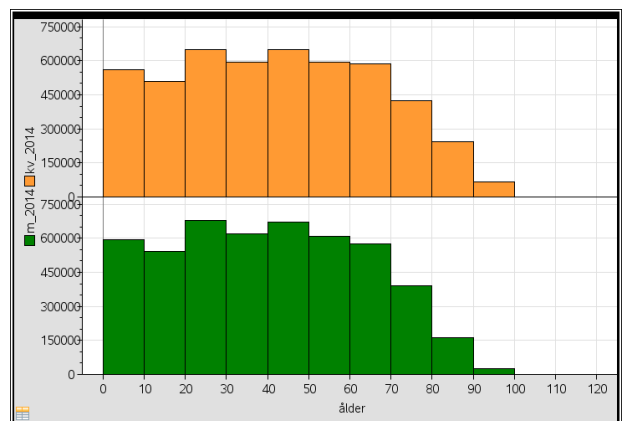
Nu kan vi rita diagram på två olika sätt.

Ålder	kv_2014	m_2014	tot_2014	intervall	antal_kv	antal_m
0.	56296.	59584.	115880.	0-9	561832.	593939.
1.	56083.	59489.	115572.	10-19	510025.	543473.
2.	56582.	59807.	116389.	20-29	647463.	679821.
3.	56429.	59395.	115824.	30-39	593277.	620469.
4.	58583.	61909.	120492.	40-49	650210.	670648.
5.	57024.	60375.	117399.	50-59	594518.	606834.
6.	56106.	59709.	115815.	60-69	584539.	576561.
7.	55641.	58891.	114532.	70-79	424216.	390122.
8.	55404.	58475.	113879.	80-89	241949.	162463.
9.	53684.	56305.	109989.	90-99	65454.	27589.
10.	53364.	56538.	109902.	100*plus...	1632.	321.

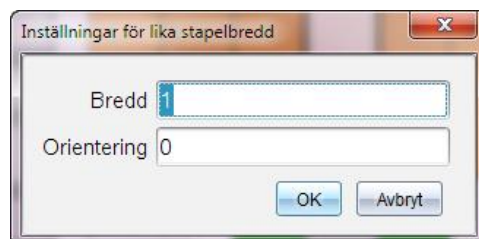
Vi börjar med att välja variabel på x-axeln. På y-axeln så väljer vi som tidigare "Lägg till Y-sammanfattningslista" och vi väljer variabelerna kv_2014 och m_2014.

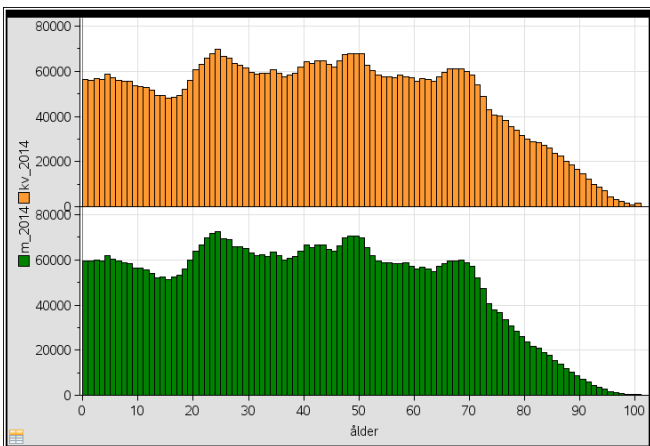


Så här blir det om vi utgår från sammanfattningsdata från kolumnerna B och C.



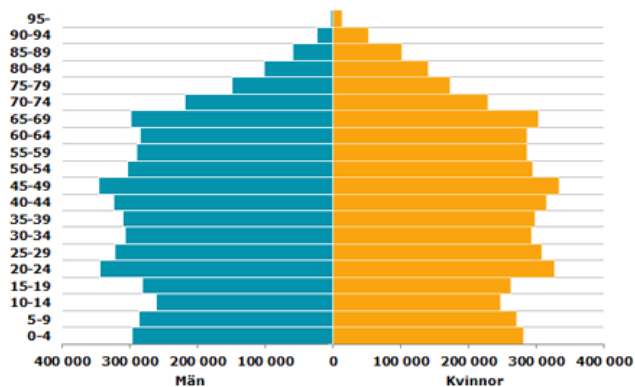
För att få en annan bredd på staplarna, t.ex. ettårs-klasser, så högerklickar vi i diagrammet och väljer Inställningar för stapelbredd.



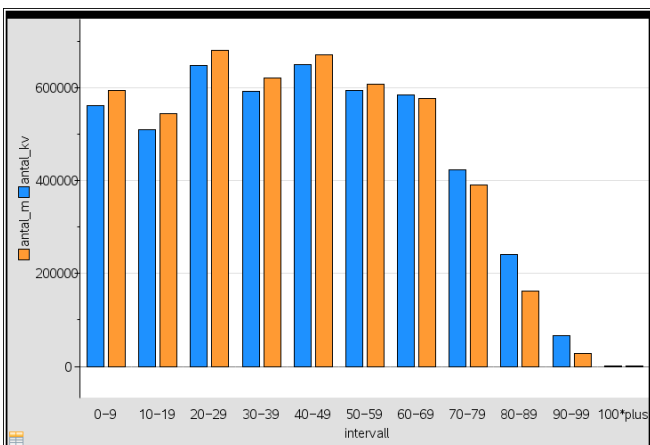


Man får en slags ålderspyramid på bredden.

Så här brukar man redovisa ålderspyramider. Här har vi 5-årsklasser. (Statistiska Centralbyrån (SCB), 31 december 2013.



På sid 4 har vi ritat ett jämförande stapeldiagram med värden för män och kvinnor inom samma åldersintervall bredvid varandra. Vi ser nu tydligt att antalet män i intervallen är större än antalet kvinnor till och med intervallet 50-59 år. Att det är så beror ju på att det i genomsnitt föds fler pojkar än flickor. Att kvinnostaplarna är högre i intervallen 60-69 och uppåt beror på att kvinnor i genomsnitt lever längre än män.



Här har vi gjort några beräkningar i appen Anteckningar. Tryck på Ctrl M för att komma i beräkningsläge.

Beräkning av medelåldern för män och kvinnor:
 $\text{mean}(\text{ålder_kv_2014}) \rightarrow 41.6836$
 $\text{mean}(\text{ålder_m_2014}) \rightarrow 39.7451$

Befolkning totalt: $\text{sum}(\text{tot_2014}) \rightarrow 9.74736\text{E}6$

andelen kvinnor (%) i befolkningen: $\frac{\text{sum}(\text{kv_2014})}{\text{sum}(\text{tot_2014})} \cdot 100 \rightarrow 50.0147$

I kalkylarket kan man också göra beräkningar förstås. I kolumn H har vi beräknat andelen i befolkningen som är yngre än 18 år och andelen i befolkningen som är äldre än 65 år (pensionärer).

	kv_2014	m_2014	tot_2014	E intervall	F antal_kv	G antal_m	H
=							
1	56296.	59584.	115880.	0-9	561832.	593939.	0.193124
2	56083.	59489.	115572.	10-19	510025.	543473.	0.196246
3	56582.	59807.	116389.	20-29	647463.	679821.	
4	56429.	59395.	115824.	30-39	593277.	620469.	
5	58583.	61909.	120492.	40-49	650210.	670648.	
6	57024.	60375.	117399.	50-59	594518.	606834.	
7	56106.	59709.	115815.	60-69	584539.	576561.	
8	55641.	58891.	114532.	70-79	424216.	390122.	
9	55404.	58475.	113879.	80-89	241949.	162463.	
10	53684.	56305.	109989.	90-99	65454.	27589.	
11							

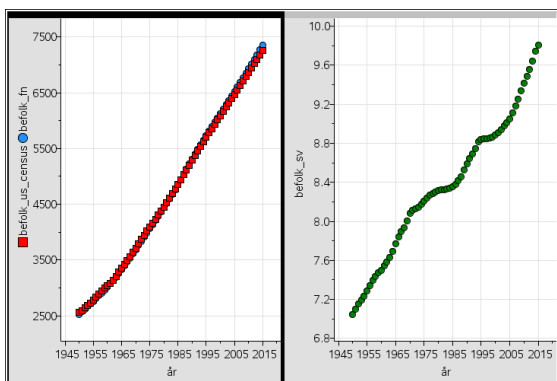
$\text{H2} = \frac{\text{sum}(d66:d101)}{\text{sum}(tot_2014)}$

Problem 4

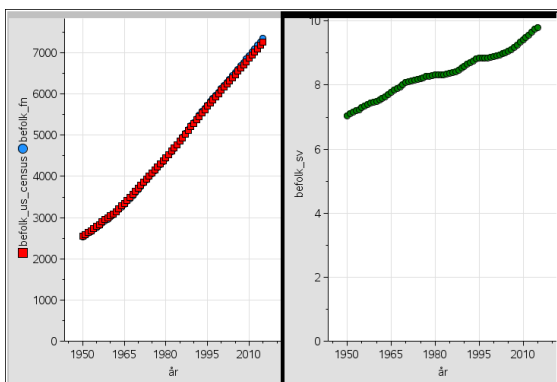
Här har vi data för befolkningen i världen från två olika källor och i Sverige (SCB).

A år	B befolk_fn	C befolk_us_ce...	D befolk_sv	E	F
=					
1	1950.	2525.15	2557.63	7.04692	
2	1951.	2571.87	2594.94	7.09874	
3	1952.	2617.94	2636.77	7.15061	
4	1953.	2664.03	2682.05	7.19232	
5	1954.	2710.68	2730.23	7.23466	
6	1955.	2758.32	2782.1	7.29011	
7	1956.	2807.25	2835.3	7.34112	
8	1957.	2857.66	2891.35	7.39287	
9	1958.	2909.65	2948.14	7.43607	
10	1959.	2963.22	3000.72	7.47135	
11	1960.	3018.34	3043.	7.49797	

Så här ser befolkningsutvecklingen ut. I det vänstra diagrammet täcker datapunkterna varandra så det ser ut som en tidsserie. En närmare undersökning av data visar att förändringen år från år för data för hela världen inte är särskilt konstant även om det ser ut så i diagrammet. För Sverige syns det tydligare.

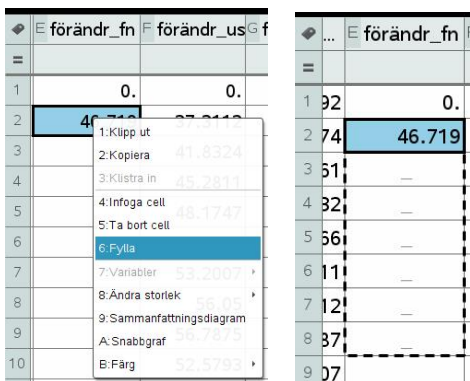


Observera att skalan på befolkningsaxeln inte börjar vid 0. Speciellt för Sverige kan förändringen se ganska dramatisk ut. Så här blir det om den lodräta axeln börjar vid 0.



För att närmare analysera befolkningsutvecklingen så gör vi beräkningar på förändringen jämfört med föregående år.

I cell E2 skriver vi $=b2-b1$ och trycker på enter. Då får vi värdet på förändringen mellan 1950 och 1951 i FN:s data. Vi högerklickar på värdet och markerar *Fylla* och drar sedan markören till den sista raden.



Man gör sedan exakt likadant för US Census-data och data för Sverige.

I nästa steg beräknar vi med exakt samma förfarande den *relativa förändringen*. Vi ser att beräkningsuttrycket i cell H2 är

$$= \frac{e2}{b1} \cdot 100$$

Där e2 är förändringen mellan 1950 och 1951 och b1 befolkningen 1950. Man använder sedan funktionen *Fylla* på samma sätt som för den absoluta förändringen.

	förändr_fn	förändr_us	förändr_sv	rel_förändr_fn	rel_förändr_us	rel_förändr_sv
1	0.	0.	0.	0.	0.	0.
2	46.719	37.3112	0.05182	1.85015	1.45882	0.7353
3	46.072	41.8324	0.051866	1.79138	1.61208	0.7306
4	46.089	45.2811	0.04171	1.76051	1.71729	0.5833
5	46.649	48.1747	0.042348	1.75107	1.79619	0.5887
6	47.637	51.8708	0.055448	1.75738	1.89987	0.7664
7	48.931	53.2007	0.05101	1.77395	1.91225	0.6997
8	50.417	56.05	0.05175	1.79596	1.97686	0.7049
9	51.988	56.7875	0.043194	1.81925	1.96405	0.5842
10	53.565	52.5793	0.035279	1.84094	1.78348	0.4744

Nu kan vi rita diagram på de relativa förändringarna. Vi ser att den relativa förändringen (i %) i princip har avtagit sedan 1970. Man kan därför inte riktigt tala om att världens befolkning ökar exponentiellt eftersom det förutsätter att den relativa förändringen är konstant.

I Sverige varierar den relativa förändringen på ett helt annat sätt.

