User-friendly presentation of statistics:

Graphs

Jan Erik Kristiansen

Senior adviser

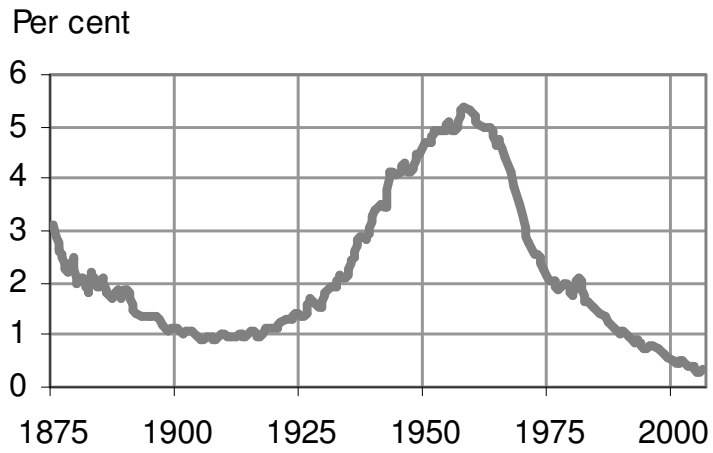
Statistics Norway

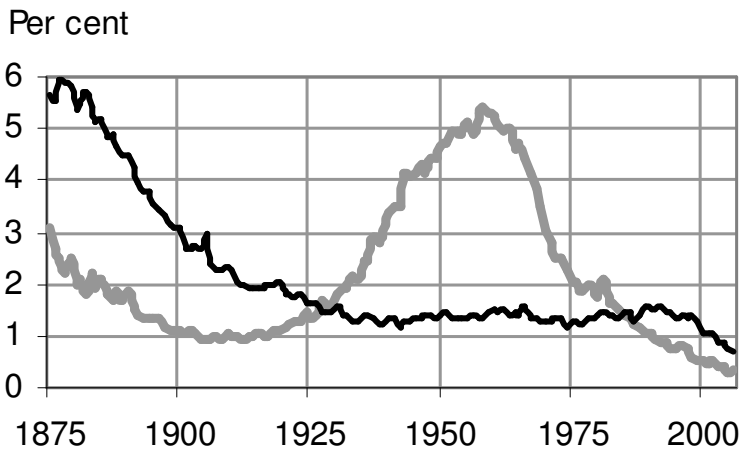
From tables to graphs:

Why?

Per cent girls given the first name *Anne* each year

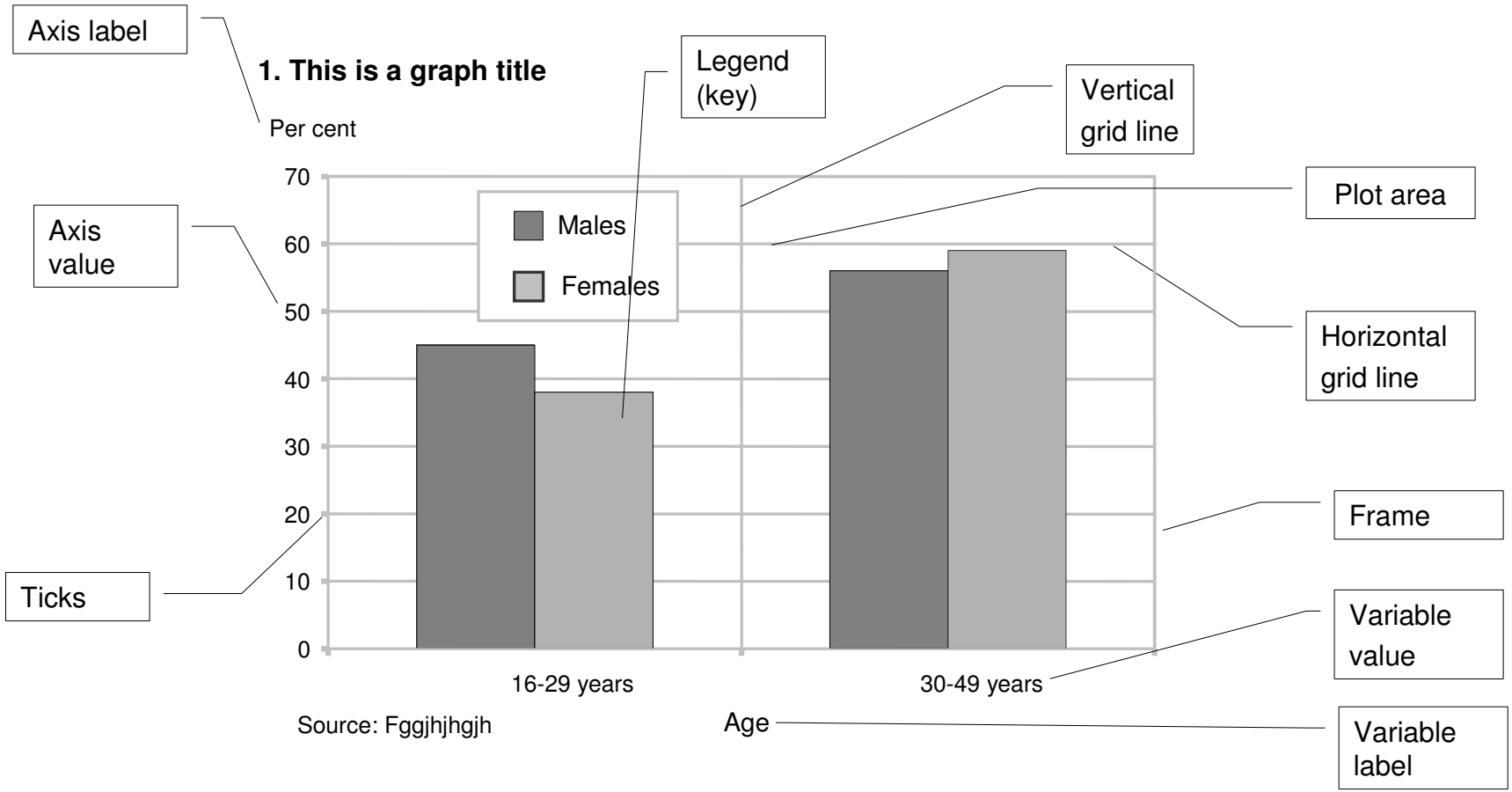
Year:	Pct.:	Year:	Pct.:	Year:	Pct.:	Year:	Pct.:	Year:	Pct.:	Year:	Pct.:
1875	3,11	**1897**	1,17	**1919**	1,12	**1941**	3,52	**1963**	4,99	**1985**	1,39
1876	2,67	**1898**	1,10	**1920**	1,16	**1942**	3,47	**1964**	4,63	**1986**	1,36
1877	2,53	**1899**	1,13	**1921**	1,24	**1943**	4,14	**1965**	4,74	**1987**	1,23
1878	2,20	**1900**	1,11	**1922**	1,31	**1944**	4,06	**1966**	4,44	**1988**	1,14
1879	2,48	**1901**	1,03	**1923**	1,31	**1945**	4,15	**1967**	4,16	**1989**	1,03
1880	1,97	**1902**	1,09	**1924**	1,43	**1946**	4,29	**1968**	3,83	**1990**	1,05
1881	2,07	**1903**	1,05	**1925**	1,36	**1947**	4,13	**1969**	3,49	**1991**	0,96
1882	1,80	**1904**	0,96	**1926**	1,41	**1948**	4,41	**1970**	3,05	**1992**	0,85
1883	2,22	**1905**	0,93	**1927**	1,70	**1949**	4,45	**1971**	2,77	**1993**	0,88
1884	1,95	**1906**	0,94	**1928**	1,61	**1950**	4,68	**1972**	2,52	**1994**	0,76
1885	2,12	**1907**	0,97	**1929**	1,55	**1951**	4,72	**1973**	2,49	**1995**	0,78
1886	1,82	**1908**	0,91	**1930**	1,81	**1952**	4,95	**1974**	2,29	**1996**	0,81
1887	1,70	**1909**	1,04	**1931**	1,91	**1953**	4,95	**1975**	2,06	**1997**	0,73
1888	1,84	**1910**	0,99	**1932**	1,95	**1954**	4,90	**1976**	2,04	**1998**	0,61
1889	1,70	**1911**	0,97	**1933**	2,16	**1955**	5,12	**1977**	1,88	**1999**	0,54
1890	1,85	**1912**	0,94	**1934**	2,10	**1956**	4,91	**1978**	1,99	**2000**	0,52
1891	1,50	**1913**	1,01	**1935**	2,42	**1957**	5,30	**1979**	1,94	**2001**	0,45
1892	1,39	**1914**	0,98	**1936**	2,47	**1958**	5,40	**1980**	1,77	**2002**	0,51
1893	1,36	**1915**	1,06	**1937**	2,91	**1959**	5,33	**1981**	2,07	**2003**	0,42
1894	1,36	**1916**	1,01	**1938**	2,81	**1960**	5,23	**1982**	1,66	**2004**	0,38
1895	1,36	**1917**	0,98	**1939**	2,93	**1961**	5,04	**1983**	1,65	**2005**	0,29
1896	1,36	**1918**	1,13	**1940**	3,40	**1962**	4,97	**1984**	1,53	**2006**	0,36





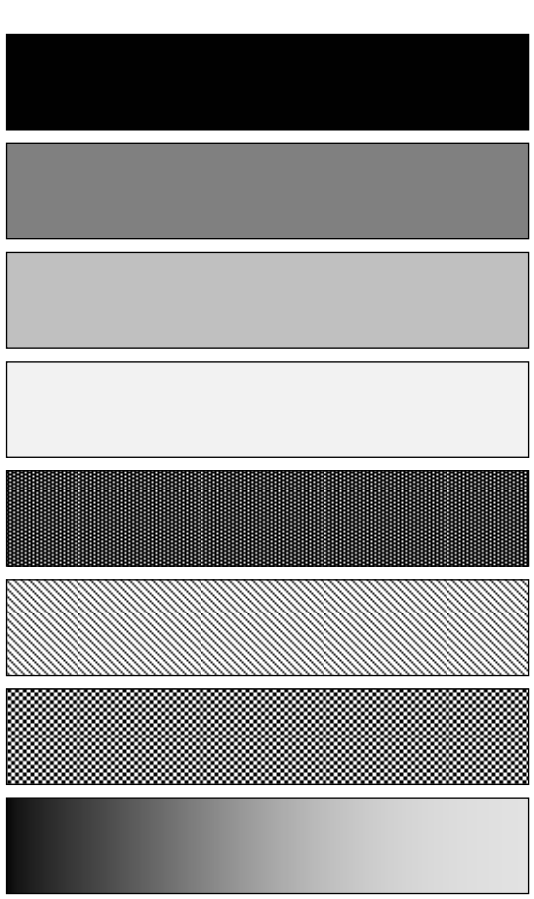
Charts speak directly to the eye!

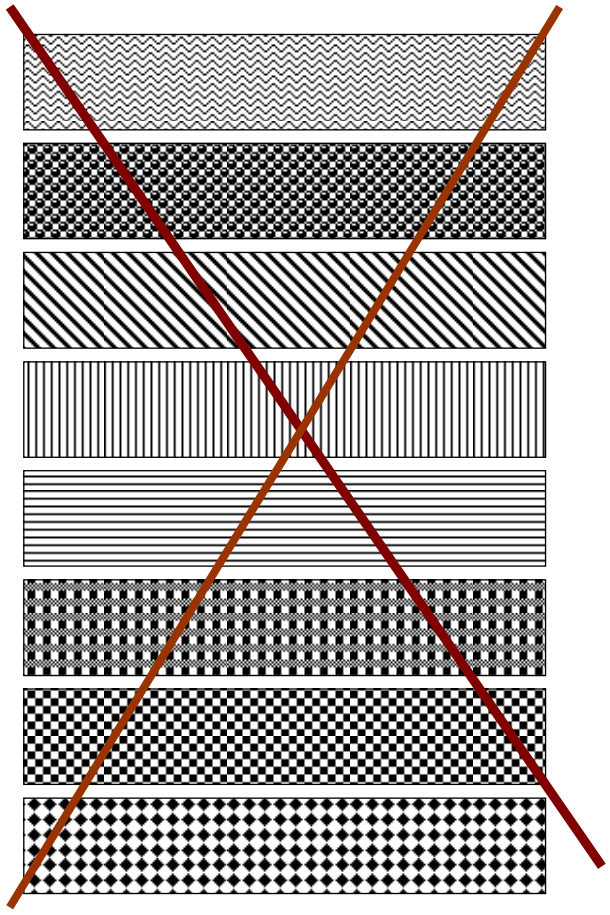
Elements of a graph:



Graphs: Issues and solutions

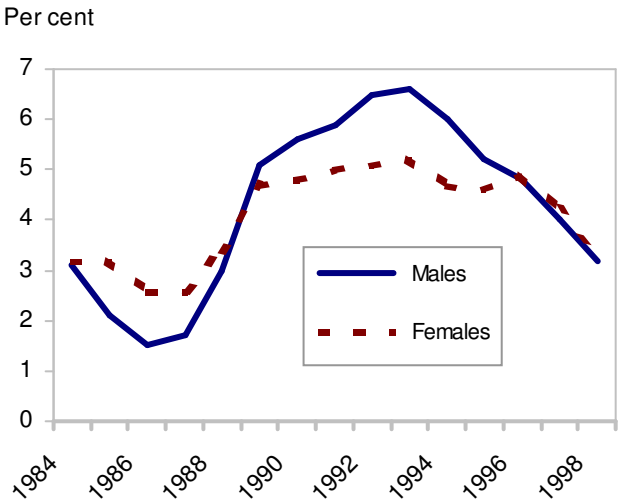
- Use of patterns

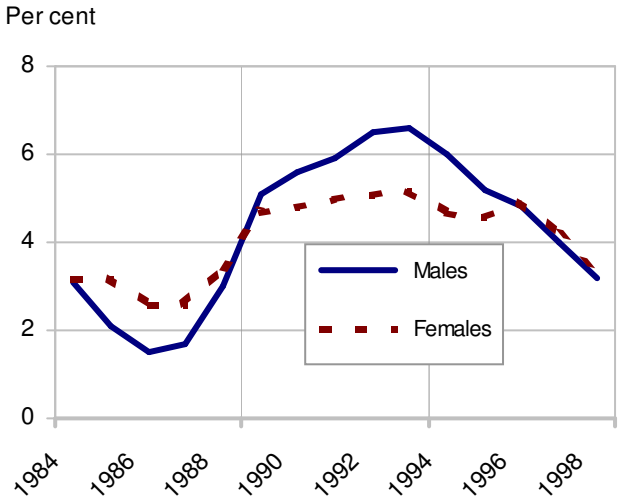




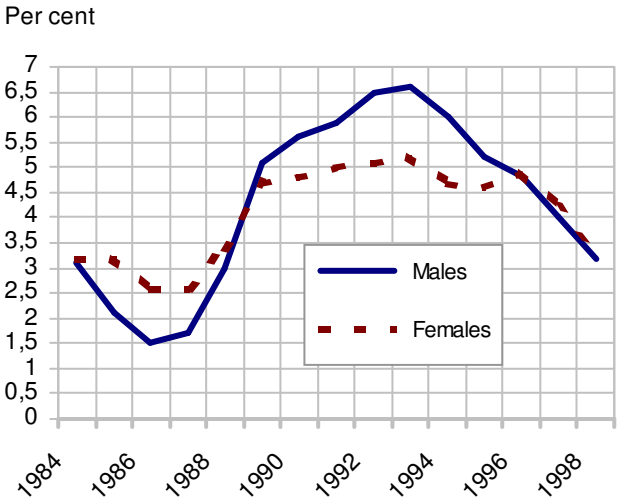
Graphs: Issues and solutions (cont.)

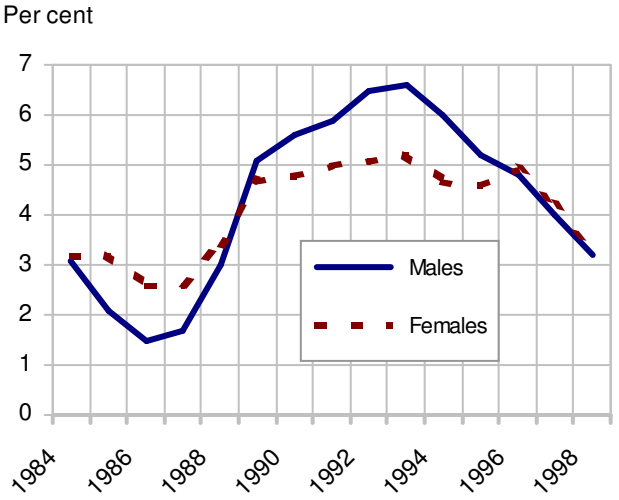
- Use gridlines to improve readability
- But how many?

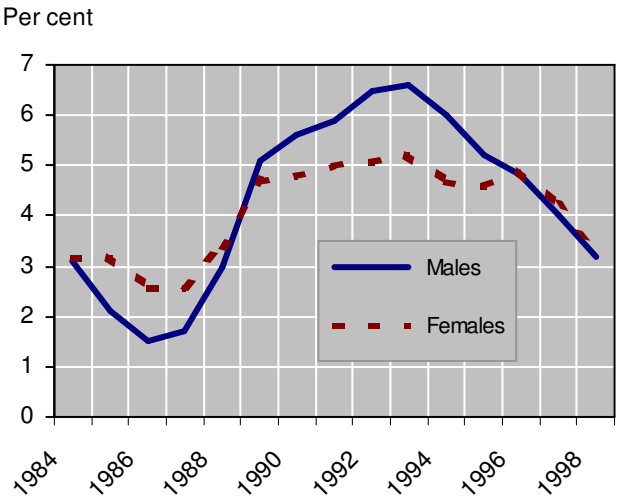


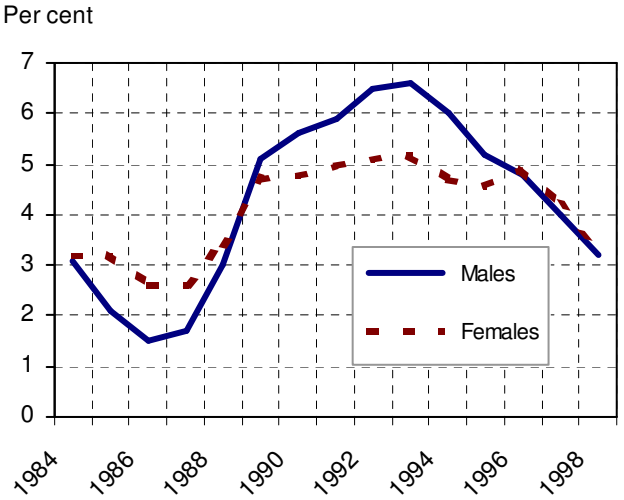


- Use of gridlines (cont.):

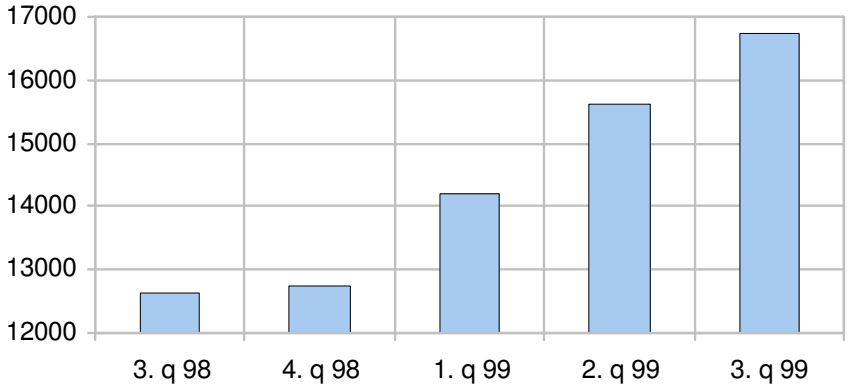




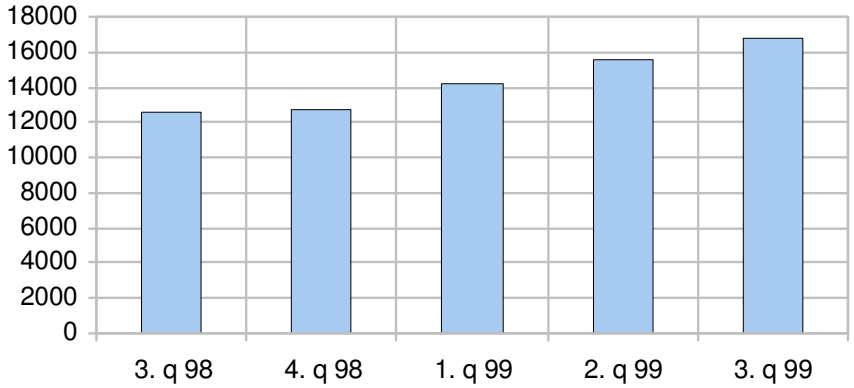




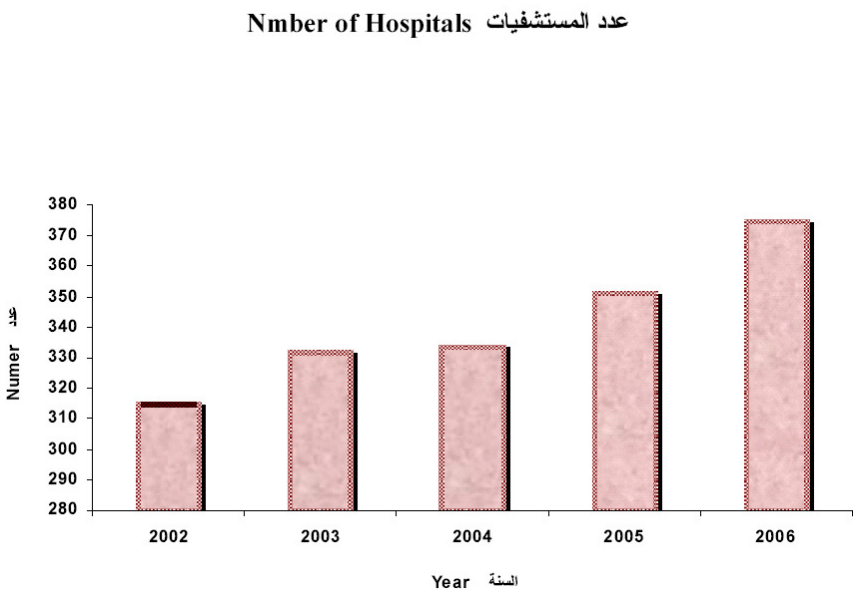
How to lie with graphs

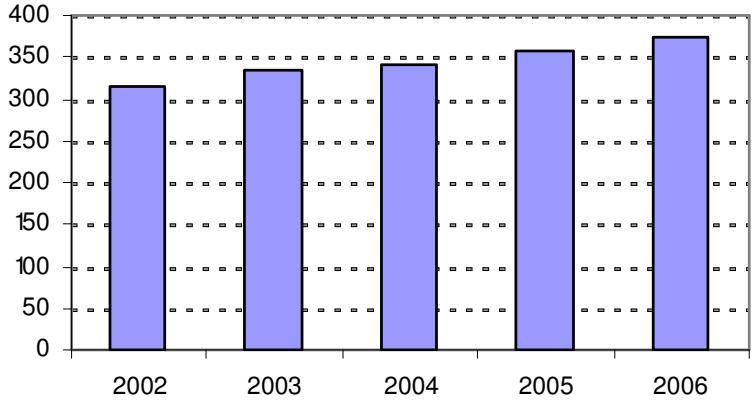


- Increase of 600 % ?!



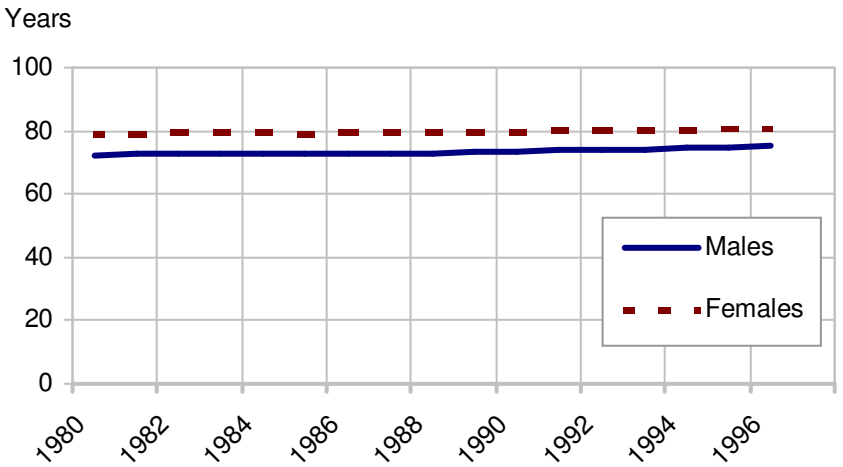
- No, only 35%!

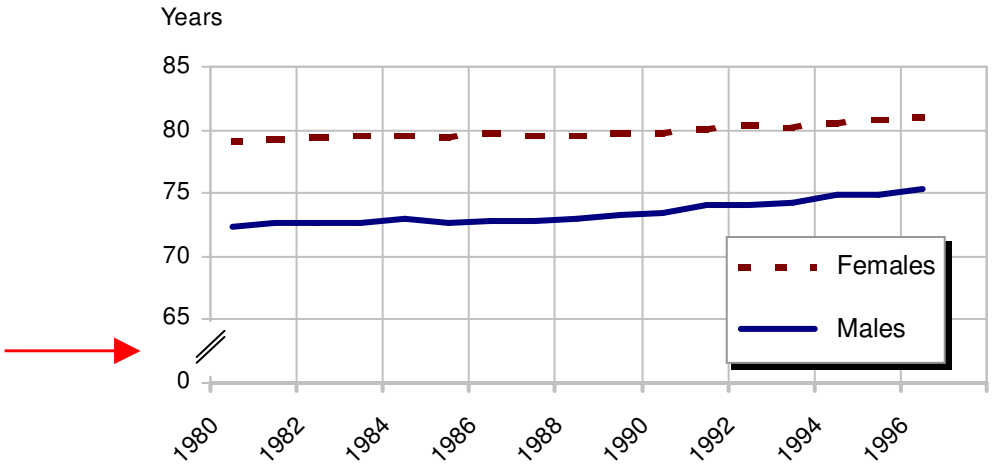




From: Sudan in Figures 2006

But:

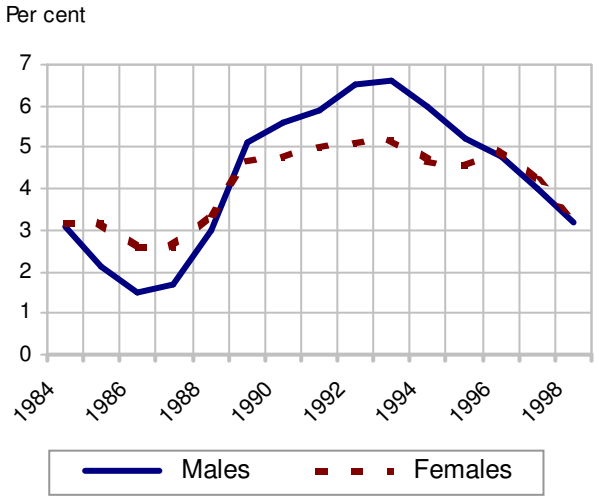


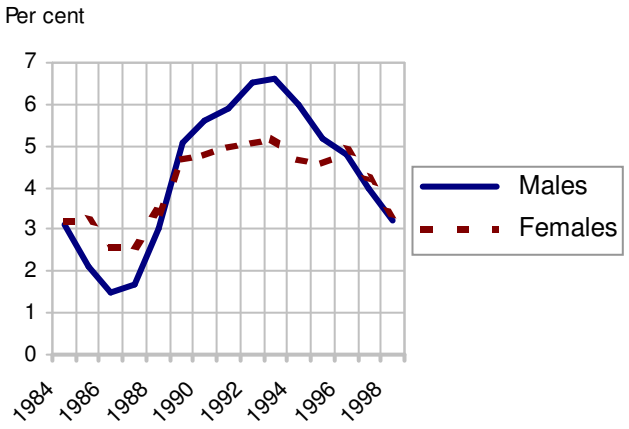


- When you compare groups…

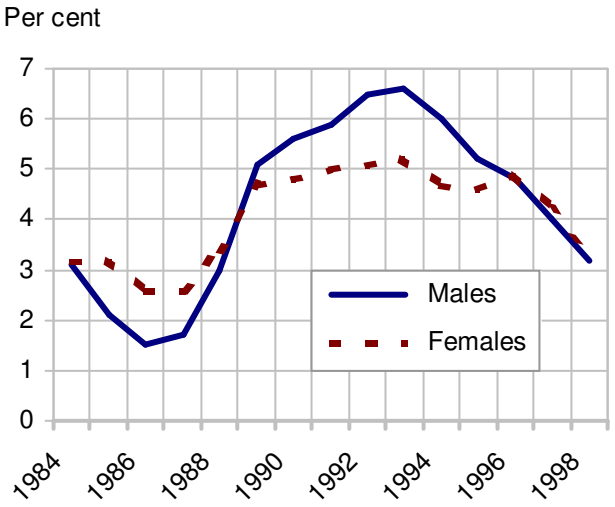
Graphs: Issues and solutions

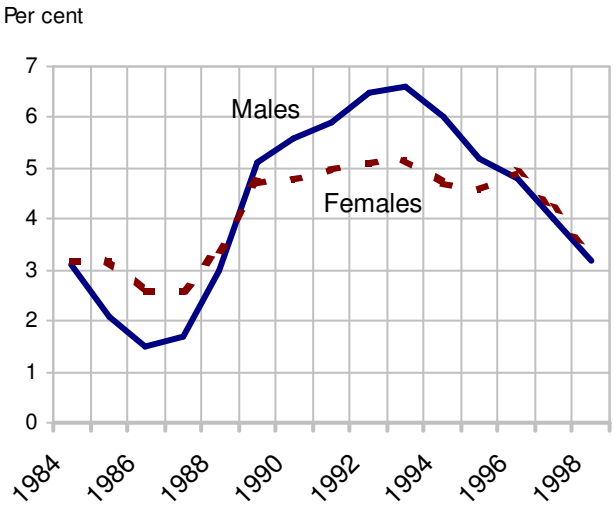
- Legend: Where to place it?





- Legend: Where to place it? (cont.)

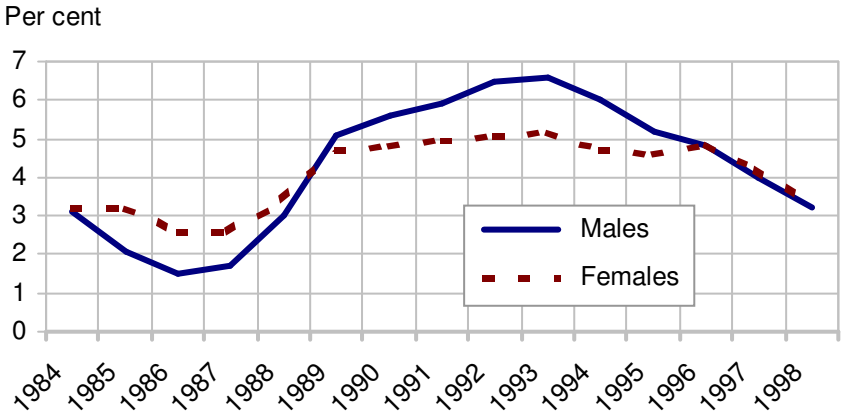




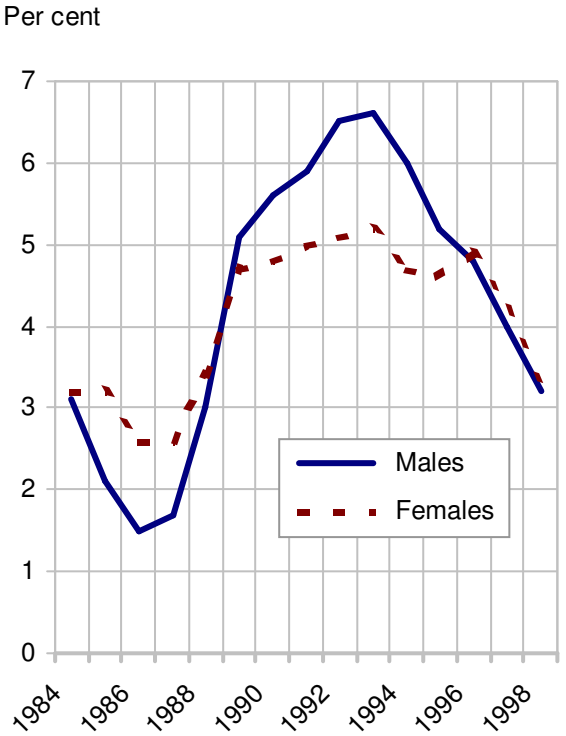
Best solution!

Graphs: Issues and solutions (cont.)

- Format/proportions

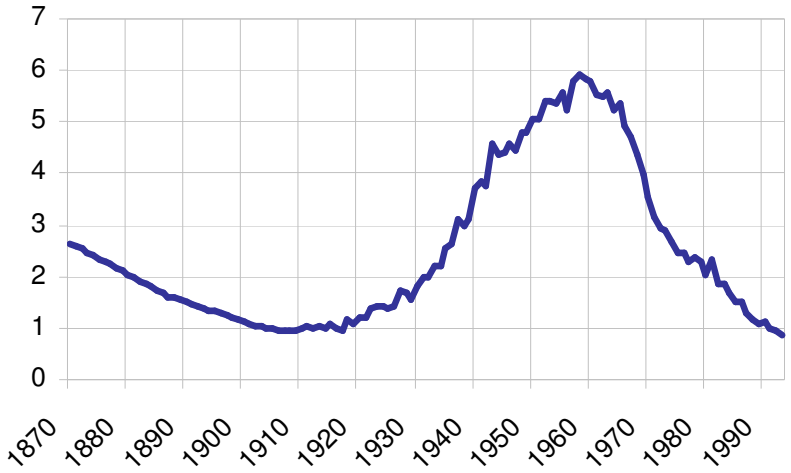


Landscape



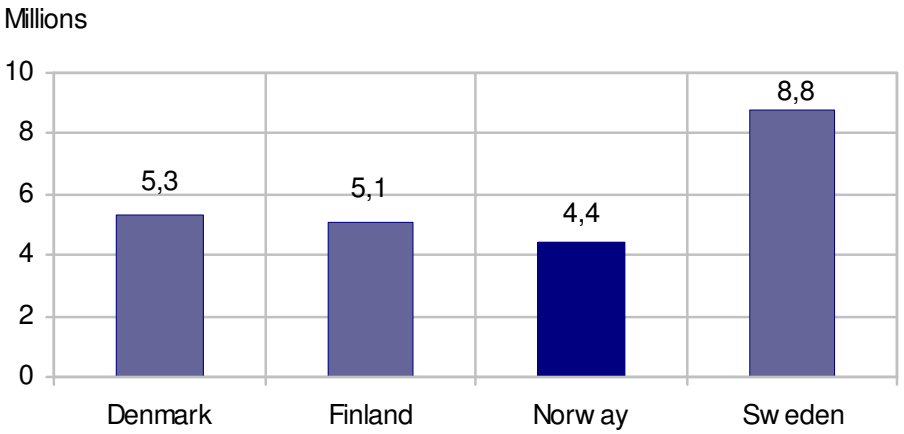
Portrait

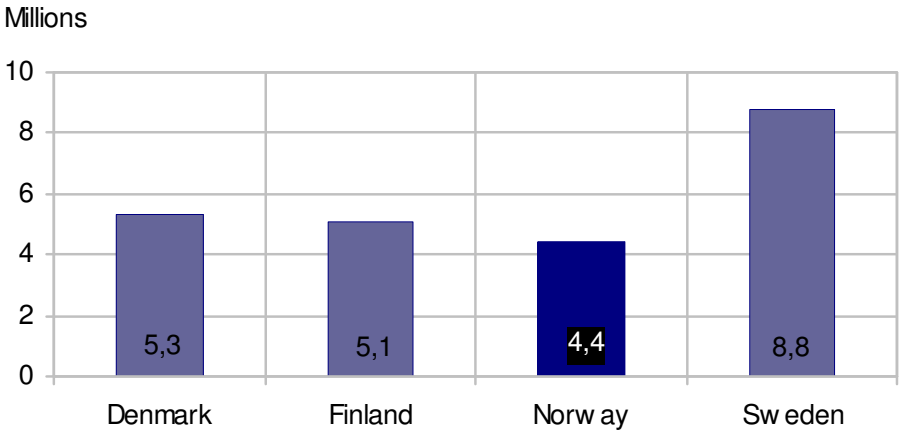
- Format/proportions
  - The most usual format is probably landscape; width x height = ca. 1.5 x 1



Graphs: Issues and solutions (cont.)

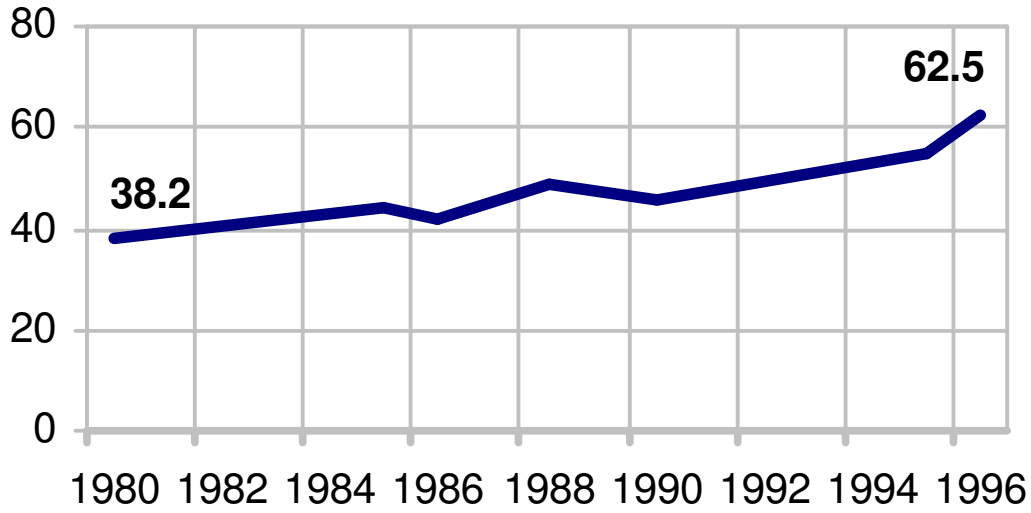
- Should values be given in the graph?





Graphs: Issues and solutions (cont.)

• In line charts, it is often useful to give the first and the latest value:



Main types of graphs:

- Bar charts (vertical)
  - Grouped
  - Stacked
- Bar charts (horisontal)
  - Grouped & stacked
- Line charts + area charts
- Pie charts
- Other types/ combinations

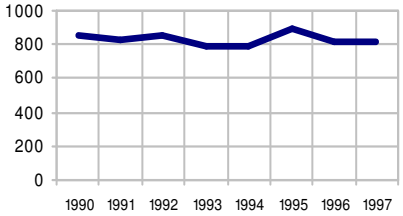
Which type of graph to use?:

Number of adopted children 1990-1997

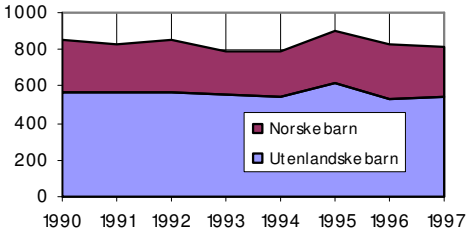
	Total				Norwegian			Foreign		
	Total	Under 3 years	3 - 11 years	12 years and over	Total	Boys	Girls	Total	Boys	Girls
1990	**855**	397	303	155	283	137	146	572	309	263
1991	**833**	418	263	152	271	131	140	562	309	253
1992	**851**	442	251	158	279	149	130	572	284	287
1993	**786**	397	246	143	236	125	111	550	312	238
1994	**788**	426	231	131	239	115	124	549	273	276
1995	**898**	490	257	151	284	141	143	614	272	342
1996	**822**	418	206	198	295	134	161	527	240	287
1997	**814**	469	189	156	272	126	146	542	244	298

Possible charts from the above table:

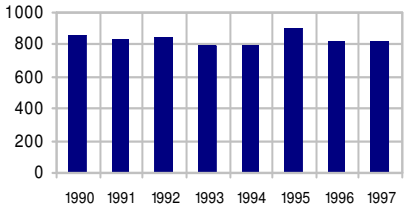
No. of adopted children 1990-1997

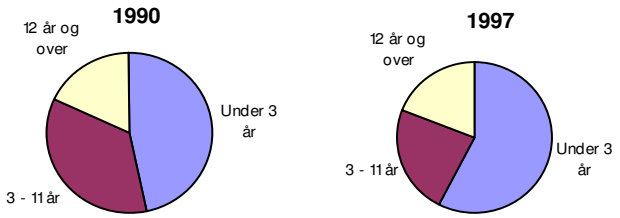


No. of adopted children. Norw egian and foreign 1990-1997

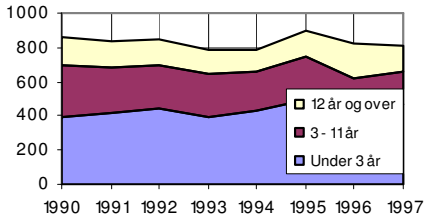


No. of adopted children 1990-1997

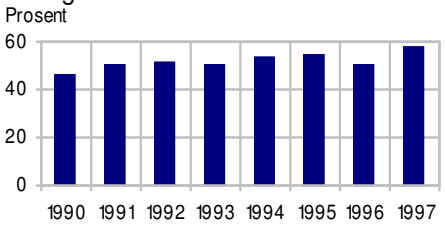




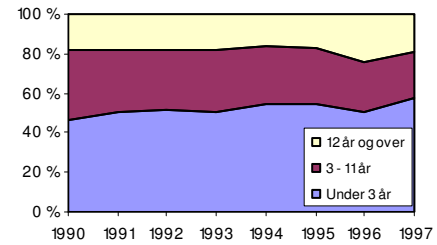
No. of adopted children by age 1990-1997



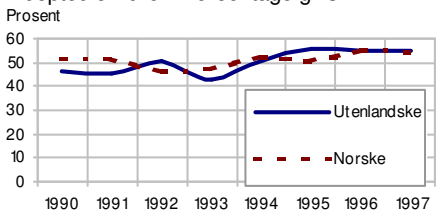
Adopted children. Percentage under the age of 3. 1990-1997



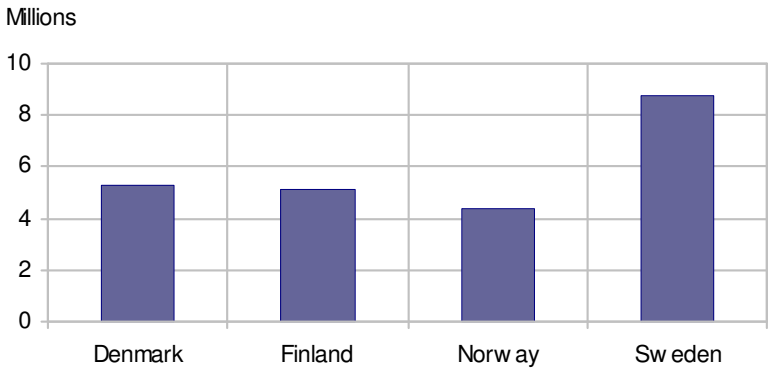
Adopted children by age 1990-1997



Adopted children. Percentage girls.

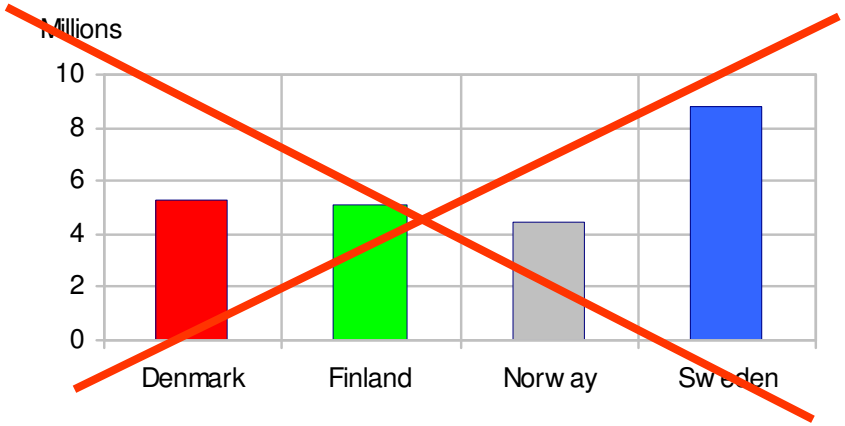


Bar charts (vertical):

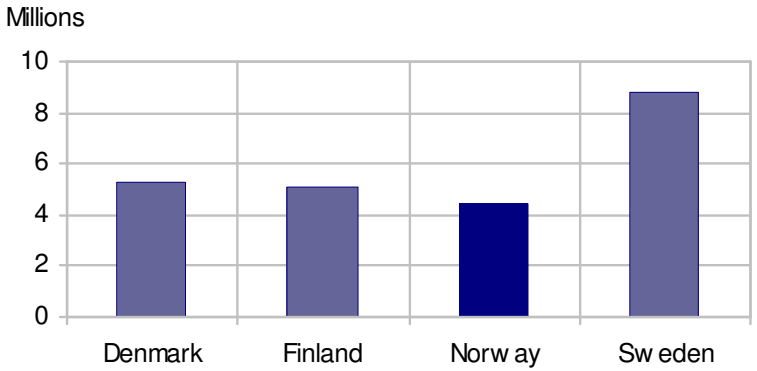


- Also called: Column charts
- The simplest and most basic chart type
- Used to show absolute or relative frequencies, percentages, averages
- Example: *Population in the Nordic countries*

Bar charts (vertical) (cont.):

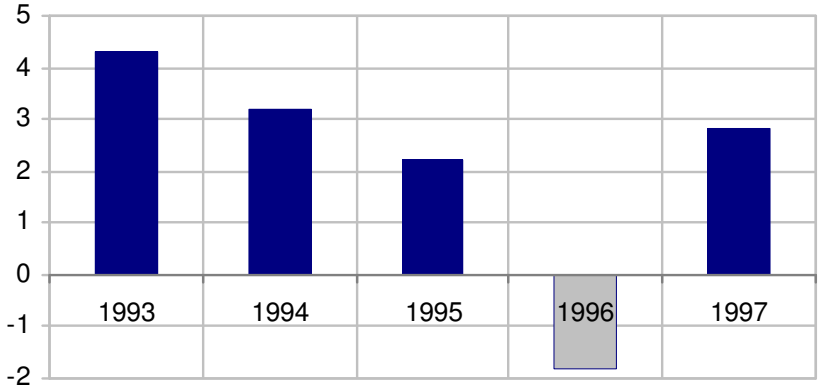


- With only one classification variable it is best to use only one colour/pattern



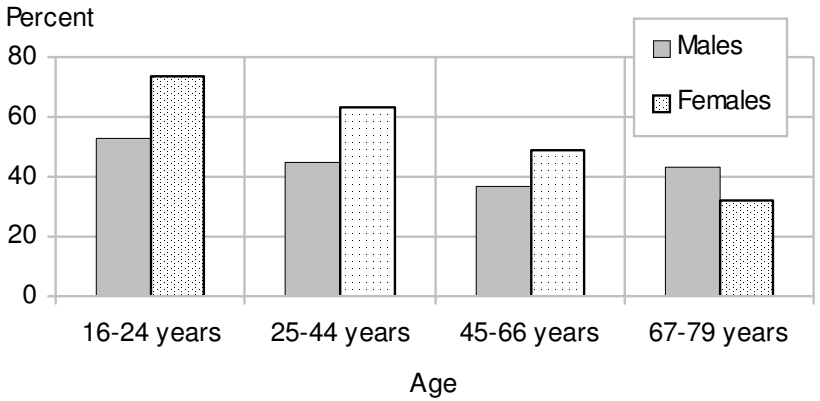
- If you want to focus on one class/group, this can be done by using a different colour/pattern

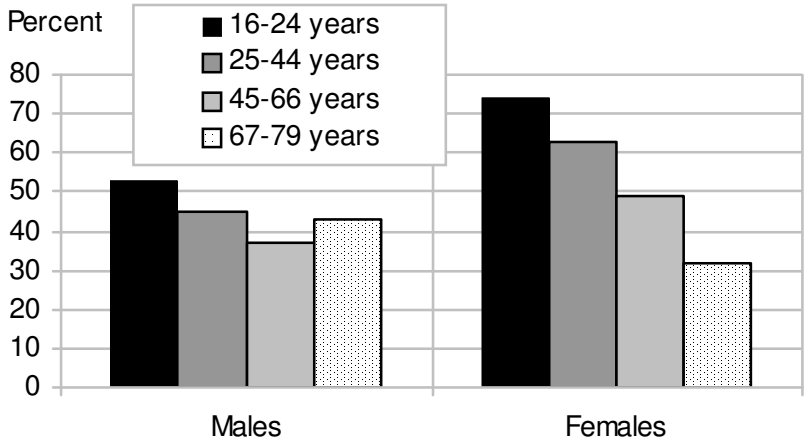
Bar charts (vertical) (cont.):



- When both positive and negative values are plotted, use different colours/ patterns

Grouped bar charts:



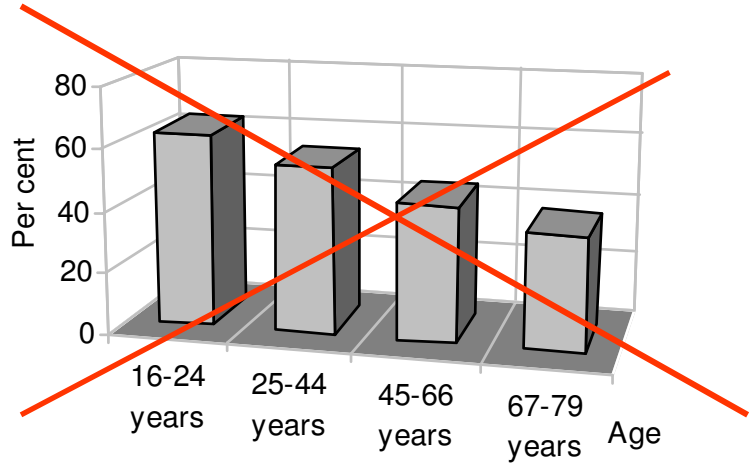


- Two (or more) categories

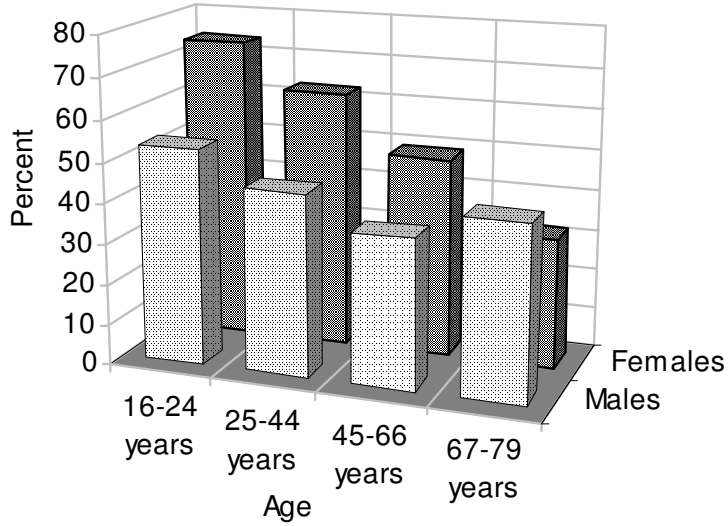
  Example: *Percent using a library last year, by sex and age*

- With two categories, we have two ways of grouping, inviting to different comparison

"3D" bar charts:



- Popular, but not recommended



- "Genuine" 3D-chart (3 variables)

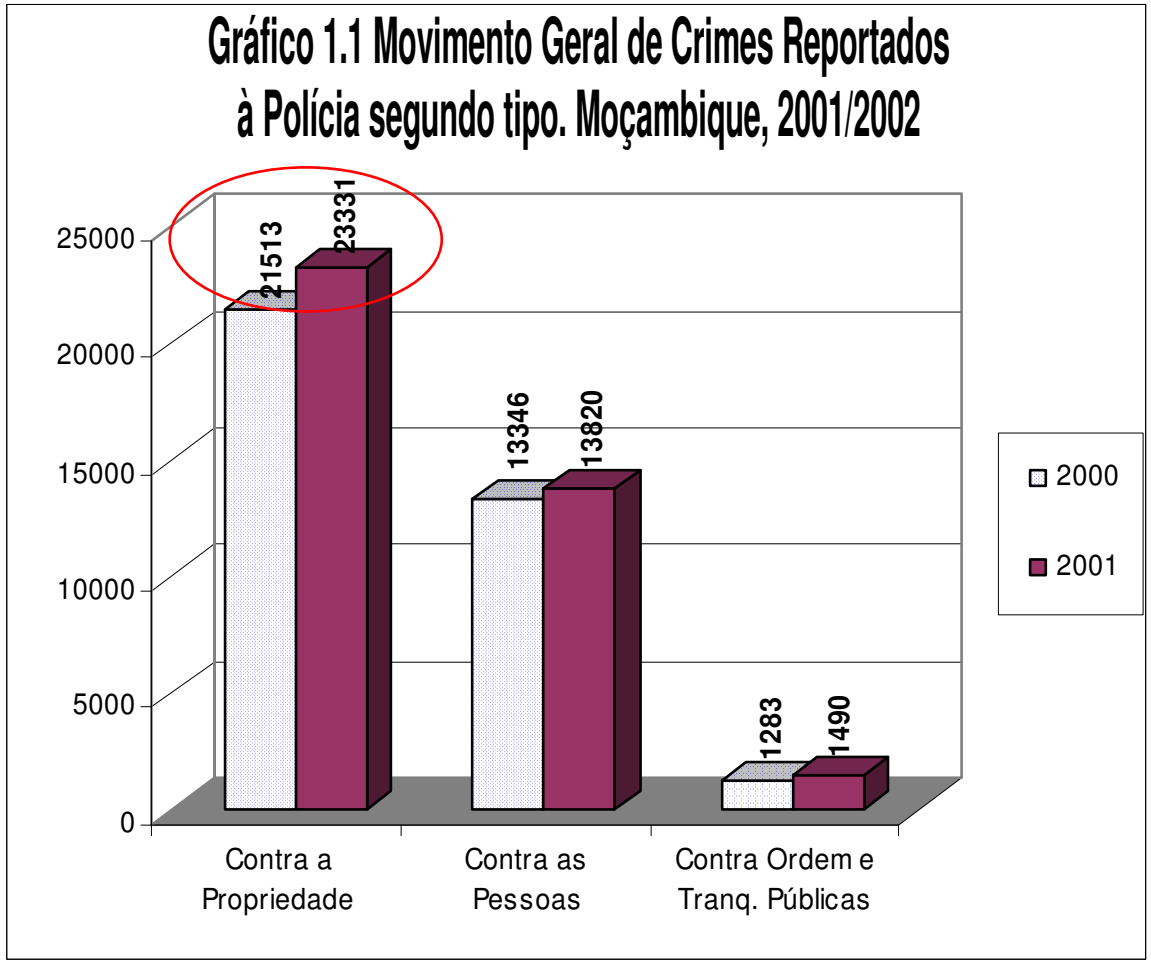
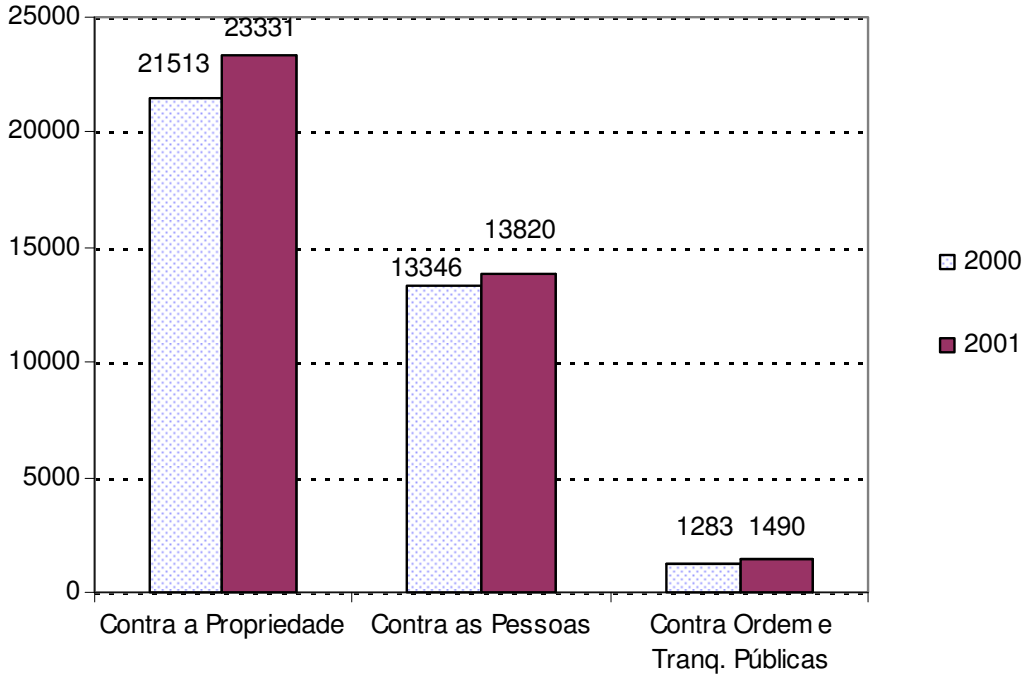
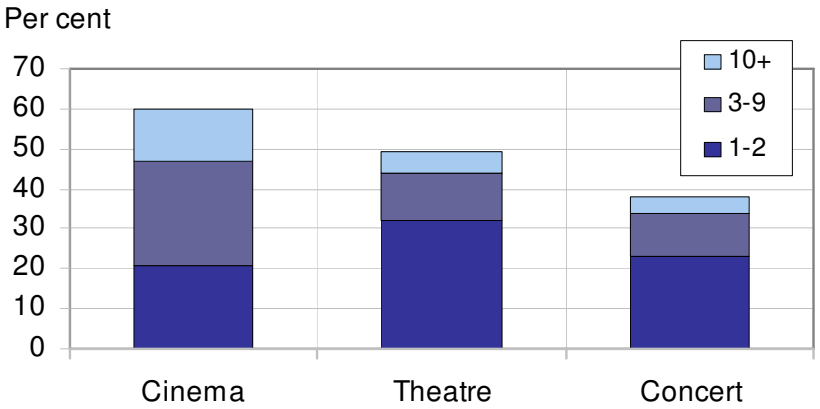


Gráfico 1.1 Movimento Geral de Crimes Reportados à Polícia segundo tipo. Moçambique, 2001/2002
25000
20000
15000
10000
5000
0
21513
23331
13346
13820
1283
1490
Contra a Propriedade
Contra as Pessoas
Contra Ordem e Tranq. Públicas
2000
2001

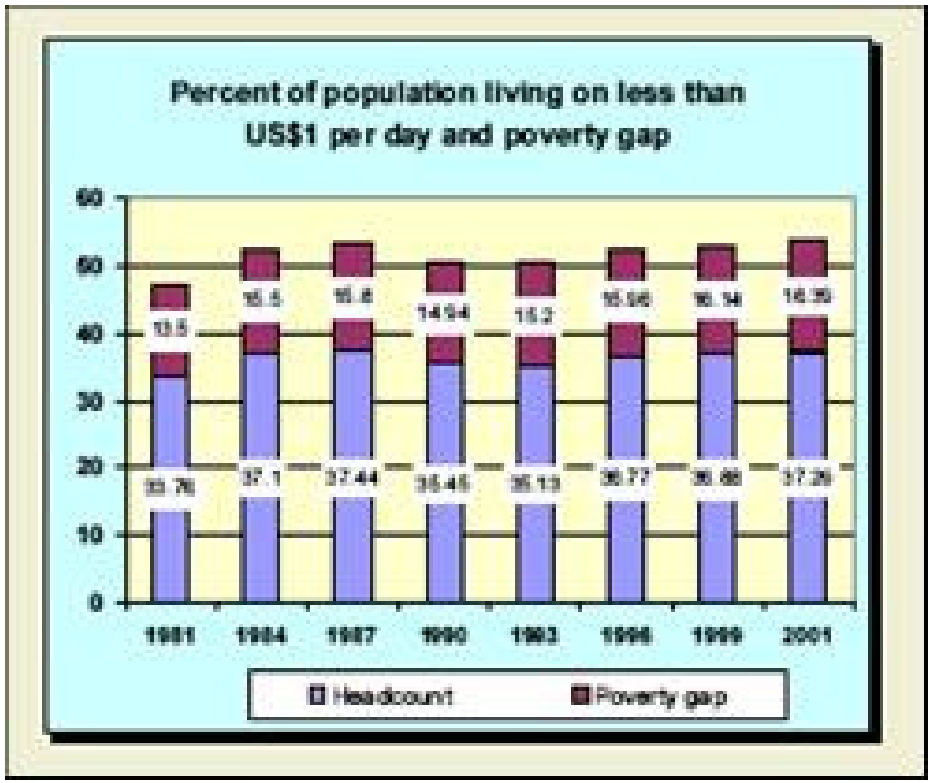
Gráfico 1.1: Crimes Reportados à Polícia segundo tipo.
2000 - 2001



Stacked bar charts

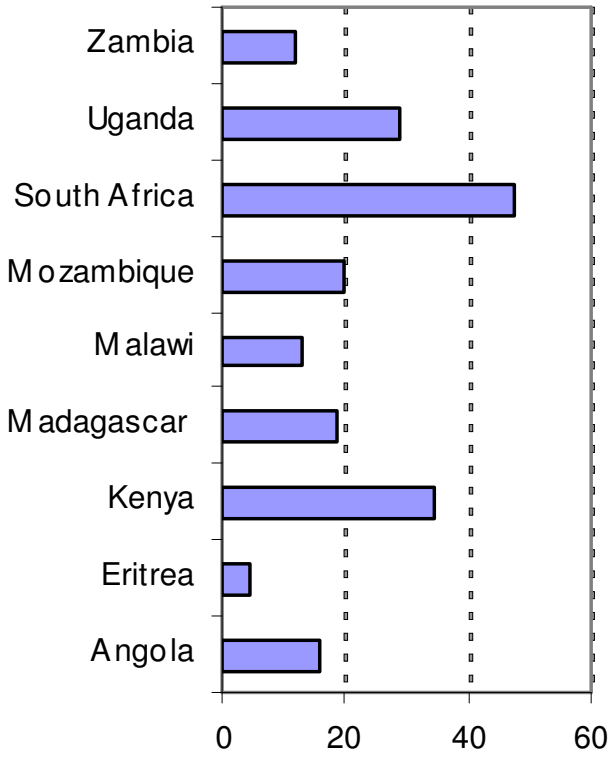


- Show total frequency and how the total is divided into different components



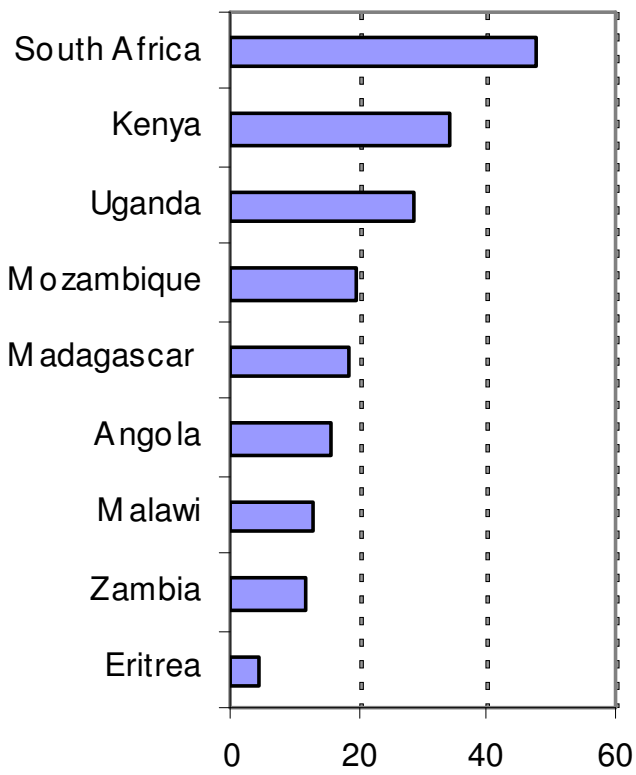
From: African Development Bank 2006. Gender, Poverty and Environmental Indicators on African Countries, page 37

Horizontal bar charts:



- Often used
  - when labels (variable names) are long
  - when there are many variables or classifications
- Here, text should be right justified

Horizontal bar charts:

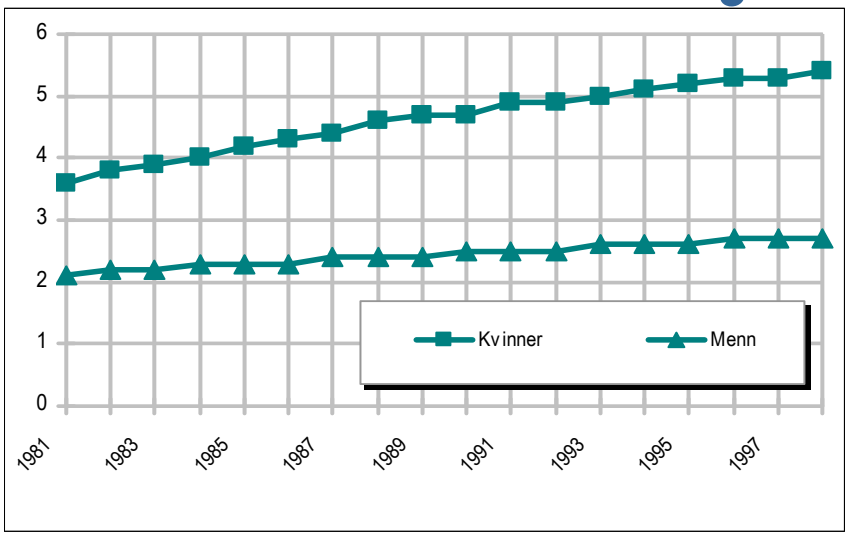


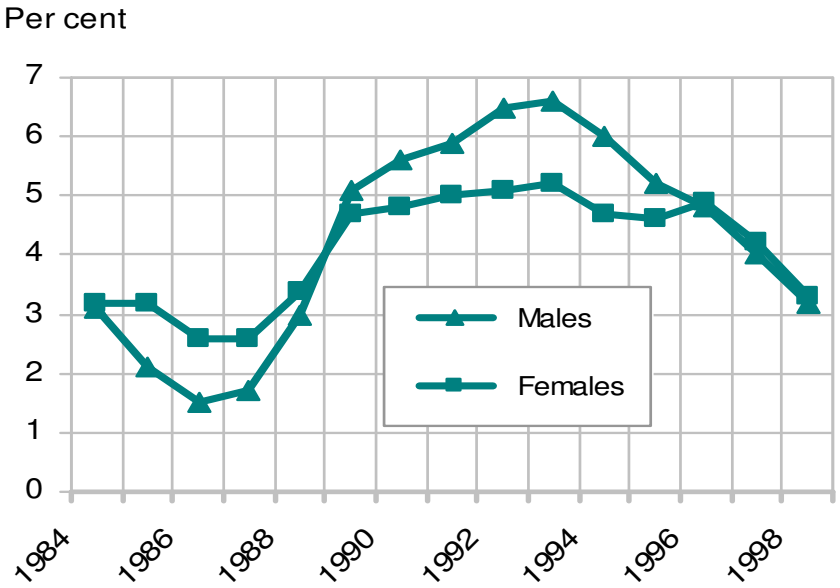
- Often it is better to sort by the value of the indicator (dependent variable)

Line charts:

- Most often used for time series
- Time = years, quarters, months, weeks, days, hours and minutes + age
- What is a time "series"? Minimum = 4 data points? Up to 4, use vertical bar chart
- The longer series, the better?!

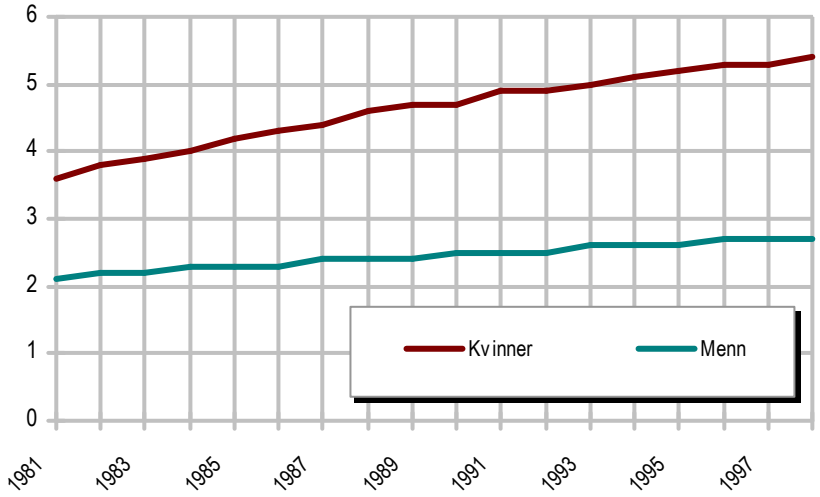
Line charts: Symbols?

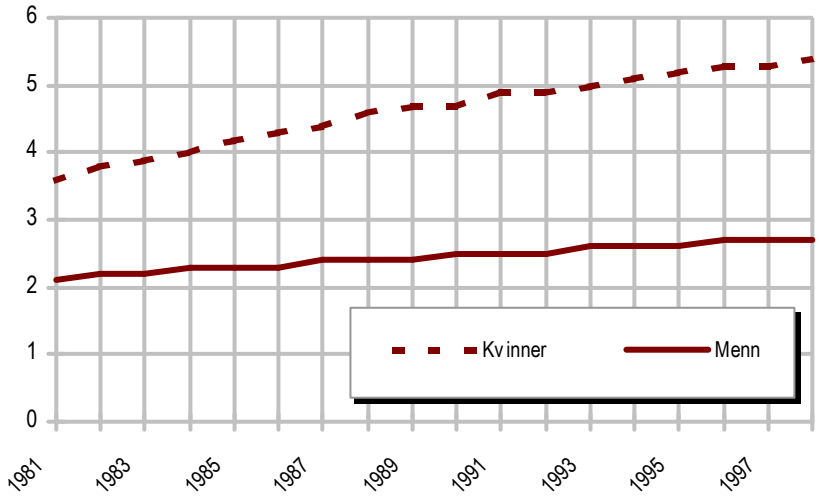




- Indicators or symbols (■ ∇) are often used to differentiate between series, but these symbols overload the chart, especially with crossing time series

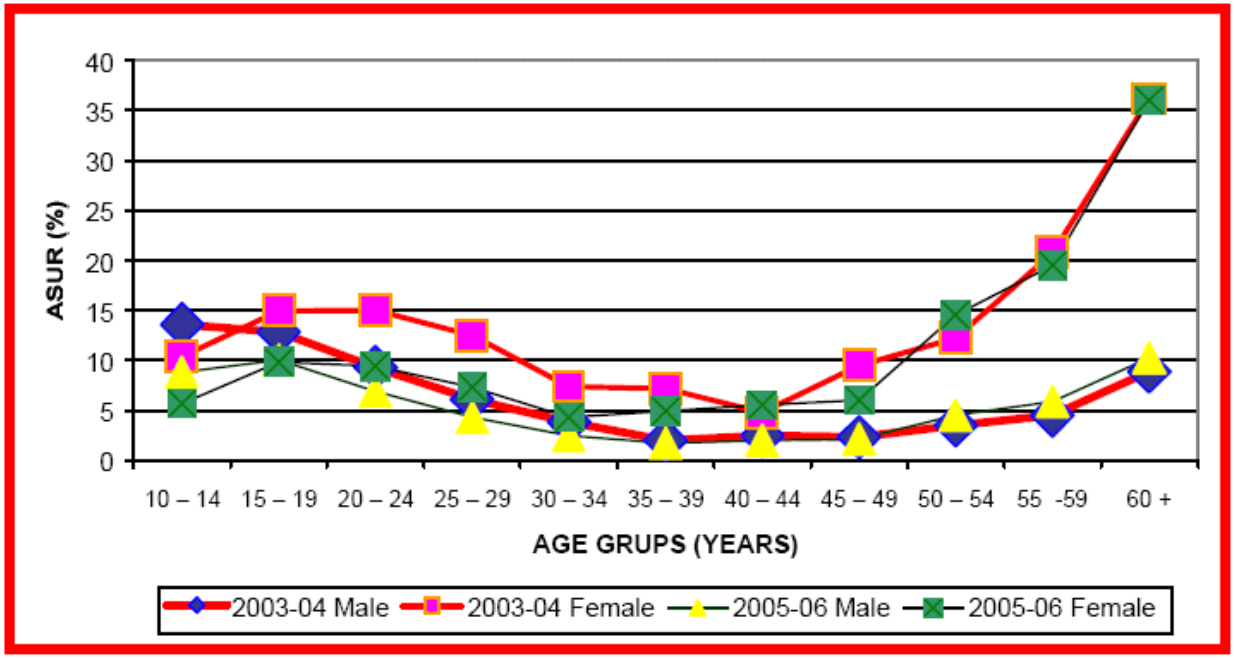
Line charts:



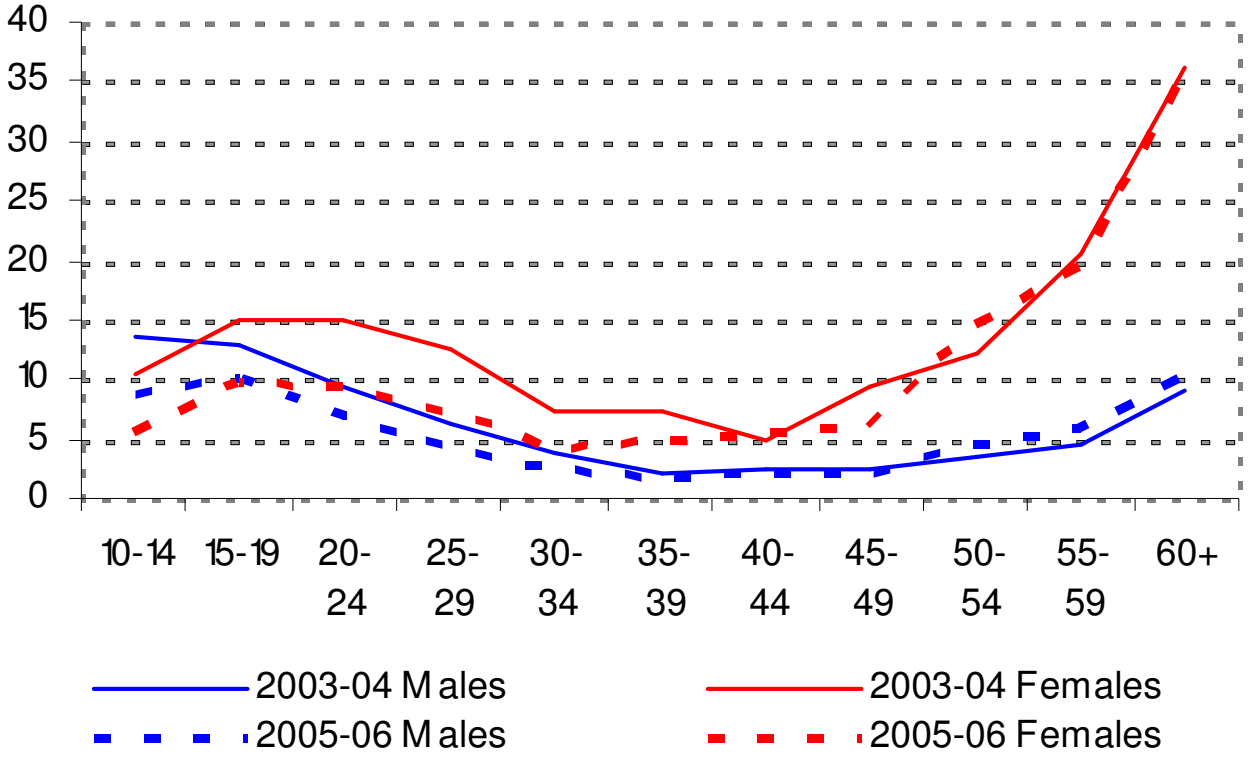


- Better to use different colours ...
- ... and/or line styles

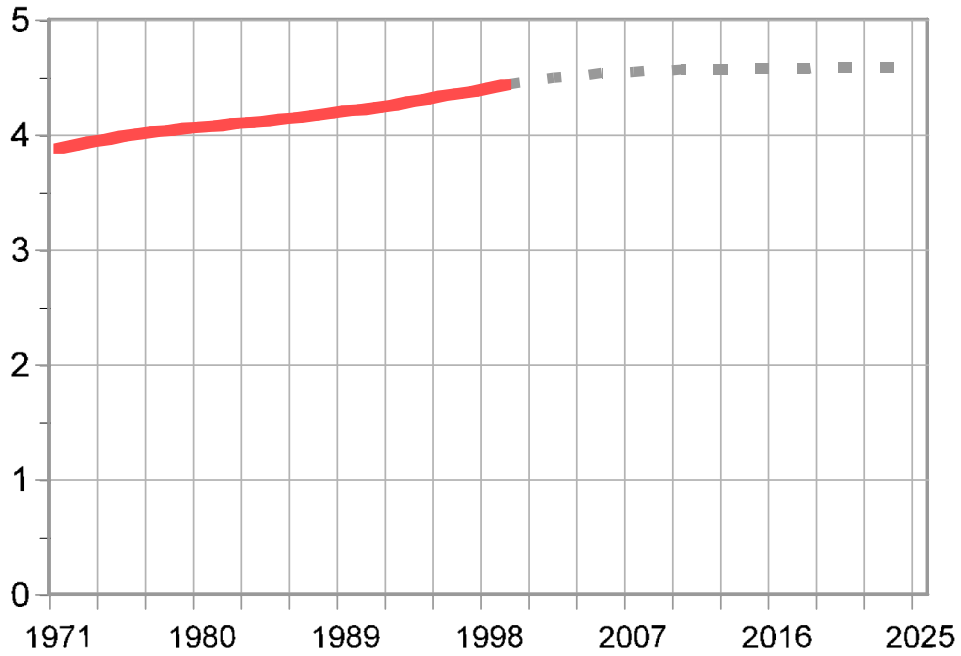
FIGURE-5: AGE SPECIFIC UNEMPLOYMENT RATES (ASUR) BY SEX FOR PAKISTAN



From: Pakistan Labour Force Survey 2005-2006

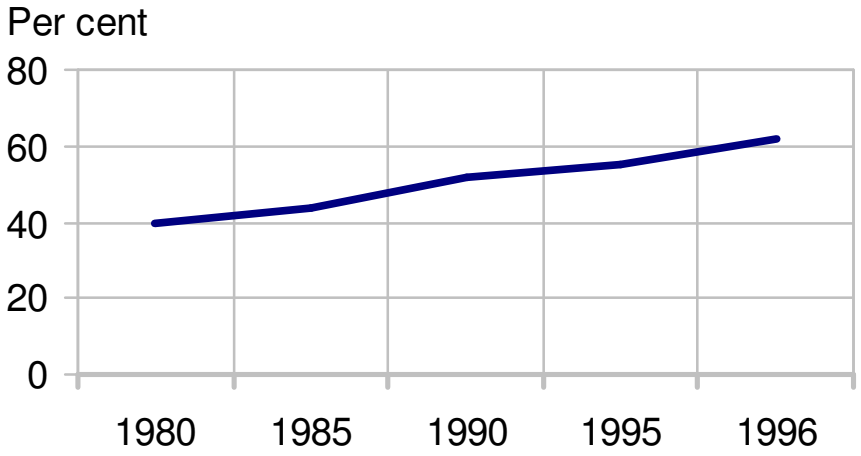

40
35
30
25
20
15
10
5
0
10-14
15-19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60+
2003-04 Males
2003-04 Females
2005-06 Males
2005-06 Females

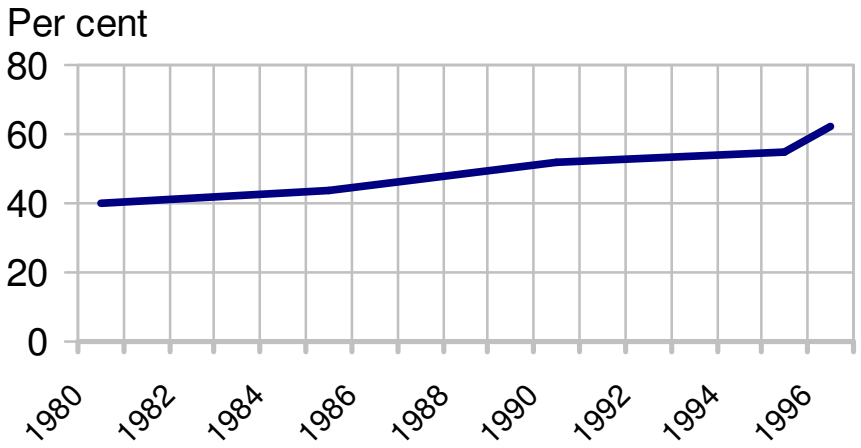
Line charts:



- When showing projections, the projection portion should be clearly differentiated graphically from the rest of the curve

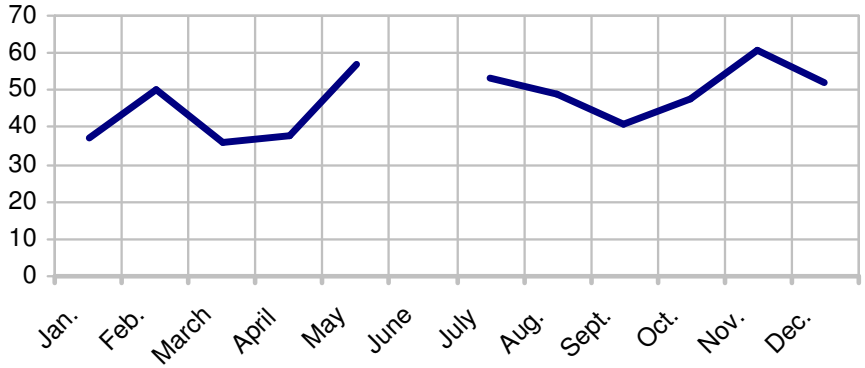
Standard line charts

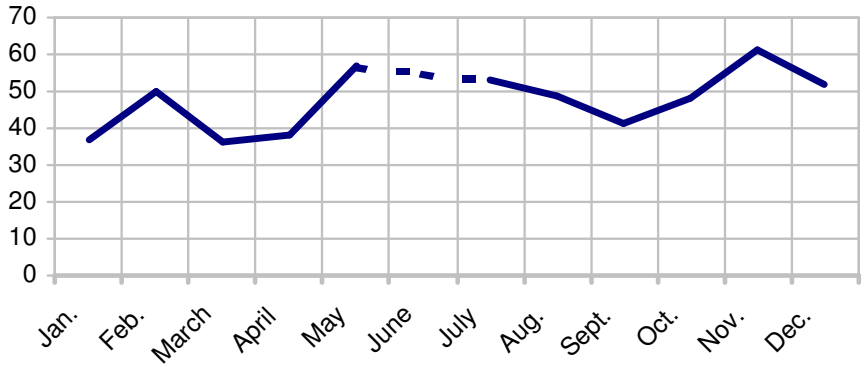




- "A time series" requires at least 4? data points (if not: use bar chart)
- Be careful when the data points are not evenly spaced; like 1980, 1985, 1990, 1995 and 1996

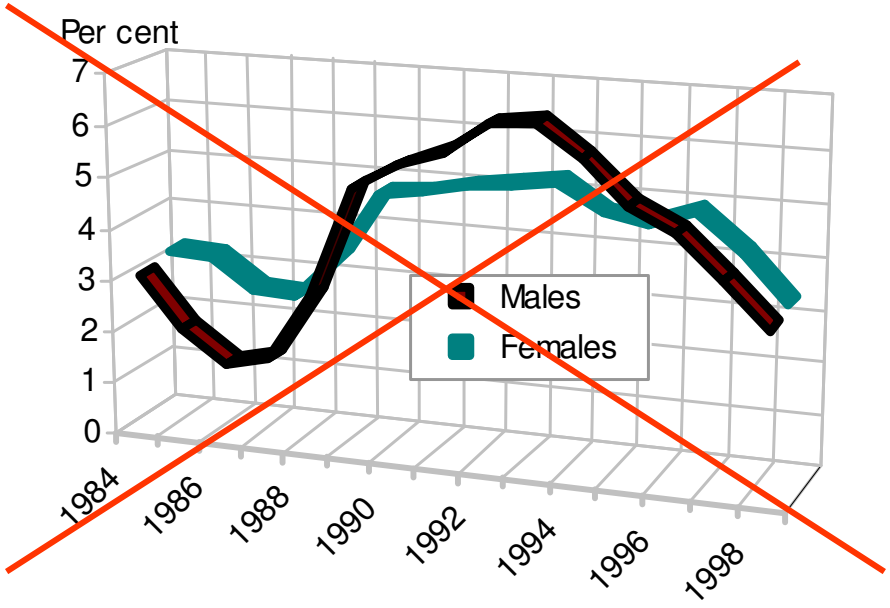
Missing data in time series





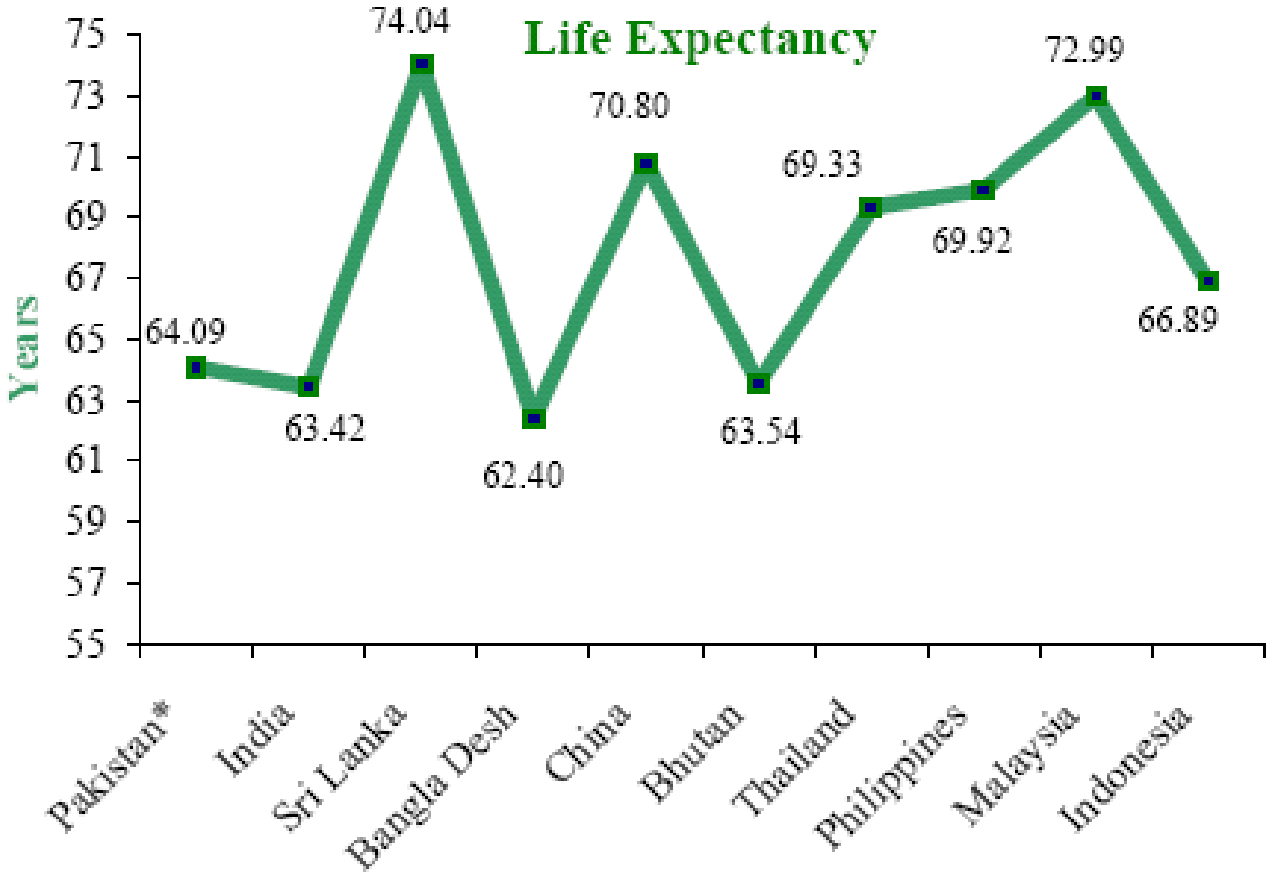
- The gap should be bridged with a dashed line

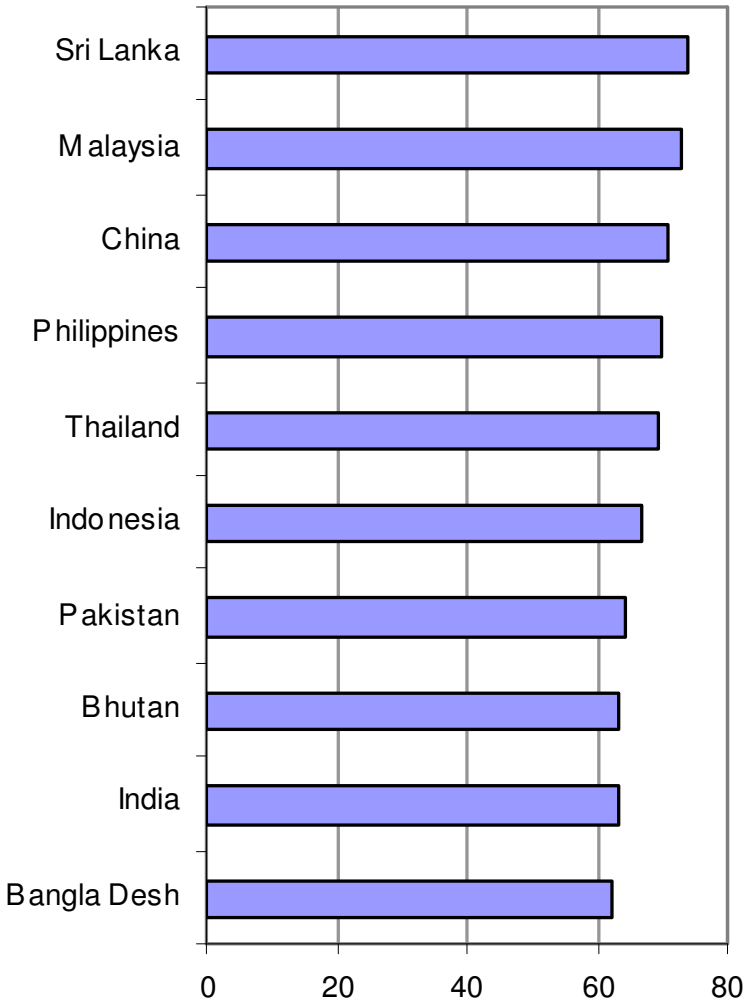
“3D” line charts?



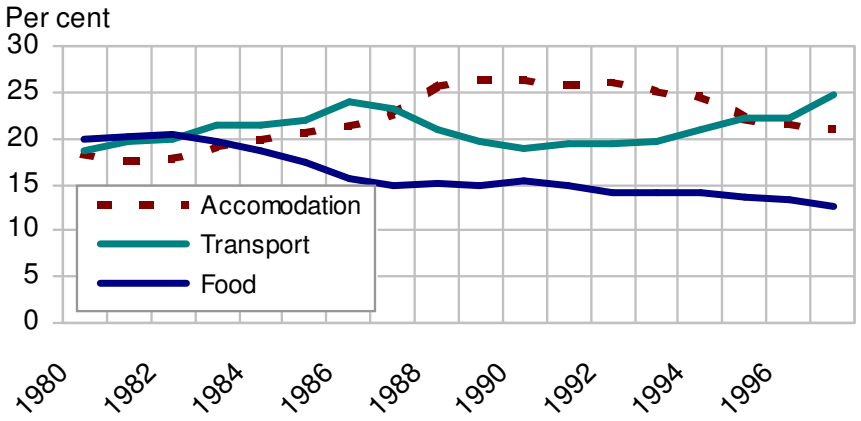
- **NO!!**

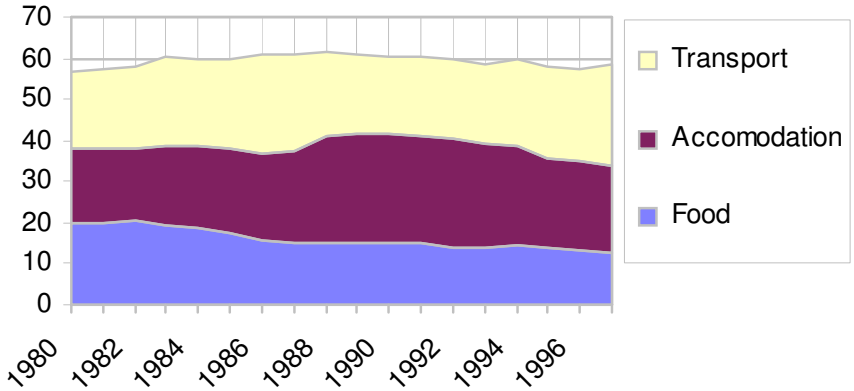
Should *never* be uses to illustrate differences between group, for instance countries




Sri Lanka
Malaysia
China
Philippines
Thailand
Indonesia
Pakistan
Bhutan
India
Bangla Desh
0
20
40
60
80

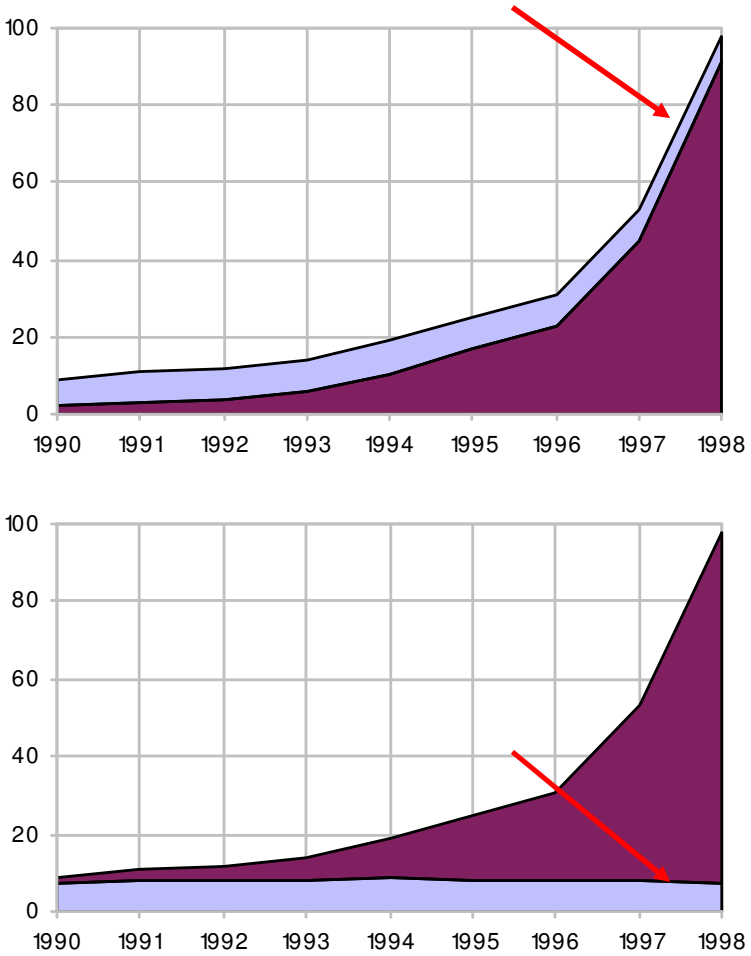
Area chart





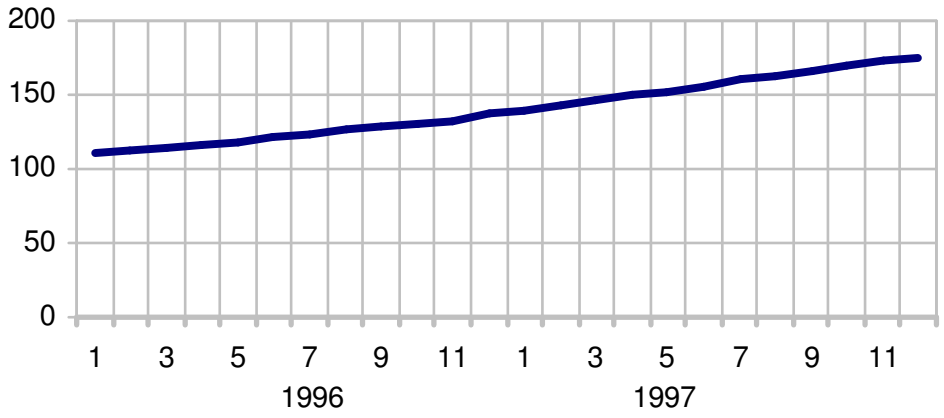
- Area charts are accumulated line charts; like stacked bars.
- Shows total and parts
- Don't use with too many groups/values

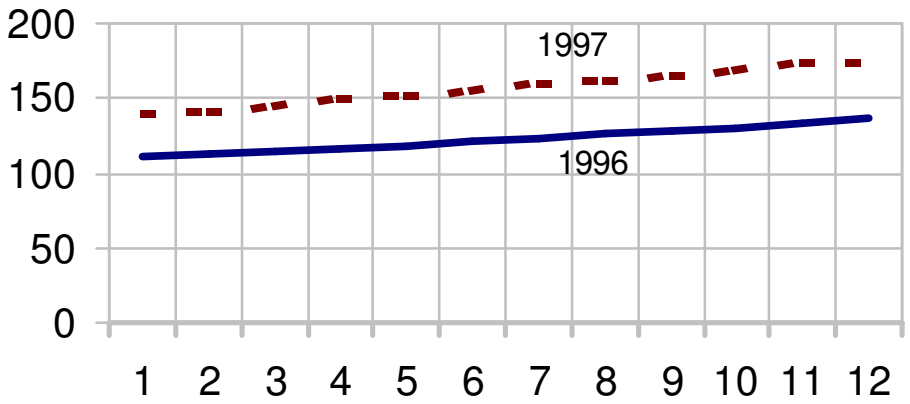
Area chart



- Order of series may be important

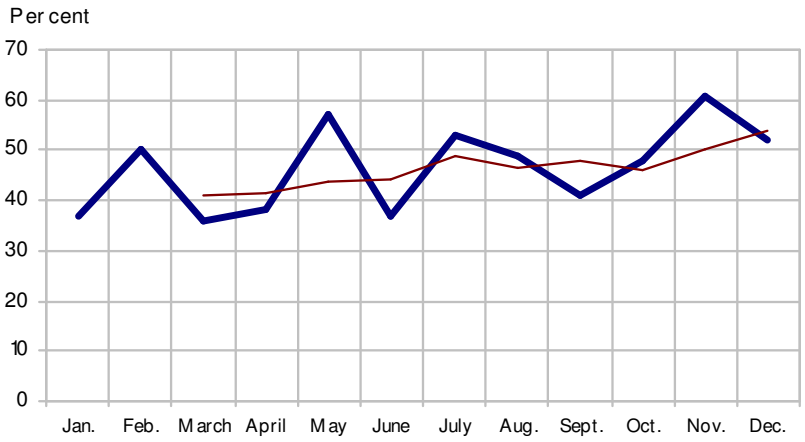
Periodical data





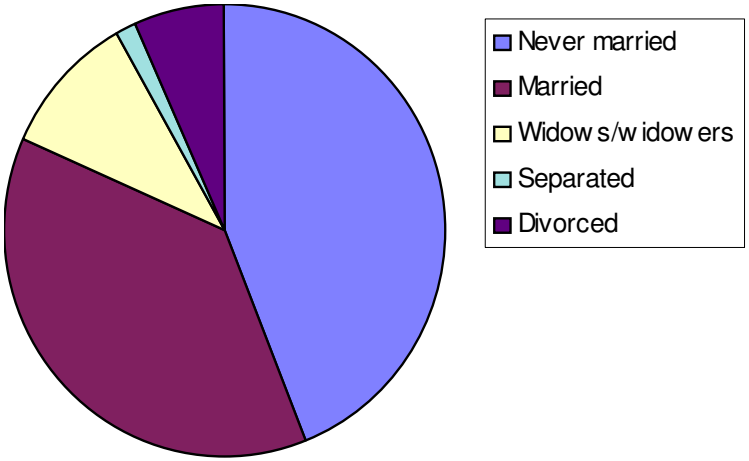
- Weekly, monthly and quarterly data. Monthly data for one year can be presented as a standard line chart
- With two or more years, we have two alternatives

Curve smoothing



- Often used with periodical data
- Moving average (here: 3 years)

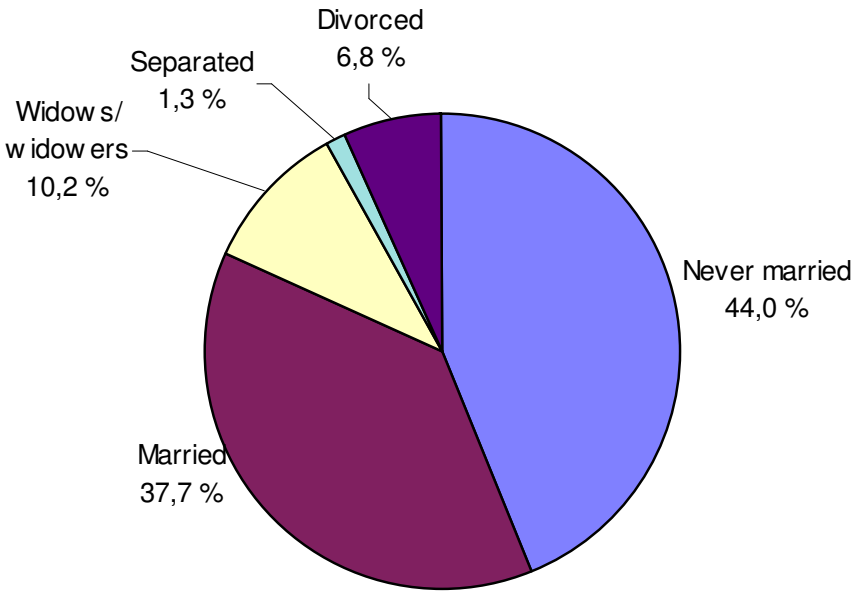
Pie chart



- Shows distribution of (qualitative) variables
- No axis
- Total area = 100%

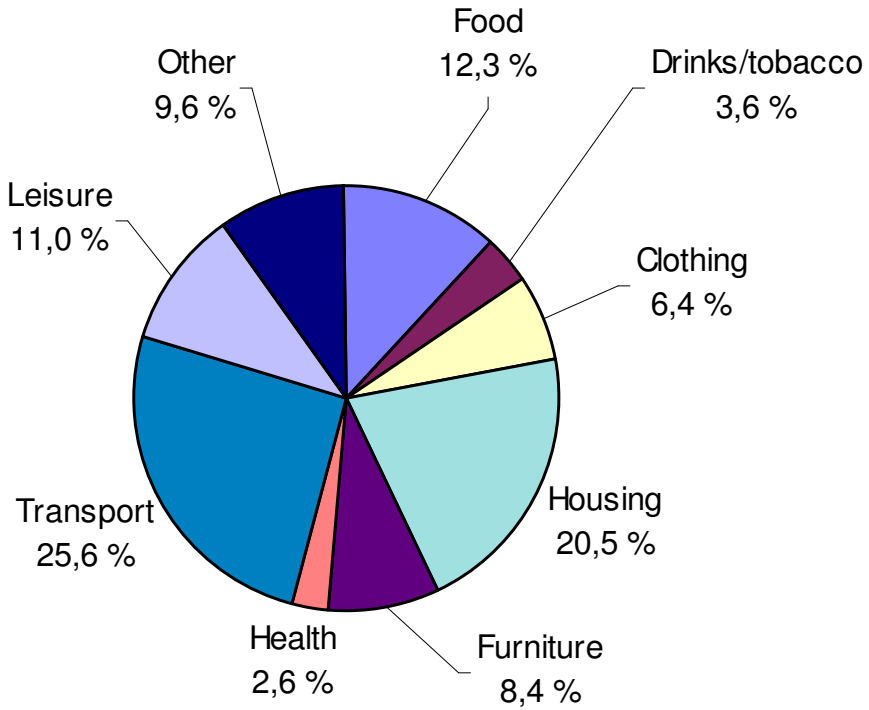
- Instead of legend....

Pie chart



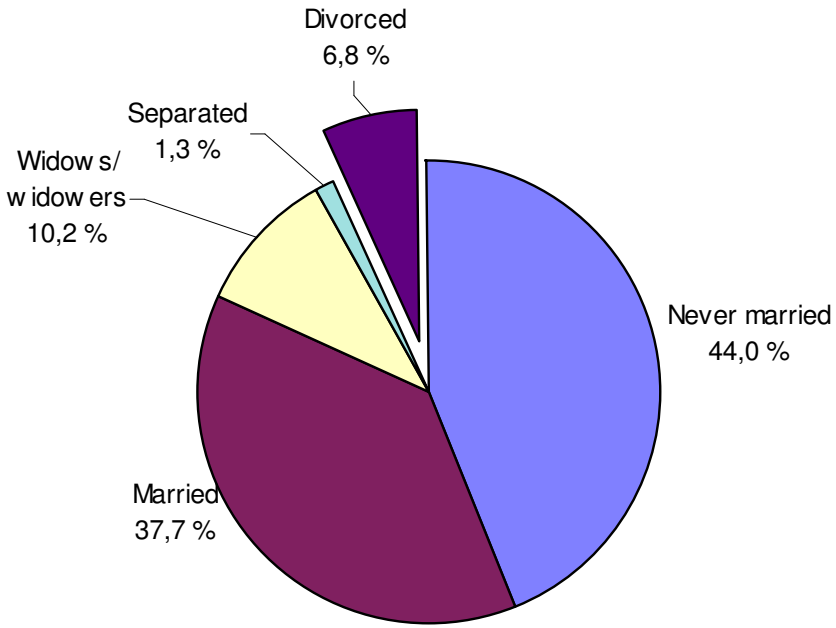
- ...use labels
- Since pie charts have no value axis, show percentage
- Maximum 5 values?

Pie chart



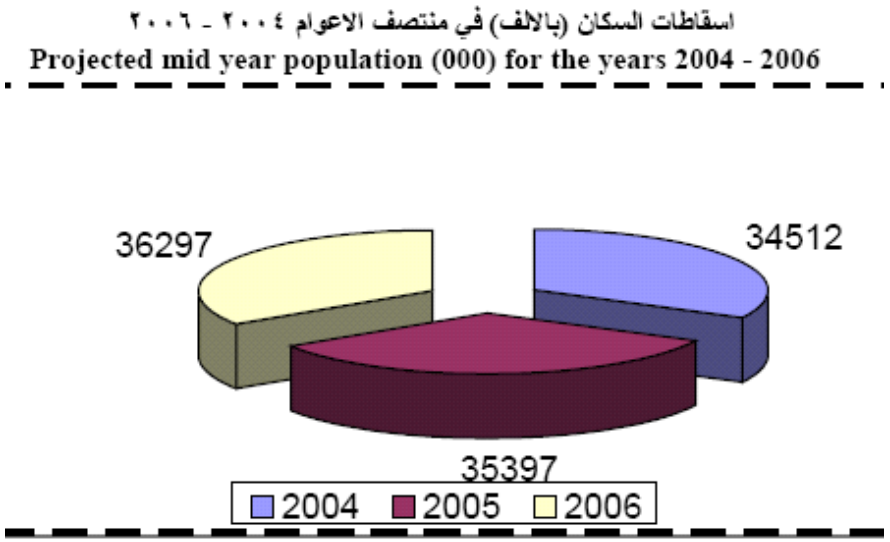
- With too many values, the reader looses interest

Pie chart



- If we want focus on or draw attention to a special group/sector, this can be done by “exploding”

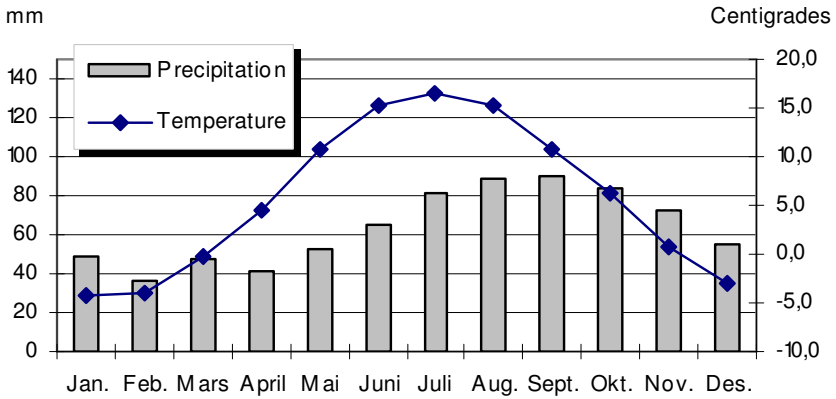
Pie chart

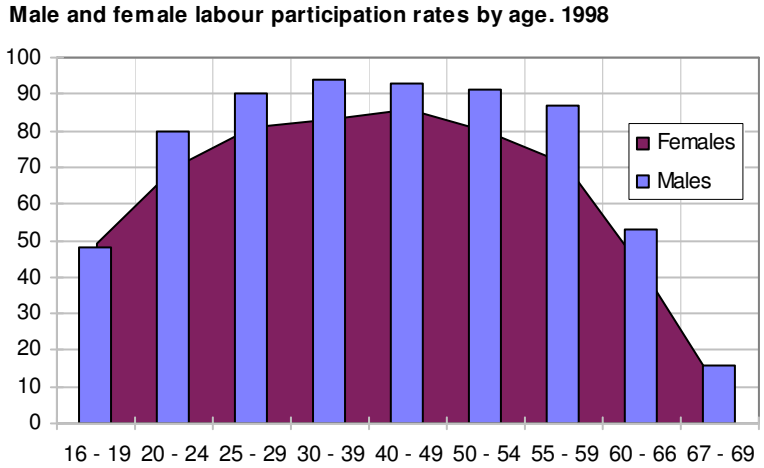


Statistical Yearbook of Sudan 2006

- Pie charts should *never* be used to illustrate time series

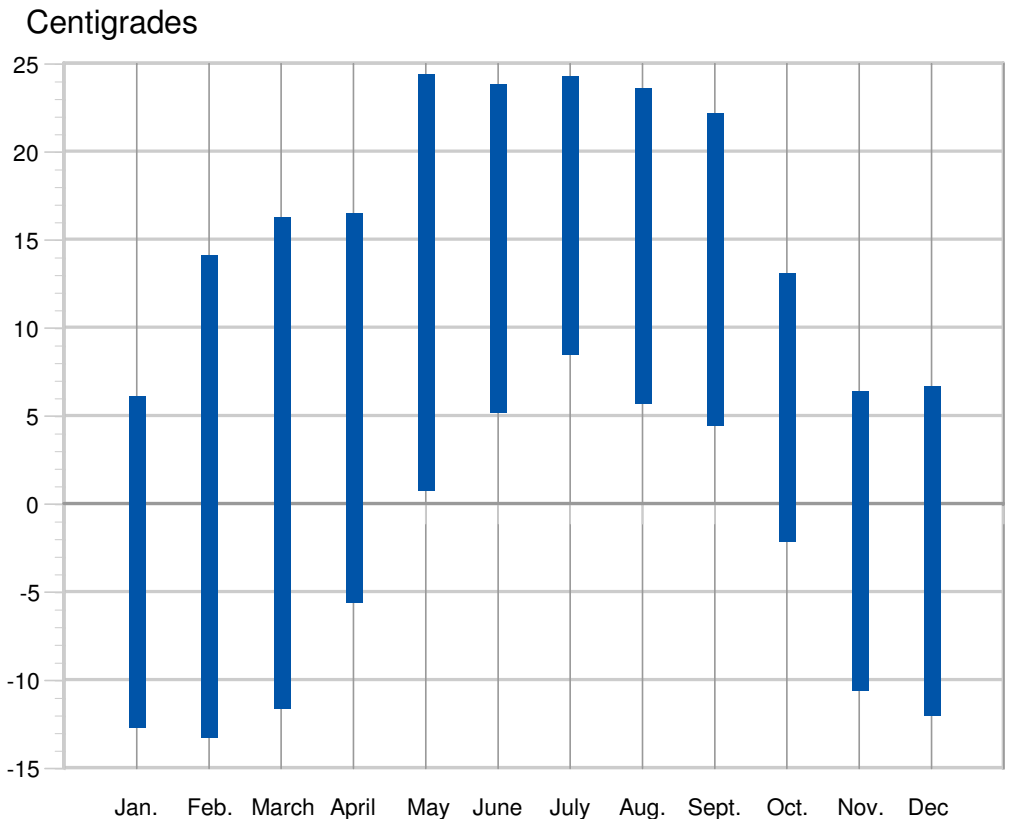
Other types





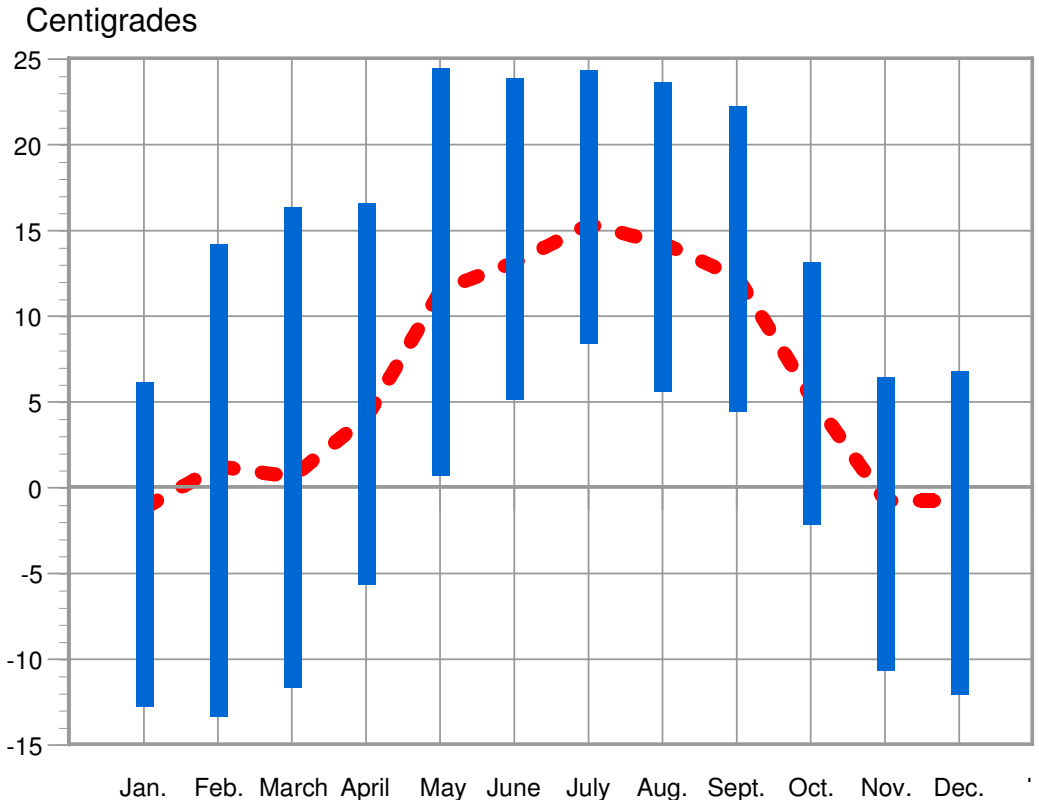
- Combination graphs (also called “overlay”):
  - Bar & line
  - Bar & area

Other types



- Min. - max. (floating column)

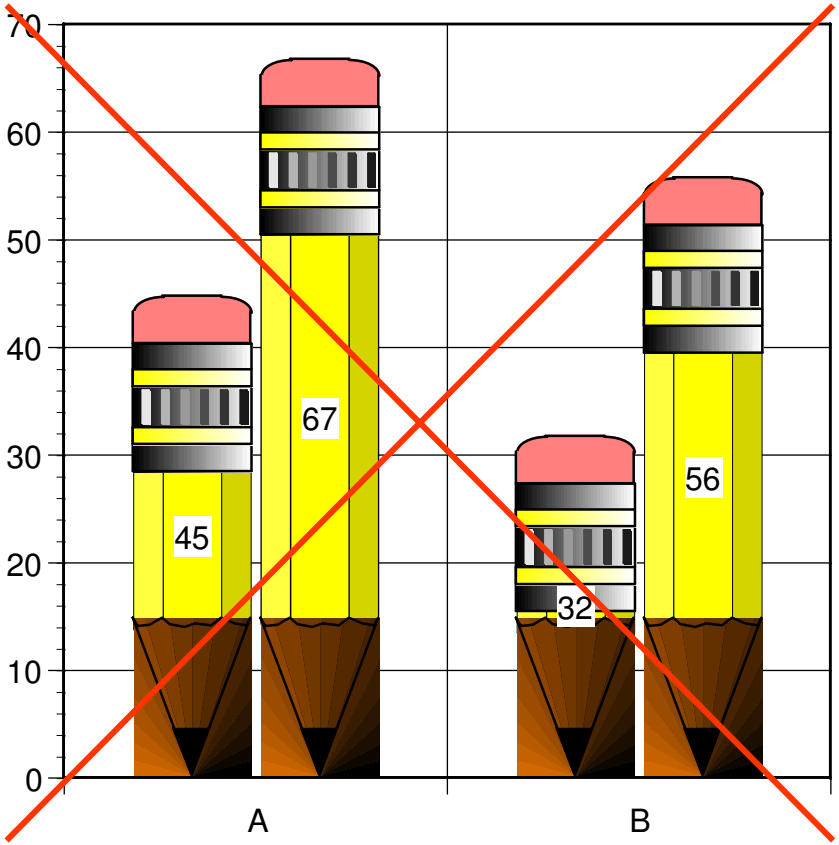
Other types



- Or: Min. - max.

  + average temperature

Other types



- Pictogram: Not recommended