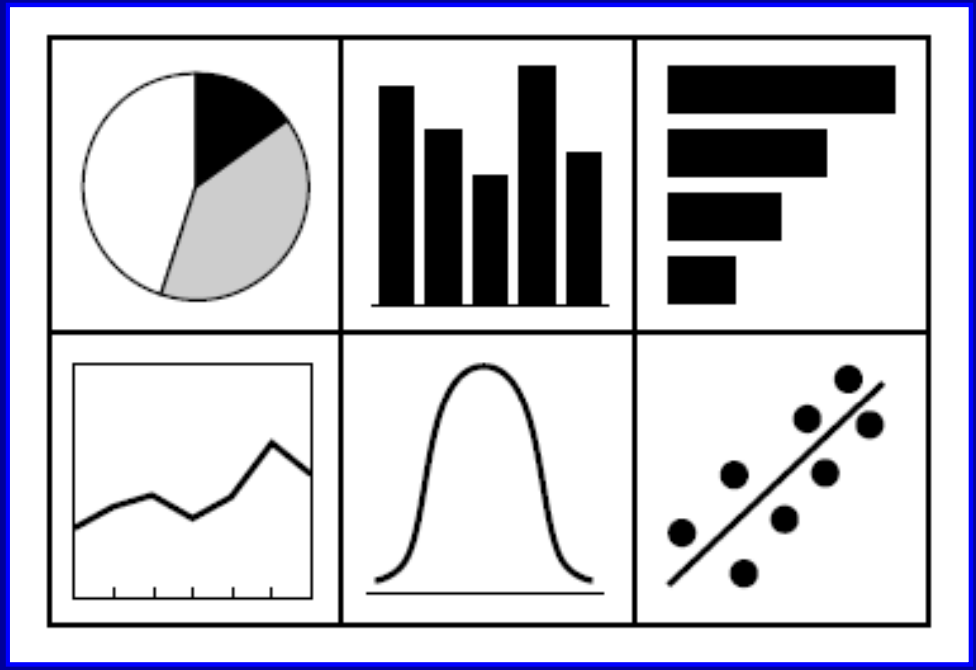
How to display your data?



Samir Haffar M.D.

Assistant Professor of Gastroenterology

Al-Mouassat University Hospital

Damascus - Syria

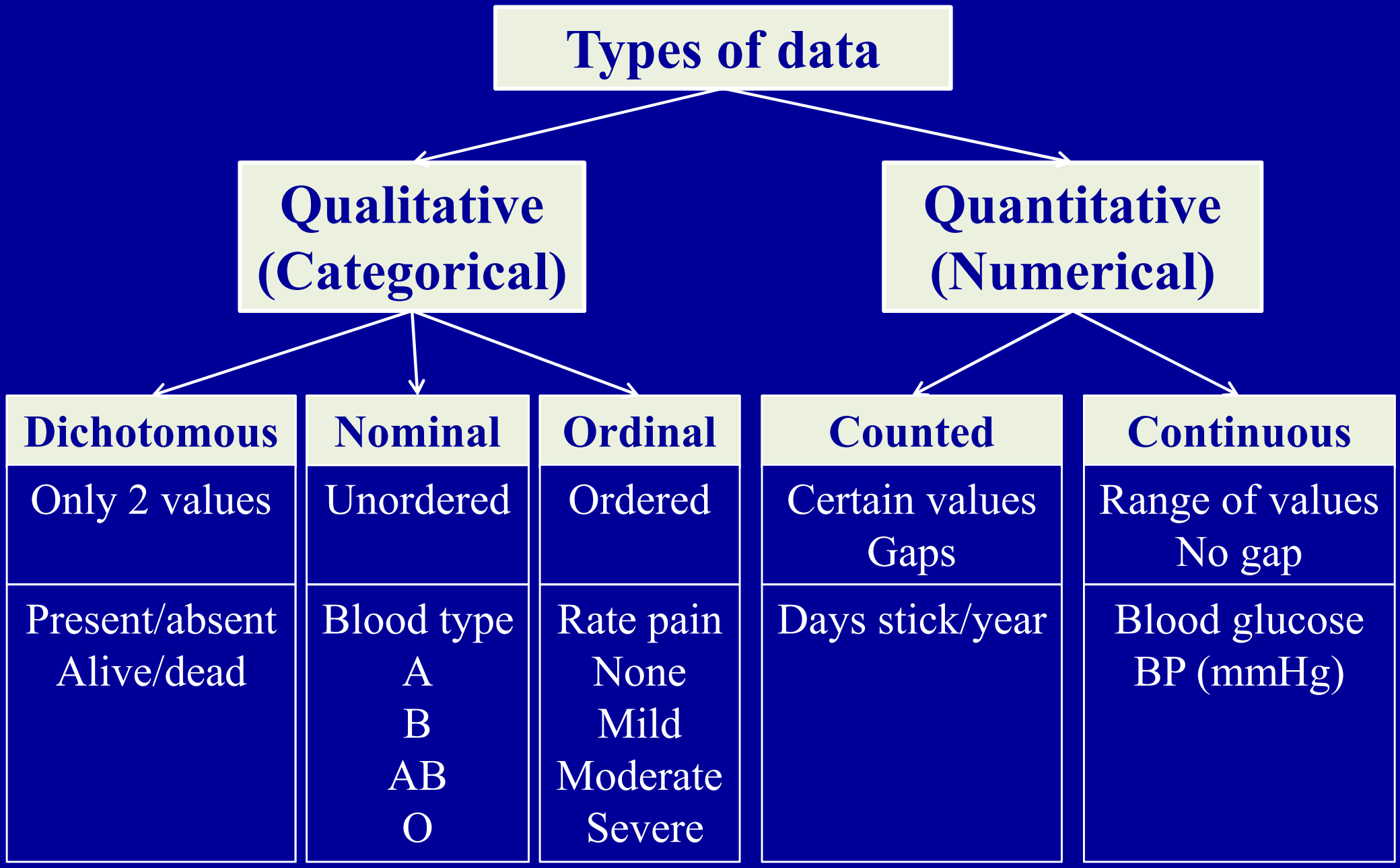
Who might benefit?

- Researchers who want to display results of their studies for **publication** in a journal
- Readers of research literature who wish to do a **critical appraisal** of a piece of work
- People who have to deliver a **presentation**

The best advice that a statistician can give a researcher is to first plot the data

Conventional statistics textbooks give only brief details on how to draw figures & display data

Freeman JV, Walters SJ, Campbell MJ. How to display data.
Blackwell Publishing, Massachusetts, USA, 1st edition, 2008.



Petrie A & Sabin C. Medical statistics at a glance.
Blackwell Publishing, Massachusetts, USA, 2nd edition , 2005.

How to present your data?

❶ Numbers

❷ Tables

❸ Graphs

❶ Displaying your data with numbers

Presenting numbers -1

- Numbers expressed in **numerals** rather than in words
- **Decimal sign** is a point preceded by 0 **[0.3 not 0,3)]**
- Use space to mark off **thousands [12 345 not 12,345]**
- **Remove surplus zeros**: 1.6×10^9 bacteria/ml
- Never use **billion**: 10^9 in USA & 10^{12} in Europe
- Use only one **slash** to express quotients of units: km/h
  Use negative exponents if >2 **[$mg.kg^{-1}.h^{-1}$ not mg/kg/h]**

Presentation of numbers - 2

Report total no of observations

- **Qualitative data**

  Use both **frequencies** & **percentages**

- **Quantitative data**

  Normal distribution **Mean & SD** (one decimal place)

  Skewed distribution **Median & IQR***

* IQR: Interquartile range

Freeman JV et all. How to display data. Blackwell Publishing, Massachusetts, 2008.

Use of percentages

Total number	Percentages & decimals
< 25	Percentages should not be used
25 – 100	Percentages without decimals **[7% not 7.2%]**
100 – 100 000	Only one decimal added **[7.2% not 7.23%]**
> 100 000	Two decimals added **[7.23% not 7.235%]**

Gustavii B. How to write & illustrate scientific papers.
Cambridge University Press, Cambridge, UK, 2nd edition, 2008.

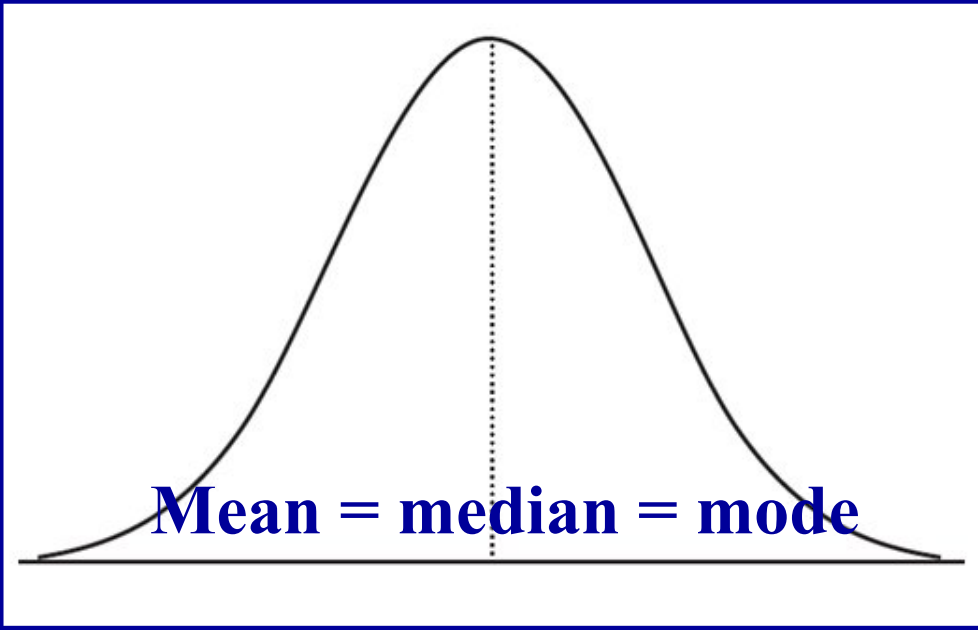
Normal distribution

Sometimes known as Gaussian distribution

Classic 'bell' shape

Peak in the middle (mean)

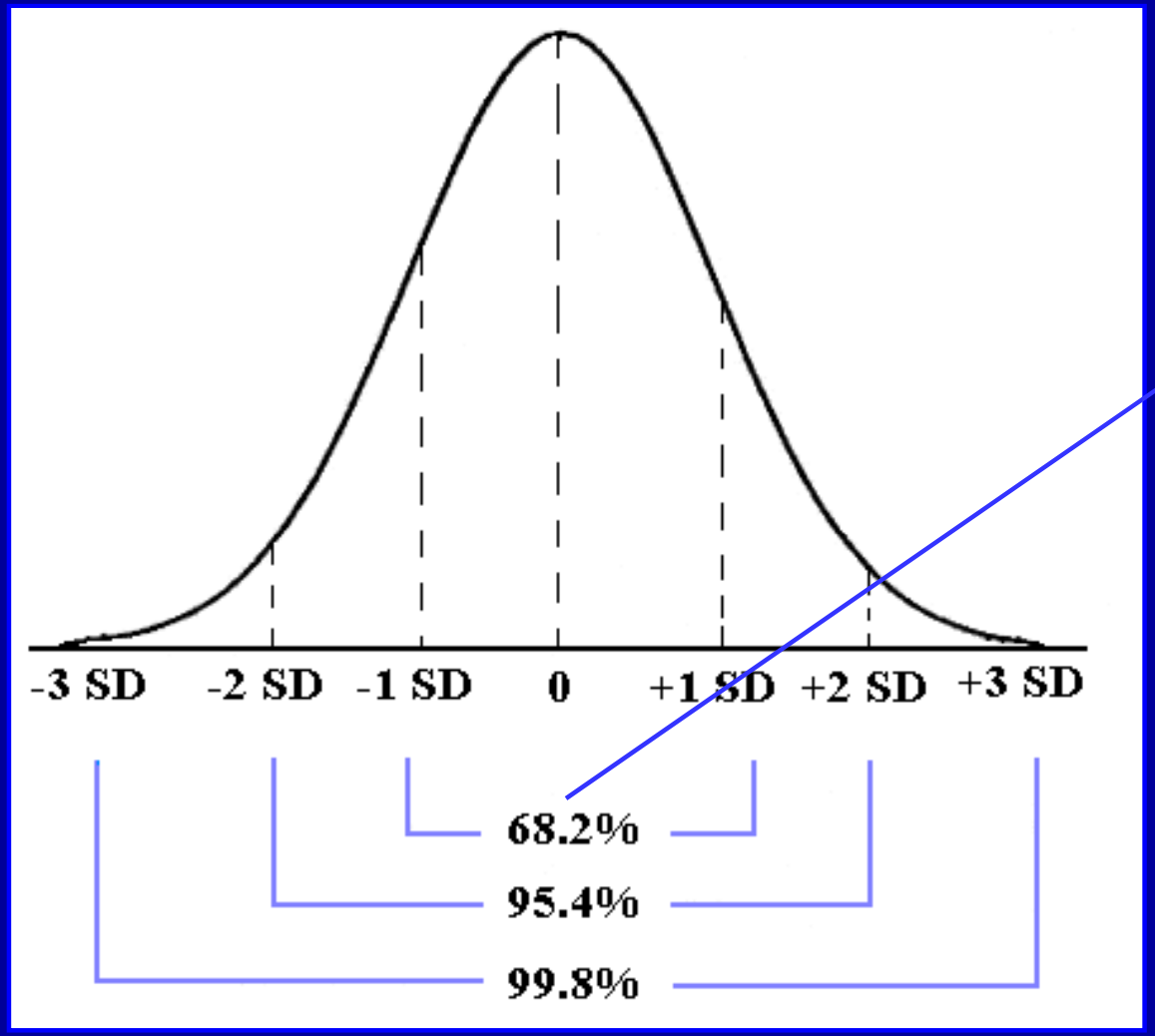
Symmetrical tails



Mean	Sum of values/number of observations
Median	Number of observations above = number below
Mode	Most frequently occurring value

Harris M Taylor G. Medical statistics made easy.
Martin Dunitz, 1st edition, London, 2003.

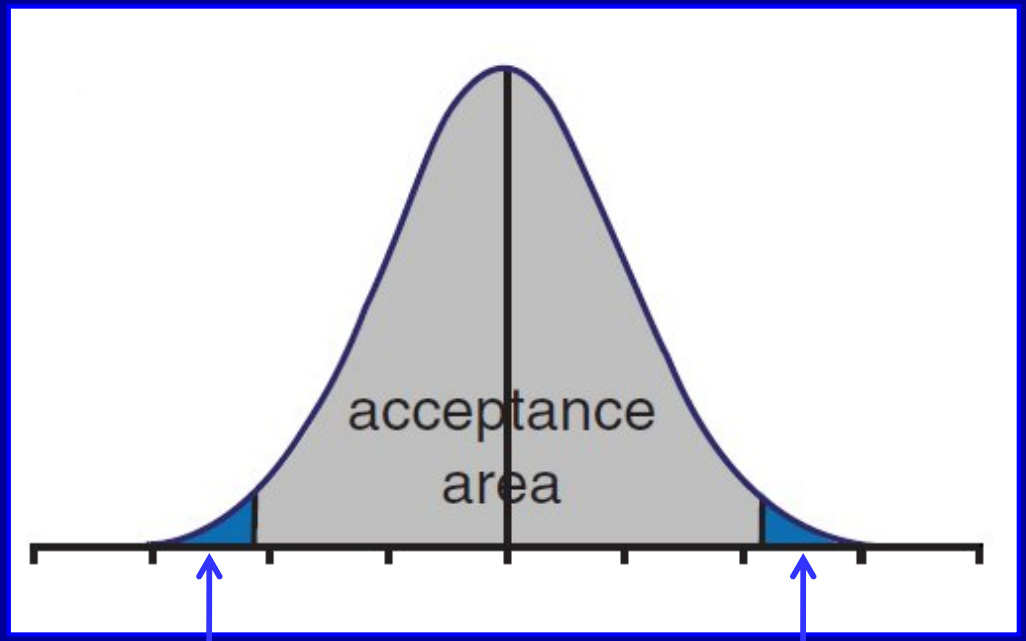
Standard normal distribution



One decimal place

Chernick MR & Friis RH. Introductory biostatistics for the health sciences. John Wiley & Sons, New Jersey, USA, 1st edition, 2003

# P < 0.05

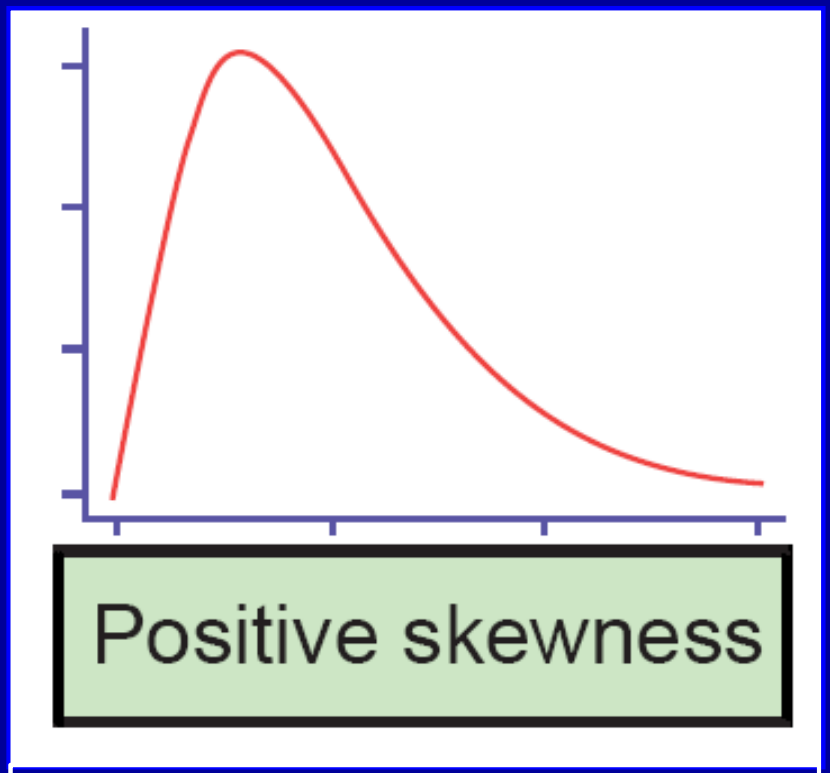


Value of test statistic P < 0.05

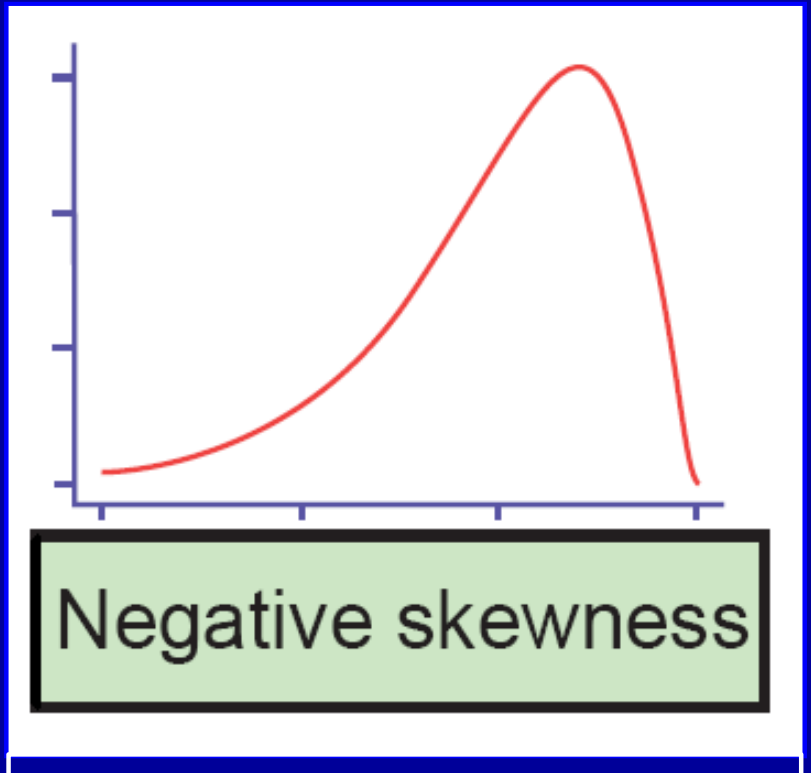
Reject the null hypothesis
Results significant at the 5% level

Perera R, Heneghan C & Badenoch D. Statistics toolkit.
Blackwell Publishing & BMJ Books, Oxford, 1st edition, 2008.

Skewness



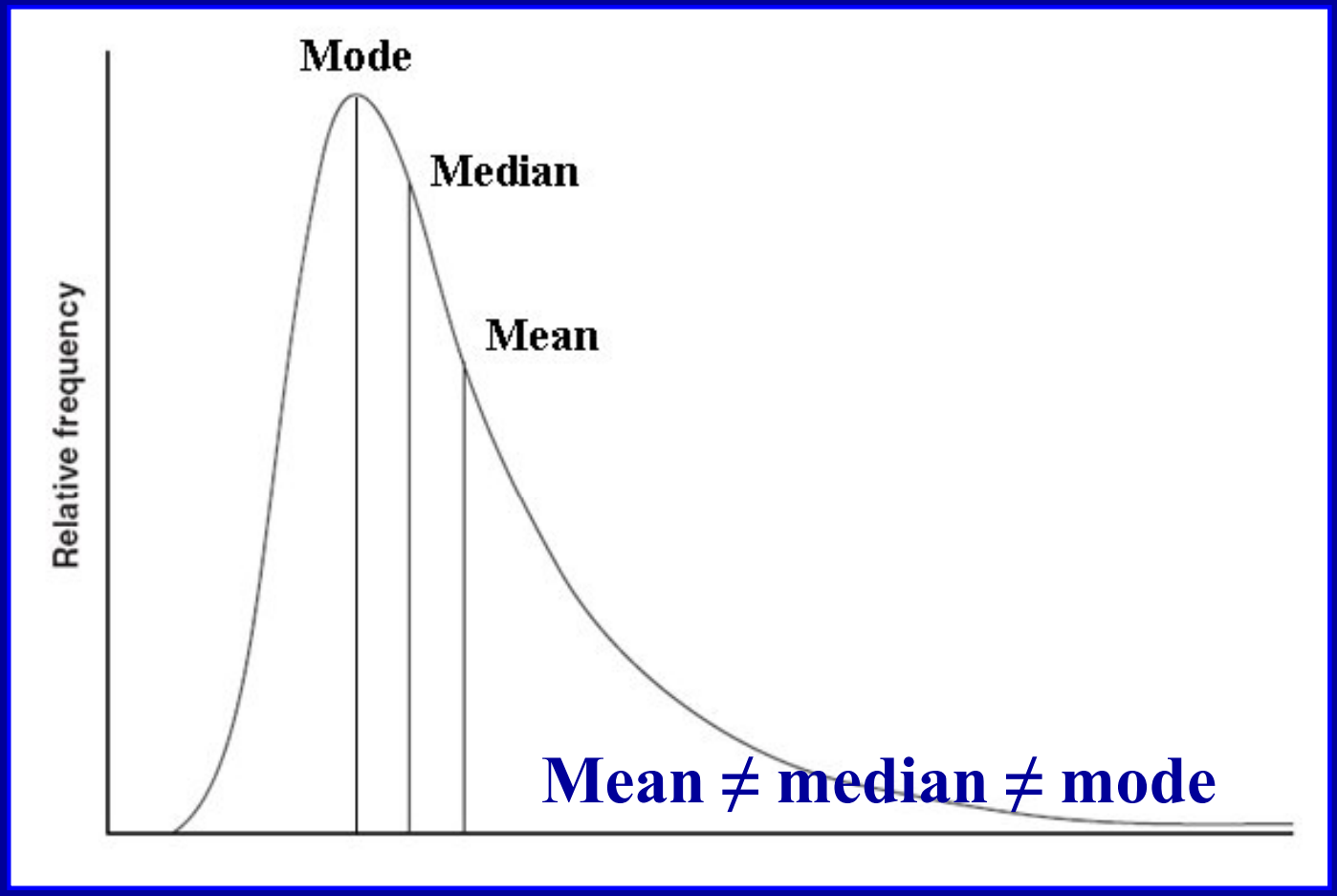
Peak at lower values
Long tail of higher values



Peak at higher values
Long tail of lower values

Perera R, Heneghan C & Badenoch D. Statistics toolkit.
Blackwell Publishing & BMJ Books, Oxford, 1st edition, 2008.

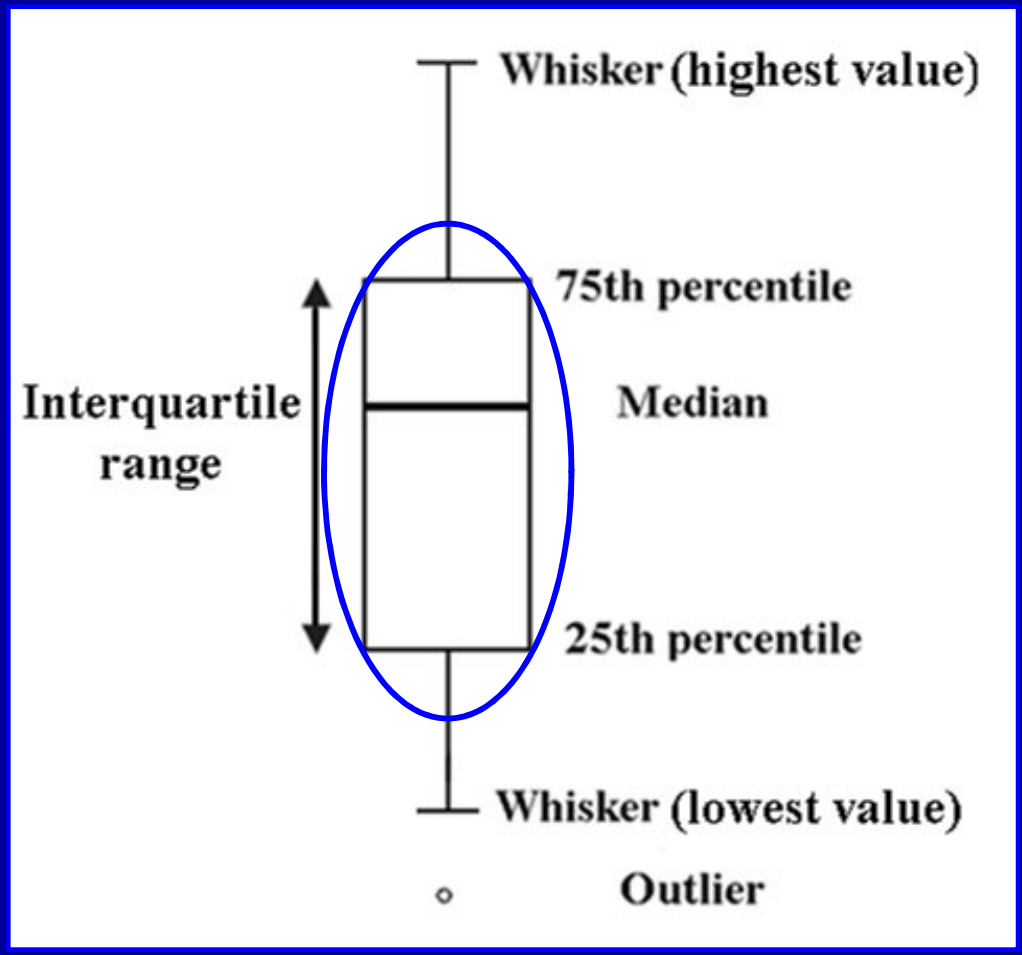
Skewed data distribution



The mean is an over-estimate of the median value

Peat JK & all. Health science research: a handbook of quantitative methods. Allen & Unwin, Rows Nest, Australia, 1st edition, 2001.

Anatomy of a box-whisker plot

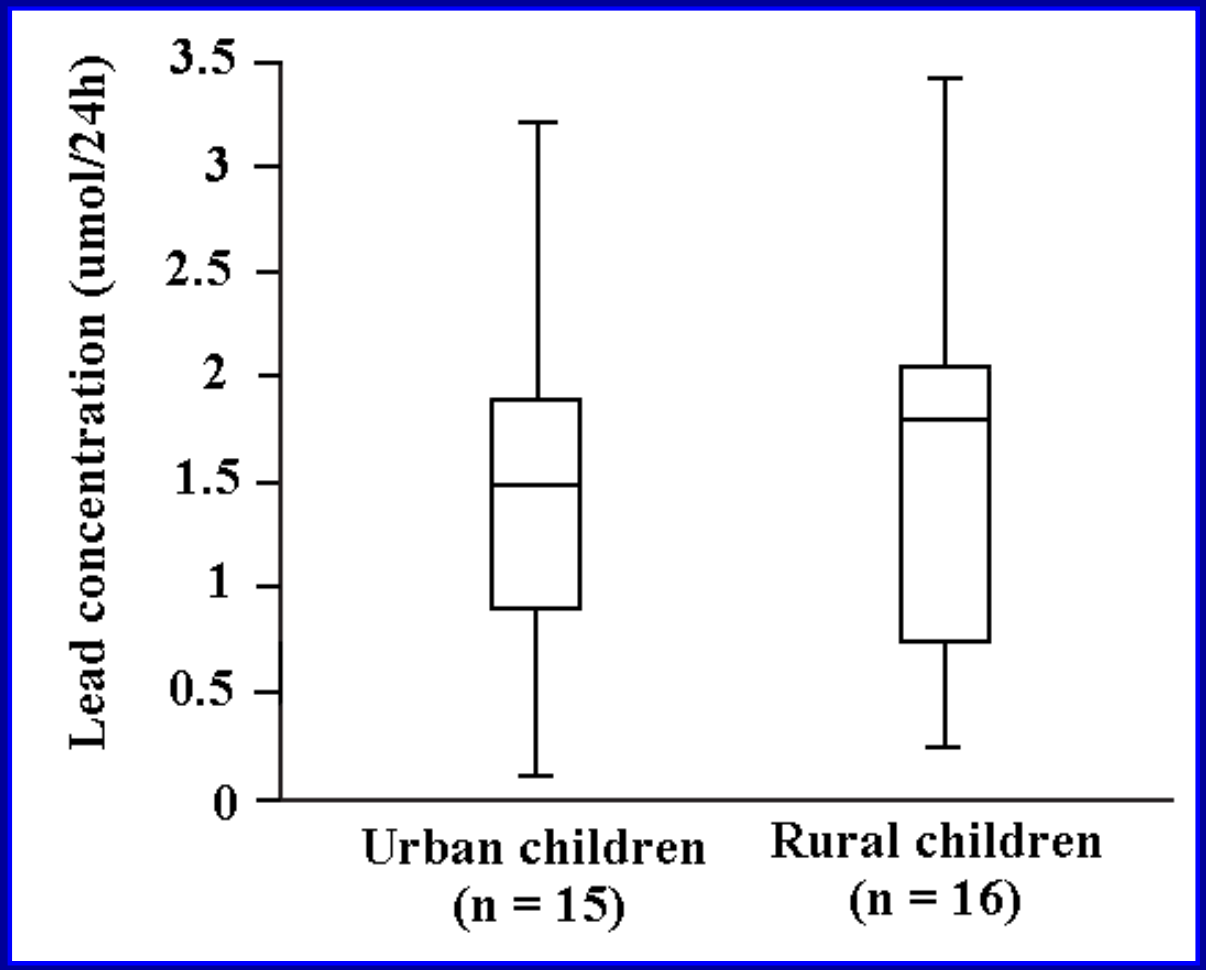


Especially good to show differences between groups

Morgan GA et all. Understanding & evaluating research in applied & clinical settings. Lawrence Erlbaum Associates, New Jersey, USA, 2006.

Box-Whisker Plot

Urinary lead concentration in urban & rural children



Swinscow TDV & Campbell MJ. Statistics at square one.
BMJ Books, London, 10th edition, 2002.

Central tendency & dispersion

- **Central tendency**

Mean	Sum of values/number of observations
Median	Number of observations above = number below
Mode	Most frequently occurring value

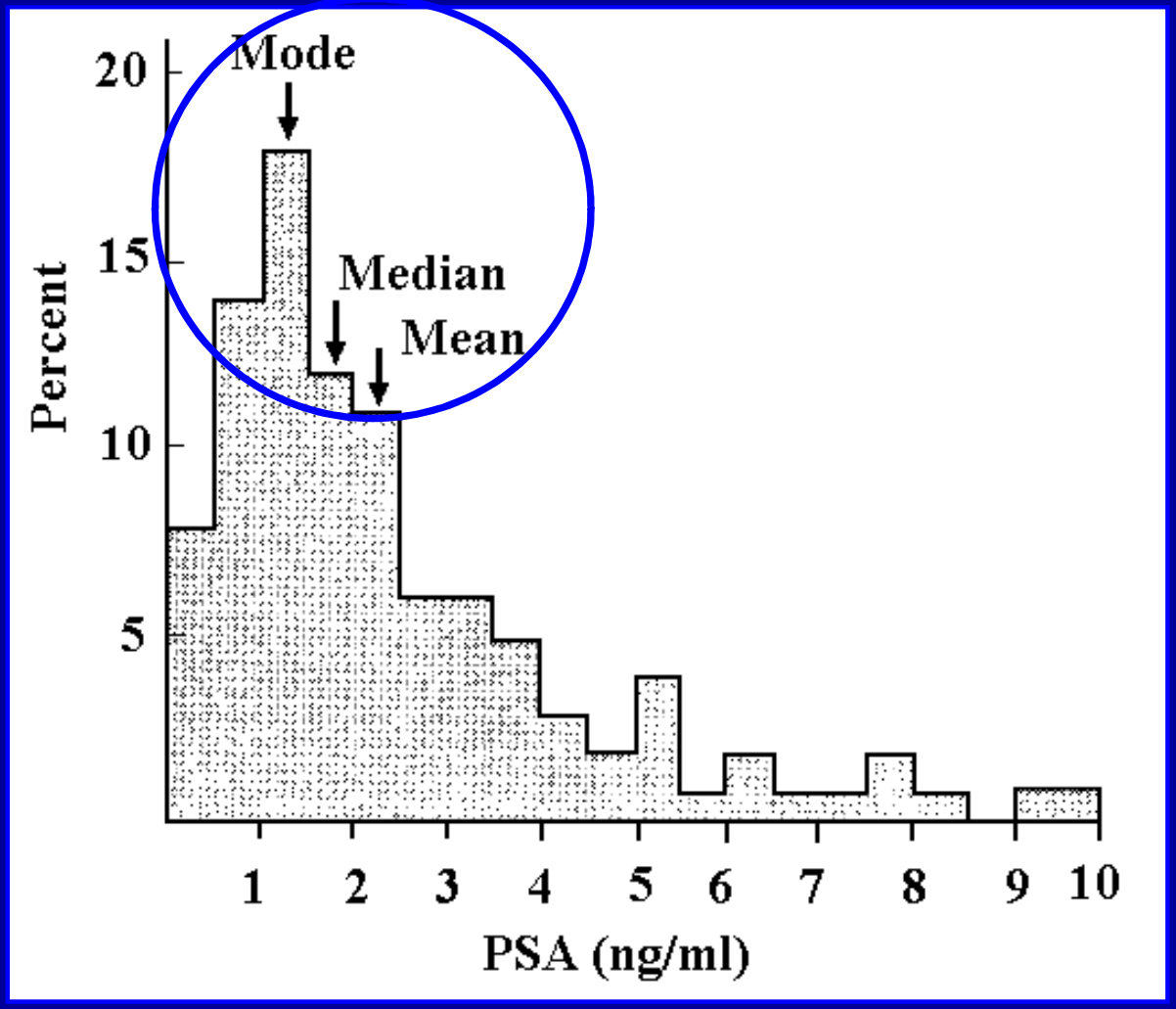
- **Dispersion**

Range	From lowest to highest value
SD	Average difference of values from mean
Quartile	% of observations falling between specif values

Fletcher R et all. Clinical epidemiology.
Williams & Wilkins, Baltimore, USA, 3rd edition, 1996.

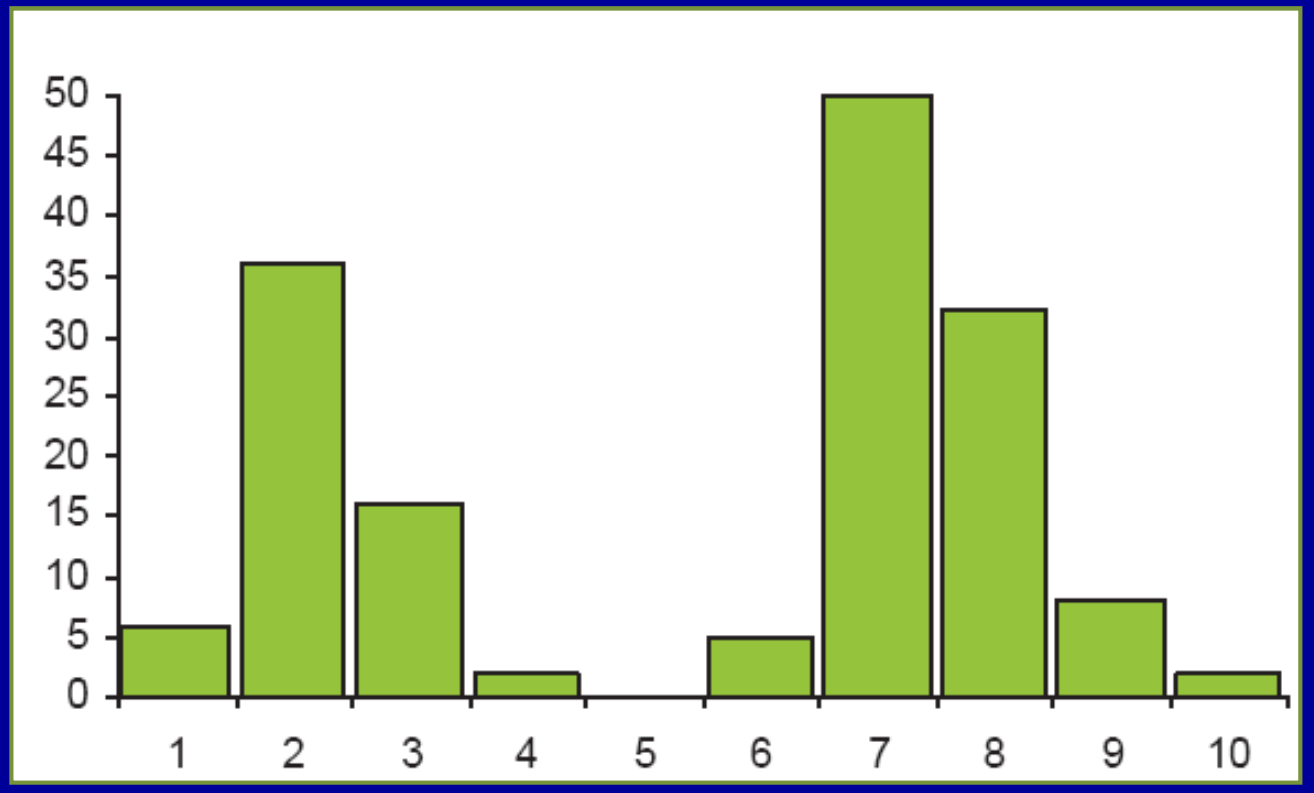
Central tendency & dispersion

Distribution of PSA in presumably normal men



Fletcher R et all. Clinical epidemiology.
Williams & Wilkins, Baltimore, USA, 3rd edition, 1996.

Bimodal distribution



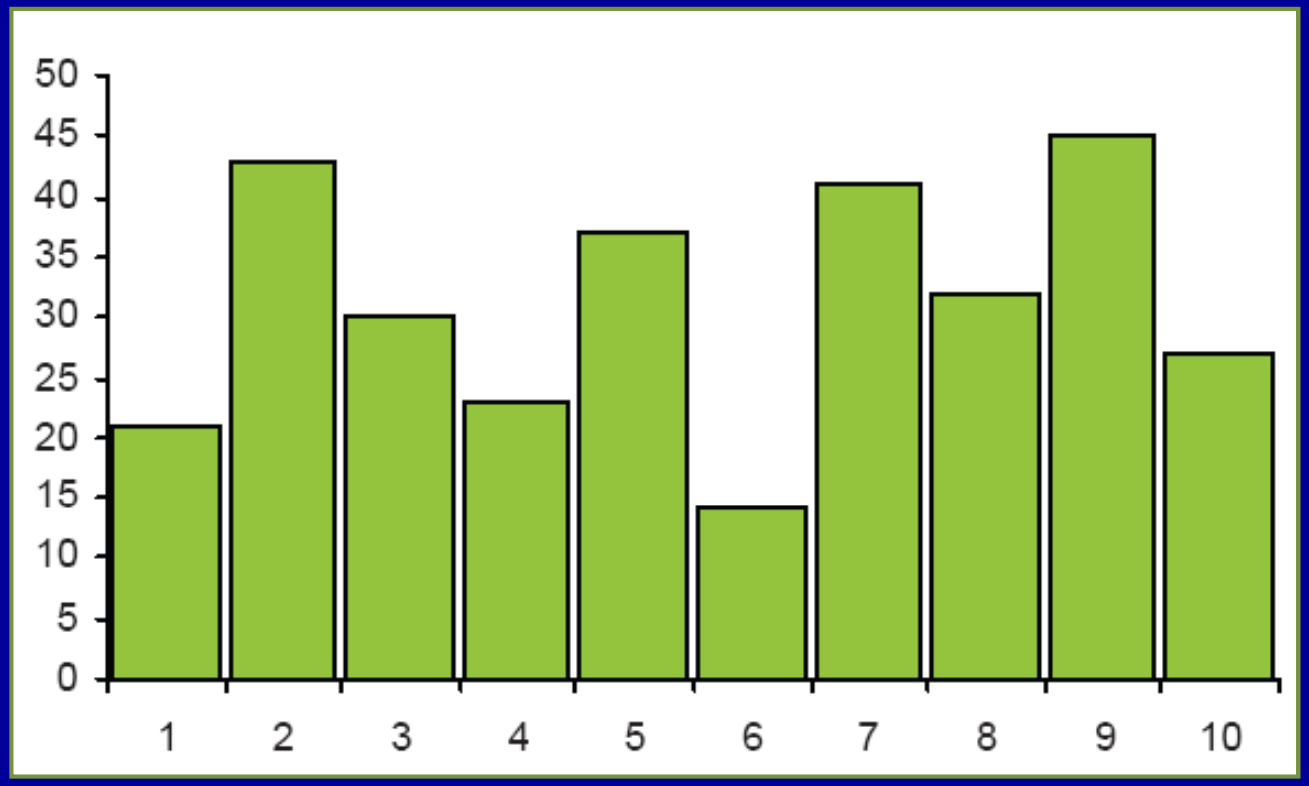
Data have 2 peaks

There may be two different populations

Each with its own central tendency

Perera R, Heneghan C & Badenoch D. Statistics toolkit. Blackwell Publishing & BMJ Books, Oxford, 1st edition, 2008.

Uniform distribution



Number of peaks

All possible values are equally likely

Central tendency measure not useful

Perera R, Heneghan C & Badenoch D. Statistics toolkit. Blackwell Publishing & BMJ Books, Oxford, 1st edition, 2008.

❷ Displaying your data with tables

Recommendations to present data in tables – 1

- Tufte's principle
- Clear title with sample size
- Solid lines kept to minimum particularly vertical ones
- Columns and rows clearly labeled
- Rows & columns ordered by size if no natural ordering

Recommendations to present data in tables – 2

- Numbers rounded to 2 effective digits
- **Qualitative data** Frequency & percentage
- **Quantitative data** Symmetrically **Mean & SD**
  Skewed **Median & IQR***

* IQR: Interquartile range

Freeman JV, Walters SJ, Campbell MJ. How to display data. Blackwell Publishing, Massachusetts, USA, 1st edition, 2008.

Tufte's principle for table & graph

Maximum amount of information for minimum amount of ink

Tufte ER. The visual display of quantitative information. Cheshire, Connecticut: Graphics Press; 1983

Marital status of 226 patients in leg ulcer study

Title

Marital status	Frequenc	nt
Married		
Widowed		38.1
Single		11.1
Divorced/ser	11	4.9
Total	226	100.0

Headings

Body

4 x 2 contingency table
4 rows x 2 columns = 8 cells

BMJ 1998 ; 316 : 1487 – 91.

Source

Marital status of 226 patients in leg ulcer study

Ordered alphabetically

Marital status	Frequency	Percent
Divorced/separated	11	4.9
Married	104	46.0
Single	25	11.1
Widowed	86	38.1
Total	226	100.0

Hard to interpret

BMJ 1998 ; 316 : 1487 – 91.

Marital status of 226 patients in leg ulcer study

Ordered by size

Marital status	Frequency	Percent
Married	104	46.0
Widowed	86	38.1
Single	25	11.1
Divorced/separated	11	4.9
Total	226	100.0

Much easier to interpret

BMJ 1998 ; 316 : 1487 – 91.

❸ Displaying your data with graphs

Table or graph?

Choice between using a table or a figure not easy

Nor is it easy to offer much general guidance

Altman D & Bland M. Presentation of numerical data. BMJ 1996 ; 312 : 572.

Table or graph?

Graph	Table
Better in presentations	Better in papers
Can only show summaries	Can often show all the data
Show only a few variables	Better for multiple variables
Trend better illustrated	Trend badly illustrated

Freeman JV, Walters SJ, Campbell MJ. How to display data. Blackwell Publishing, Massachusetts, USA, 1st edition, 2008.

Table or graph?

Trend badly illustrated with a table

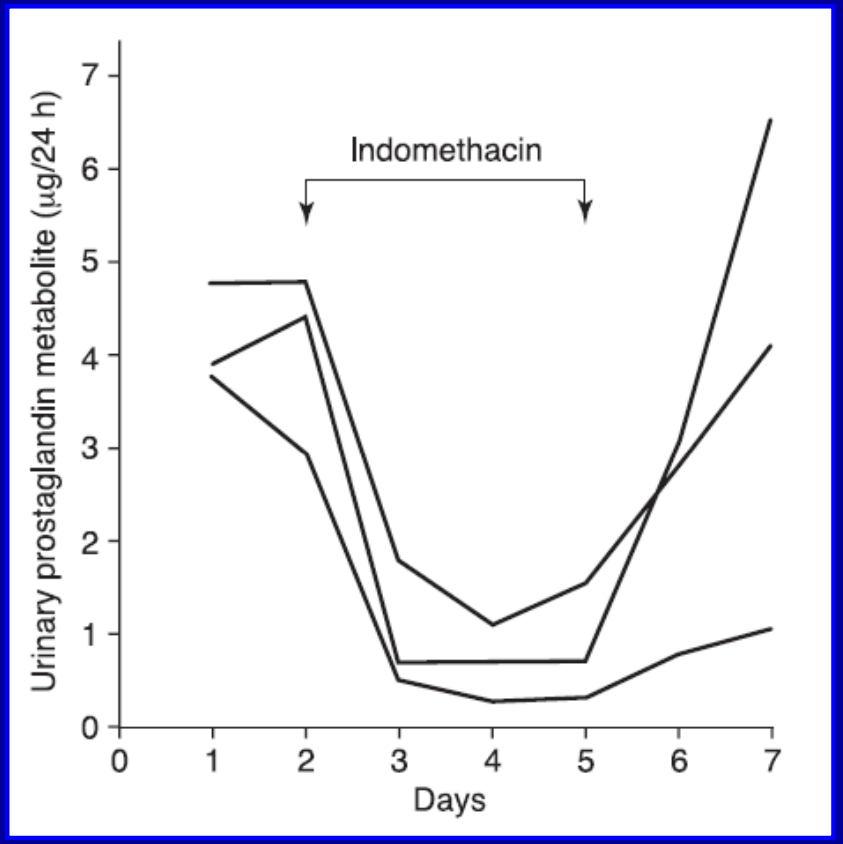
Urinary excretion of PG metabolite after indomethacin administration

Subject	Urinary prostaglandin metabolite (mg/24 h)						
	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
I	4.8	4.8	1.8*	1.1*	1.5*	2 .7	4.1
II	3.9	4.4	0.7*	0.7*	0.7*	3.1	6.5
III	3.8	3.0	0.5*	0.3*	0.3*	0.8	1.1

Subject taking indomethacin 4 x 50 mg/24 h
Hamberg 1972

Table or graph?

Trend better illustrated with a graph



Urinary excretion of a prostaglandin metabolite decreased following indomethacin administration in three humans

Hamberg 1972

Why use of graphs in presentation?

- You need to get your audience's attention
- Many people respond better to visual cues than to straight text or lists of numbers
- Effective graph can help drive home your point

Software for graphs

- No single package can draw all graphs to display data
- Simple graphs can be drawn in **Microsoft Excel**

- More complex graphs

  Major statistical packages: **SPSS, STATA, SAS**

  S-Plus for superimposing several graphs into single figure

- Packages change regularly

Types of graph

- Bar/column graph & variants
- Pie graph
- Dot plot
- Stem & leaf plot
- Histogram
- Box-whisker plot
- Line graph
- Spider or radar plot
- Pictogram
- Venn diagram

Types of data

- **Qualitative (categorical)**

  Dichotomous — Only 2 values

  Nominal — Unordered

  Ordinal — Ordered

- **Quantitative (numerical)**

  Counted — Gaps

  Continuous — No gaps

Displaying qualitative data

- **Bar/column graph**
- **Grouped column graph**
- **Segmented column graph**
- **Pie graph**

Recommendations for construction of graph

- Tufte's principle
- Clear title with sample size
- Labeled axes
- Gridlines kept to a minimum
- Categories ordered by size
- No three-dimensional graphs

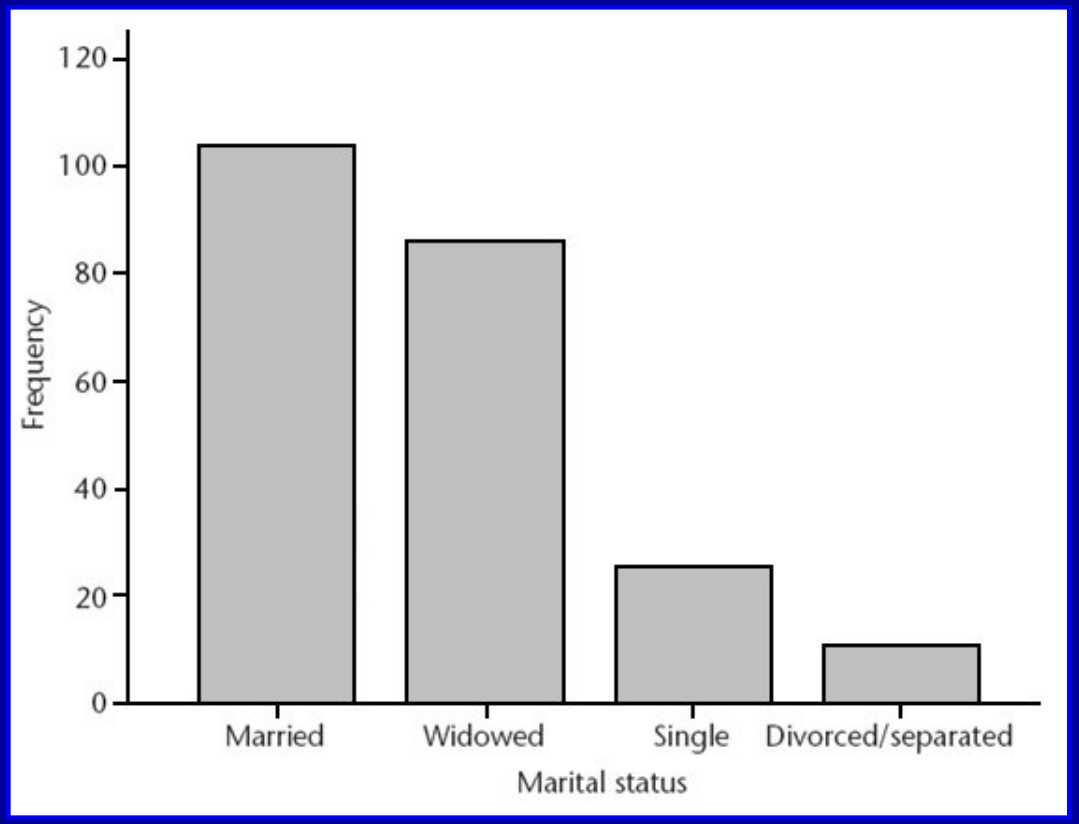
Tufte's golden rule

Maximum amount of information for minimum amount of ink

Tufte ER. The visual display of quantitative information. Cheshire, Connecticut: Graphics Press; 1983.

Column chart

Marital status for 226 patients in leg ulcer study



Columns wider than spaces between them

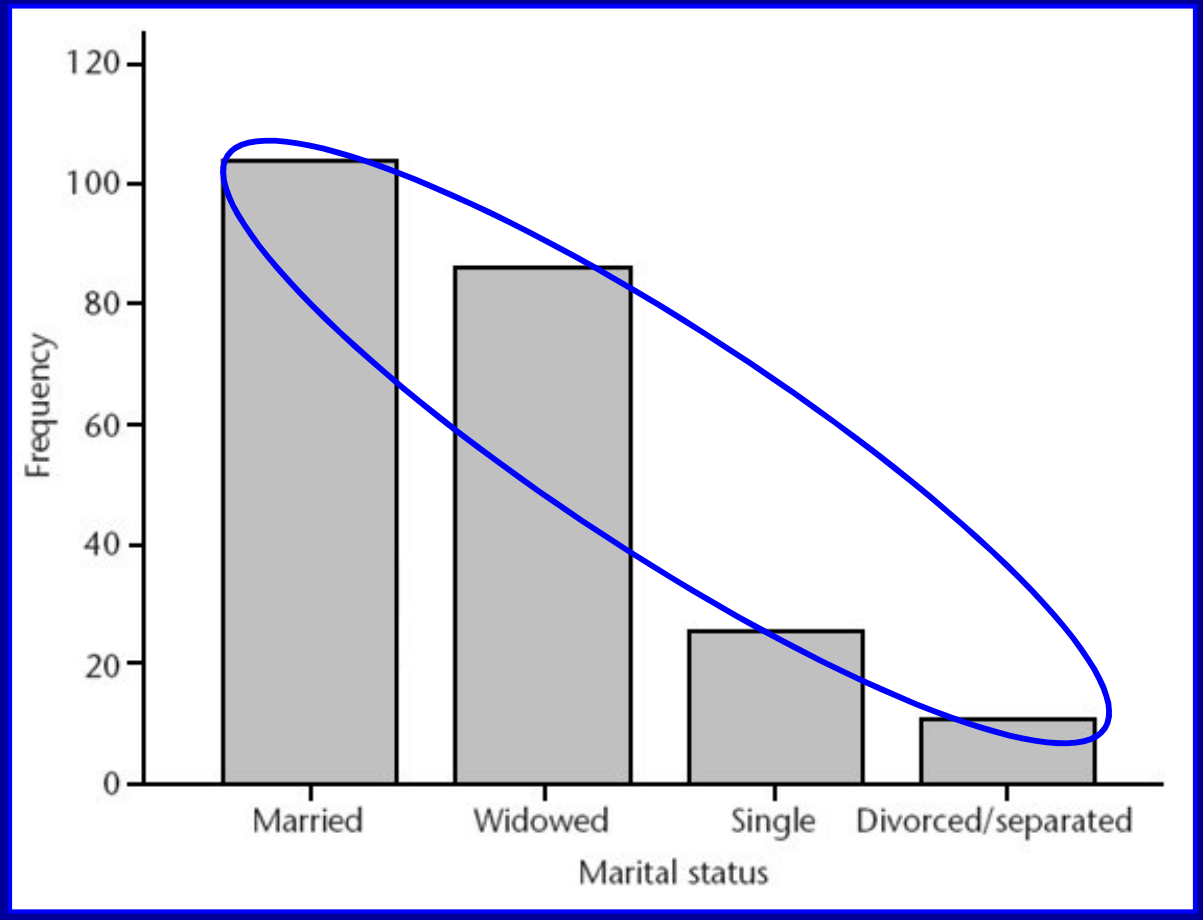
Bars have gray tone which is more pleasing to the eye

Vertical axis doesn't extend beyond what the graph demands

BMJ 1998 ; 316 : 1487 – 91.

Column chart

Marital status for 226 patients in leg ulcer study

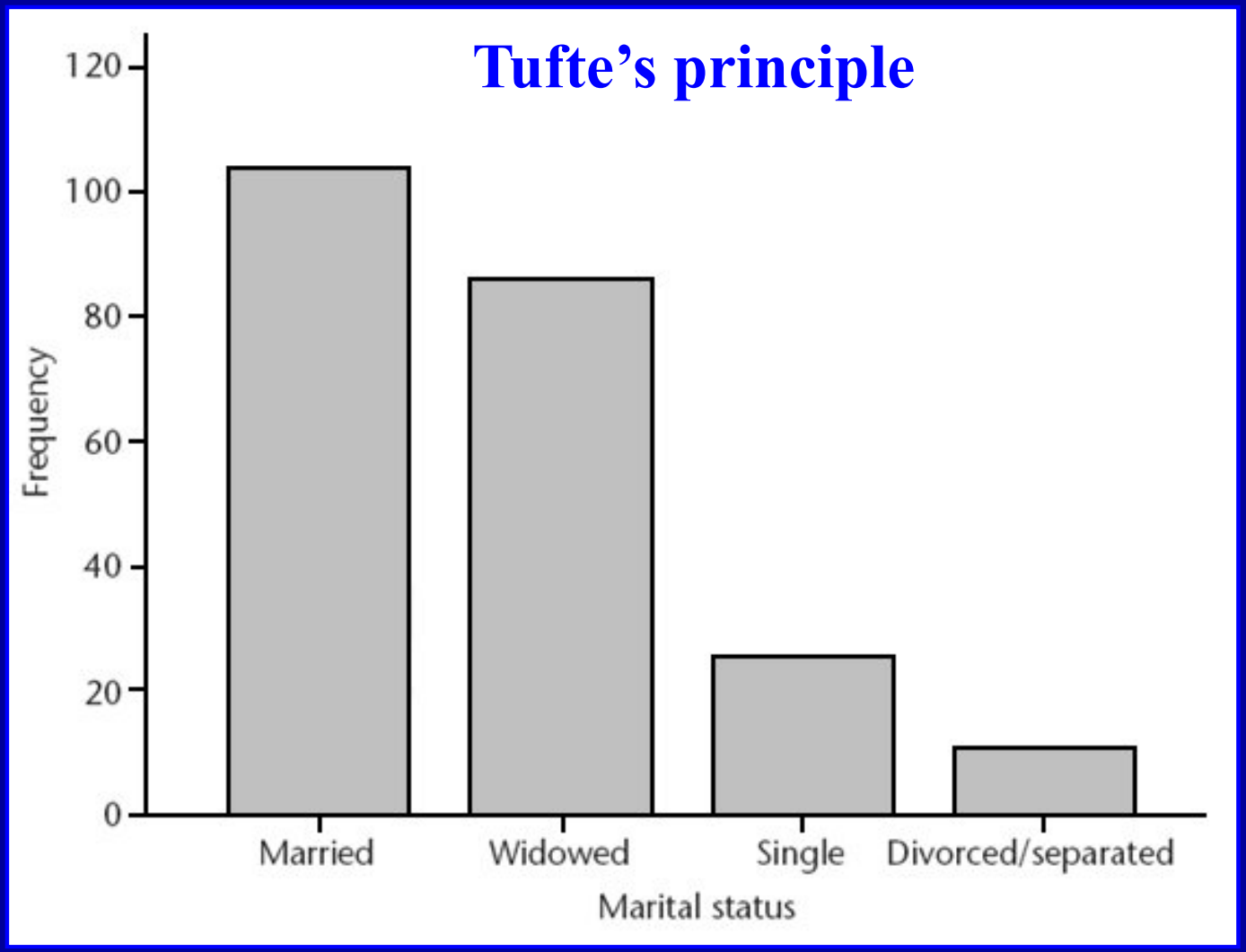


Only the height of columns presents the data of interest

BMJ 1998 ; 316 : 1487 – 91.

Column chart

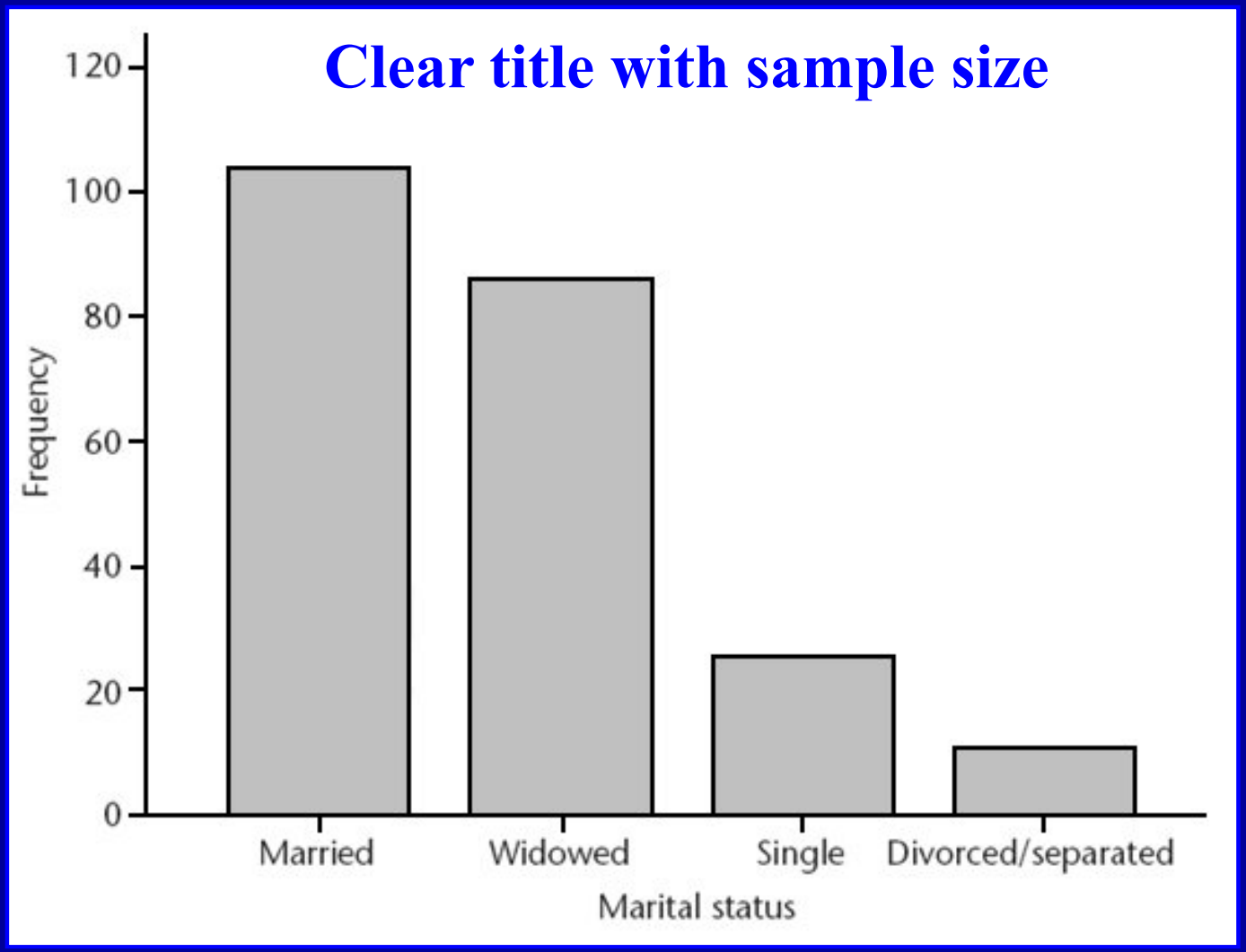
Marital status for 226 patients in leg ulcer study



BMJ 1998 ; 316 : 1487 – 91.

Column chart

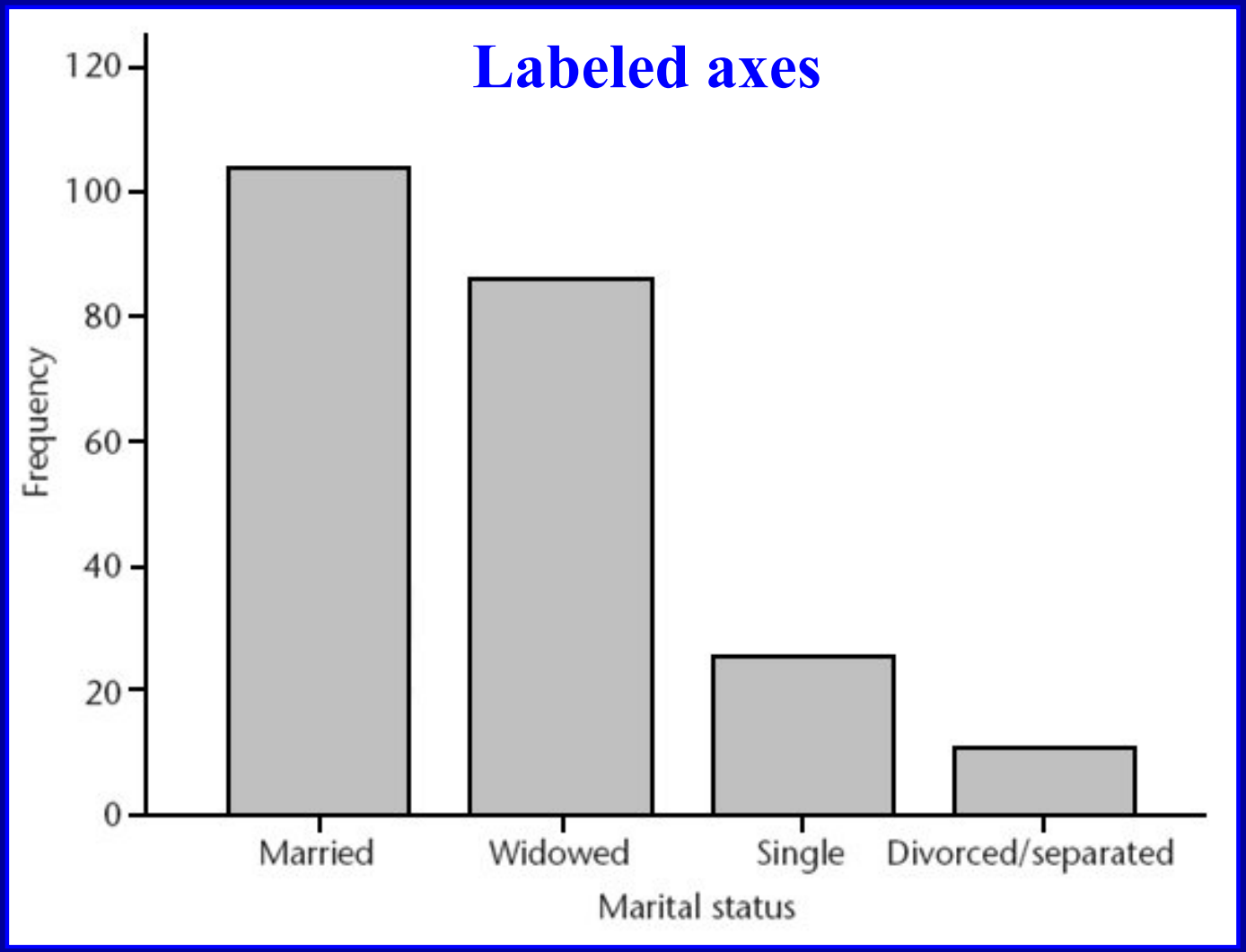
Marital status for 226 patients in leg ulcer study



BMJ 1998 ; 316 : 1487 – 91.

Column chart

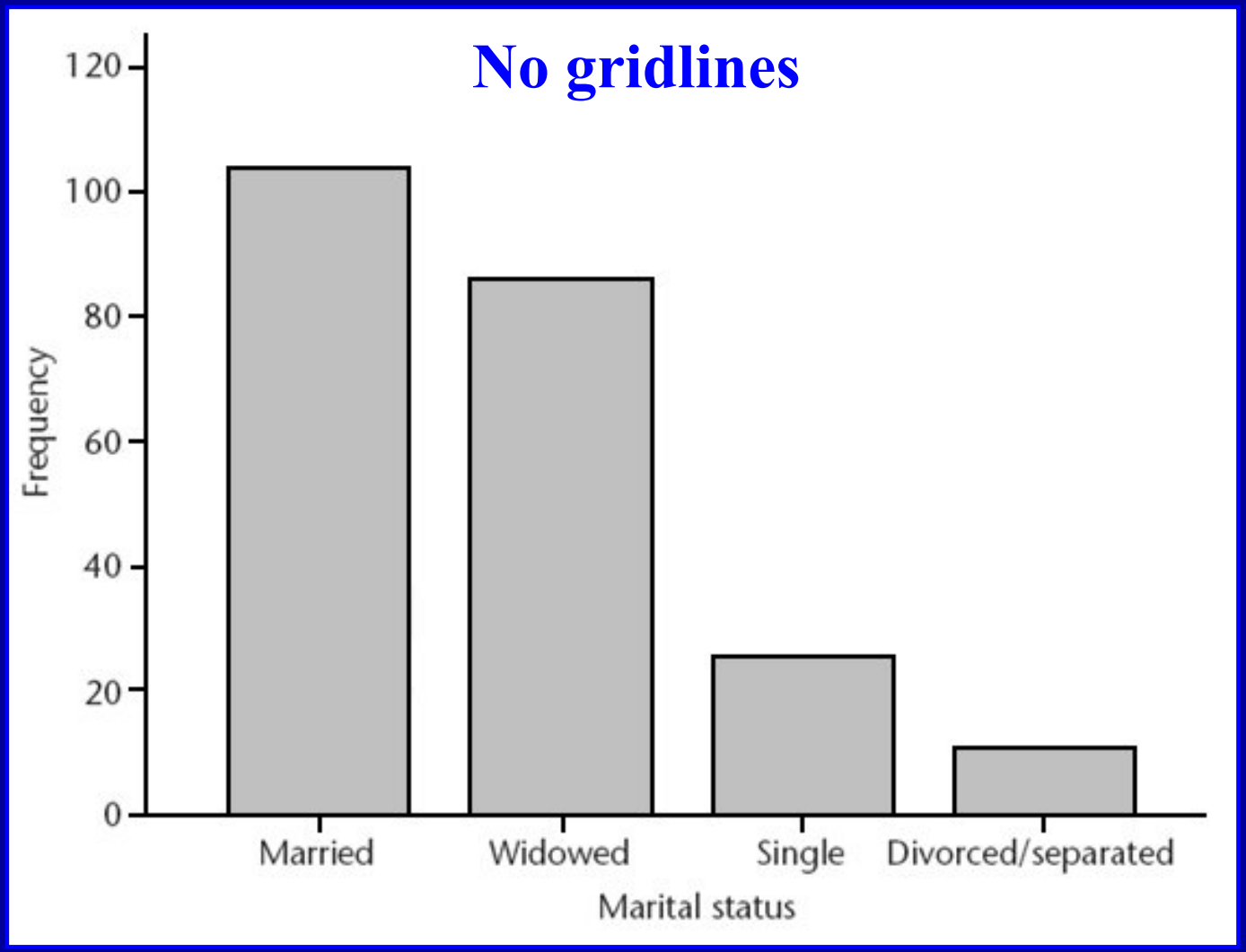
Marital status for 226 patients in leg ulcer study



BMJ 1998 ; 316 : 1487 – 91.

Column chart

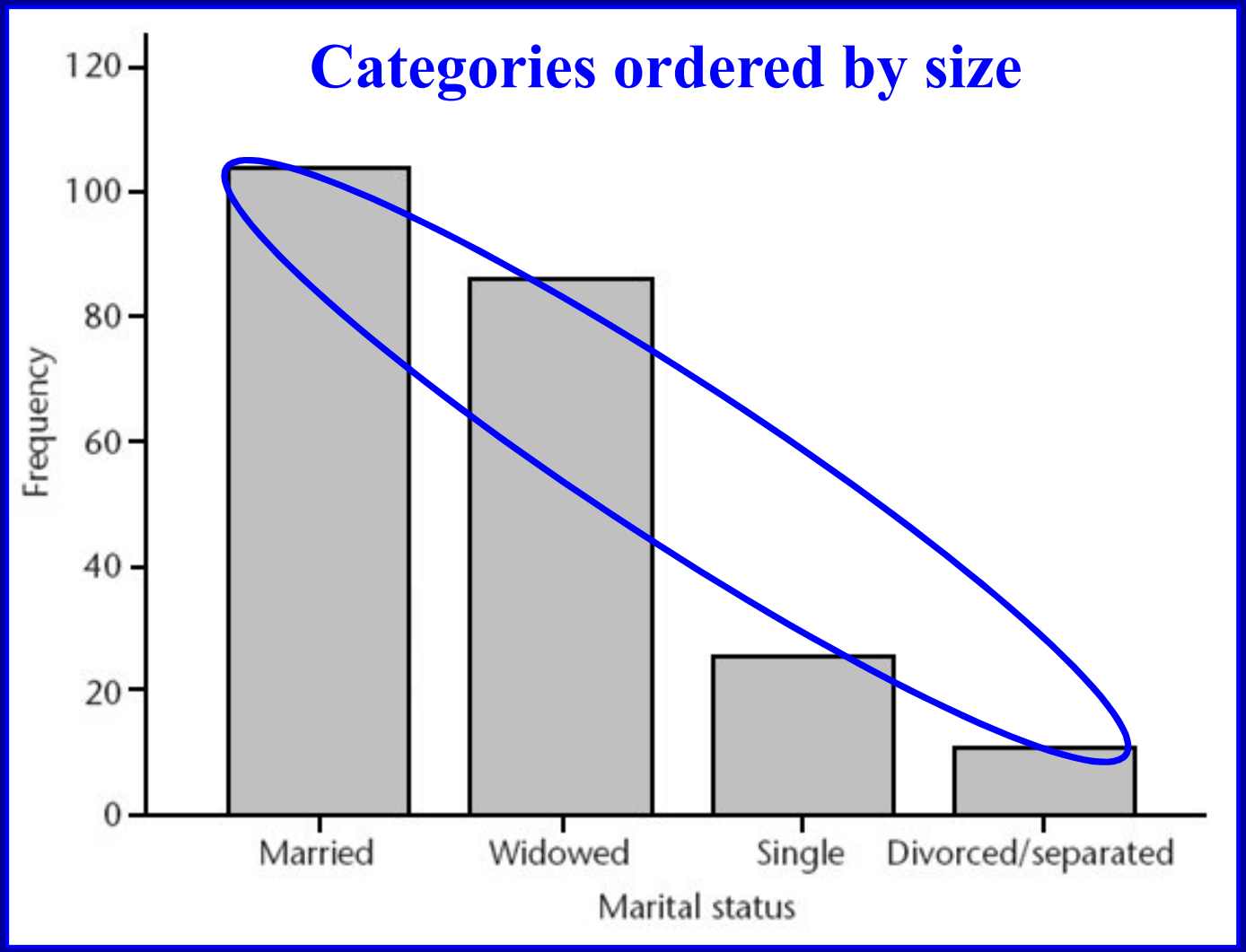
Marital status for 226 patients in leg ulcer study



BMJ 1998 ; 316 : 1487 – 91.

Column chart

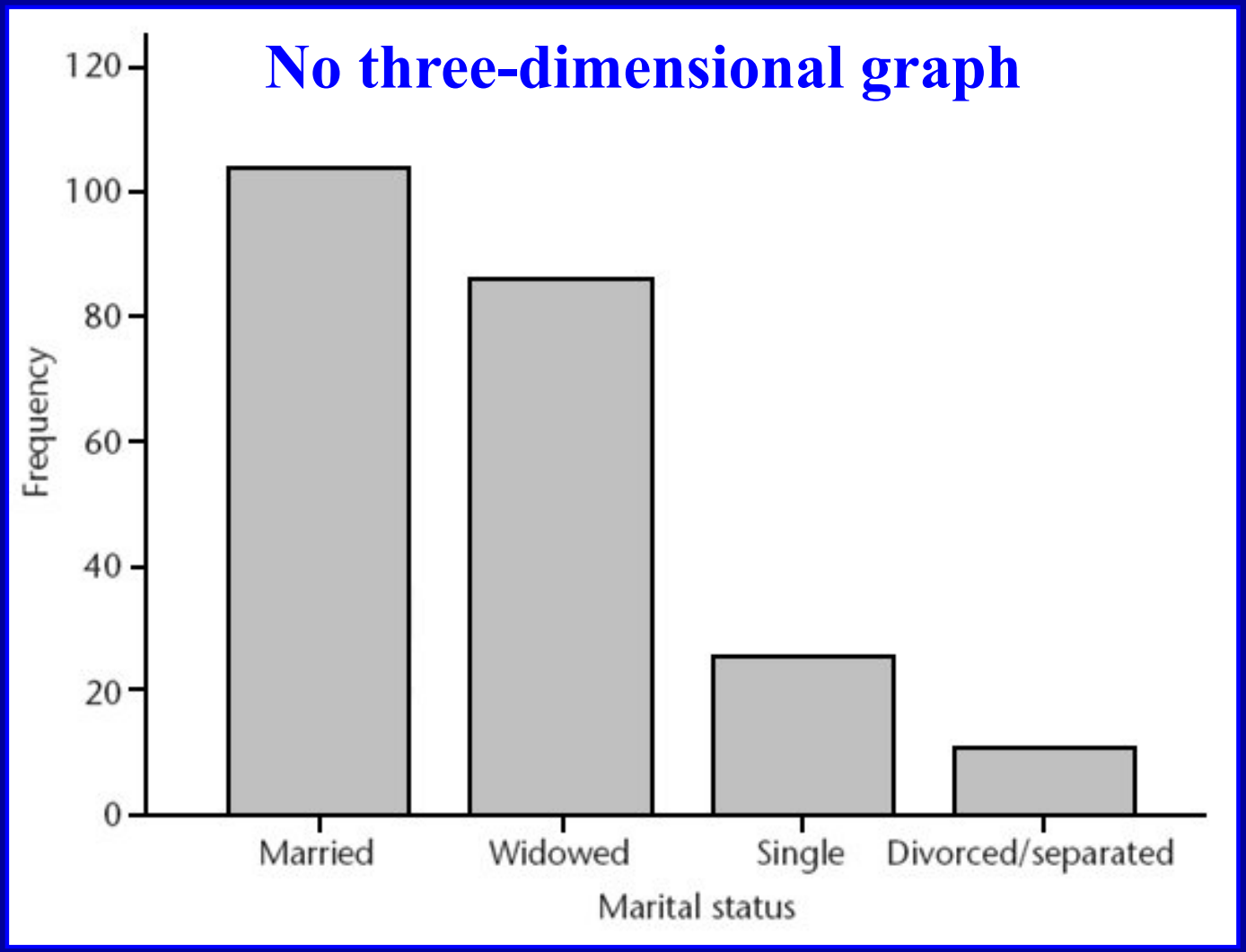
Marital status for 226 patients in leg ulcer study



BMJ 1998 ; 316 : 1487 – 91.

Column chart

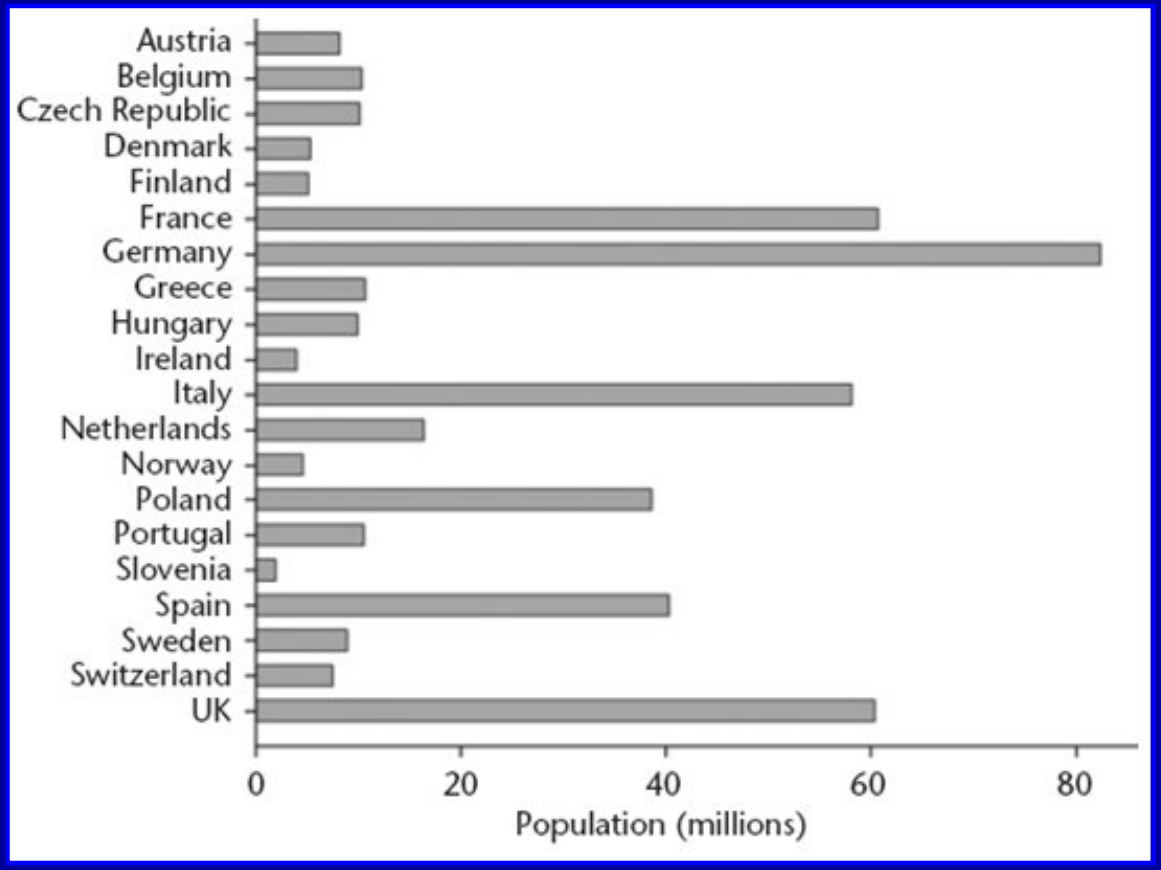
Marital status for 226 patients in leg ulcer study



BMJ 1998 ; 316 : 1487 – 91.

Bar chart ordered alphabetically

Population for 20 European countries in 2004



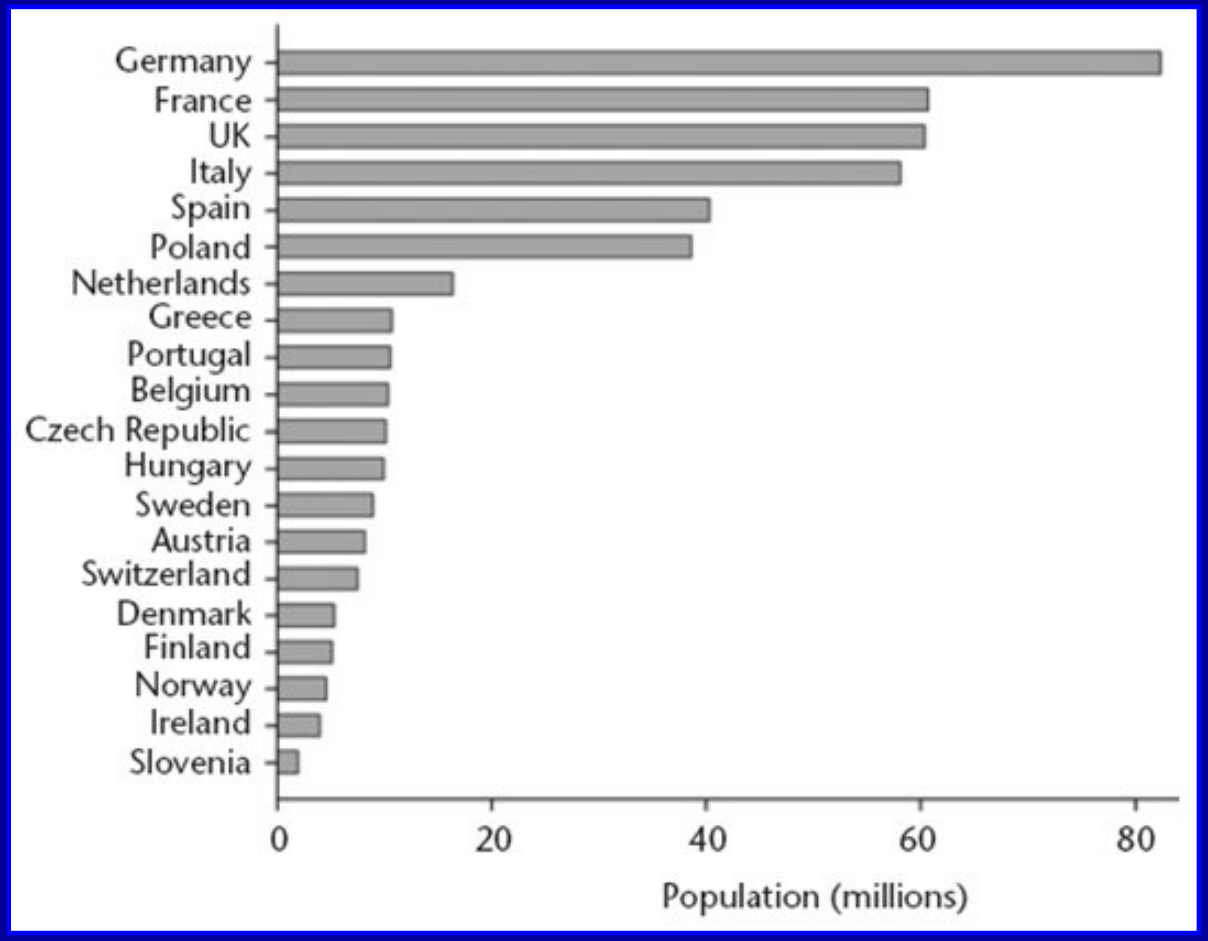
The most populous country, Germany, can be readily seen

It's not obvious for France, Italy, & UK which has largest population

Schott B. Schott's almanac. London: Bloomsbury; 2006.

Bar chart ordered by size

Population for 20 European countries in 2004

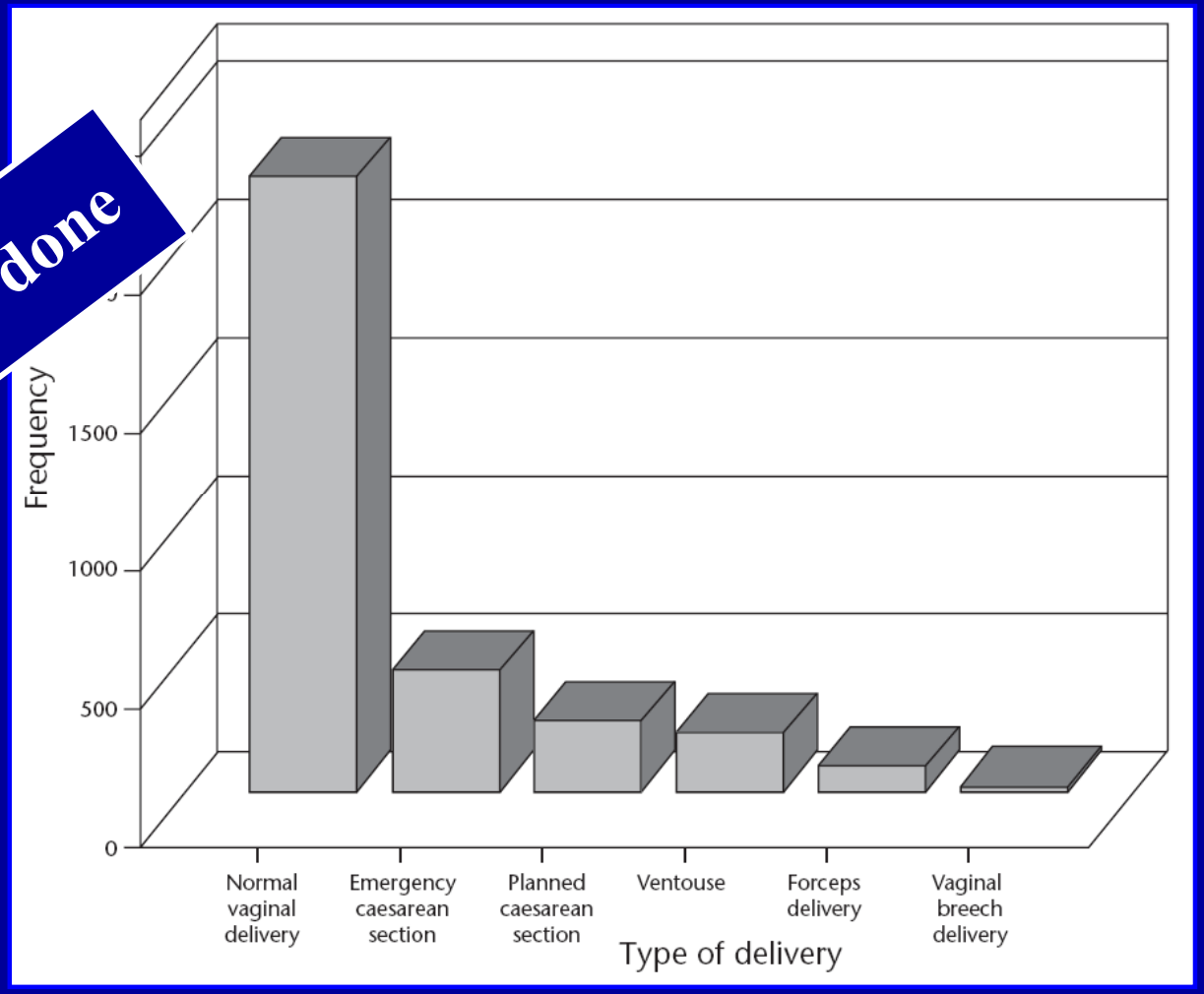


It is clear now how each country relates to others for population size

Schott B. Schott's almanac. London: Bloomsbury; 2006.

Three-dimensional column charts

Self-reported type of delivery for all new mothers (N: 3 321)



Should never be done

Data have only two dimensions

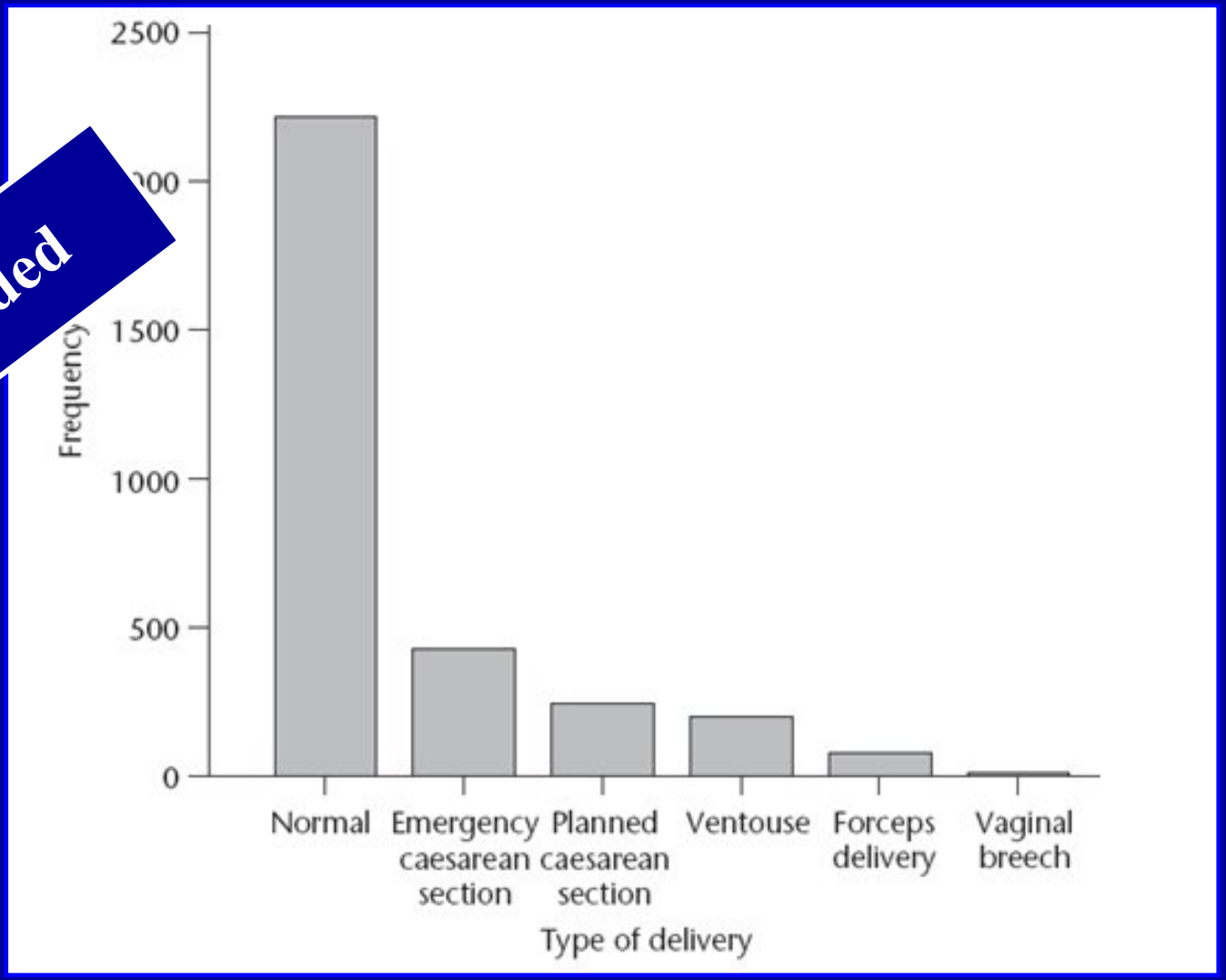
A third dimension is falsely introduced

BMJ 2002 ; 324 : 643 – 6.

Two-dimensional column charts

Self-reported type of delivery for all new mothers (N: 3 321)

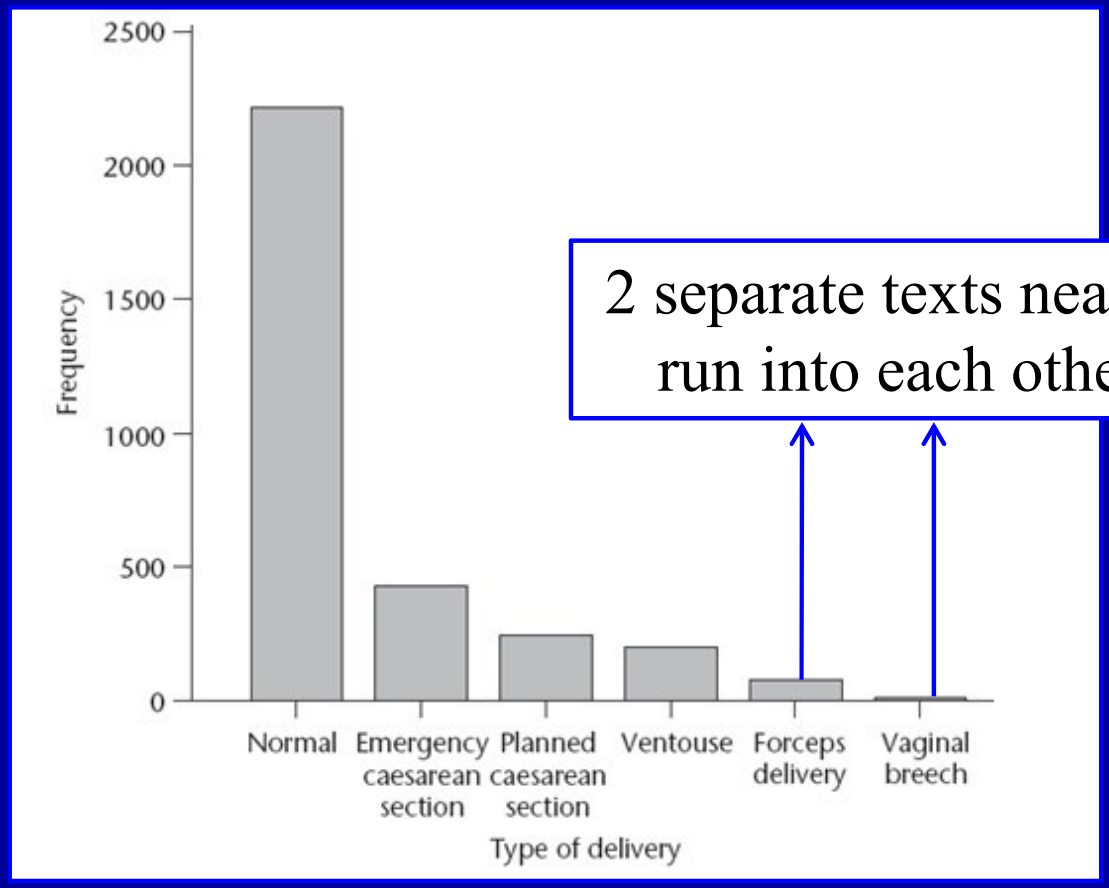
Recommended



BMJ 2002 ; 324 : 643 – 6.

Two-dimensional column charts

Self-reported type of delivery for all new mothers (N: 3 321)



Space of a journal column (**8 cm**)

Chart on the verge of being **overcrowded**

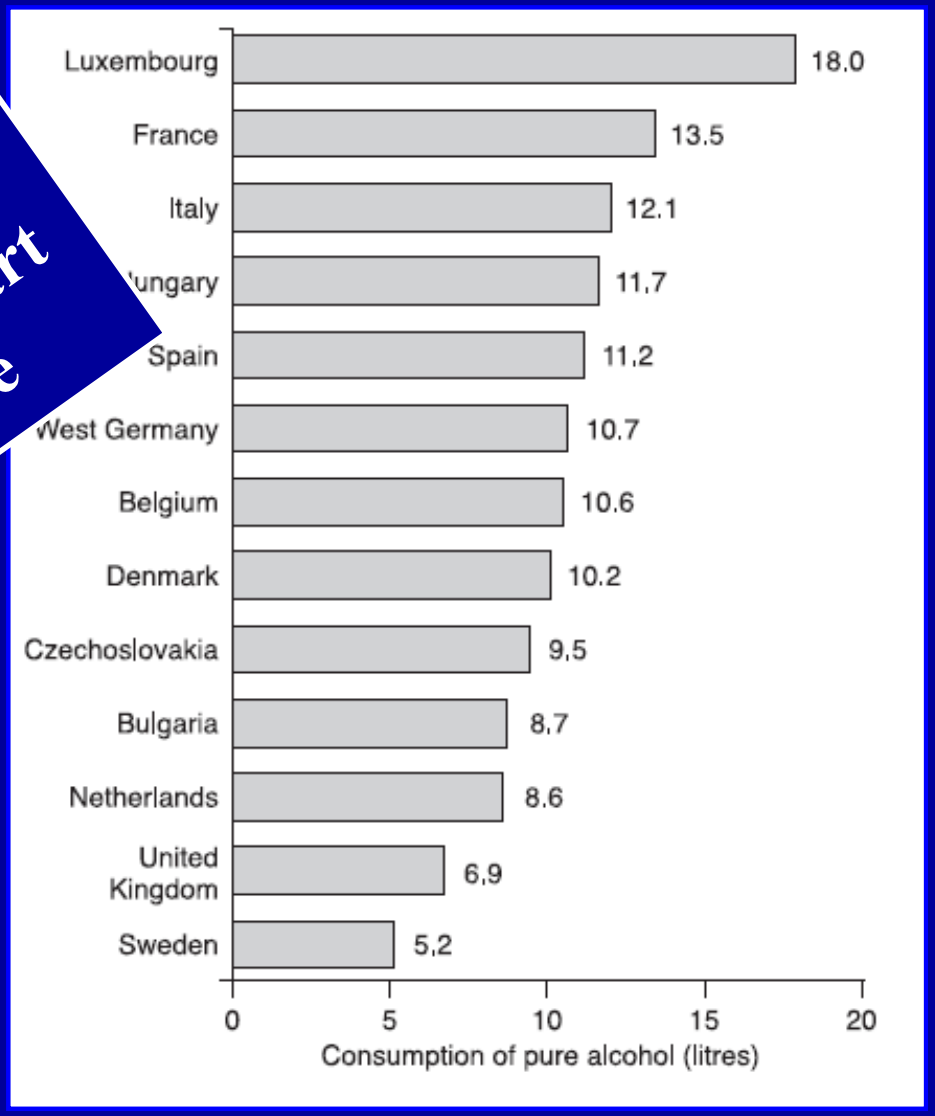
Problem overcome with use of **bar (horizontal) chart**

BMJ 2002 ; 324 : 643 – 6.

Bar chart

Annual alcohol consumption per inhabitant in Europe

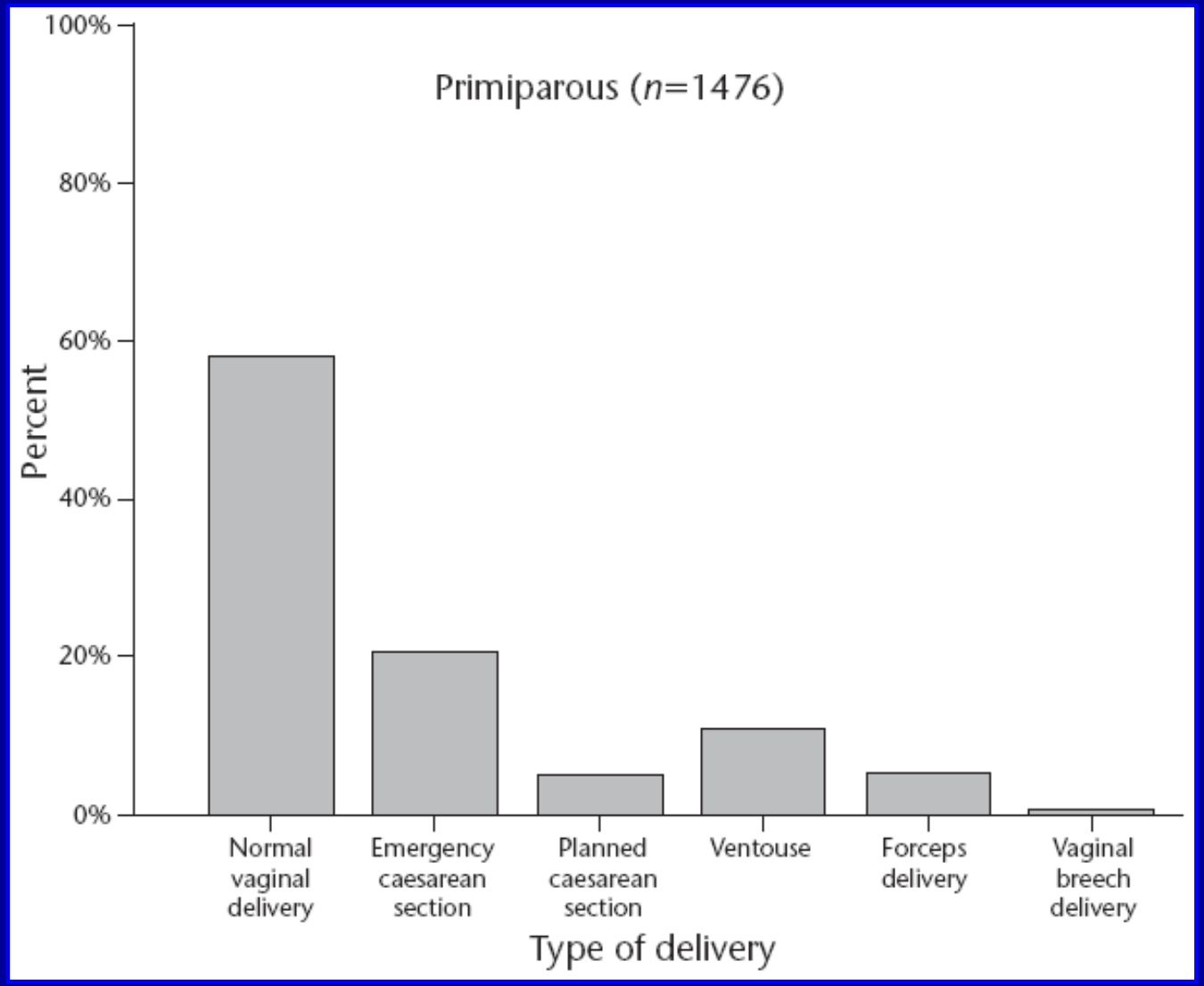
Good example
Bar (horizontal) chart
Ordered by size



Gustavii B. How to write & illustrate scientific papers.
Cambridge University Press, Cambridge, UK, 2nd edition, 2008.

Two-dimensional column charts

Self-reported type of delivery for all new mothers (N: 3 321)

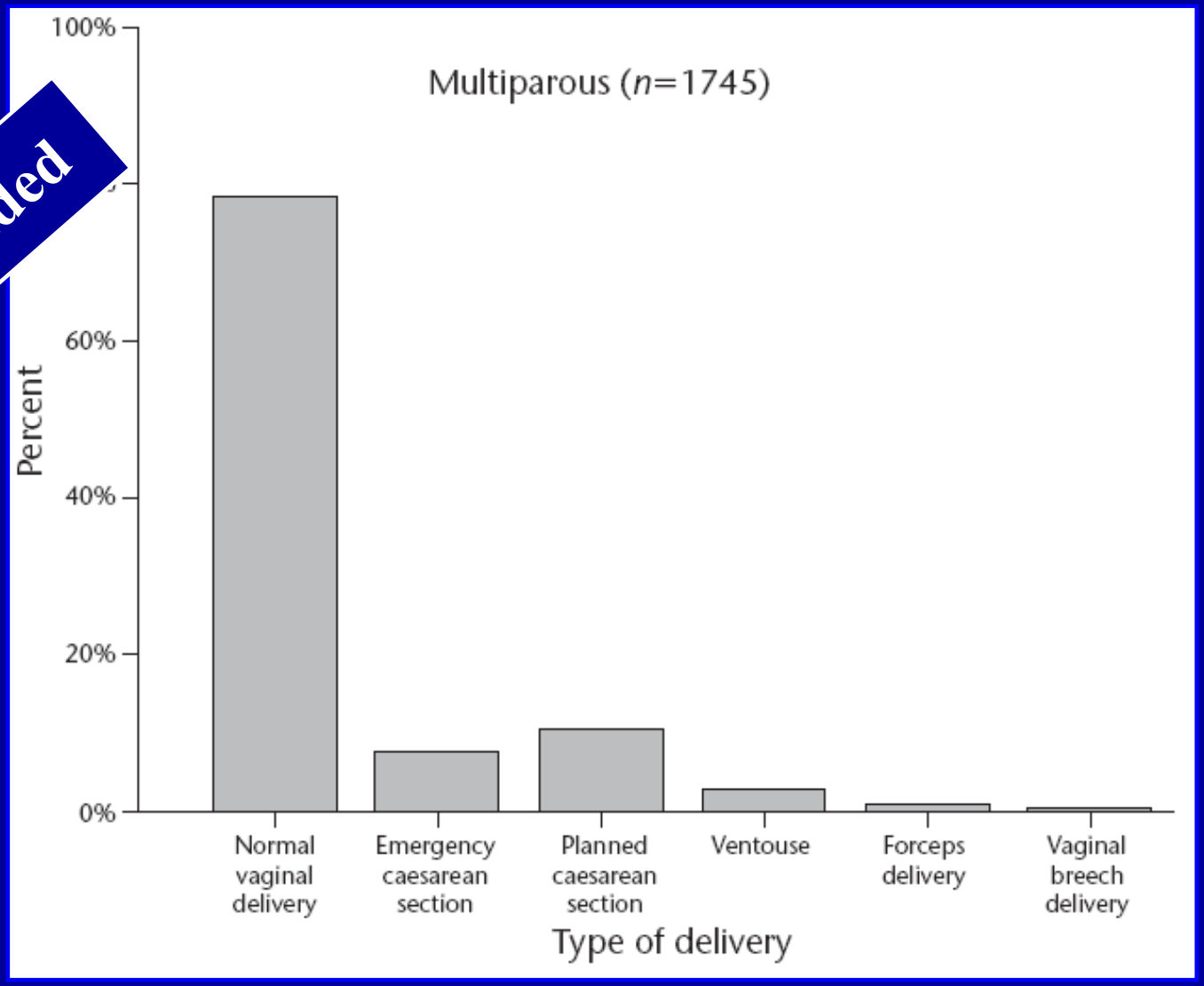


BMJ 2002 ; 324 : 643 – 6.

Two-dimensional column charts

Self-reported type of delivery for all new mothers (N: 3 321)

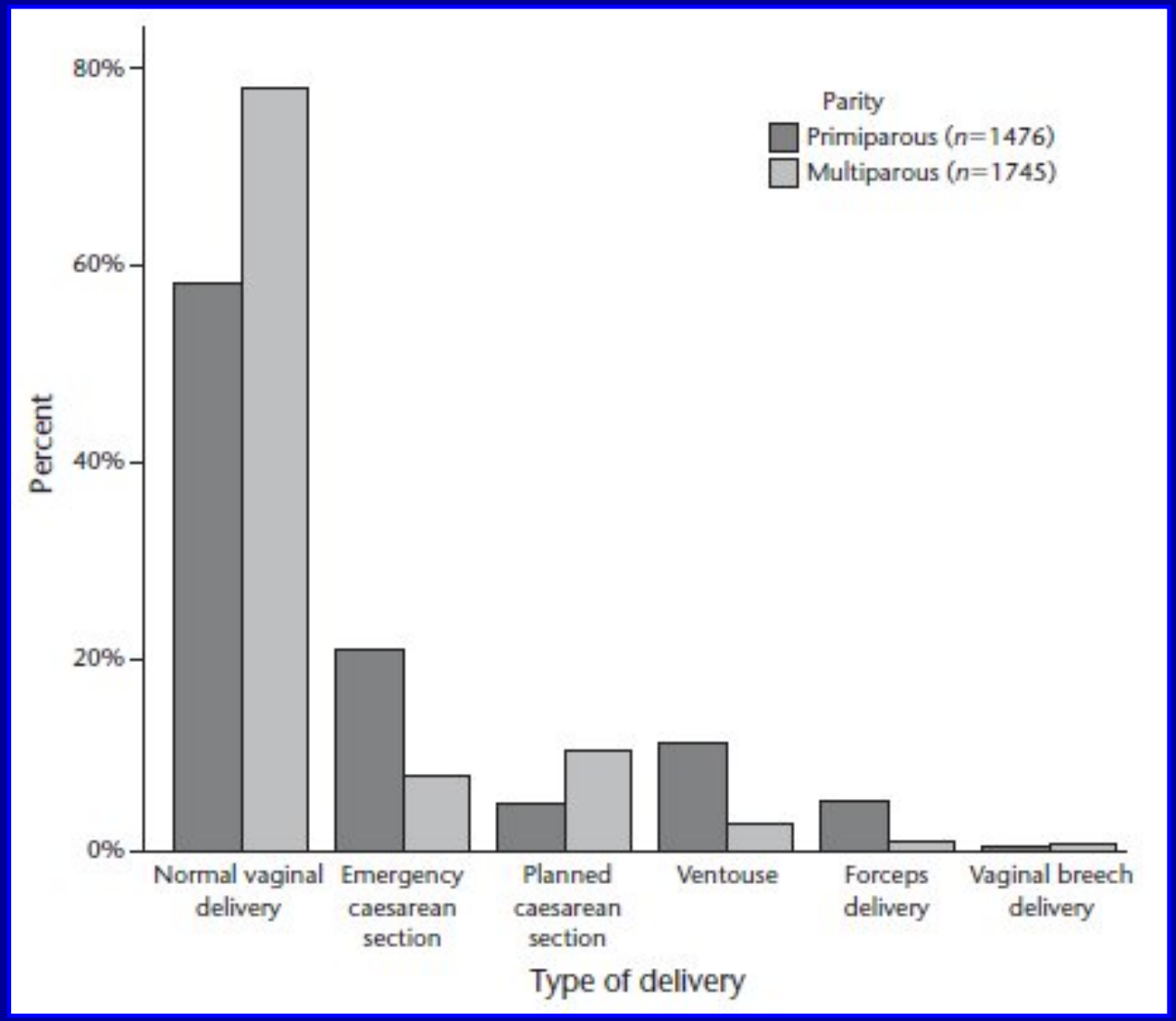
Not recommended



BMJ 2002 ; 324 : 643 – 6.

Grouped column graph

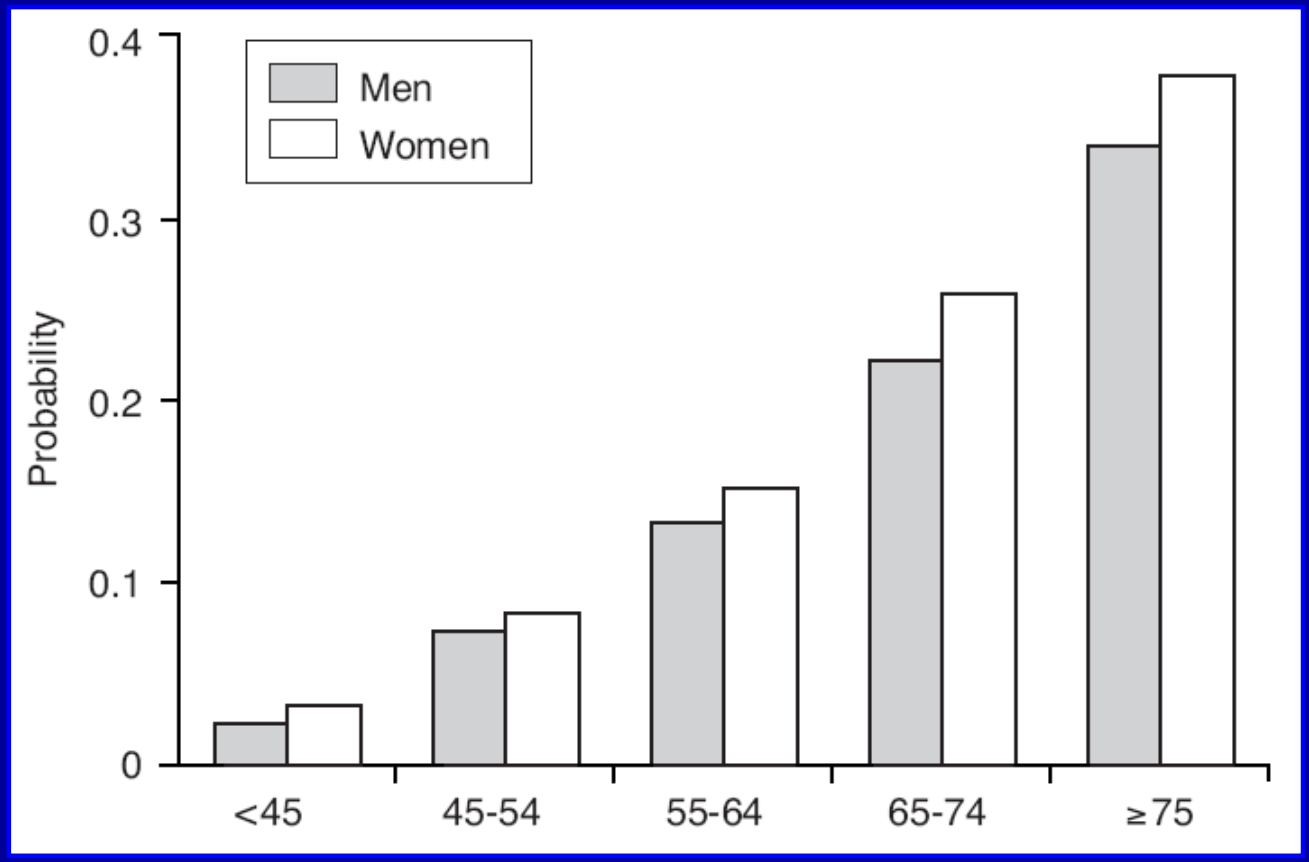
Self-reported type of delivery for all new mothers (N: 3 321)



BMJ 2002 ; 324 : 643 – 6.

Grouped column graph

Probability of dying in ICU after admission with AMI



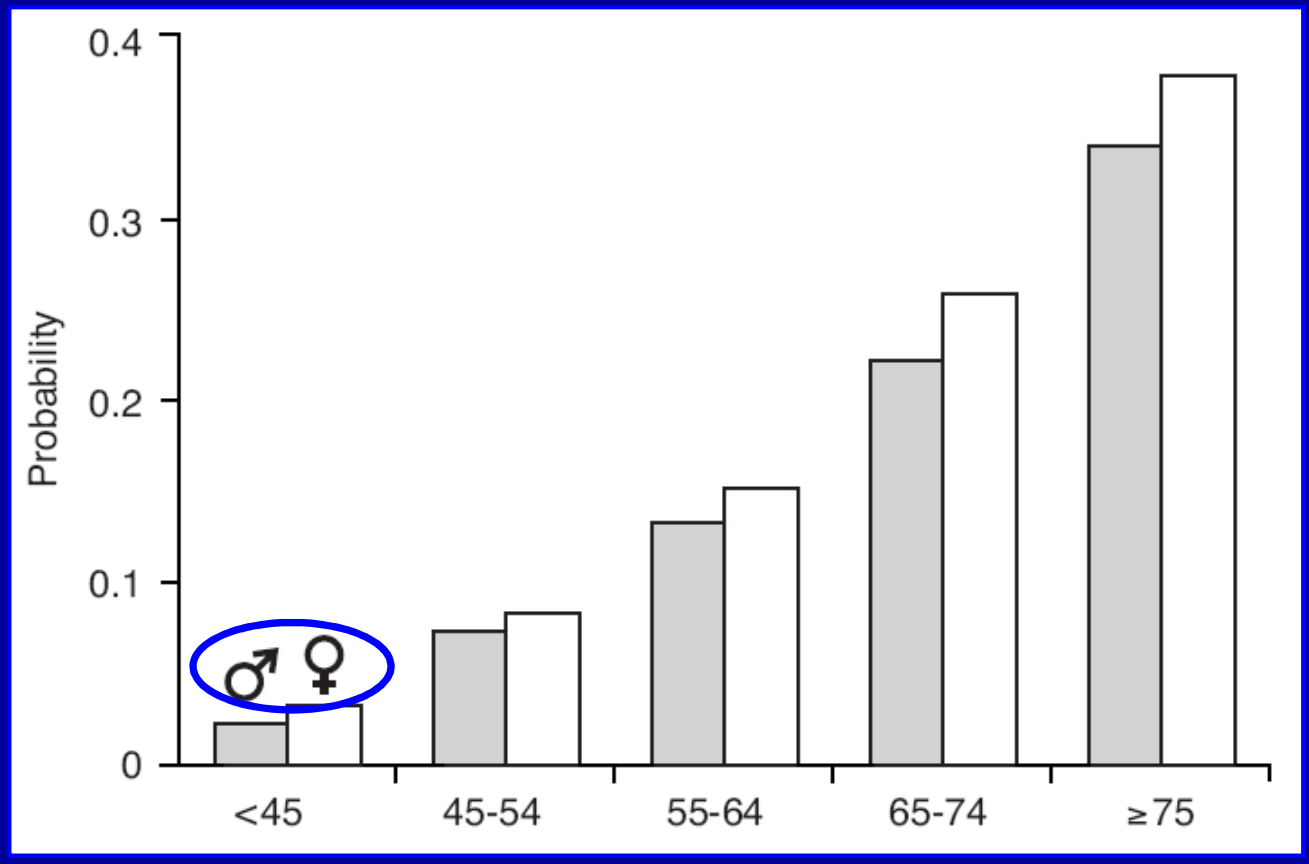
2 – 3 categories in each group should be the maximum

Remove the keys

Gustavii B. How to write & illustrate scientific papers.
Cambridge University Press, Cambridge, UK, 2nd edition, 2008.

Grouped bar chart

Probability of dying in ICU after admission with AMI



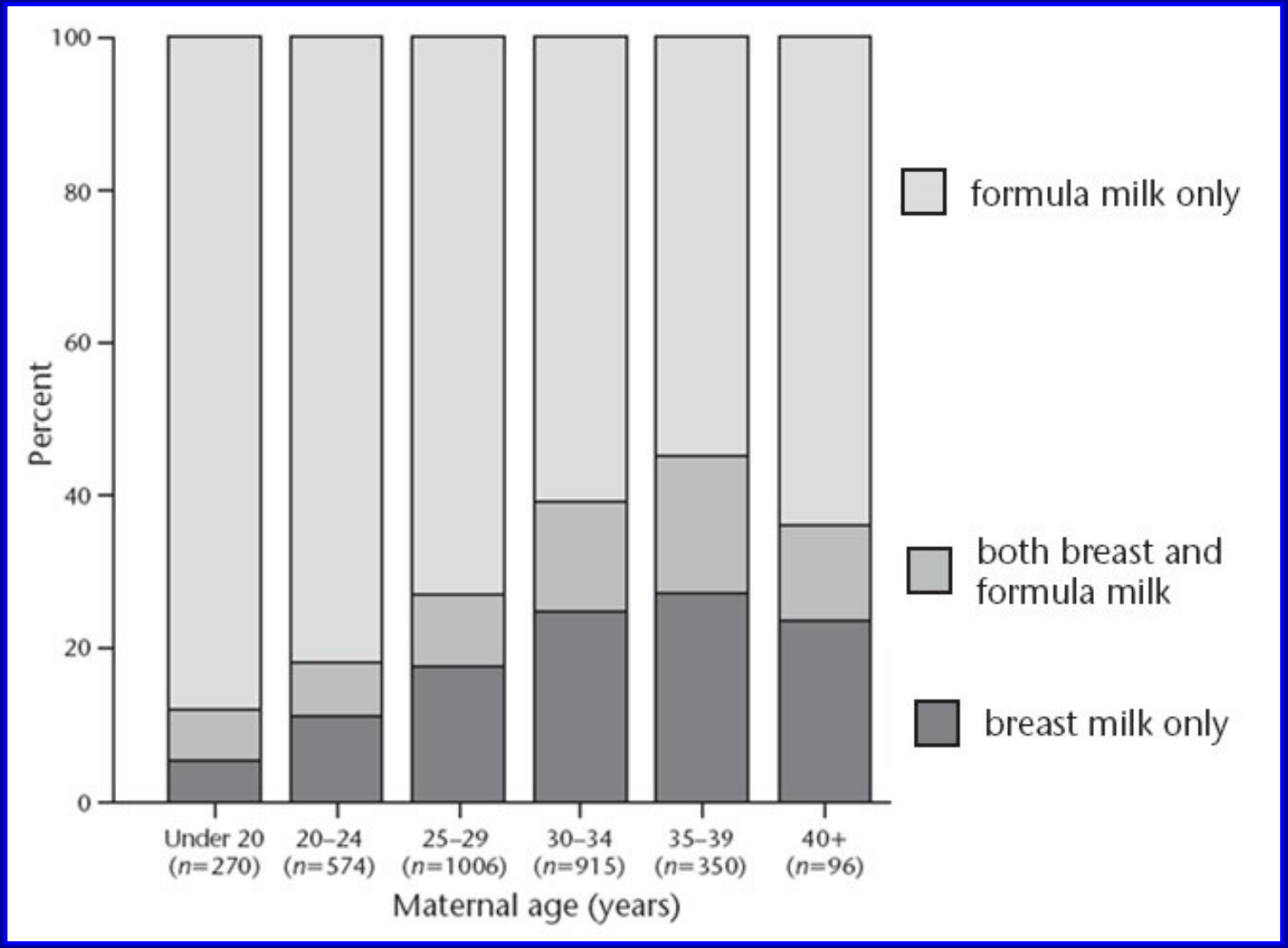
Remove the keys

One way to remove the keys is to label first group directly

Gustavii B. How to write & illustrate scientific papers.
Cambridge University Press, Cambridge, UK, 2nd edition, 2008.

Segmented column charts

Feeding method by maternal age for all women

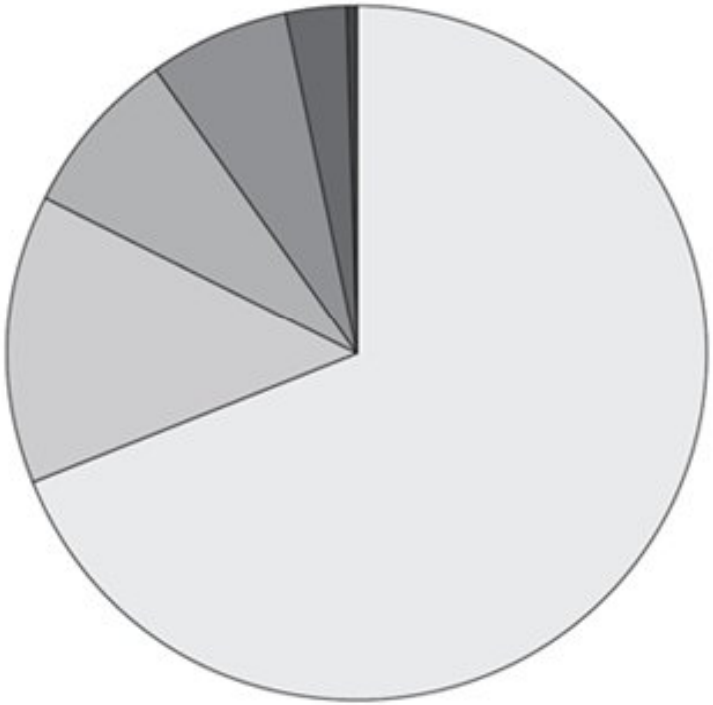


BMJ 2002 ; 324 : 643 – 6.

Pie chart

Appropriate usage in a magazine article

- Large segment begins at 12 o'clock
- Proceed in clockwise direction
- Ordered by size
- No of observations & percentages
- Number of segments ≤ 5
- Color employed with caution

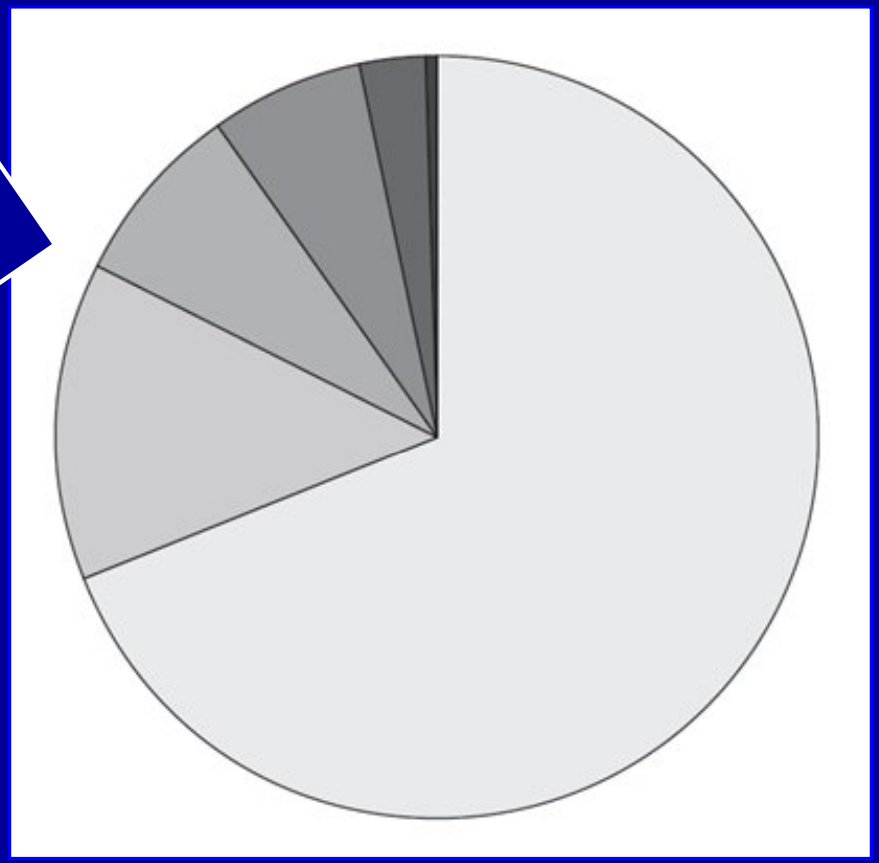


Freeman JV, Walters SJ, Campbell MJ. How to display data. Blackwell Publishing, Massachusetts, USA, 1st edition, 2008.

Pie chart

Self-reported type of delivery for all new mothers (N: 3 321)

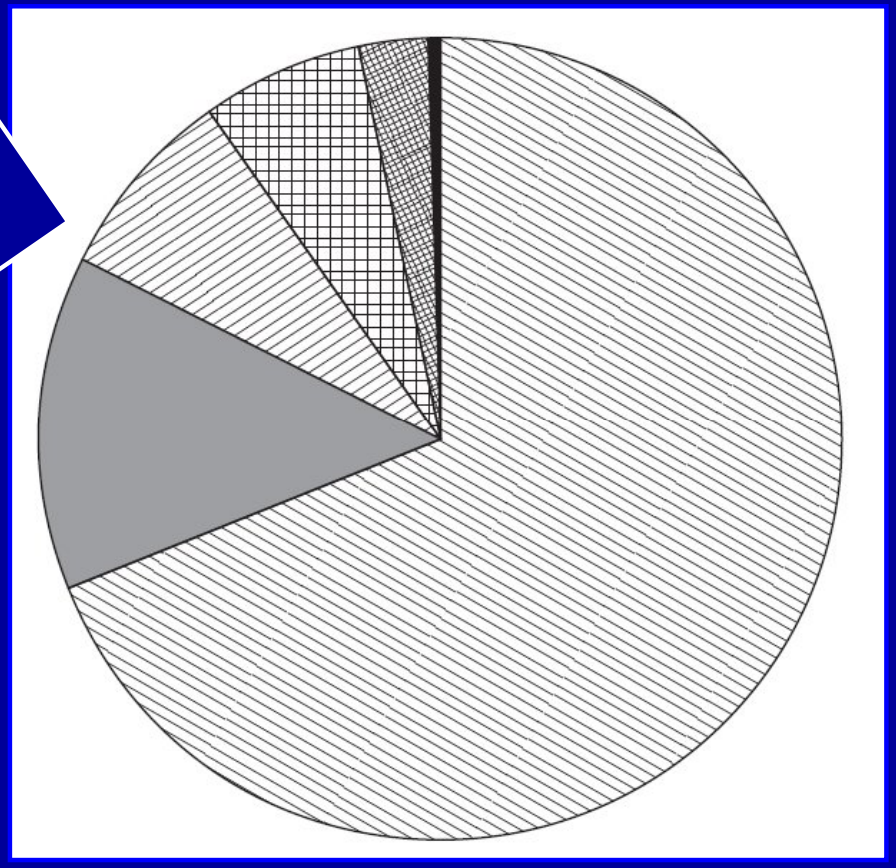
Different shadings of the same color



Pie chart

Self-reported type of delivery for all new mothers (N: 3 321)

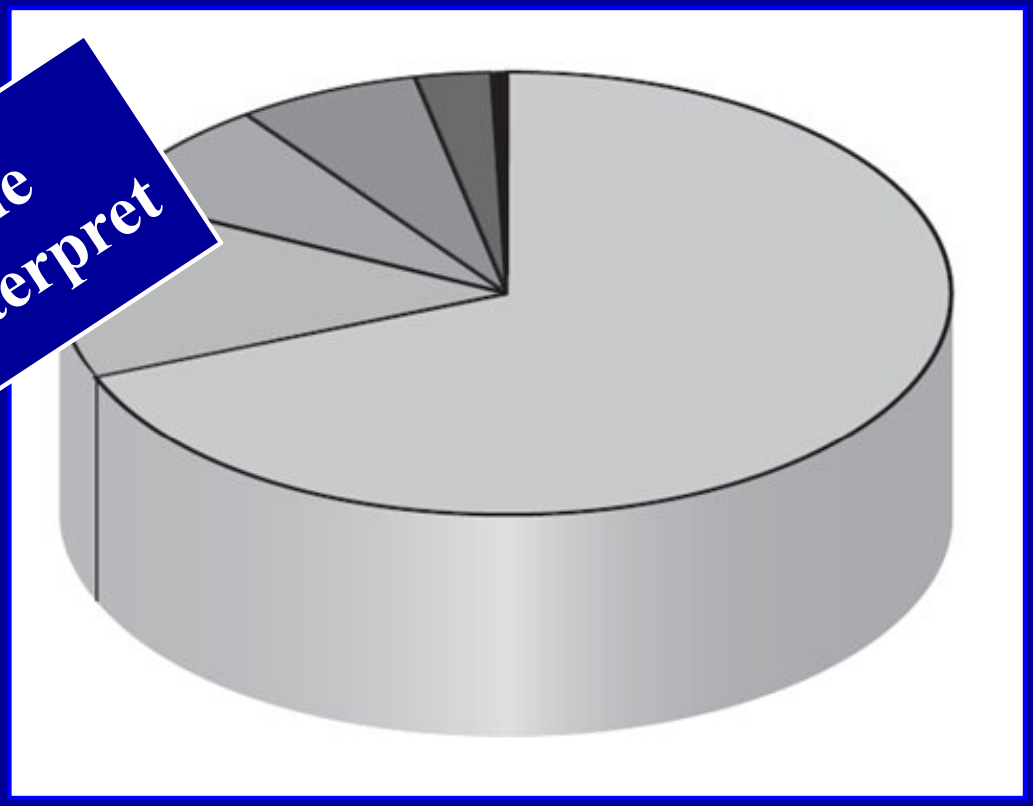
Different patterns
Chart look very busy



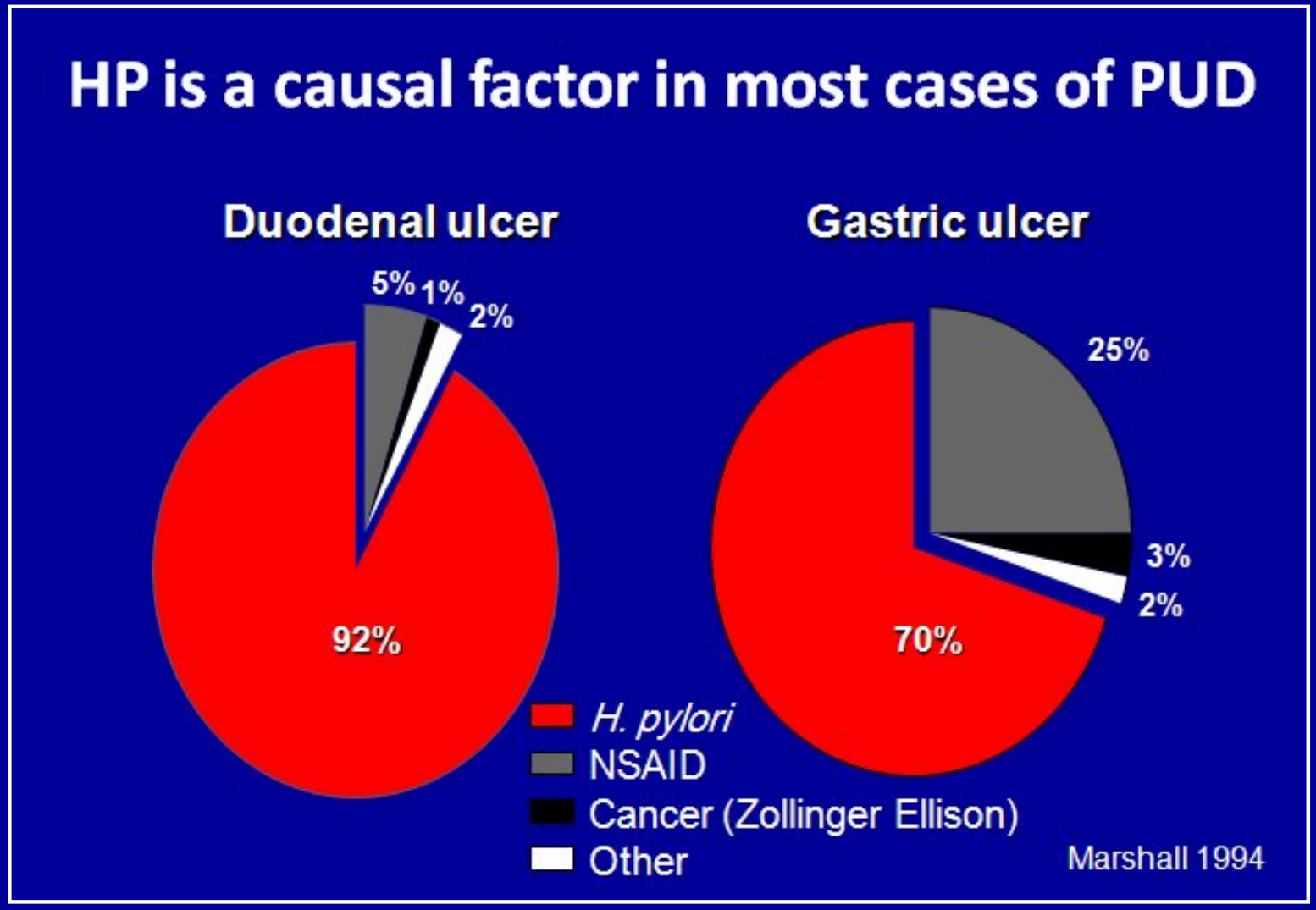
Three-dimensional pie charts

Self-reported type of delivery for all new mothers (N: 3 321)

Should never be done
Difficult to read & interpret



Pie charts



Pull out the slice you want to highlight

“The only worse design than a pie chart is several of them”

Tufte ER. The visual display of quantitative information.
Cheshire, Connecticut: Graphics Press; 1983

Types of data

- **Qualitative (categorical)**

Dichotomous	Only 2 values
Nominal	Unordered
Ordinal	Ordered

- **Quantitative (numerical)**

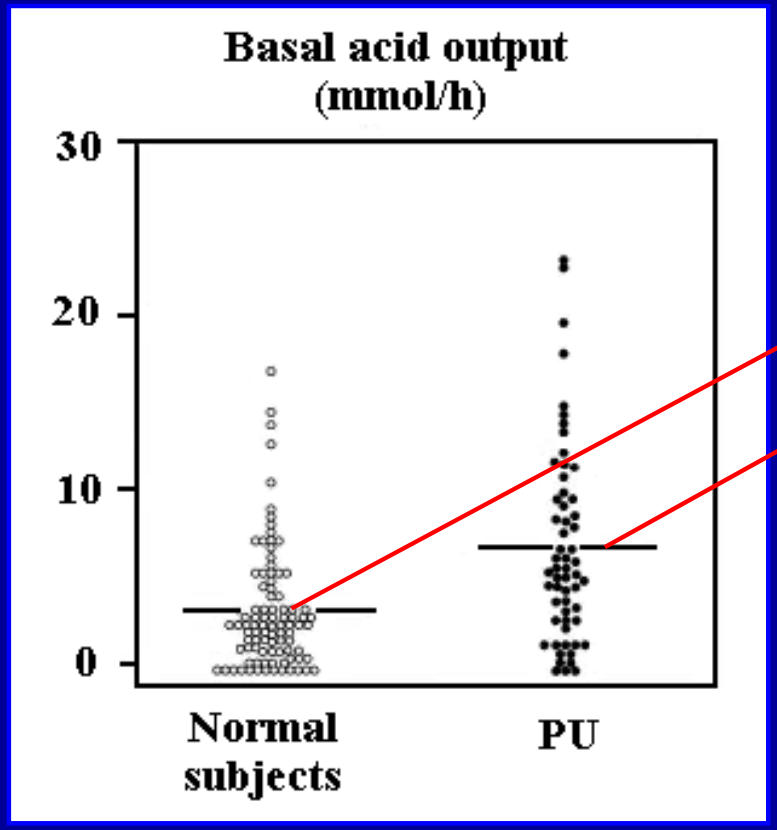
Counted	Gaps
Continuous	No gaps

Display quantitative data

- **Counted (gaps)** — **Bar chart**
- **Continuous (no gaps)** — **Dot plot**
  Stem & leaf plot
  Histograms
  Box-whisker plot

Dot plot

BAO in normal subjects & PU patients



Use dot plot if sample size per group is low (<100)

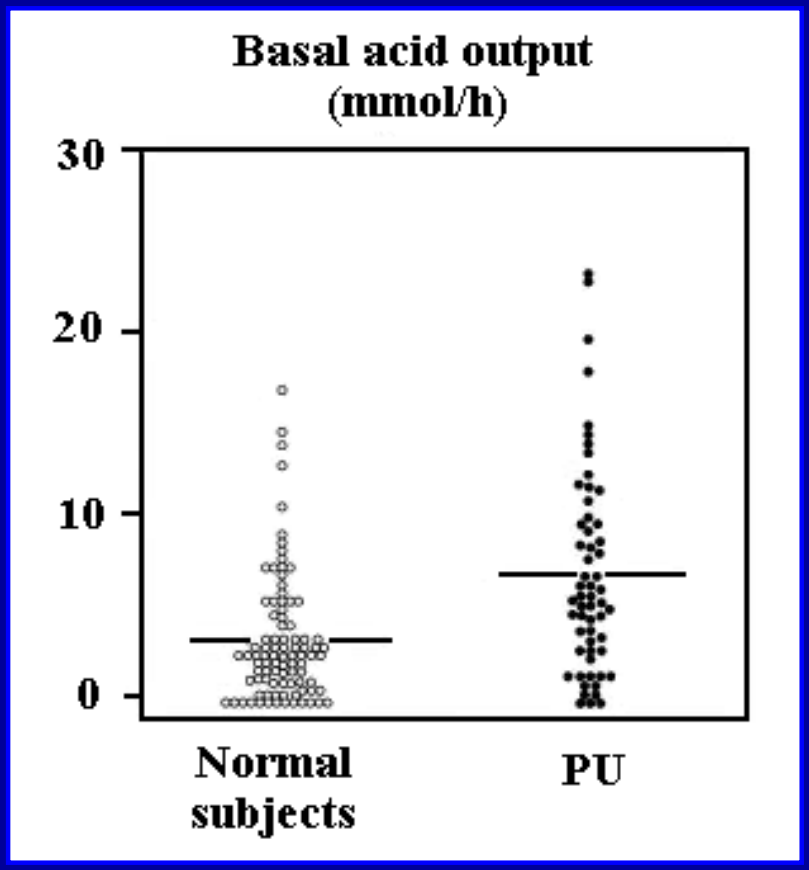
Each point represents a value for a single individual

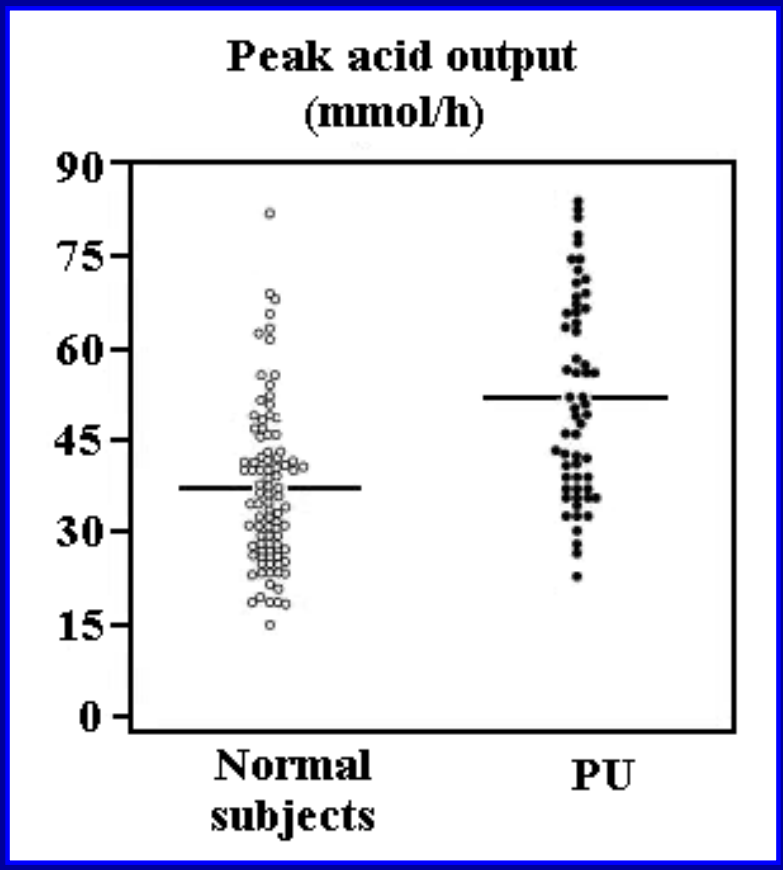
Horizontal lines indicate mean values

Blair AJ et al. J Clin Invest 1987 ; 79 : 582.

Dot plot

BAO & PAO in normal subjects & PU patients





Substantial overlap in values among individuals in the groups

Blair AJ et al. J Clin Invest 1987 ; 79 : 582.

Stem and leaf plot

Height of male in leg ulcer patients (n: 77)

Frequency	Stem	Leaf
1	1.55-	7
3	1.60-	333
4	1.65-	5588
18	1.70-	000000333333333333
24	1.75-	555558888888888888888888
15	1.80-	000000003333333
10	1.85-	5555888888
1	1.90-	13

Median

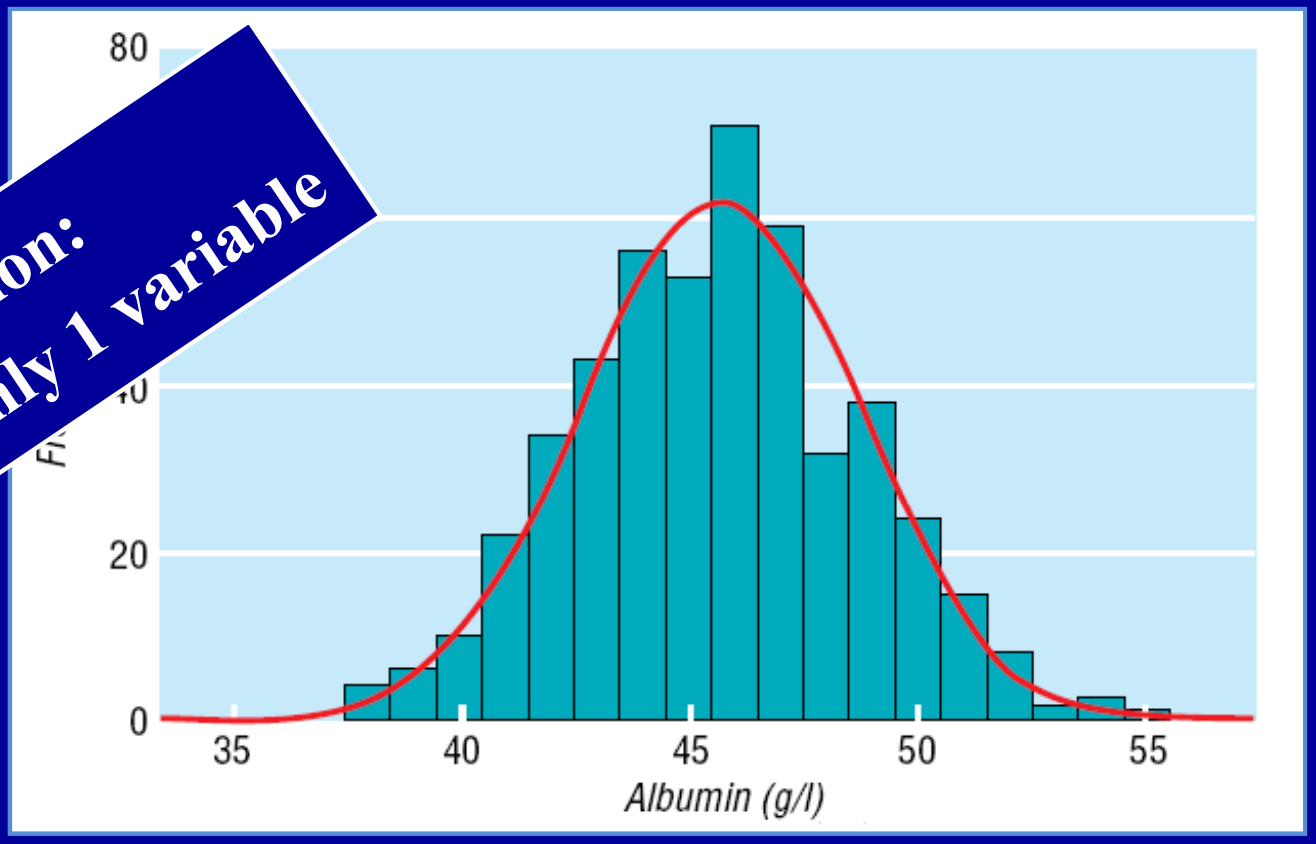
Each data is divided into 2 parts: leaf (last digit) & stem (other part)

Separate line for each different stem value

Stem on left of plot & leaves on right

Histogram

Serum albumin in 481 white men aged over 20



Limitation:
Represents only 1 variable

No gaps between columns (**continuous data**)

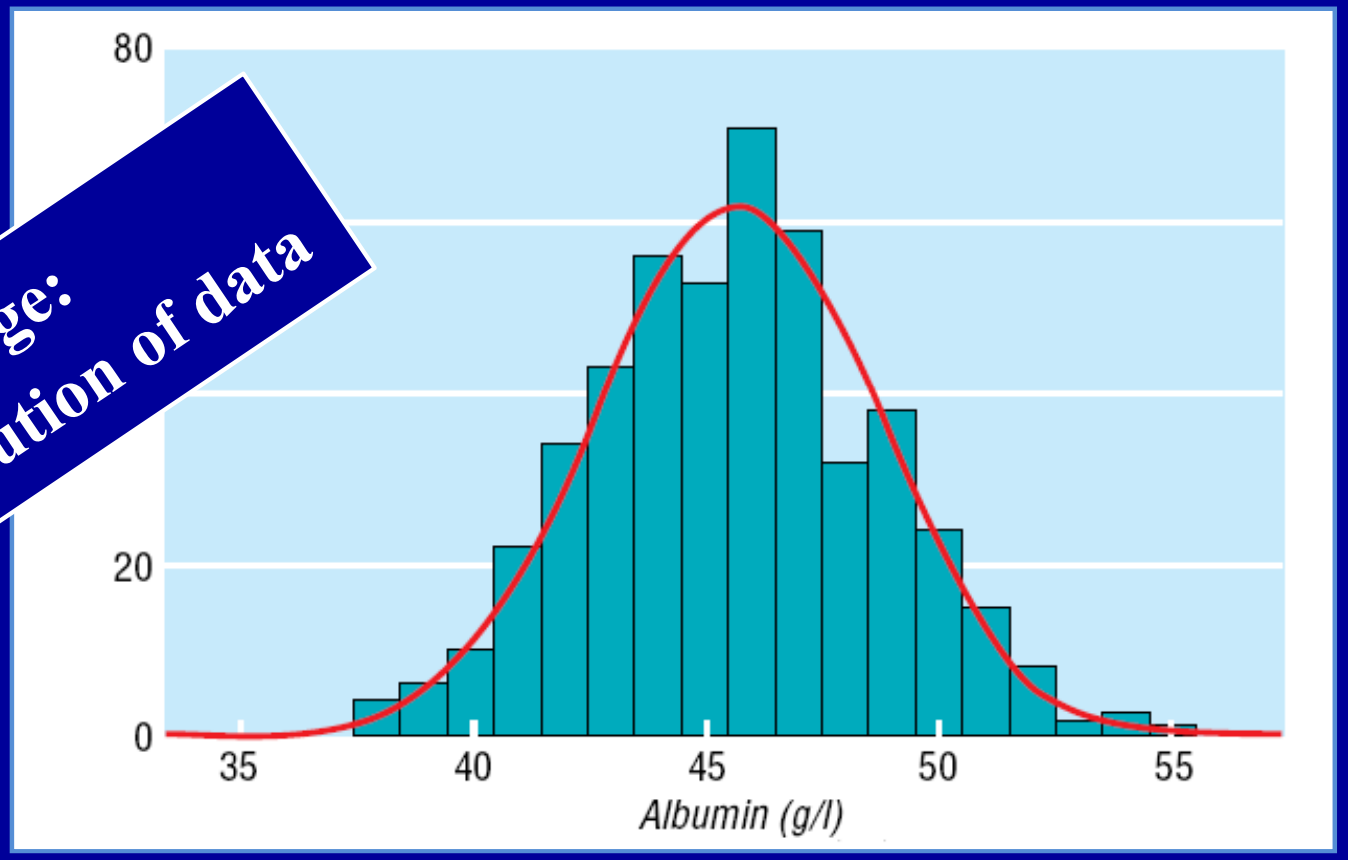
Keep same width of each group (**bin width**)

Columns labeled by using midpoint, or better **start or end** of interval

BMJ 1999 ; 318 : 1667.

Histogram – Normal distribution

Serum albumin in 481 white men aged over 20



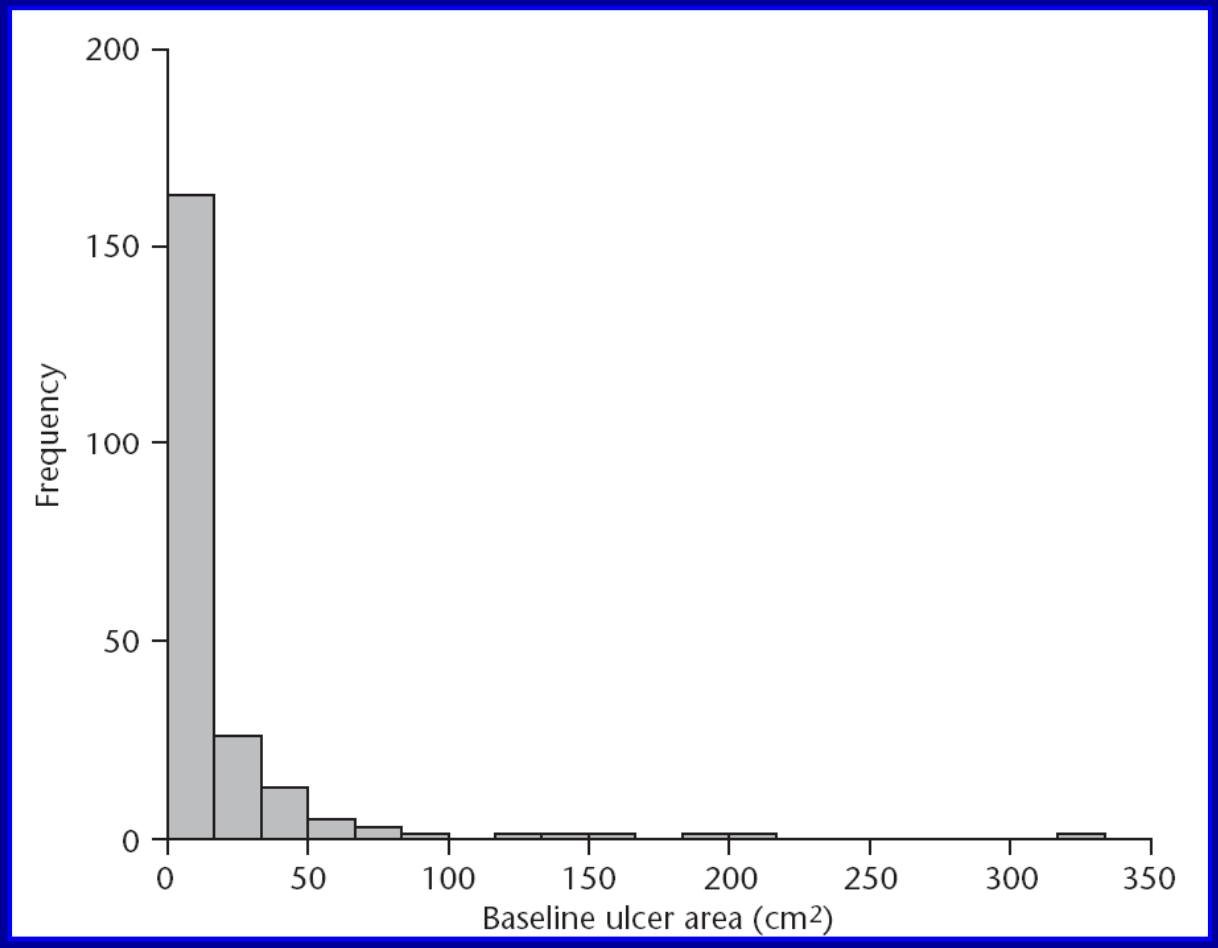
Advantage:
Assess distribution of data

Mean: 46.14 g/l – SD: 3.08 g/l

BMJ 1999 ; 318 : 1667.

Histogram – Positively skewed data

Baseline ulcer area from the leg ulcer trial (n: 233)

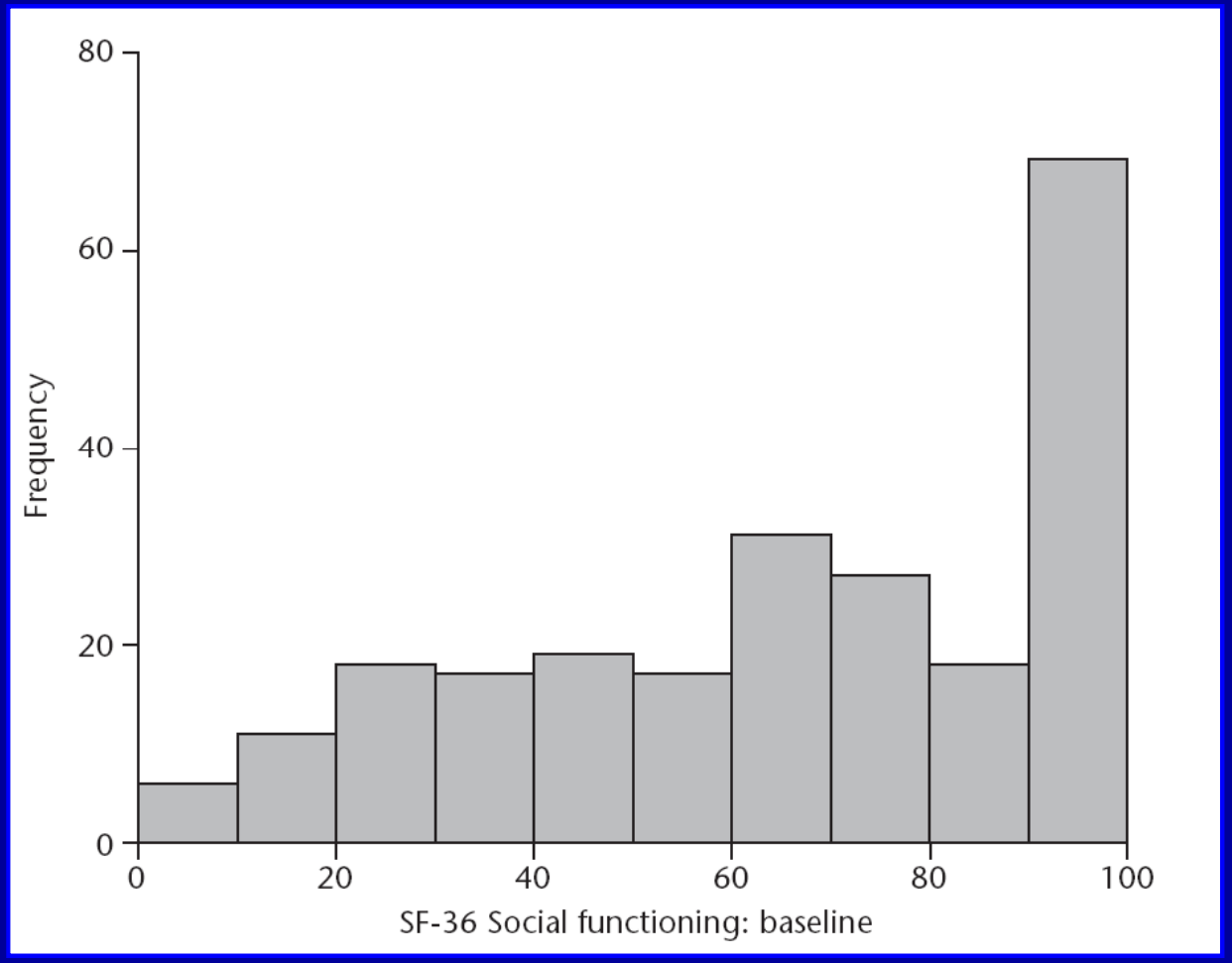


Peak at lower values & a long tail of higher values

BMJ 1998 ; 316 : 1487 – 91.

Histogram – Negatively skewed data

Baseline social functioning in leg ulcer trial (n: 233)



Long left tail of lower values & peak at higher values

BMJ 1998 ; 316 : 1487 – 91.

Number of categories in a histogram

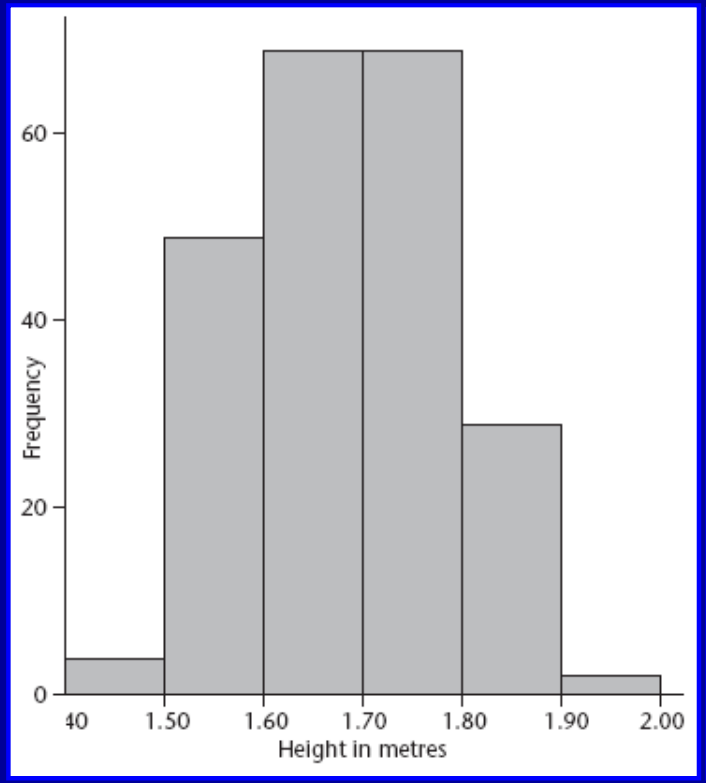
No hard & fast rules about appropriate number

- **Too few** Much important information lost
- **Too many** Patterns obscured by too much detail
- **Usually** **5 – 15 categories** will be enough

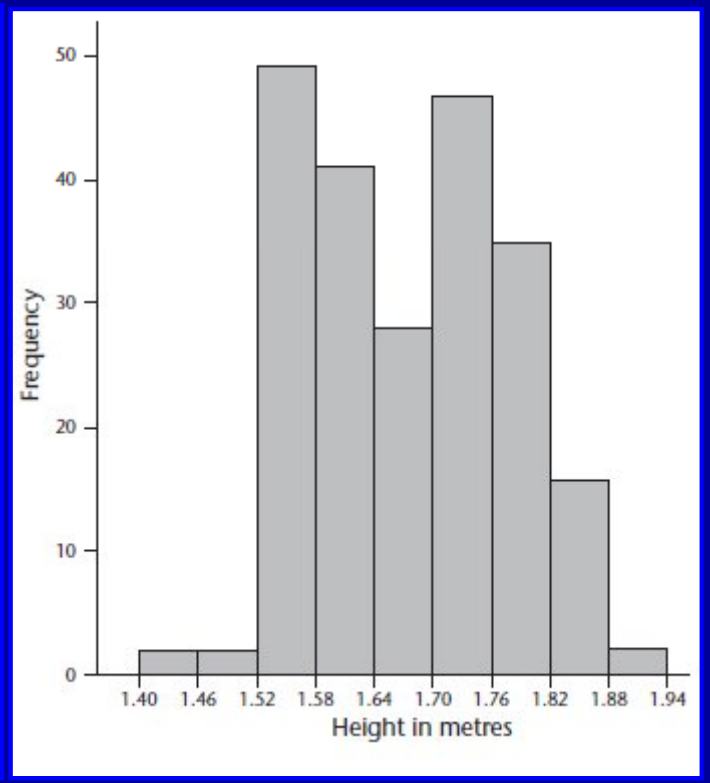
Number of categories in a histogram

Height for leg ulcer patients (n 233)

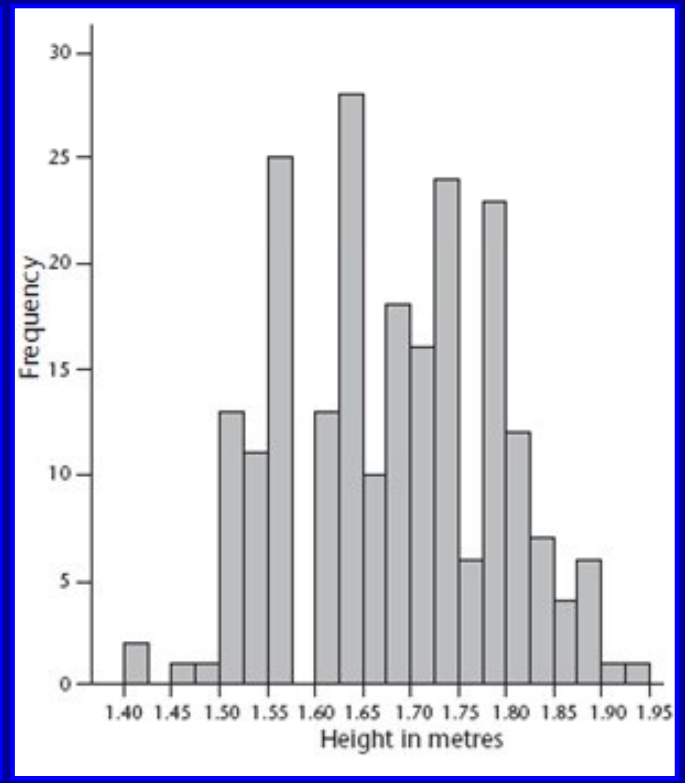
Too few (6 categories)



Good (9 categories)

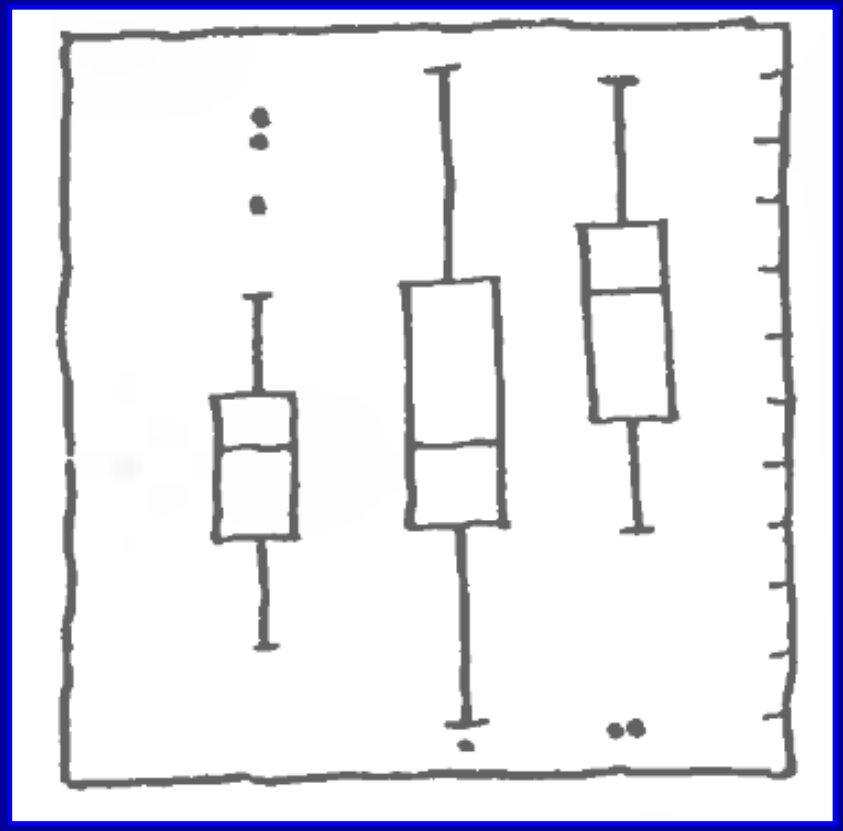


Too many (22 categories)



Freeman JV et all. How to display data. Blackwell Publishing, MA, USA, 2008.

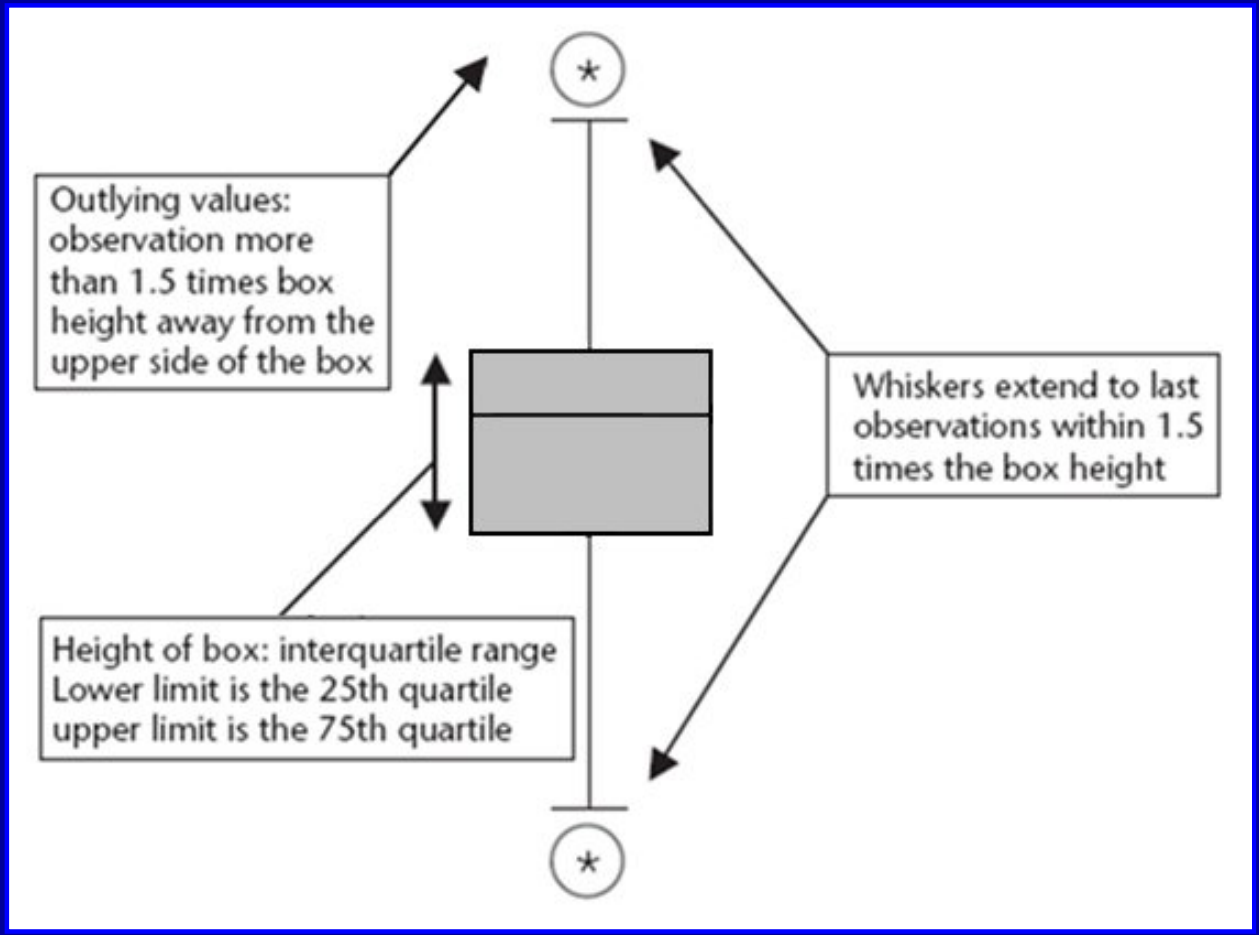
Box-and-whiskers plots



Especially good to show differences between groups

Gonick L & Smith W. The cartoon guide to statistics.
HarperCollins Publishers, New York, USA, 1st edition, 1993

Box-whisker plot

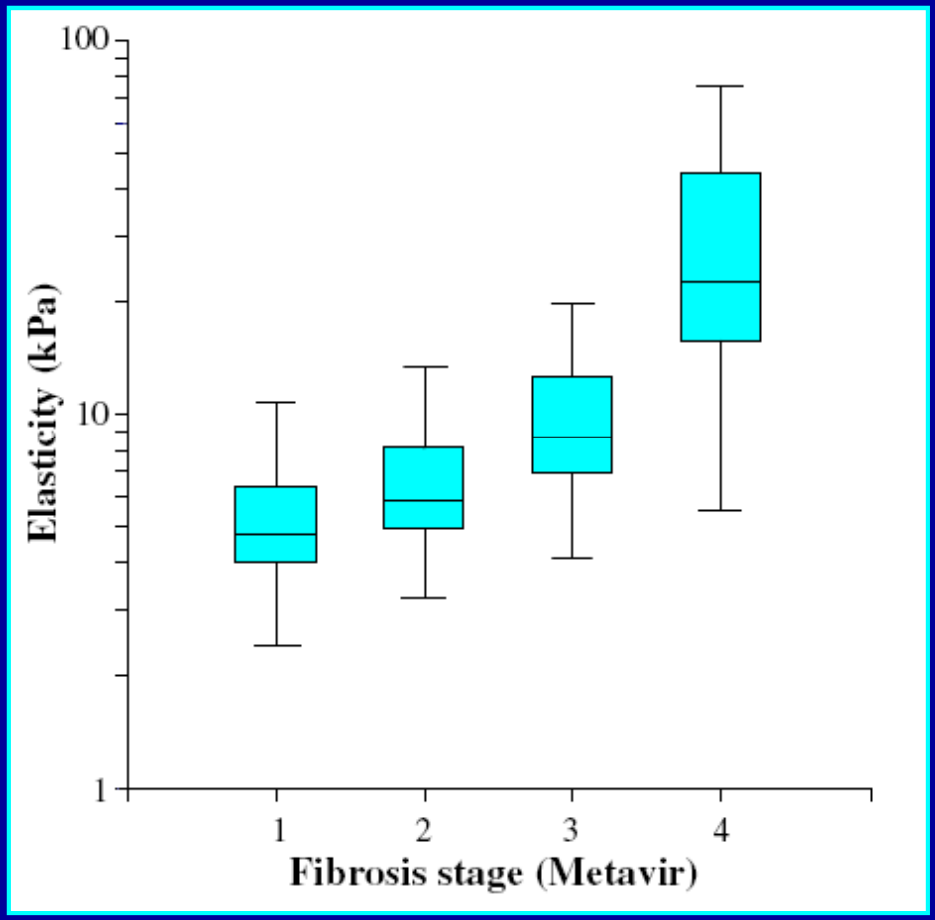


As there are many variations, you have to explain details of the plot

Freeman JV et all. How to display data. Blackwell Publishing, Massachusetts, 2008.

Box-and-whiskers plots

Liver stiffness for each Metavir stage in CHC



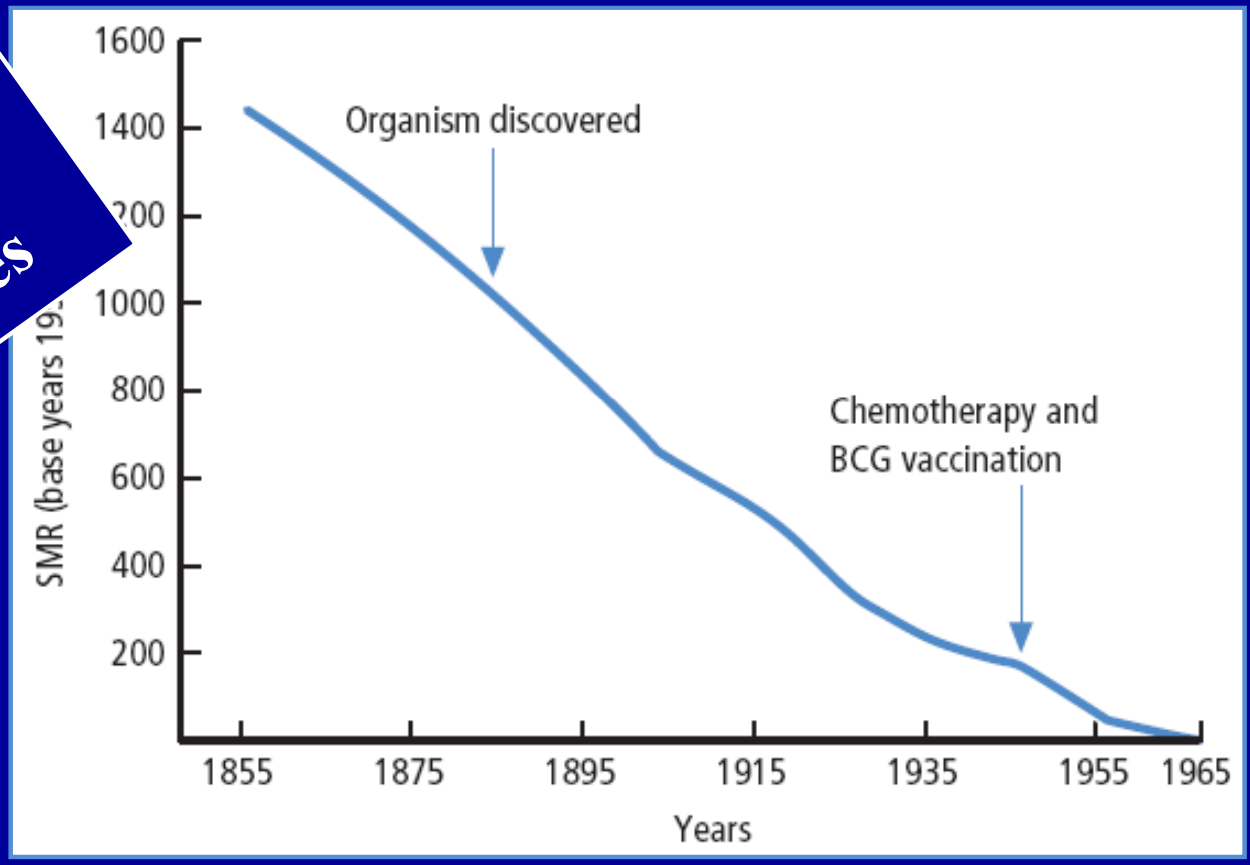
Vertical axis is in logarithmic scale (wide range of F4 values)

Gastroenterology 2005 ; 28 : 343 – 350.

Line graph

TB mortality in England &Wales

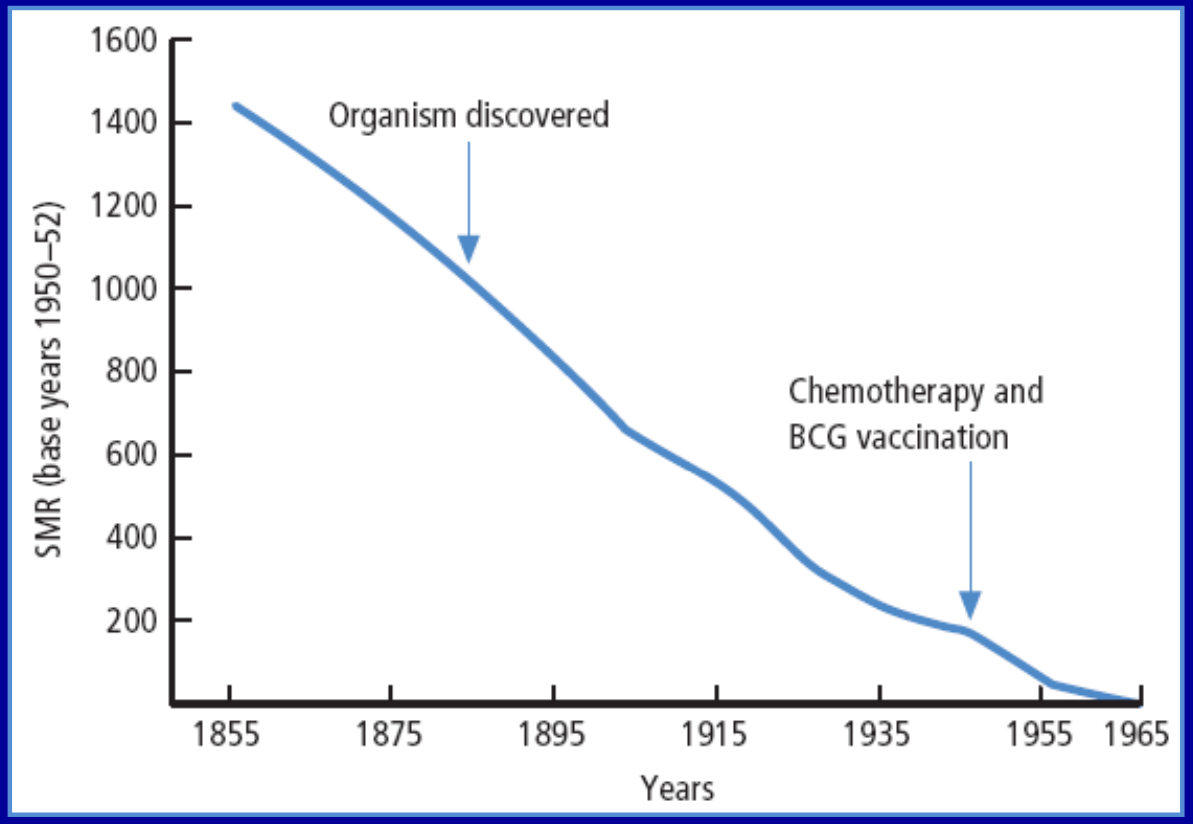
Main role:
Show time series



Farmer R Lawrenson R. Lecture Notes: Epidemiology & public health medicine. Blackwell Publishing, Oxford, 5th edition, 2004

Line graph – Arithmetic scale

TB mortality in England &Wales



Mortality seems hardly affected by the events

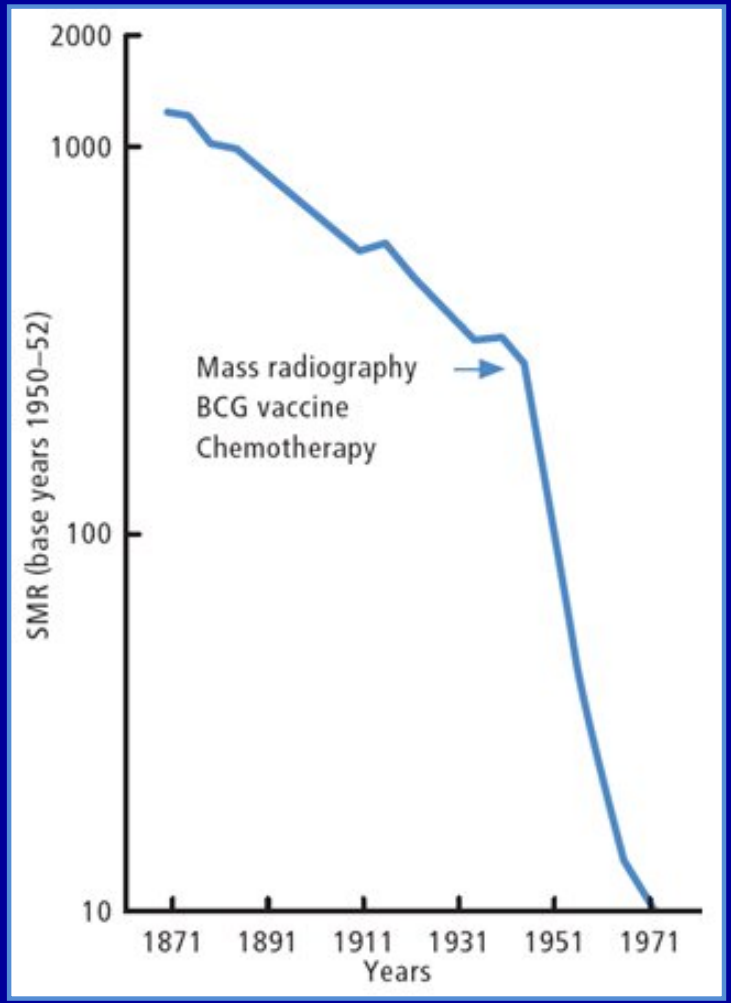
They played little part in mortality decline

Farmer R Lawrenson R. Lecture Notes: Epidemiology & public health medicine. Blackwell Publishing, Oxford, 5th edition, 2004

Line graph – Logarithmetic scale

TB mortality in England &Wales

Introduction of BCG vaccine & chemotherapy was associated with acceleration in established decline in mortality



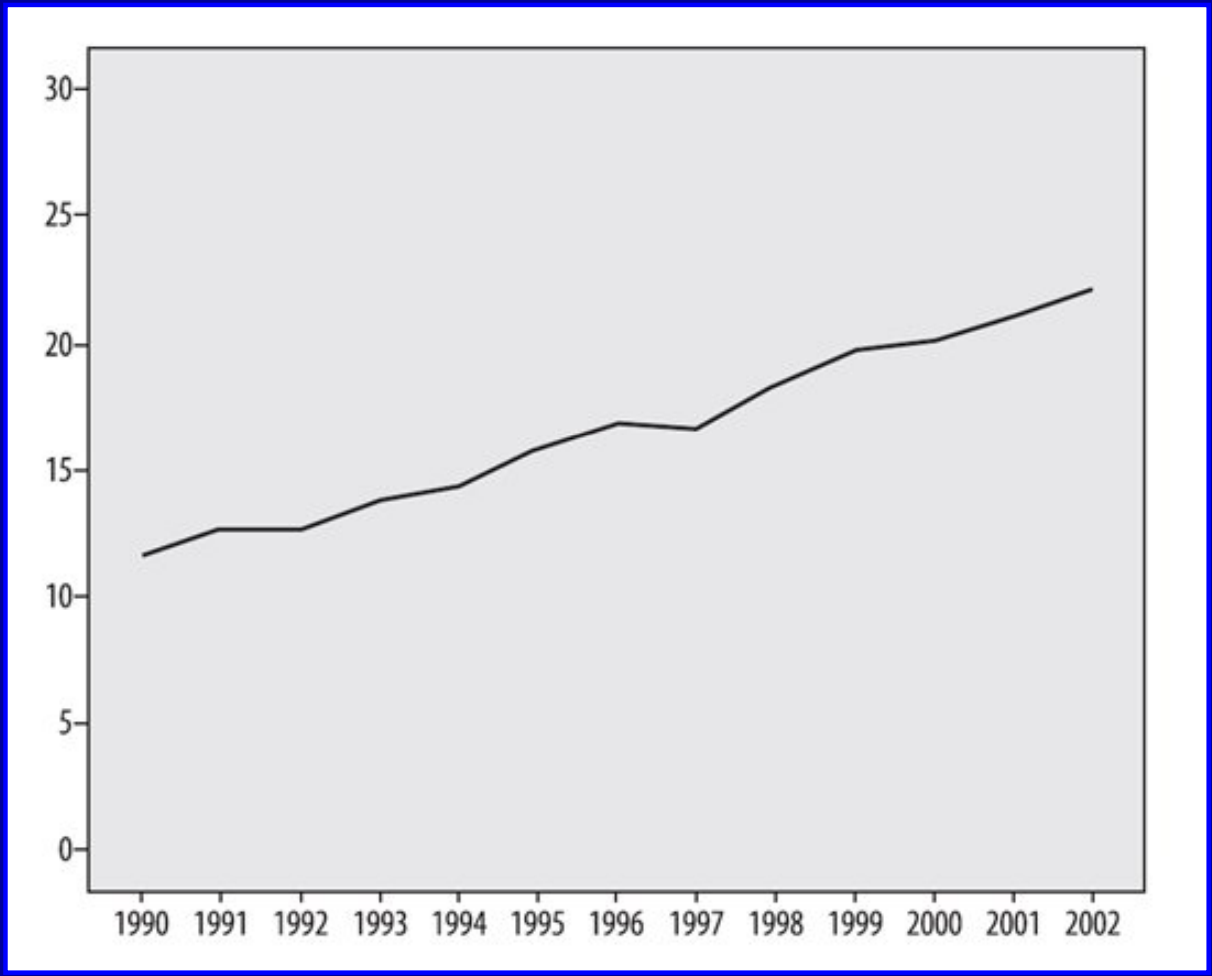
Farmer R, Lawrenson R. Lecture Notes: Epidemiology & public health medicine. Blackwell Publishing, Oxford, 5th edition, 2004

It is frequently necessary to examine secular trends both as changes in rates (arithmetic scale) and as rates of change (logarithmic scale) if the nature of a trend is to be fully appreciated

Farmer R Lawrenson R. Lecture Notes: Epidemiology & public health medicine. Blackwell Publishing, Oxford, 5th edition, 2004

Line chart

Obesity among adults from 1990 – 2002 (US-CDC)

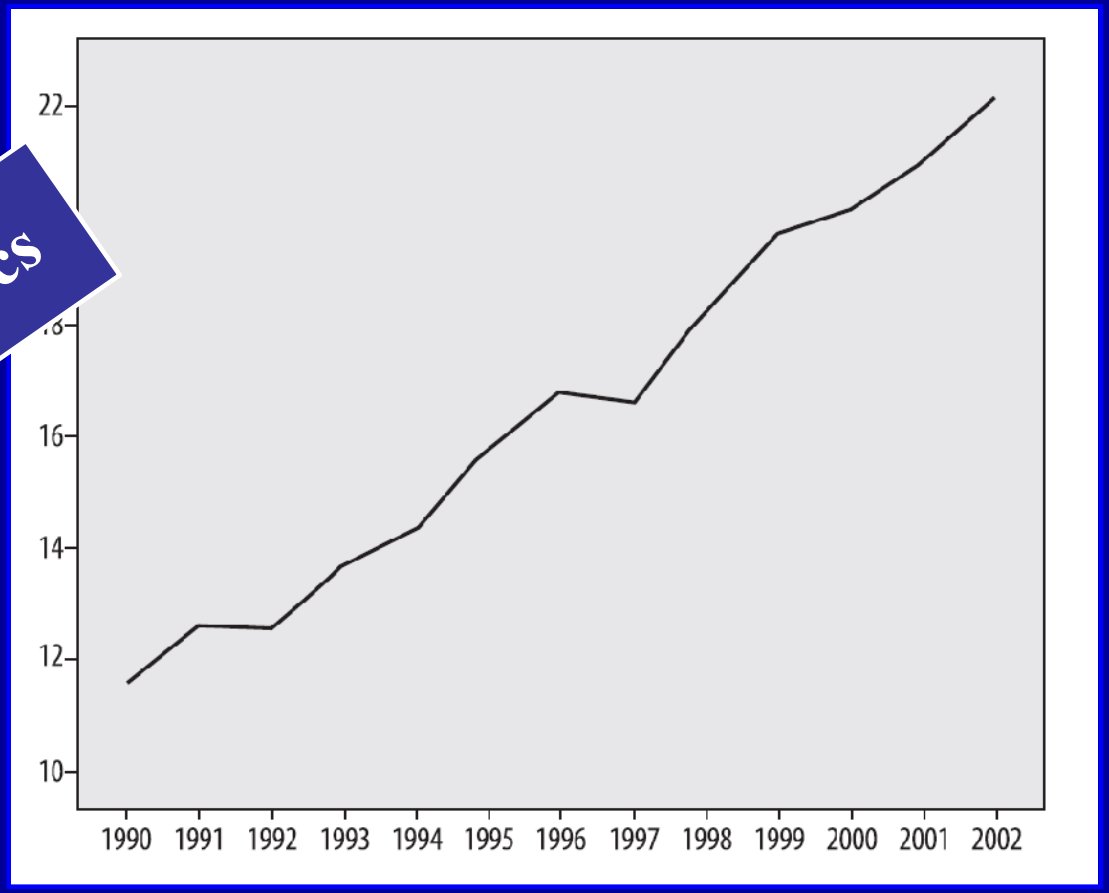


Boslaugh S & Watters PA. Statistics in a nutshell. O'Reilly Media, California, USA, 1st edition, 2008.

Line chart

Obesity among adults from 1990 – 2002 (US-CDC)

Lie with statistics



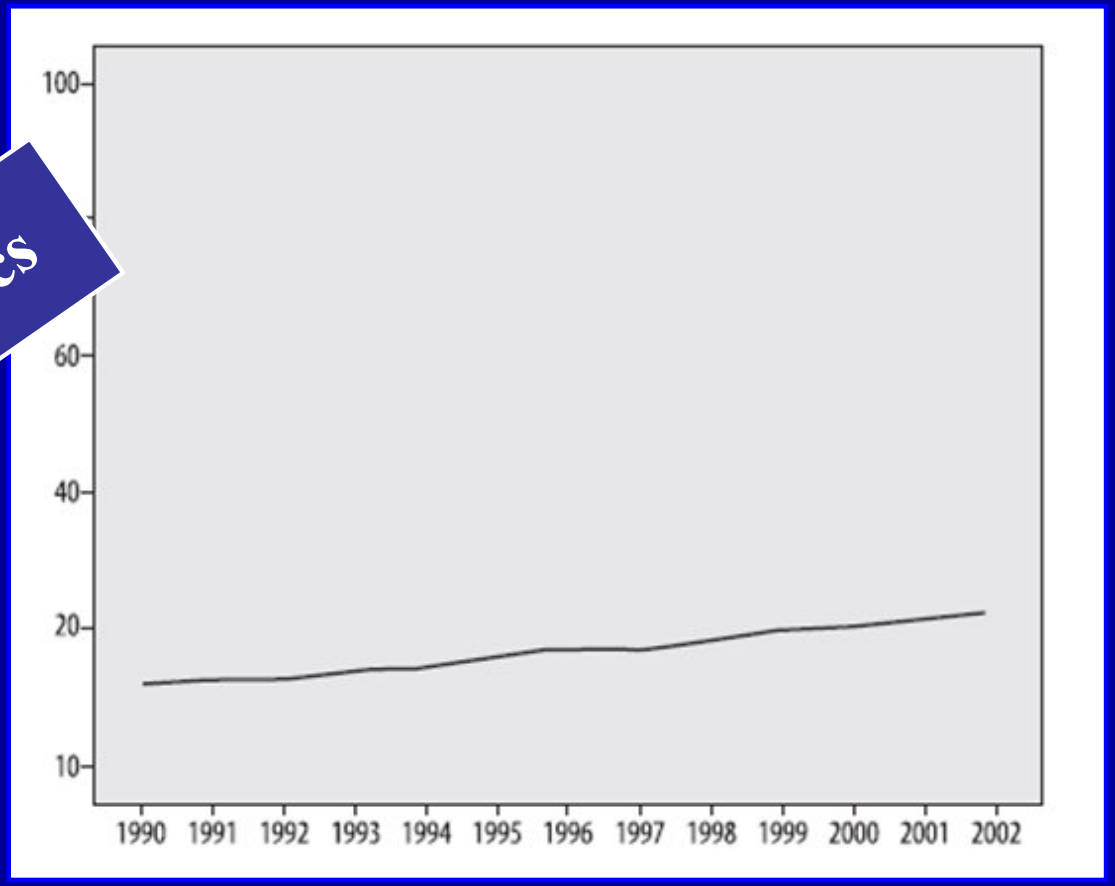
Smaller range for y-axis increases visual impact of the trend

Boslaugh S & Watters PA. Statistics in a nutshell.
O'Reilly Media, California, USA, 1st edition, 2008.

Line chart

Obesity among adults from 1990 – 2002 (US-CDC)

Lie with statistics



Wider range for the y-axis decreases visual impact of the trend

Boslaugh S & Watters PA. Statistics in a nutshell.
O'Reilly Media, California, USA, 1st edition, 2008.

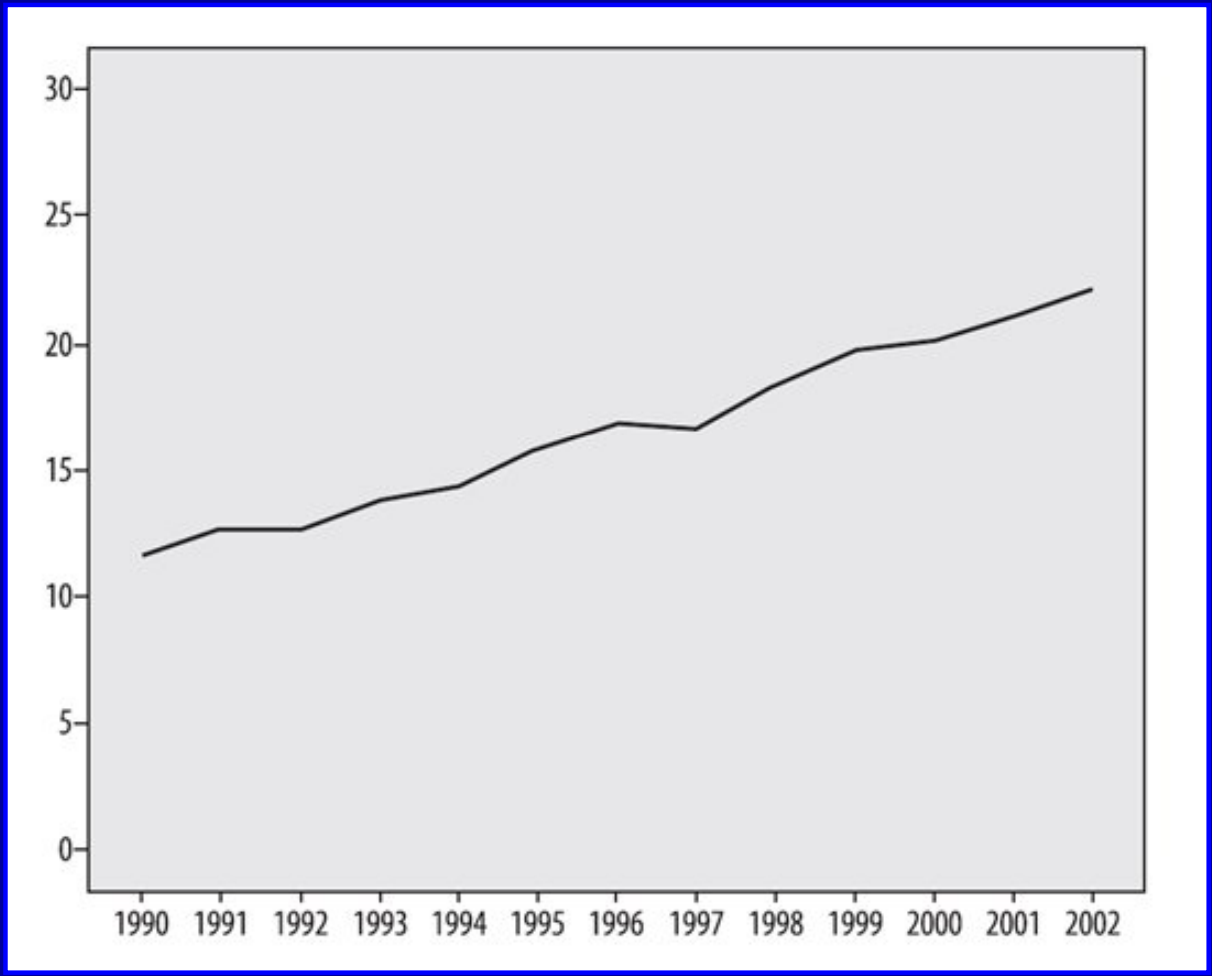
Which scale should be chosen?

- **No perfect answer to this question**
  All present the same information
  None strictly speaking are incorrect
- In this case, the scale would be **the first**
  It shows **true floor for data** (0%, lowest possible value)
  It includes **reasonable range above highest data point**

Boslaugh S & Watters PA. Statistics in a nutshell.
O'Reilly Media, California, USA, 1st edition, 2008.

Line chart

Obesity among adults from 1990 – 2002 (US-CDC)

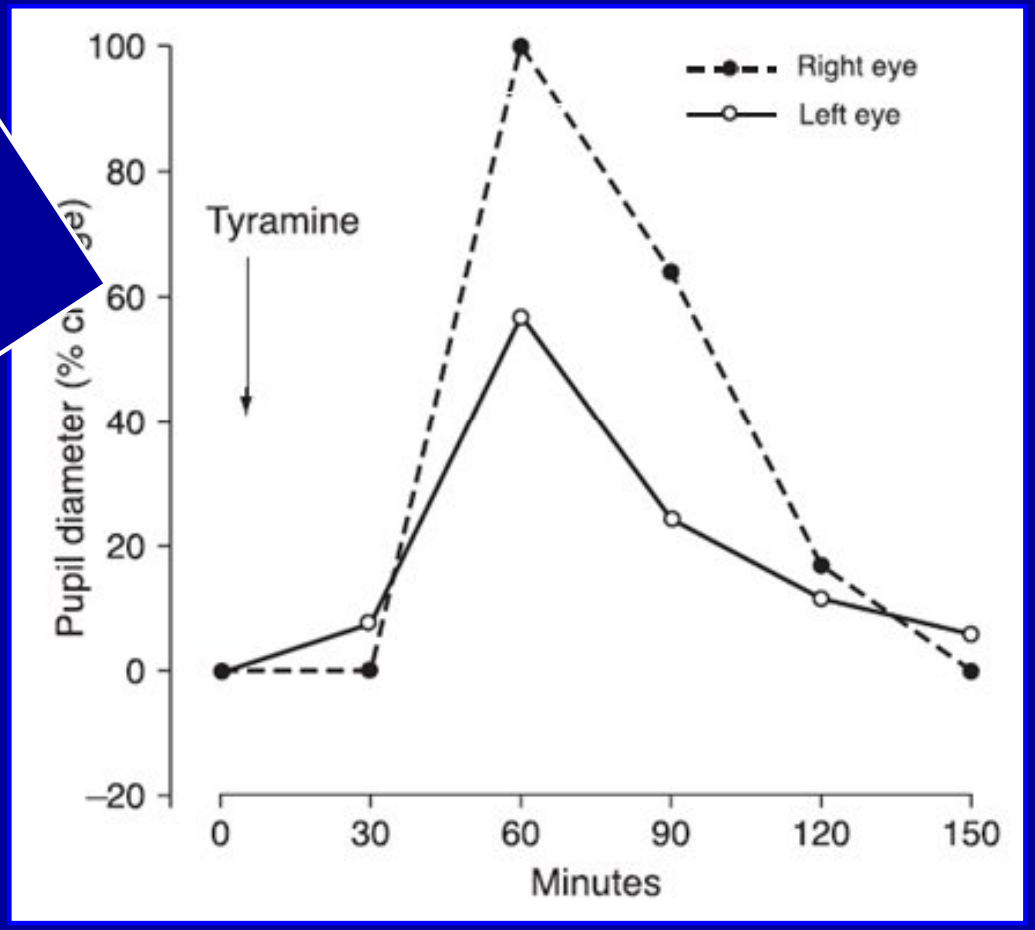


Boslaugh S & Watters PA. Statistics in a nutshell. O'Reilly Media, California, USA, 1st edition, 2008.

Line graph

Effect of tyramine solution on pupillary size

Seemingly excellent line graph



Gustavii B. How to write & illustrate scientific papers. Cambridge University Press, Cambridge, UK, 2nd edition, 2008.

Line graph

Effect of tyramine solution on pupillary size

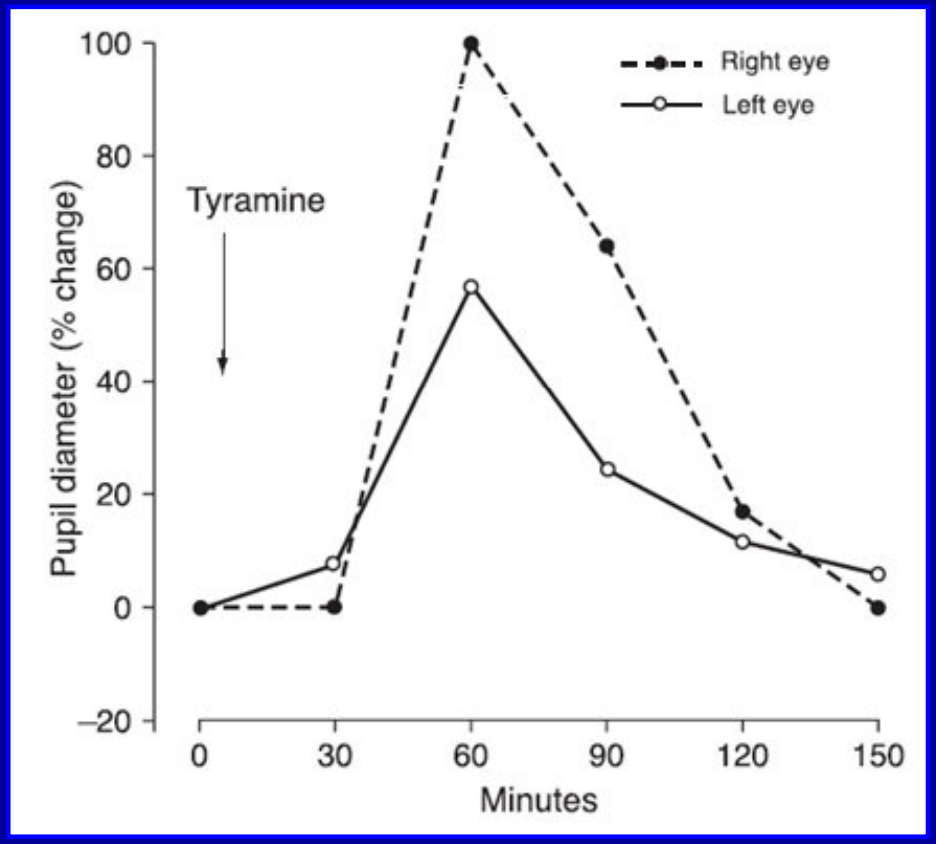
Two common defects:

1- Curves distinguished both by:

- Type of line
- Type of data-point symbol

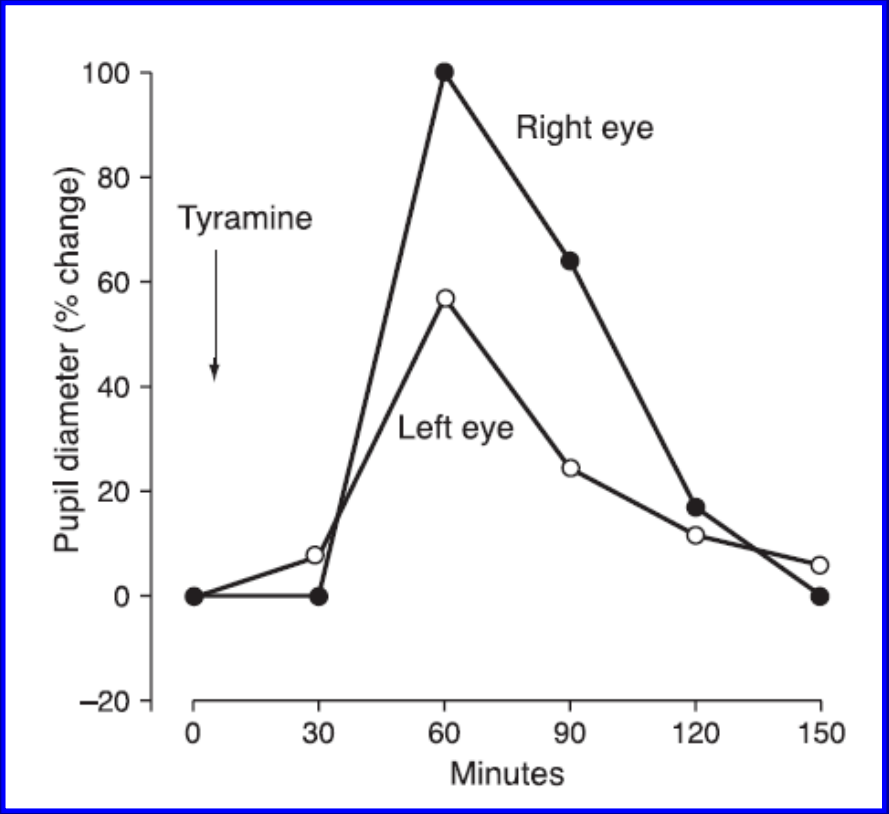
2- Curves identified by separate key

Reader scan back & forth to the key to see what they represent



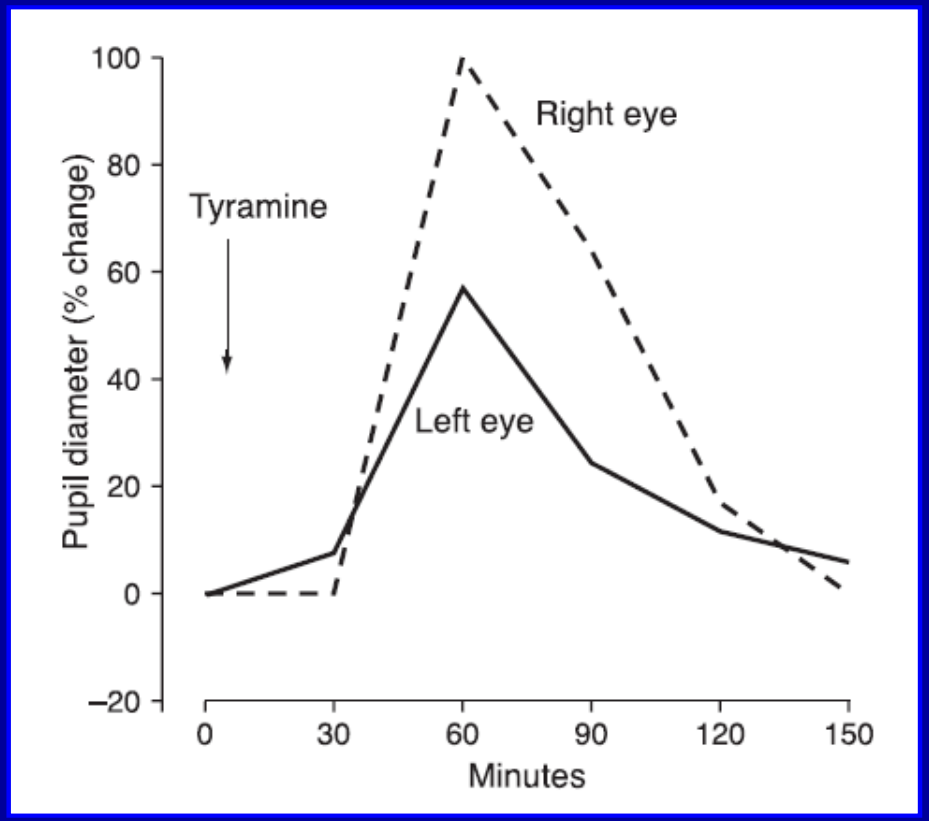
Gustavii B. How to write & illustrate scientific papers.
Cambridge University Press, Cambridge, UK, 2nd edition, 2008.

Redrawn line graphs



Type of data-point symbol

Labeled directly



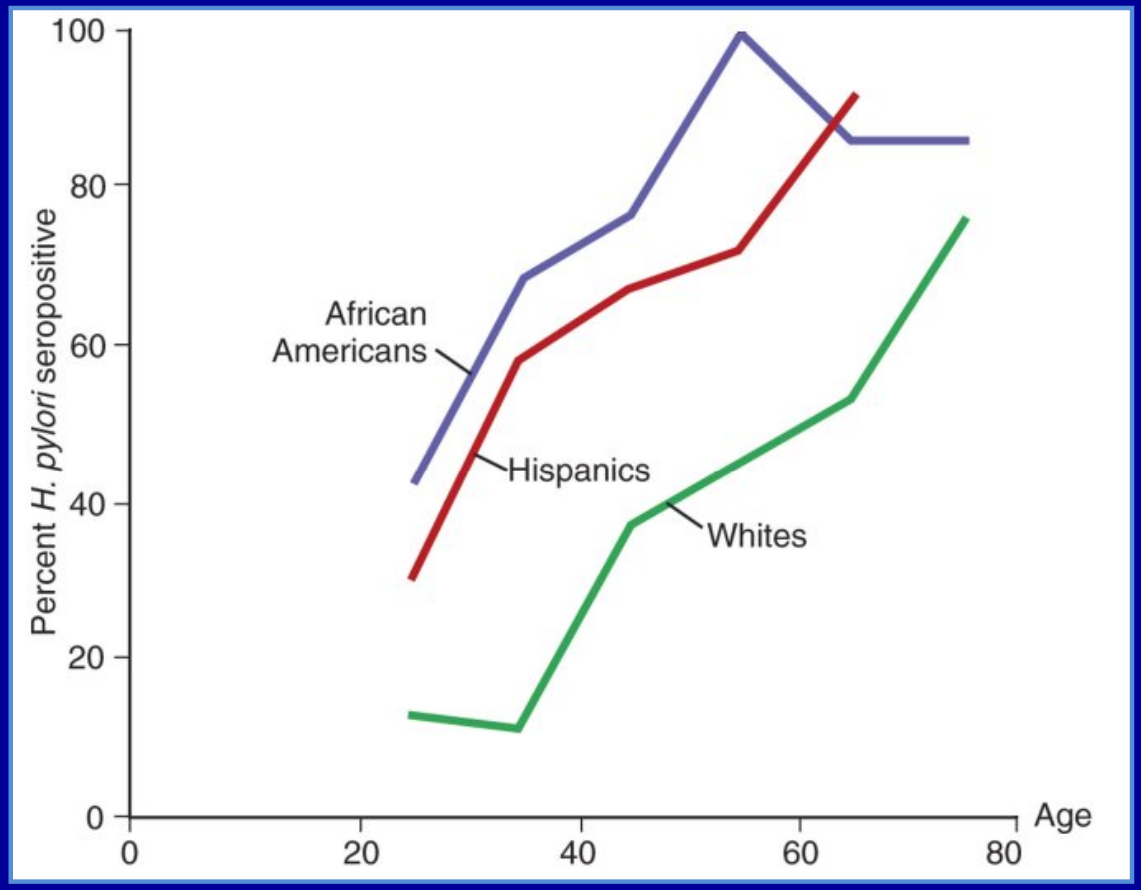
Type of line

Labeled directly

Gustavii B. How to write & illustrate scientific papers.
Cambridge University Press, Cambridge, UK, 2nd edition, 2008.

Line graph

HP seroprevalence in USA in function of age & race

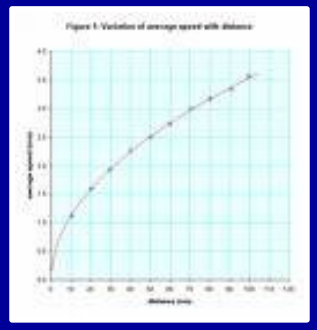


Making trend lines thick for easy visibility

Maximum: 3 – 4 lines

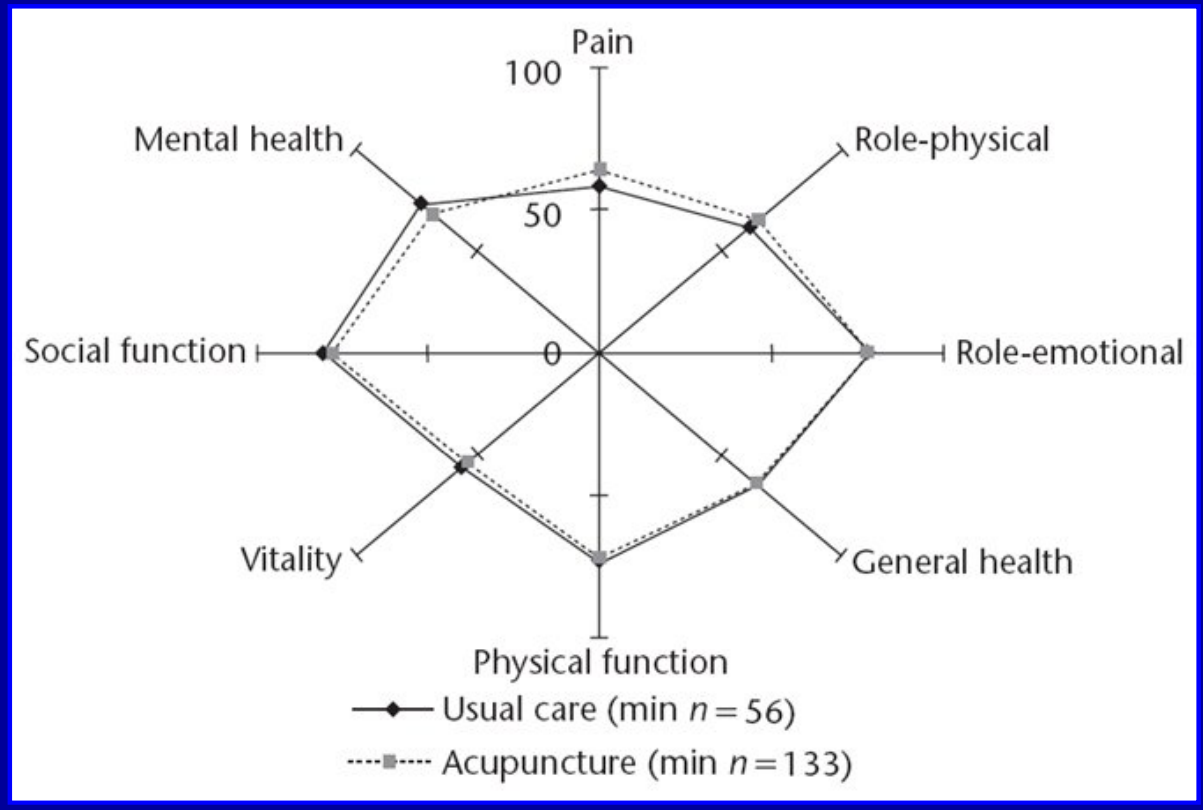
Gastroenterology 1992 ; 103 : 813.

Characteristics of some graphs

Pie graph		Good for showing Percentages
Bar/column graph		Good for showing separate unrelated pieces of data
Line graph		Good for showing how data changes over time

Spider or radar plot

Acupuncture vs usual care in persistent non-specific back pain



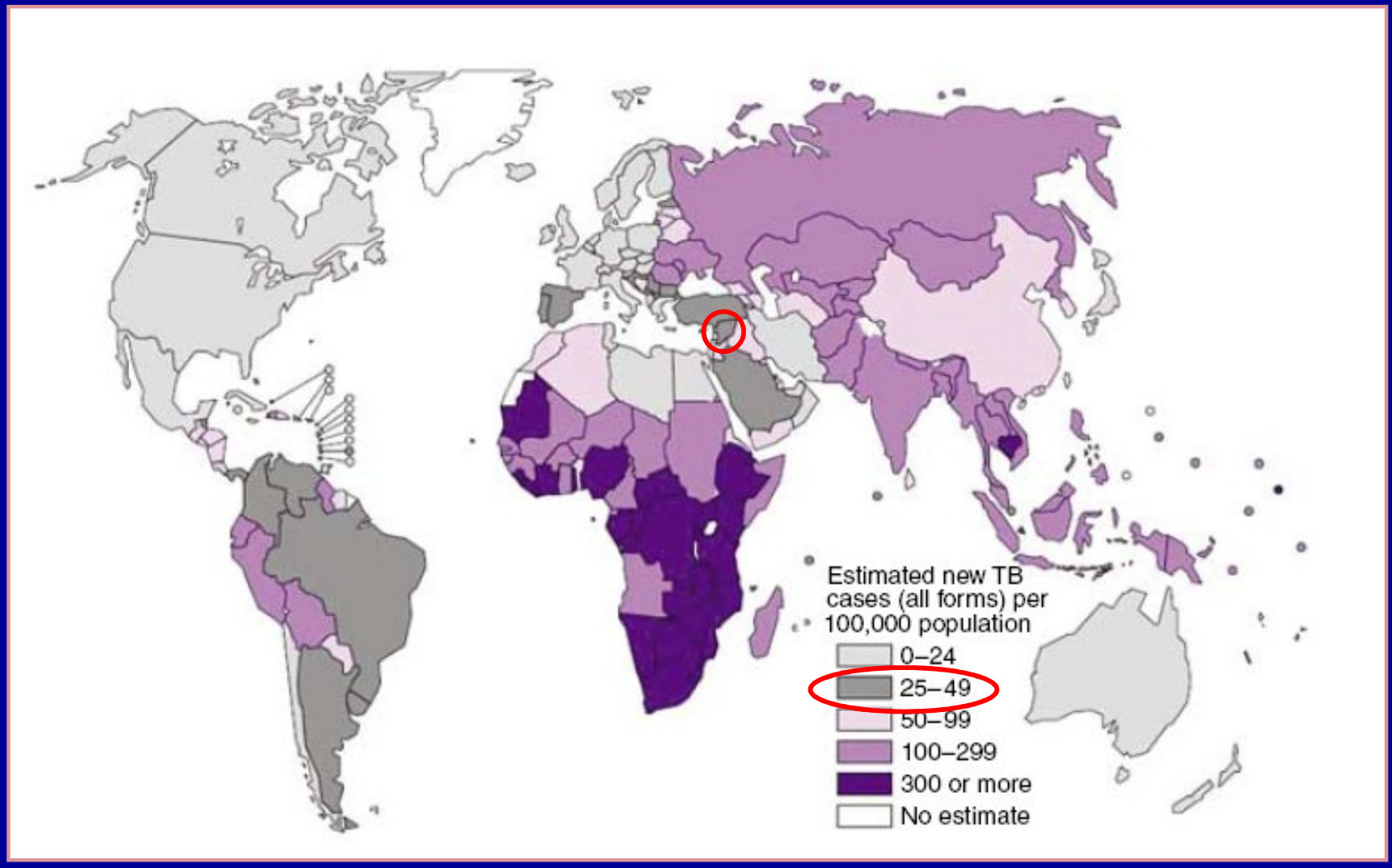
HRQol assessed over 12 months by **SF-36**

SF-36 dimensions scored on a **0 (poor)** to **100 (good)** health scale

BMJ 2006 ; 333 : 623 – 6.

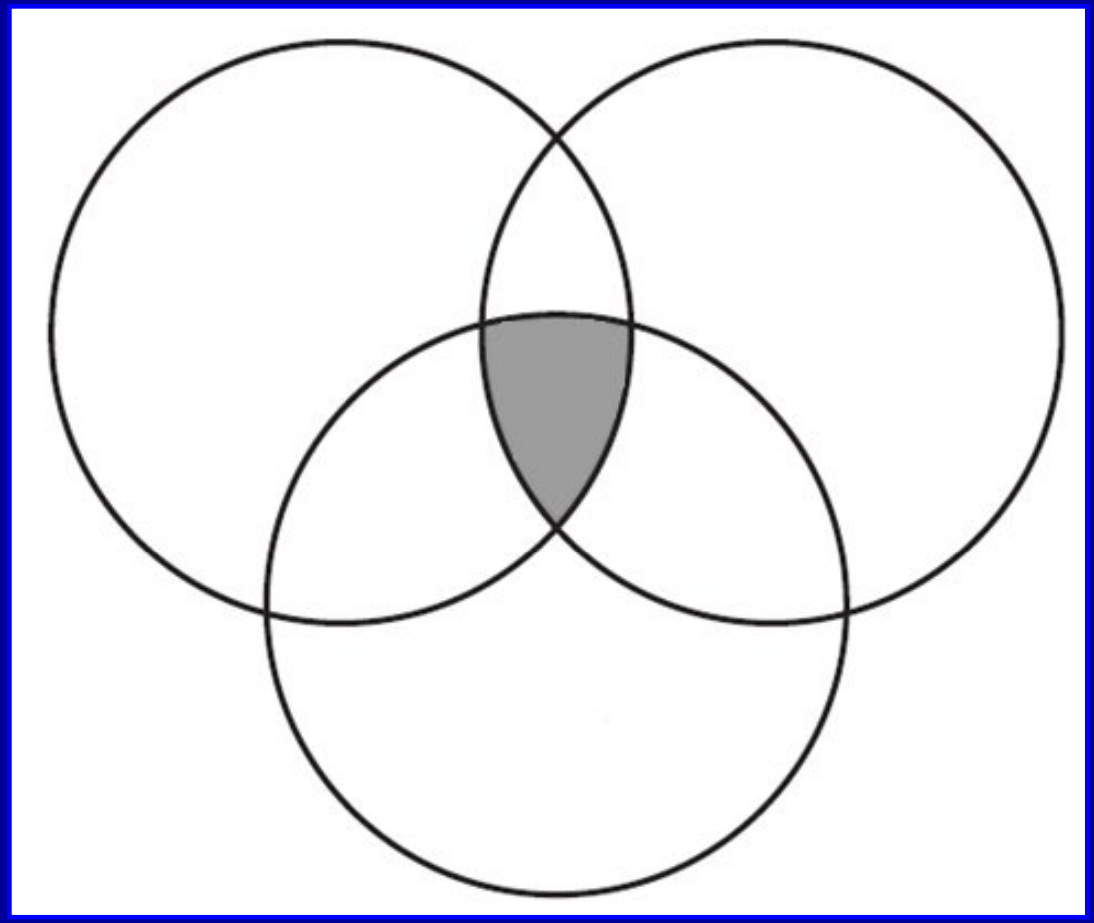
Pictogram

Estimated annual incidence of TB in 2006



Global tuberculosis control: surveillance, planning, financing
WHO report 2008

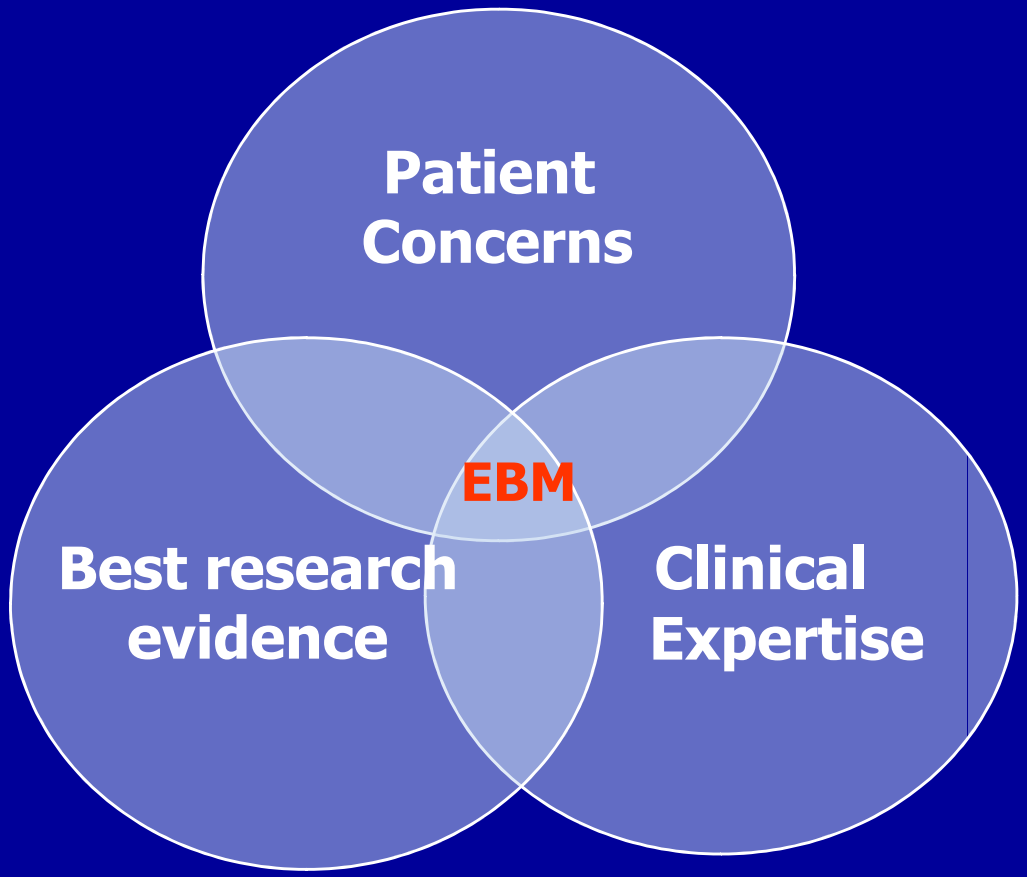
Venn diagram



Any number of overlapping circles in theory

When > 3 – 4 circles, the diagram becomes rather cluttered

The 3 components of EBM



"EBM is the integration of best research evidence with clinical expertise & patient values"

- *David Sackett*

Place of graphs in your study

Essentials you need to get started	**Types of data Qualitative - Quantitative**
Null hypothesis & alternative hypothesis	**H0 & H1**
What type of test?	**Choose the right type of test**
Is it significant?	**Compute the value of the statistic & compare with the critical value**
Software tools	**Start with Excel (simple graphs) & then move on to SPSS & then STATA, SAS or R**

Perera R et al. Statistics Toolkit. Blackwell Publishing, MA, USA, 1st edition, 2008

Useful questions to ask when considering how to display your data

- **What do you want to show?**

  Type of data – Normal or skewed distribution

- **What methods are available for this?**

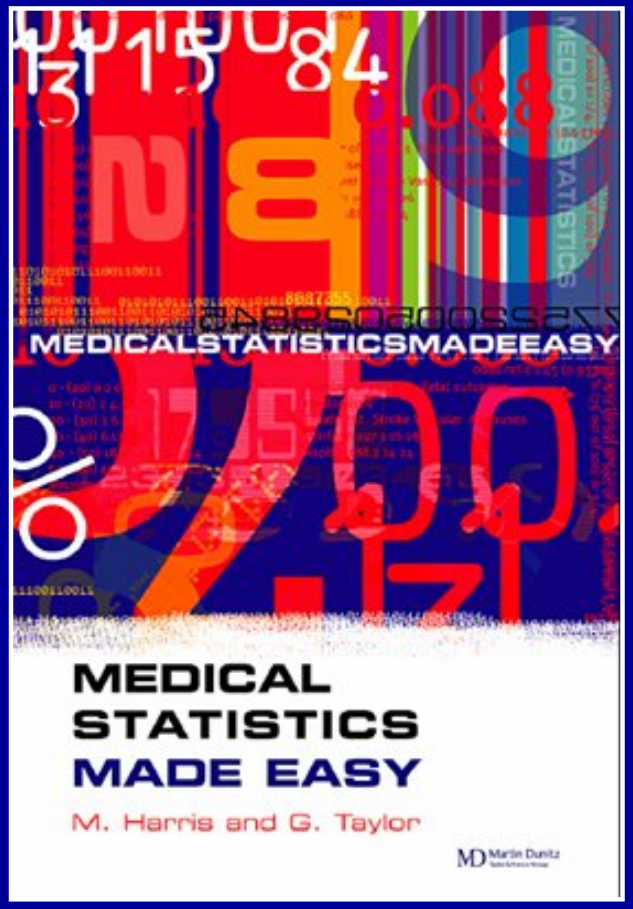
  Table – Graph – Type of graph

- **Is the method chosen the best?**

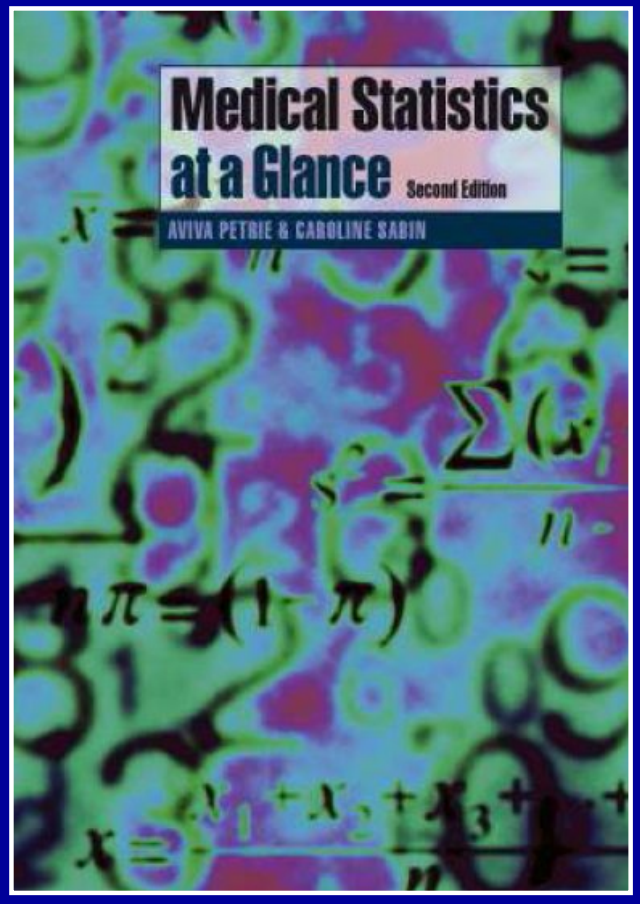
  Would another have been better?

Freeman JV et al. How to display data. Blackwell Publishing, MA, USA, 1st edition, 2008.

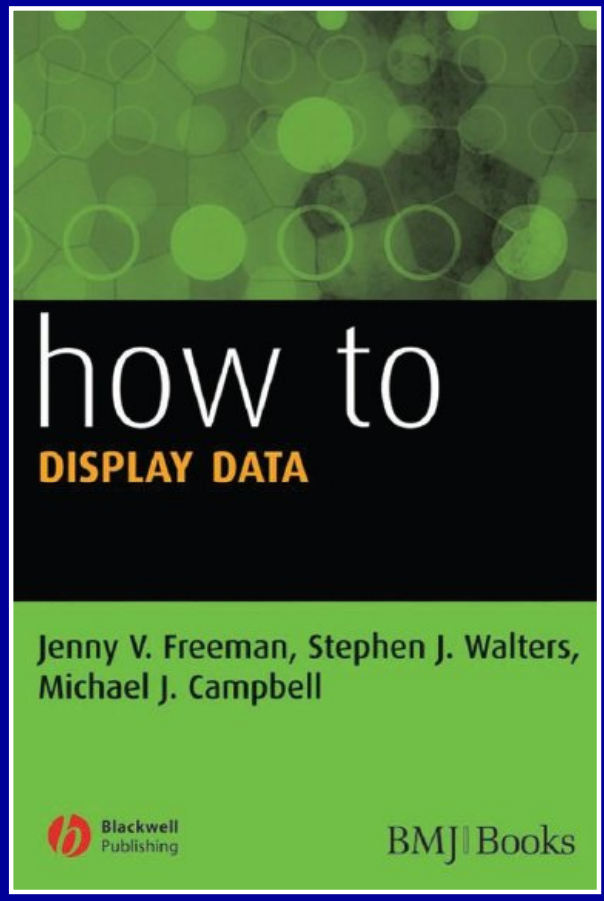
Suggested readings



Martin Dunitz
1st edition – 2003



Blackwell Publishing
2nd edition – 2005



Blackwell Publishing
1st edition – 2008

Thank You