Statistics 120
Good and Bad Graphs

The Plan

- In this lecture we will try to set down some basic rules for drawing good graphs.
- We will do this by showing that violating the rules produces bad graphs.
- Later in the course we see that there is a solid perceptual basis for some of these rules.

Data Content

- It does not make sense to use graphs to display very small amounts of data.
- The human brain is quite capable of grasping one two, or even three values.

The Company Cafeteria was used by 9 Out of 10 Employees during the Fiscal Year 1949

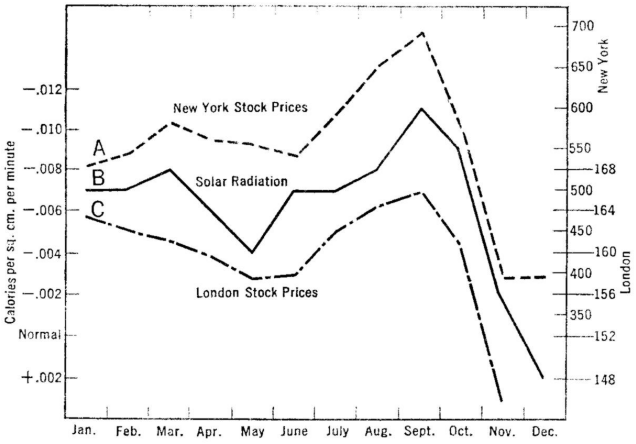
0 100%

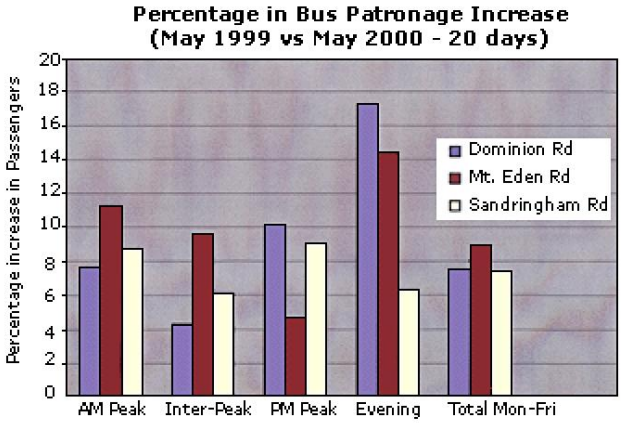
Source: COMPANY REPORTS


Class Gender Breakdown
60
50
40
30
20
10
0
Percentage
Male
Female
Gender

Data Relevance

- Graphs are only as good as the data they display.
- No amount of creativity can produce a good graph from dubious data.


New York Stock Prices
Solar Radiation
London Stock Prices
A
B
C
Calories per sq. cm. per minute
−.012
−.010
−.008
−.006
−.004
−.002
Normal
+.002
Jan.
Feb.
Mar.
Apr.
May
June
July
Aug.
Sept.
Oct.
Nov.
Dec.
New York
700
650
600
550
500
450
400
350
168
164
160
156
152
148
London



Auckland City Council: *City Scene*

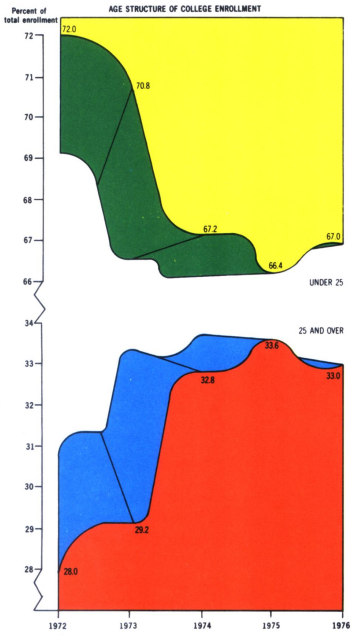
Complexity

- Graphs should be no more complex than the data which they portray.
- Unnecessary complexity can be introduced by
  - irrelevant decoration
  - colour
  - 3d effects
- These are collectively known as “chartjunk.”

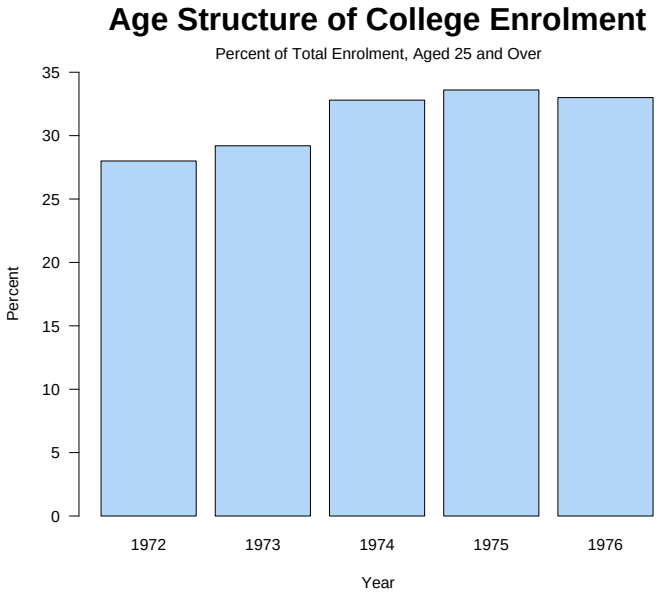
Age Structure of College Enrolment (1972-1976)

This graph presents five values. It uses six colours, unnecessary perspective and a split axis to do so.

American Education Magazine.



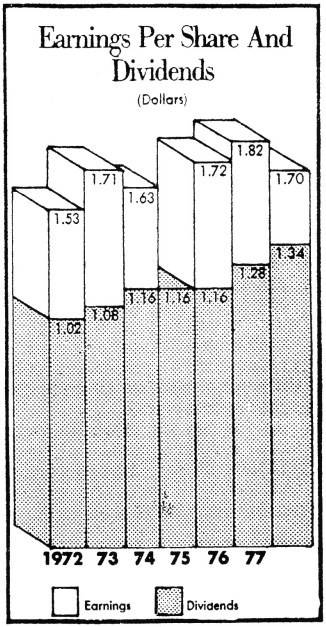
Age Structure of College Enrolment


Age Structure of College Enrolment
Percent of Total Enrolment, Aged 25 and Over
Percent
0
5
10
15
20
25
30
35
1972
1973
1974
1975
1976
Year

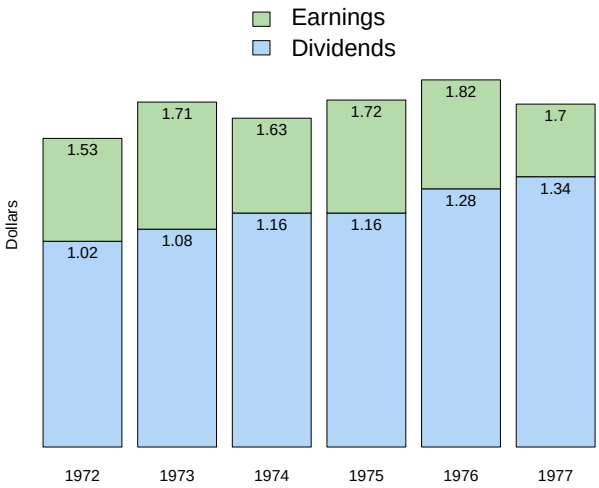
Share Results

The extra dimension used in this graph has confused even the person who created it.

The Washington Post, 1979.



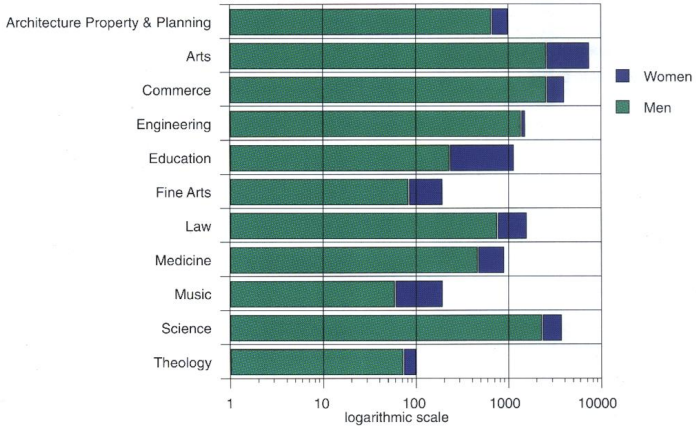
Earnings Per Share and Dividends

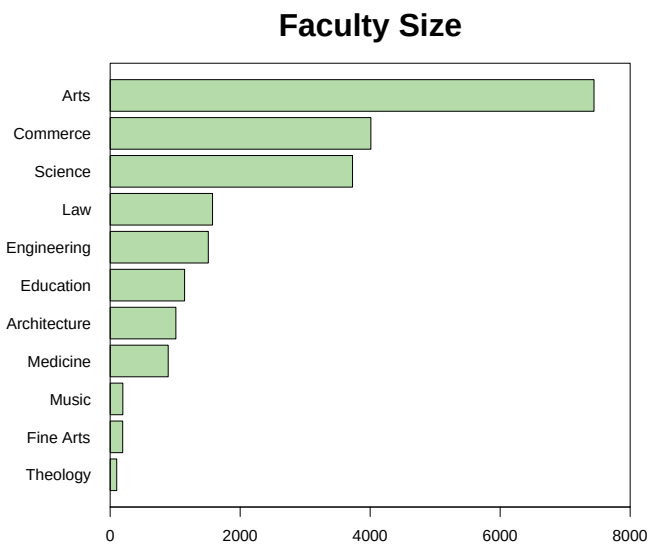


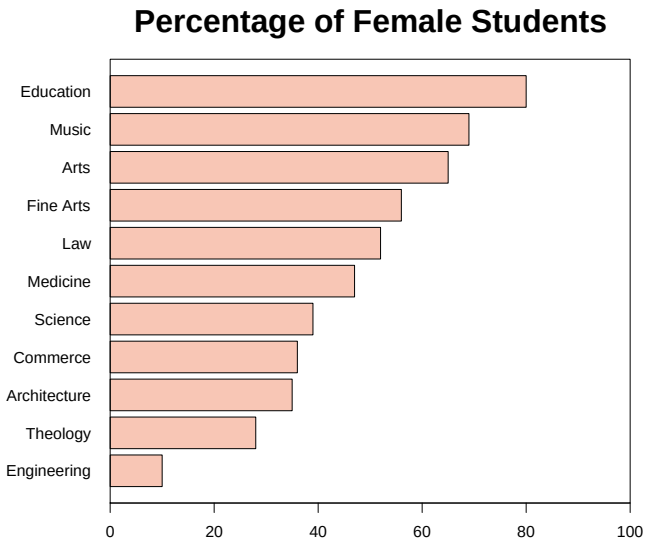
Distortion

- Graphs should not provide a distorted picture of the values they portray.
- Distortion can be either deliberate or accidental.
- (Of course, it can be useful to know how to produce a graph which bends the truth.)

FACULTIES


Architecture Property & Planning
Arts
Commerce
Engineering
Education
Fine Arts
Law
Medicine
Music
Science
Theology
Women
Men
1
10
100
1000
10000
logarithmic scale


Faculty Size
Arts
Commerce
Science
Law
Engineering
Education
Architecture
Medicine
Music
Fine Arts
Theology
0
2000
4000
6000
8000

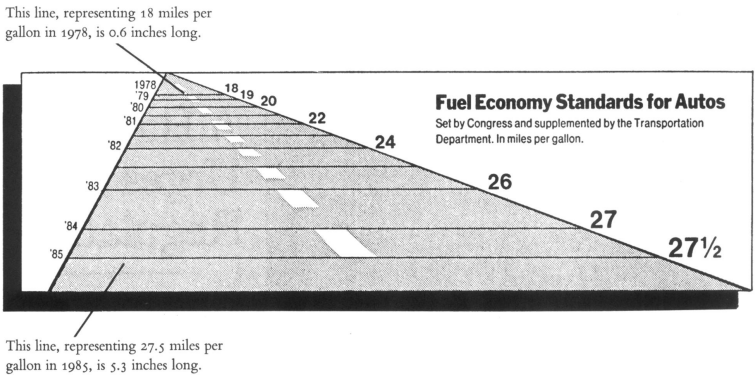

Percentage of Female Students
Education
Music
Arts
Fine Arts
Law
Medicine
Science
Commerce
Architecture
Theology
Engineering
0
20
40
60
80
100

The “Lie Factor”

- Ed Tufte of Yale University has defined a “lie factor” as a measure of the amount of distortion in a graph. (Don’t take this too seriously – i.e don’t learn it for the exam).
- The lie factor is defined to be:

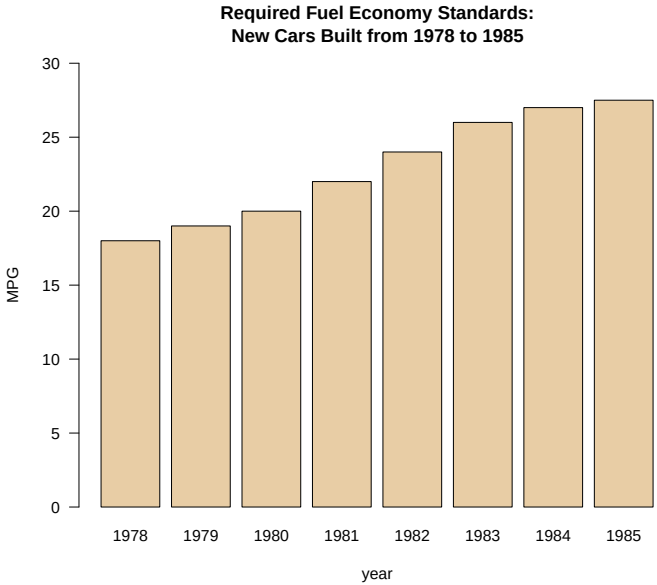
$$\text{Lie Factor} = \frac{\text{size of effect shown in graphic}}{\text{size of effect shown in data}}$$

- If the lie factor of a graph is greater than 1, the graph is exaggerating the size of the effect.



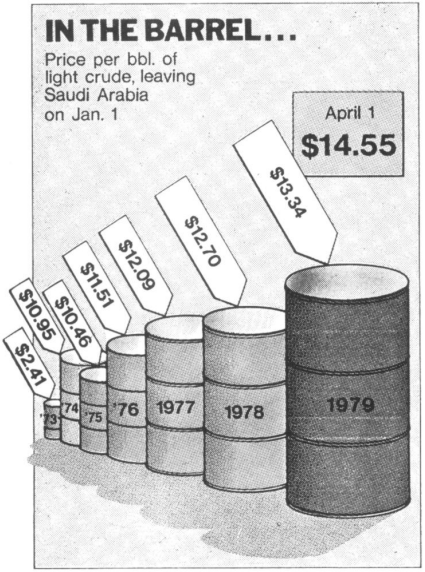
$$\text{Data Effect} = \frac{27.5 - 18}{18} = 0.53, \qquad \text{Graph Effect} = \frac{5.3 - .6}{.6} = 7.83,$$

$$\text{Lie Factor} = 14.8$$


Required Fuel Economy Standards:
New Cars Built from 1978 to 1985
MPG
30
25
20
15
10
5
0
1978
1979
1980
1981
1982
1983
1984
1985
year

Common Sources of Distortion

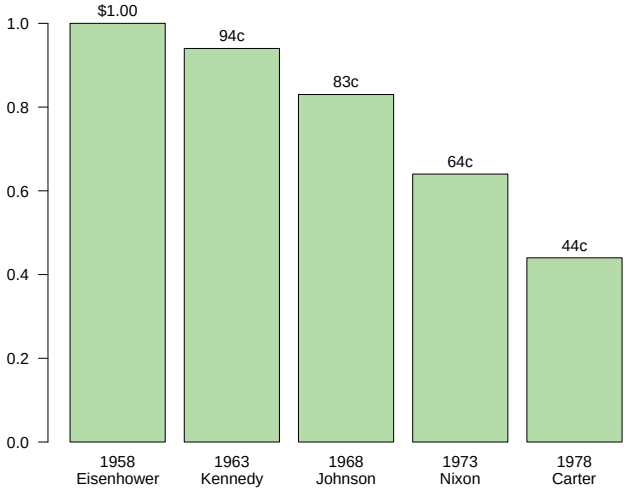
- The use of 3 dimensional “effects” is a common source of distortions in graphs.
- Another common source is the inappropriate use of linear scaling when using area or volume to represent values.



Lie Factor = 9.4

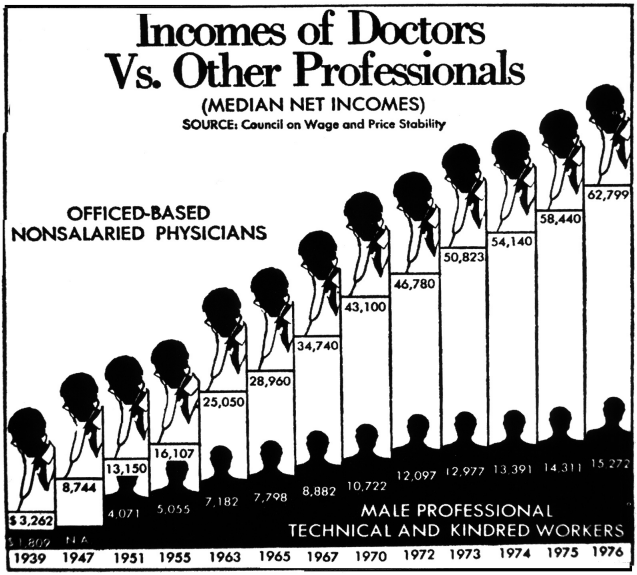
Is the bottom dollar note roughly half the size of the top one?

Purchasing Power of the Diminishing Dollar

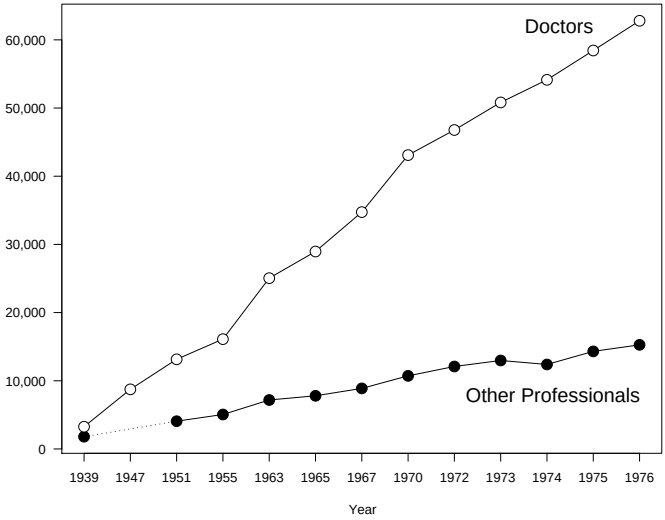


Deliberate Distortion

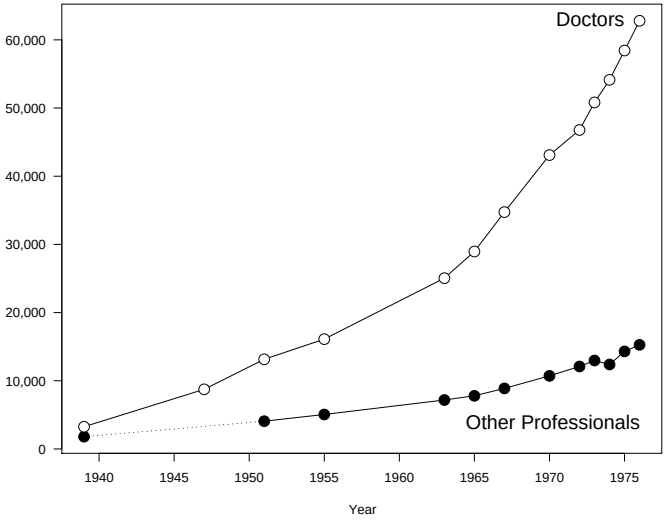
- Sometimes graphs contain deliberate distortions.
- Usually these are an attempt to hide some feature of the data.


Incomes of Doctors
Vs. Other Professionals
(MEDIAN NET INCOMES)
SOURCE: Council on Wage and Price Stability
OFFICED-BASED
NONSALARIED PHYSICIANS
$3,262
8,744
13,150
16,107
25,050
28,960
34,740
43,100
46,780
50,823
54,140
58,440
62,799
$1,809
4,071
5,055
7,182
7,798
8,882
10,722
12,097
12,977
13,391
14,311
15,272
MALE PROFESSIONAL
TECHNICAL AND KINDRED WORKERS
1939
1947
1951
1955
1963
1965
1967
1970
1972
1973
1974
1975
1976

Median Net Incomes



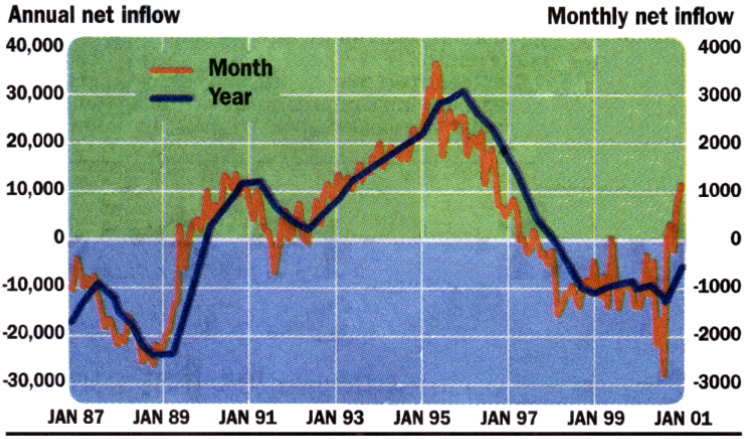
Median Net Incomes



Errors

- Sometimes distortions occur because of errors.
- This can be because the artist tried to be clever and sometimes because they don’t understand statistics.

Brain gain



Source: DB Global Markets Research, Statistics NZ. HERALD GRAPHIC

Drawing Good Graphs

- If the “story” is simple, keep it simple.
- If the “story” is complex, make it look simple.
- Tell the truth – don’t distort the data.