# Statistics 120
# Good and Bad Graphs


## The Plan

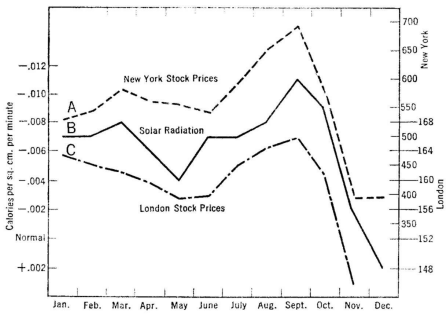
- In this lecture we will try to set down some basic rules for drawing good graphs.
- We will do this by showing that violating the rules produces bad graphs.
- Later in the course we see that there is a solid perceptual basis for some of these rules.

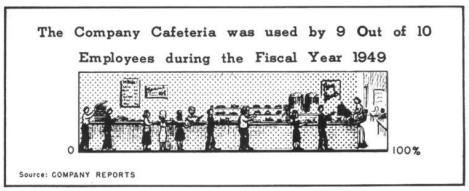
## Data Relevance

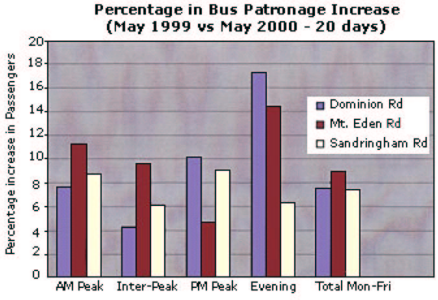
- Graphs are only as good as the data they display.
- No amount of creativity can produce a good graph from dubious data.

## 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.




Auckland City Council: *City Scene*

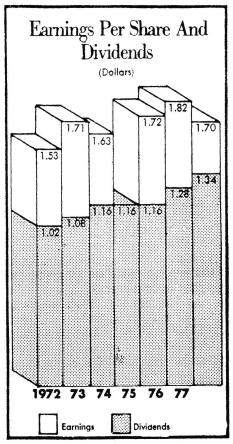
## Complexity

- 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.”

**Share Results**

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

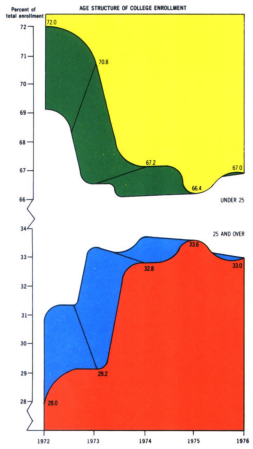
*The Washington Post*, 1979.

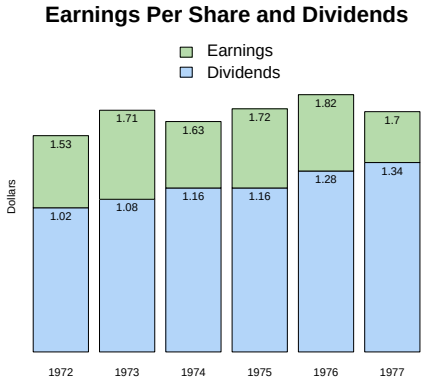


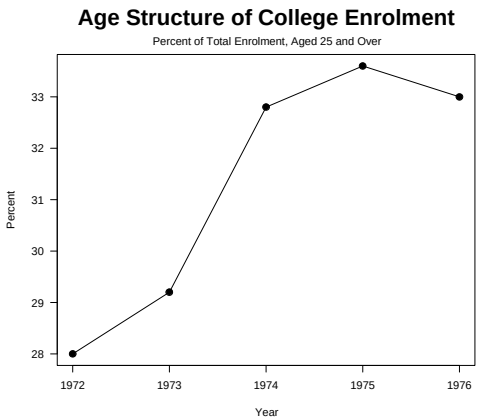

**Age Structure of College Enrolment (1972-1976)**

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

*American Education Magazine.*

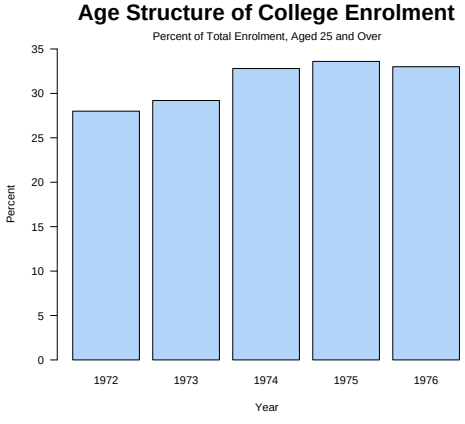


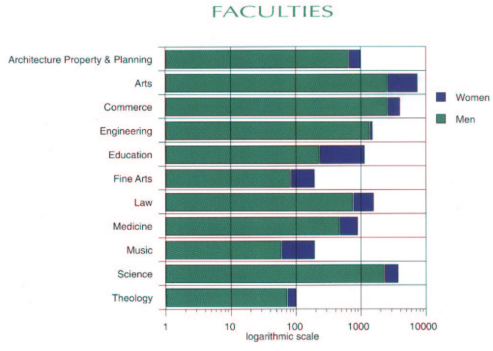



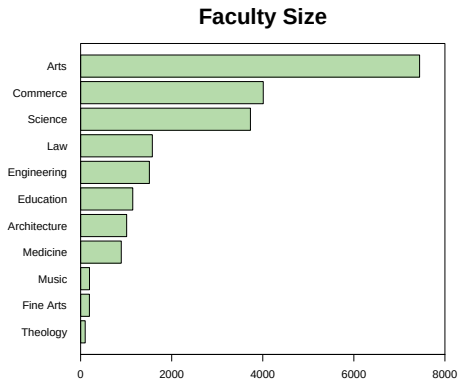



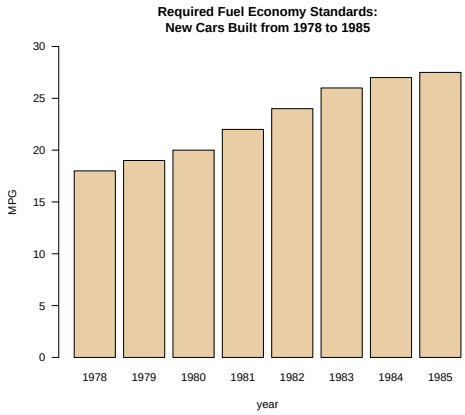

## Distortion

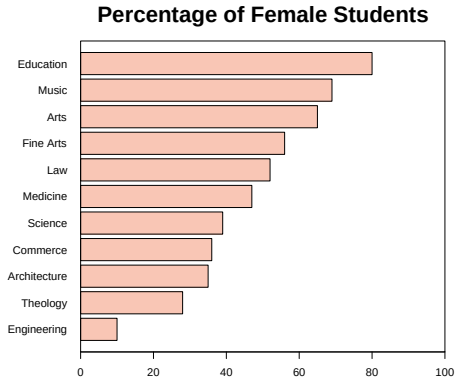
- Graphs should not provide a distorted picture of the values they portray.
- Distortion can be either deliberate or accidental.
- (Of course, it can be useful to know how to produce a graph which bends the truth.)






## Common Sources of Distortion

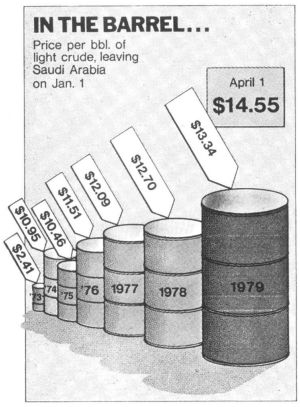
- The use of 3 dimensional "effects" is a common source of distortions in graphs.
- Another common source is the inappropriate use of linear scaling when using area or volume to represent values.

## The "Lie Factor"

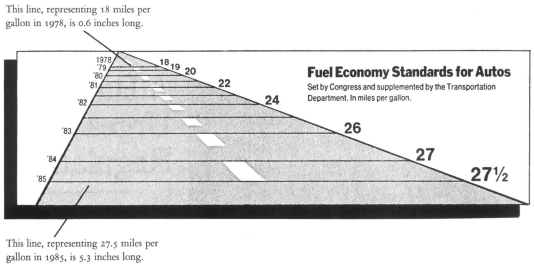
- Ed Tufte of Yale University has defined a "lie factor" as a measure of the amount of distortion in a graph. (Don't take this too seriously – i.e don't learn it for the exam).
- The lie factor is defined to be:

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

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

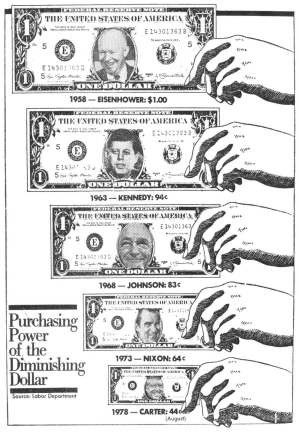



Lie Factor = 9.4

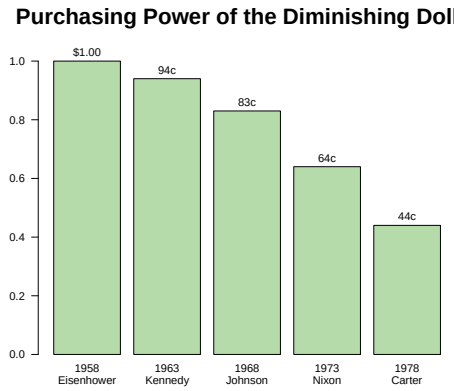


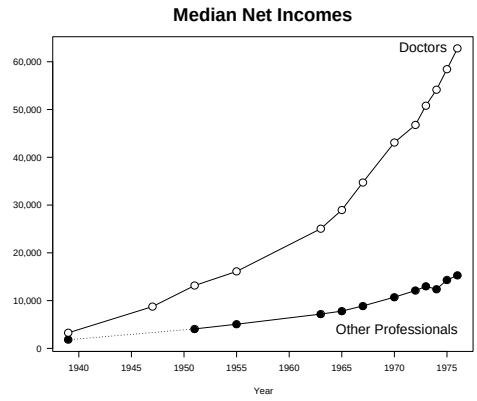

$$\text{Data Effect} = \frac{27.5 - 18}{18} = 0.53, \qquad \text{Graph Effect} = \frac{5.3 - .6}{.6} = 7.83,$$

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




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



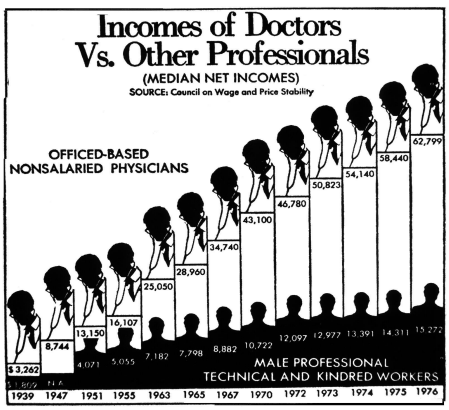



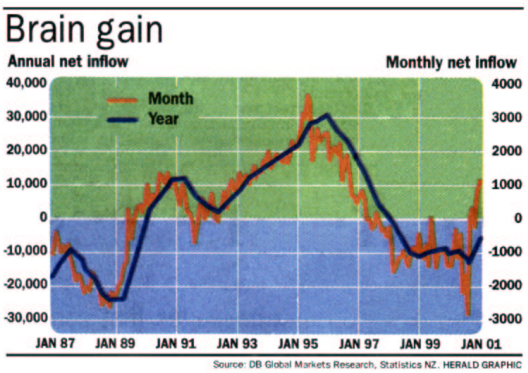

## Deliberate Distortion

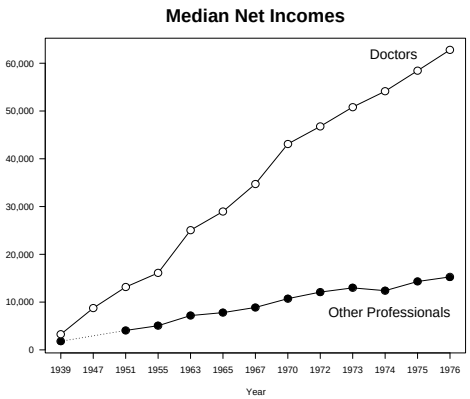
- Sometimes graphs contain deliberate distortions.
- Usually these are an attempt to hide some feature of the data.

## 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.




## 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.