

Sample surveys

representative/ biased samples

selection bias

non-response bias

simple random sample

stratified sample

cluster sample

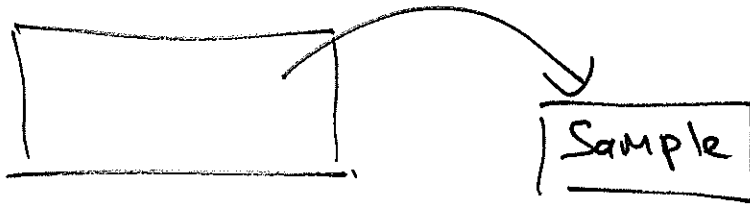
looking at data

pie charts

bar charts

histograms

Population.



Designing the survey that generates the sample.

- so that the sample is representative of the population
- use the sample to infer parameters of the population

Selection Bias.

1936 Election Poll.

			<u>result</u>
<u>predicted</u>	Roosevelt	43%	62%
	Landon	57%	38%

10 million surveys mailed out.

But addresses taken from phone book.  
club membership lists.

In 1936 (Great depression) this misses poor people

Taking a large sample doesn't help  
if the procedure is biased.

## Response Bias.

were the 2.4 million replies  
representative of the 10 million people  
who received the survey?

- probably not - class bias.

98% of respondents were dissatisfied  
with their marriage

75% had extramarital affairs

4% response rate

93% satisfied.

7%

# VARSITY

Friday October 31 2008

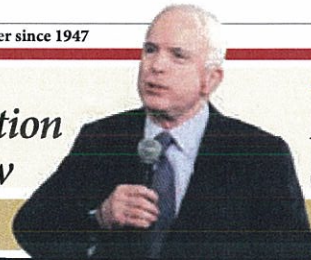
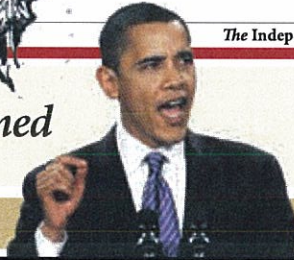
The Independent Cambridge Student Newspaper since 1947

Issue No 681 | varsity.co.uk

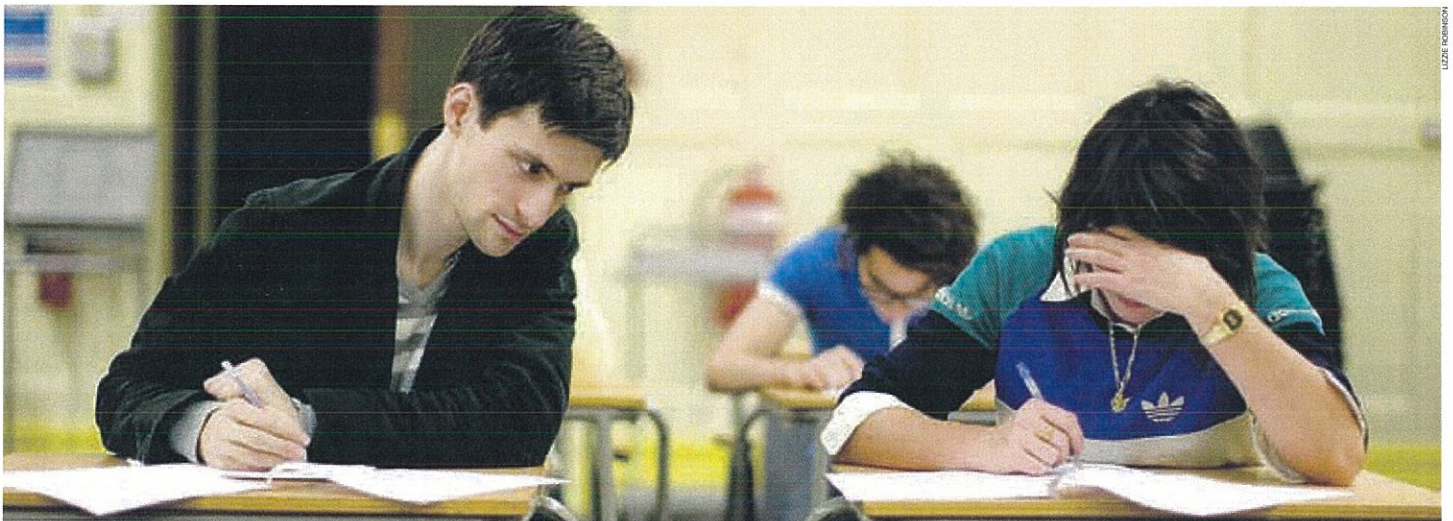
»p14 Arts  
**Halloween-themed comic strip**

»p12 Features  
**American Election Special Preview**

»p34 Sport  
**Legendary writer Simon Barnes**



## '1 in 2' admits to plagiarism



» *Varsity survey reveals that 49% per cent of Cambridge students have committed some form of plagiaristic act whilst at the University*  
» *Students criticise unrealistic plagiarism guidelines; others blame rigour of Cambridge degrees*

Michael Stothard  
Chief News Editor

Half of Cambridge students have committed plagiarism as defined by the university, according to a *Varsity* survey.

49 per cent of students admitted that they have plagiarised work, although this differed radically between subjects and colleges. Ironically, students of the Law faculty plagiarised the most out of any subject, with 62 per cent of them breaking the university rules. The sec-

ond highest was the Archaeology and Anthropology department with 59 per cent.

"It is a depressing set of statistics," said Robert Foley, a Professor in Biological Anthropology at King's College.

The college at the bottom of the Tompkins Table, St Edmund's, had the highest proportion of plagiarising students, with 67 per cent admitting to breaking the university rules. Selwyn, at the top of the Tompkins table, had the fewest number of plagiarising students.

"It stands to reason that those students who are performing less well will resort to more underhand means to get by," said a member of the University Council, the principal executive and policy making body of the university.

It is perhaps not surprising that 80 per cent of students said that the university is doing enough to punish plagiarism. "You can see why students, a great number of whom are frequently breaking the rules to their own benefit, would be keen to uphold the impression that the system is working," said

a member of the General Board, the body responsible for education policy at the University.

"Sometimes when I am really fed up," said Land Economy student at Pembroke, "I Google the essay title, copy and throw everything on to a blank word document and jiggle the order a bit. They usually end up being the best essays." 100 per cent of Land Economy students admitted to plagiarism, but the results should be taken lightly because less than five per cent of the student population replied to the survey. *Continued on page 5*

# 49%

Percentage of survey respondents who admitted to some form of plagiarism

# 62%

Percentage of respondents studying Law who admitted to plagiarism

## Cambridge application levels hit record high of 15,000

Vicky Woolley  
News Reporter

The University has announced that the number of applicants for undergraduate courses has reached a record high this year. For the past four years,

around 13,500 students have applied to study at Cambridge, whereas this year the figure is set to be closer to 15,000. This equates to an increase of around 12 per cent, compared with an average rise of only 6.5 per cent across all UK universities.

Some subjects have fared better than others. Computer Science has seen a 53 per cent rise in applications, with the next biggest rise being in Philosophy which is up by 43 per cent. Theology and ASNaC saw applications drop by 18.5 and 10 per cent respectively.

This rise in applicant numbers comes at the end of a year in which many changes have been made to the admissions process. The requirement for a foreign language GCSE has been removed because of fears it disadvantaged applicants from state schools,

where foreign language teaching has declined rapidly in recent years. The Cambridge Application Form was also abolished: the form, which students had to fill in alongside their UCAS application, was felt to be off-putting for some students. *Continued on page 3*

»p16 Richard Eyre, former Director of the National Theatre, on the ADC »p22 Goldfrapp reviewed

## How to choose a good sample?

Simple random sampling.

- removes selection bias.

1/ decide how big a sample is needed

2/ from the list of the entire population

- draw an element at random

- delete that element from the list

- repeat until you have the desired number in the sample

but often hard to do.

so sometimes pretend it's what we're doing - it makes the math easier.

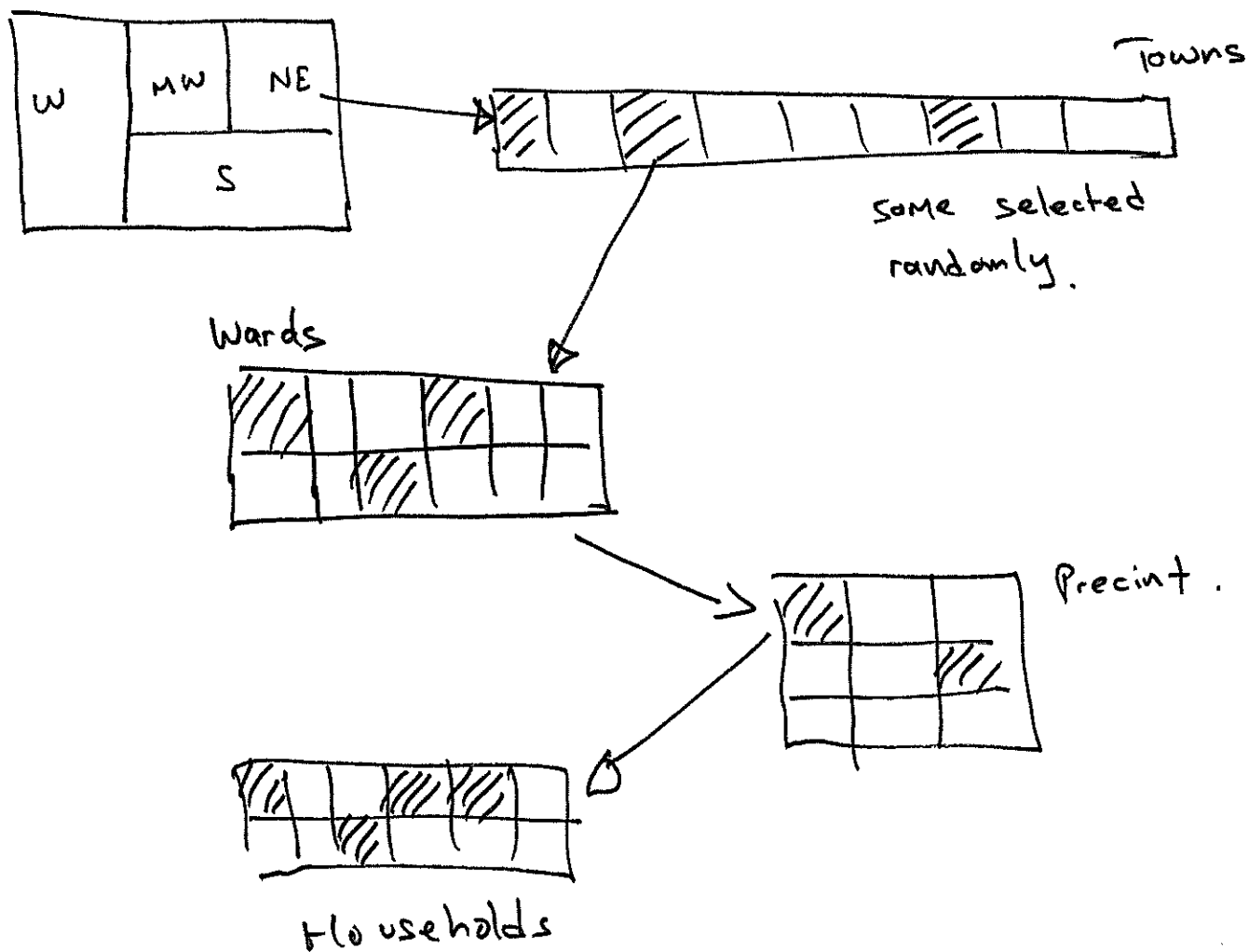
## Quota / stratified sampling.

- try to make the sample ~~with~~ match the population

- if the choice of who to interview is left to human judgement, selection bias can come back.

## Cluster Sampling.

Multiple stages - each subject to chance



The person to be interviewed is determined in advance. (eg youngest person over 18)

Gallup poll. - A sample of a  
few 1000 gives  
errors  $\pm 3\%$

Random dial telephone surveys.

## Looking at Data.

graphical representations that can aid understanding the distribution of a data set.

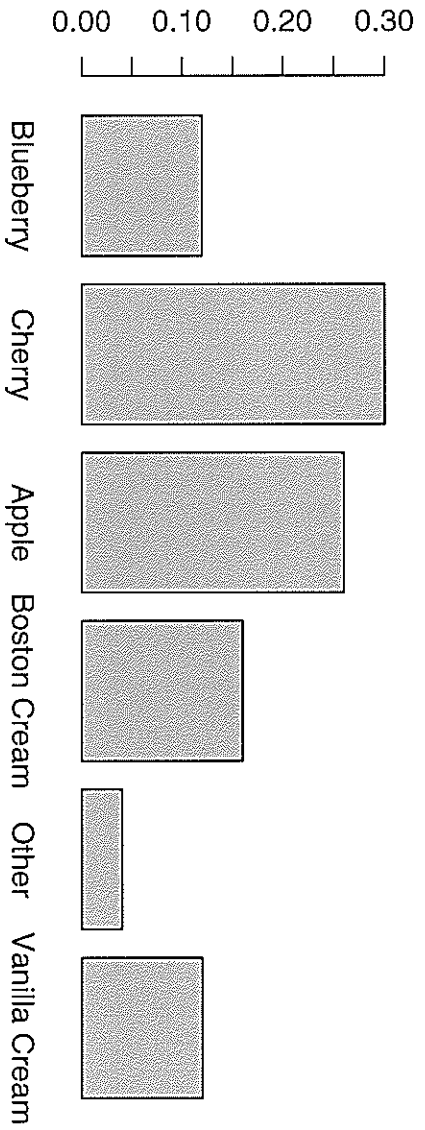
- range of values
- most likely values
- symmetry
- uni/multimodality
- tail behavior

### Bad way

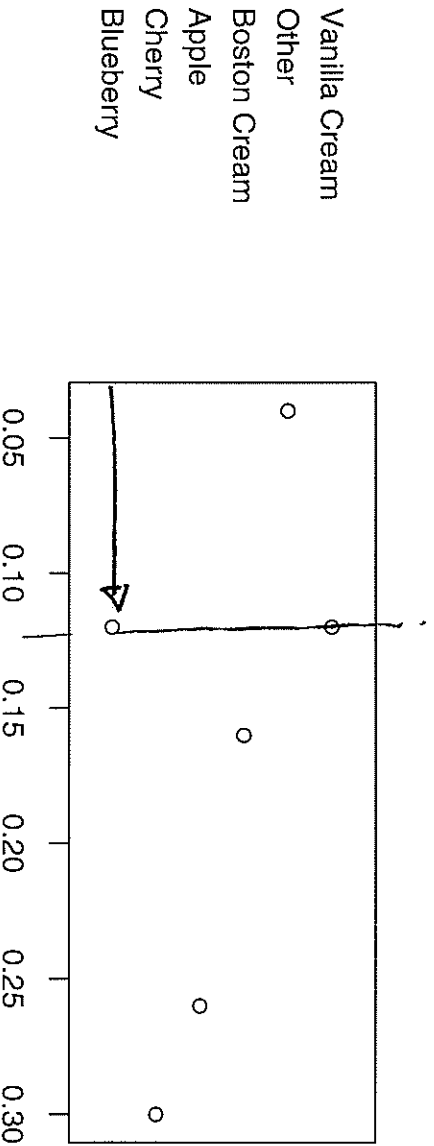
Pie chart - humans are good at judging linear measures (distances)

not angles.

Bar Chart.



Dot Chart.



Better for displaying categorical data  
than pie charts.

For continuous variables

- Histogram.

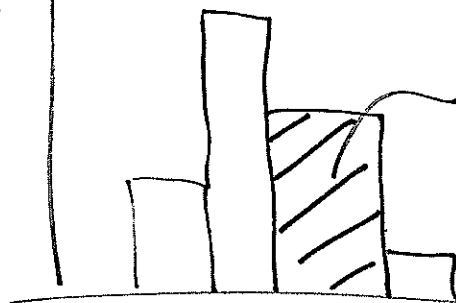
visual display of variability.

Constructing a histogram.

table.

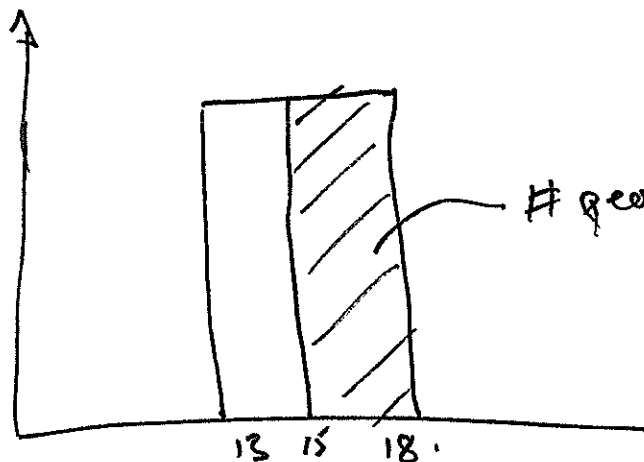
<u>class interval</u>	<u>frequency</u>	<u>%</u>	<u>height</u> : $\% \div \text{length of class interval}$ .
-----------------------	------------------	----------	---

% per chip



AREA represents frequency.

# chips.



# people with 15-18 chips.

# Exploratory Data Analysis.