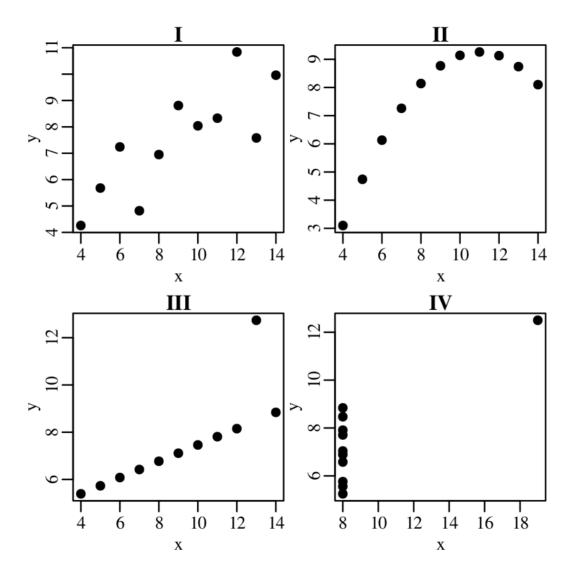
Statistics for geoscientists

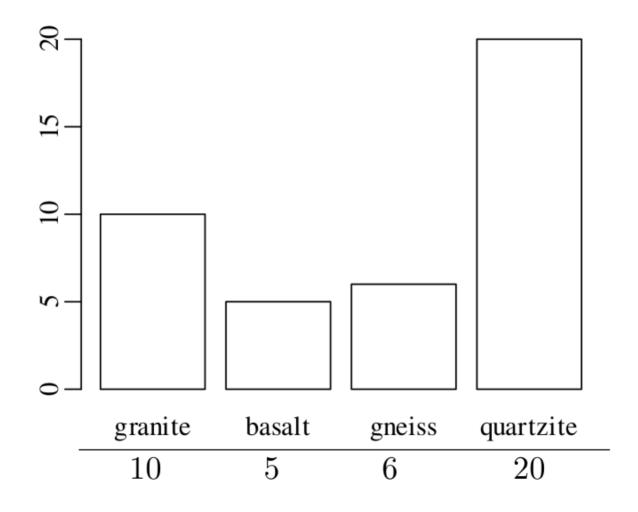
Plotting data

		_			- ·		
y	x	y	x	y	x	y	
8.04	10.0	9.14	10.0	7.46	8.0	6.58	- the mean of $x$ is 9
6.95	8.0	8.14	8.0	6.77	8.0	5.76	
7.58	13.0	8.74	13.0	12.74	8.0	7.71	- the variance of $x$ is 11
8.81	9.0	8.77	9.0	7.11	8.0	8.84	
8.33	11.0	9.26	11.0	7.81	8.0	8.47	- the mean of $y$ is $7.50$
9.96	14.0	8.10	14.0	8.84	8.0	7.04	
7.24	6.0	6.13	6.0	6.08	8.0	5.25	- the variance of $y$ is $4.125$
4.26	4.0	3.10	4.0	5.39	19.0	12.50	
10.84	12.0	9.13	12.0	8.15	8.0	5.56	- the correlation coefficient of $x$ and $y$ is 0.816
4.82	7.0	7.26	7.0	6.42	8.0	7.91	the heat 6t line is given by a - 2.00 t 0.500m
5.68	5.0	4.74	5.0	5.73	8.0	6.89	- the best fit line is given by $y = 3.00 + 0.500x$
	8.04 6.95 7.58 8.81 8.33 9.96 7.24 4.26 10.84 4.82	8.04 10.0   6.95 8.0   7.58 13.0   8.81 9.0   8.33 11.0   9.96 14.0   7.24 6.0   4.26 4.0   10.84 12.0   4.82 7.0	8.04 10.0 9.14   6.95 8.0 8.14   7.58 13.0 8.74   8.81 9.0 8.77   8.33 11.0 9.26   9.96 14.0 8.10   7.24 6.0 6.13   4.26 4.0 3.10   10.84 12.0 9.13   4.82 7.0 7.26	8.04 10.0 9.14 10.0   6.95 8.0 8.14 8.0   7.58 13.0 8.74 13.0   8.81 9.0 8.77 9.0   8.33 11.0 9.26 11.0   9.96 14.0 8.10 14.0   7.24 6.0 6.13 6.0   4.26 4.0 3.10 4.0   10.84 12.0 9.13 12.0   4.82 7.0 7.26 7.0	8.04   10.0   9.14   10.0   7.46     6.95   8.0   8.14   8.0   6.77     7.58   13.0   8.74   13.0   12.74     8.81   9.0   8.77   9.0   7.11     8.33   11.0   9.26   11.0   7.81     9.96   14.0   8.10   14.0   8.84     7.24   6.0   6.13   6.0   6.08     4.26   4.0   3.10   4.0   5.39     10.84   12.0   9.13   12.0   8.15     4.82   7.0   7.26   7.0   6.42	8.04 10.0 9.14 10.0 7.46 8.0   6.95 8.0 8.14 8.0 6.77 8.0   7.58 13.0 8.74 13.0 12.74 8.0   8.81 9.0 8.77 9.0 7.11 8.0   8.33 11.0 9.26 11.0 7.81 8.0   9.96 14.0 8.10 14.0 8.84 8.0   7.24 6.0 6.13 6.0 6.08 8.0   4.26 4.0 3.10 4.0 5.39 19.0   10.84 12.0 9.13 12.0 8.15 8.0   4.82 7.0 7.26 7.0 6.42 8.0	8.04   10.0   9.14   10.0   7.46   8.0   6.58     6.95   8.0   8.14   8.0   6.77   8.0   5.76     7.58   13.0   8.74   13.0   12.74   8.0   7.71     8.81   9.0   8.77   9.0   7.11   8.0   8.84     8.33   11.0   9.26   11.0   7.81   8.0   8.47     9.96   14.0   8.10   14.0   8.84   8.0   7.04     7.24   6.0   6.13   6.0   6.08   8.0   5.25     4.26   4.0   3.10   4.0   5.39   19.0   12.50     10.84   12.0   9.13   12.0   8.15   8.0   5.56     4.82   7.0   7.26   7.0   6.42   8.0   7.91

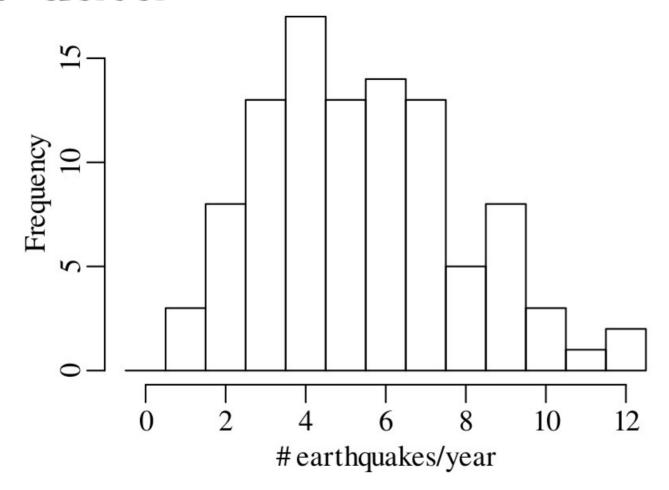
III IV



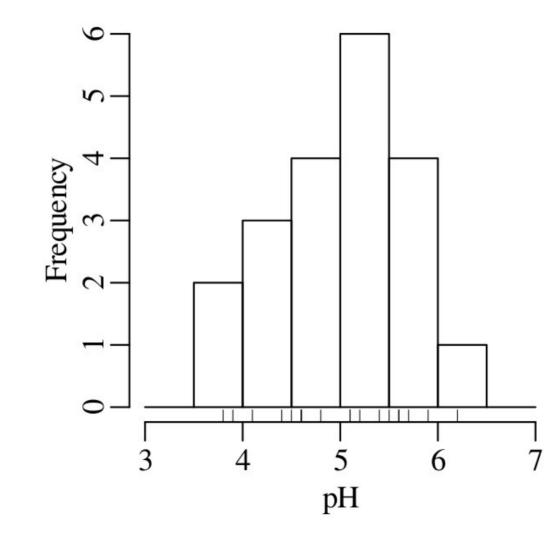
## Categorical data



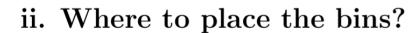
## Count data

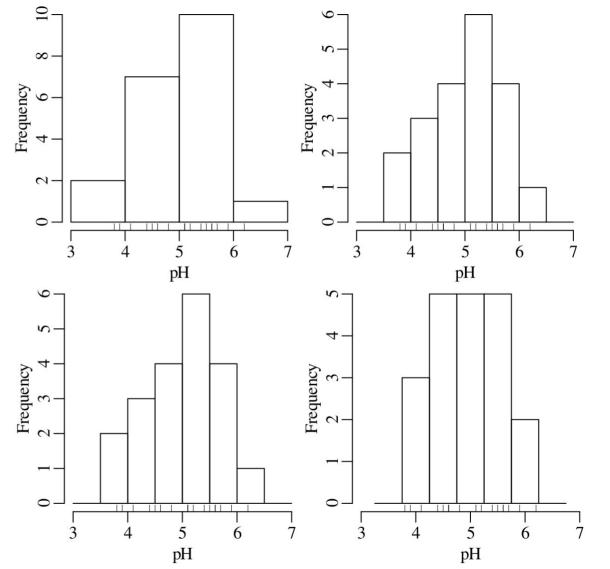


### Continuous data



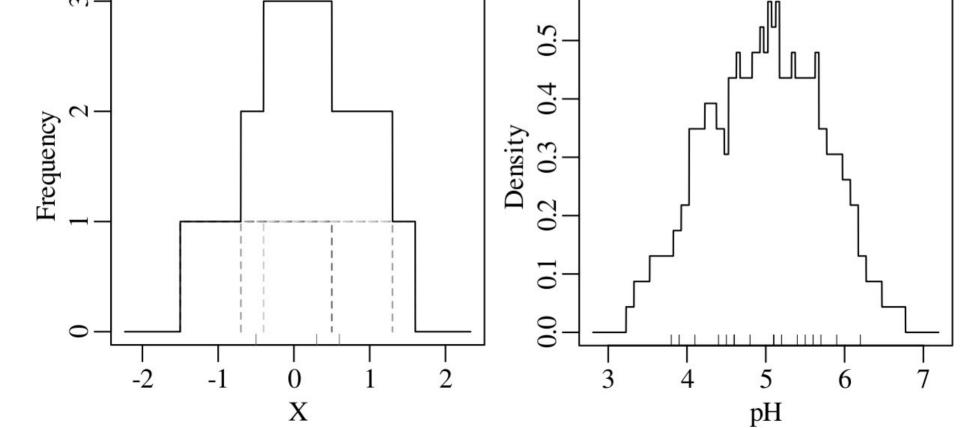
i. How many bins?



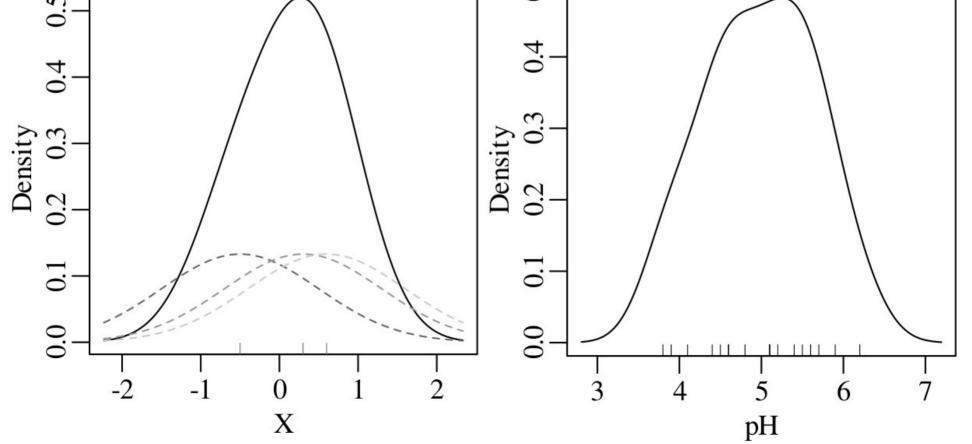


Kernel Density Estimate

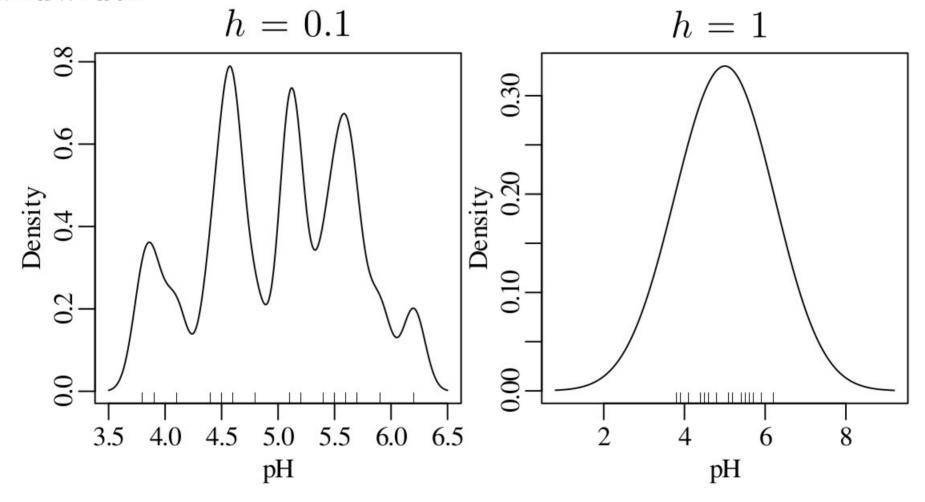
mate  $KDE(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$ 

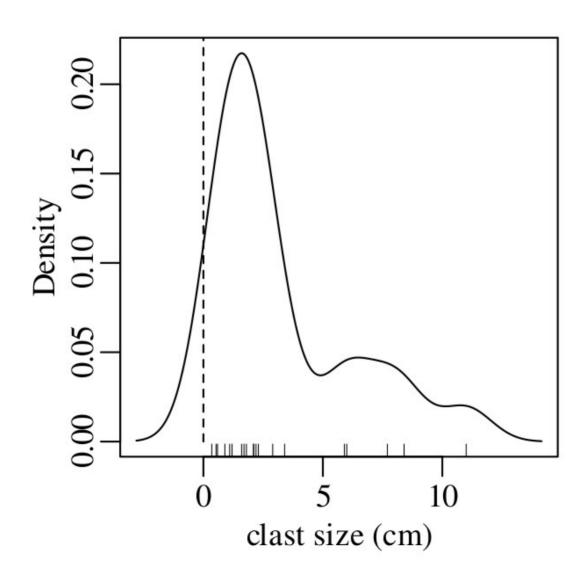


 $-\frac{1}{\sqrt{2\pi}}\exp^{1}$ Gaussian kernel:

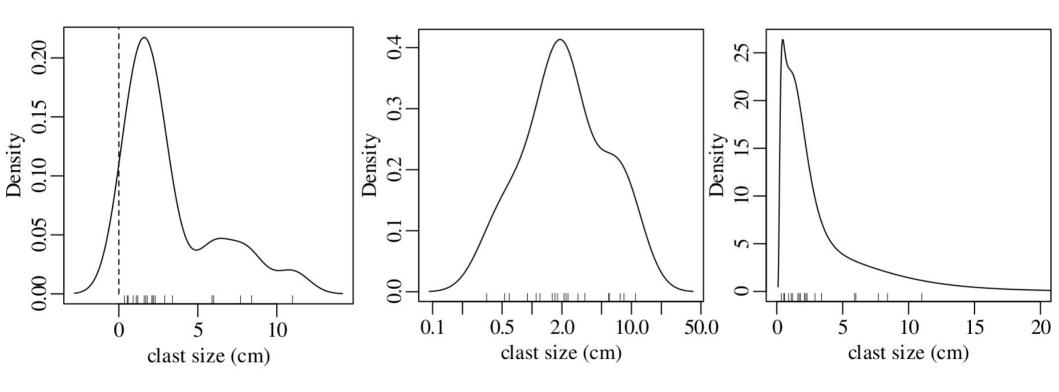


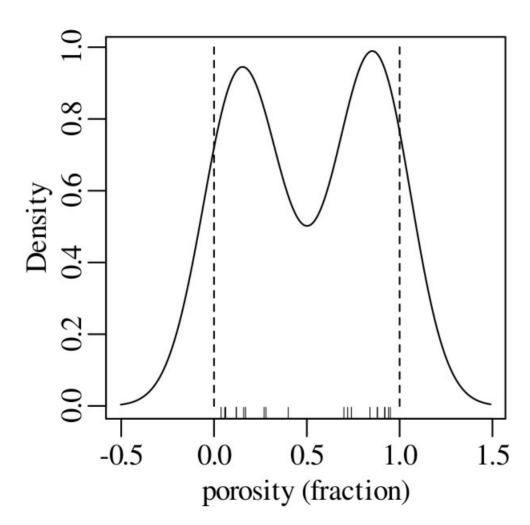
bandwidth





#### logarithmic transformation





logistic transformation

8.0

Density 0.4 0.6

0.2

0.0

-0.5

0.0

0.5

porosity (fraction)









Density

1.5

1.0

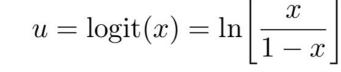
0.00

0.01









$$x = \operatorname{logit}^{-1}(u) = \frac{\exp[u]}{\exp[u] + 1}$$

80

9

20

0.0

0.2

0.4

porosity (fraction)

0.8

0.6

1.0

Density

0.99





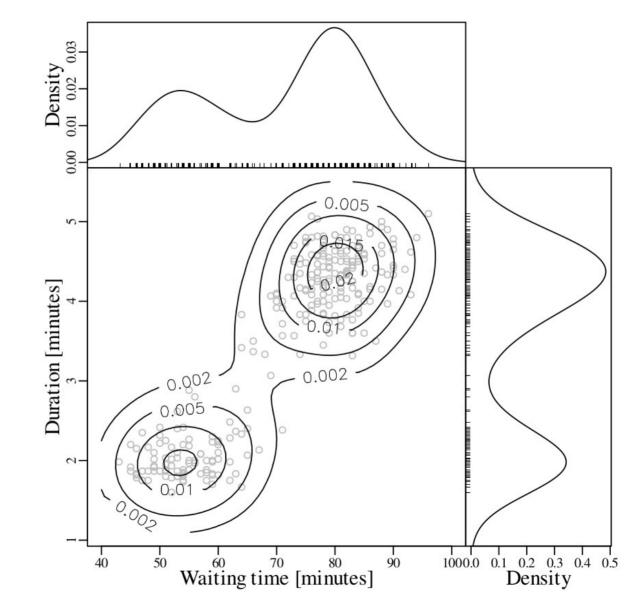
logit[porosity]

0.5

porosity (fraction)

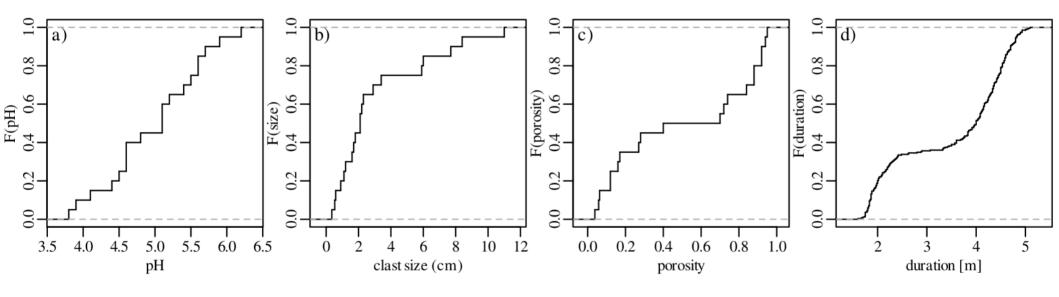
0.9

# Multivariate distributions



#### Empirical cumulative distribution fuctions

$$F(x) = \sum_{i=1}^{n} 1(x_i < x)/n$$



Statistics for geoscientists An introduction to R