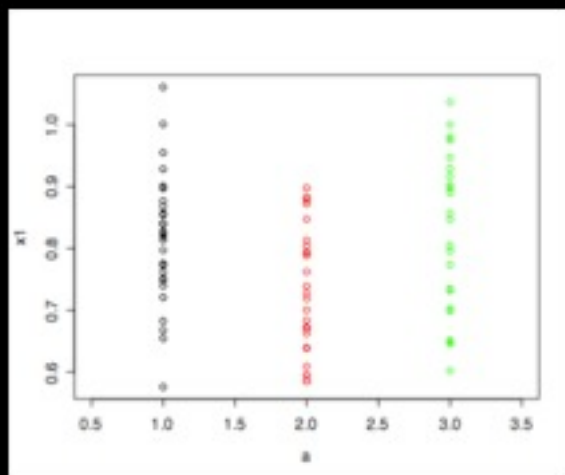


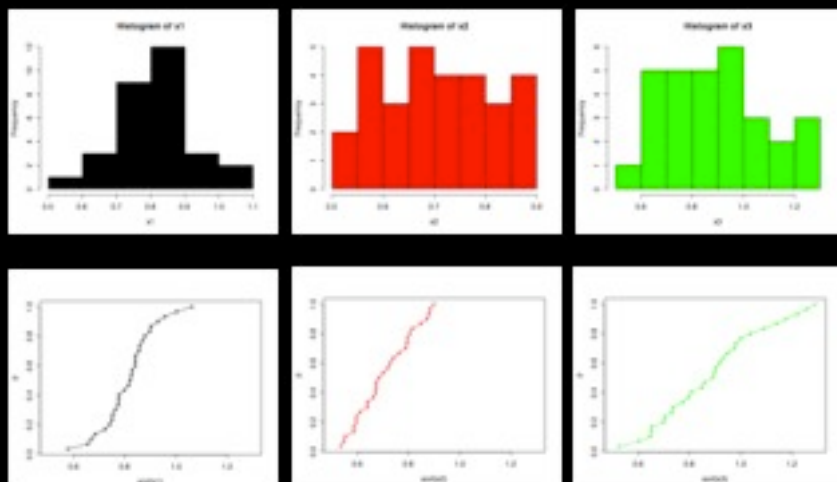
Tuesday, November 18, 2014

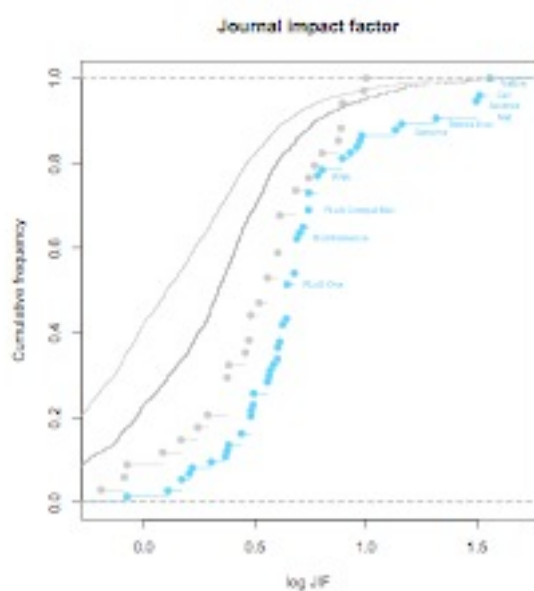
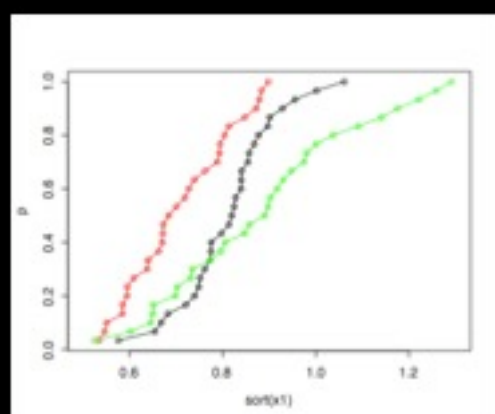
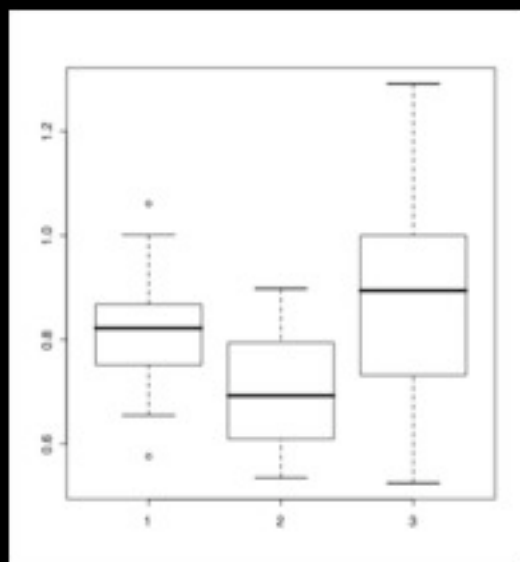
## Data exploration and representation

### Informal data analysis

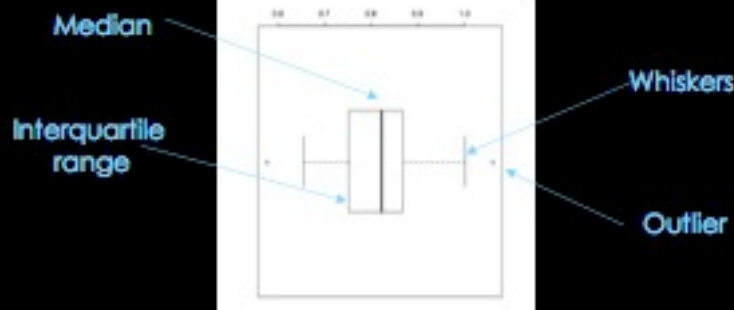
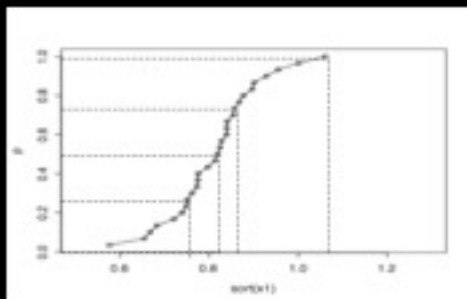
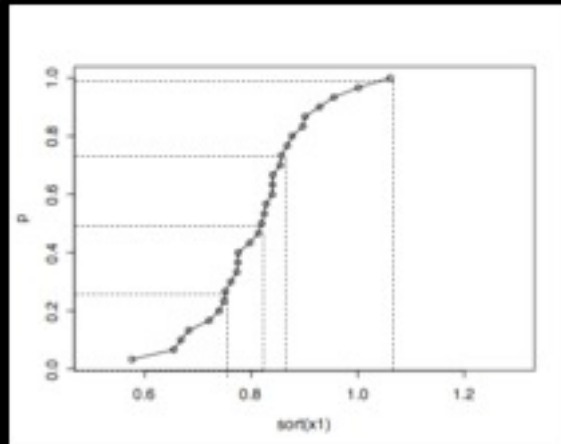


```
x1<-rnorm(30,0.8,0.1)  
x2<-rnorm(30,0.7,0.1)  
x3<-rnorm(30,0.8,0.2)
```





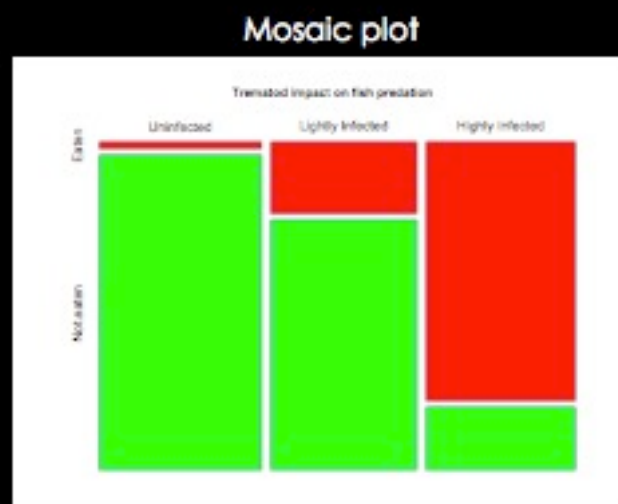
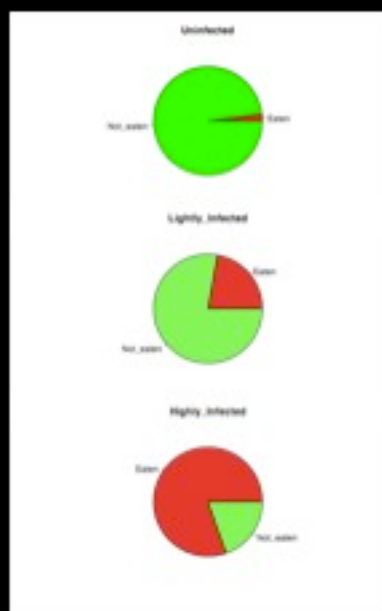
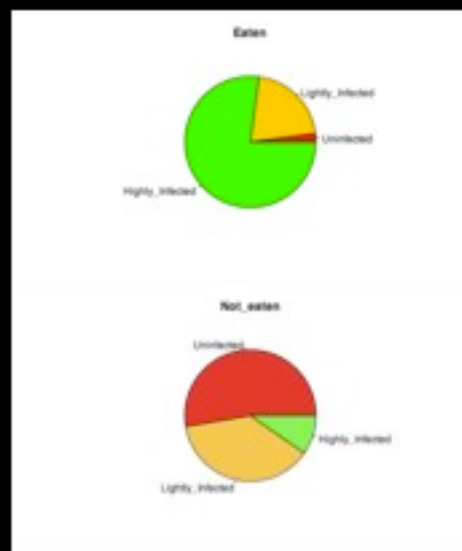
# Percentiles and quantiles



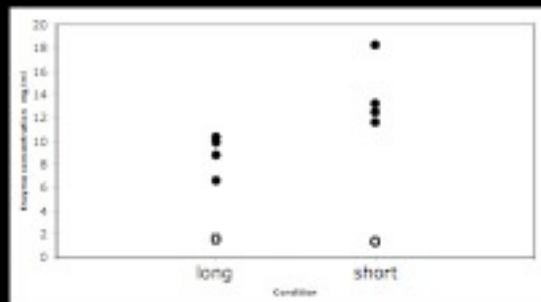
Trematodes of the species *Euhaplorchis californiensis* use three hosts during their life cycle. The worm matures in birds where it lays eggs that pass out in feces. The horn snail *Cerithidea californica* eats the eggs, which hatch and grow to another life form in the snails. When an infected snail is eaten by the California killifish *Fundulus parvipinnis* the parasite develops to the next life stage encysted in the brain case of the fish. Finally, closing the life cycle, when the fish is eaten by a bird, it becomes mature and produces new eggs.

The following table displays the observed frequencies of fish eaten or not eaten by birds as a function of the degree of trematode infection.

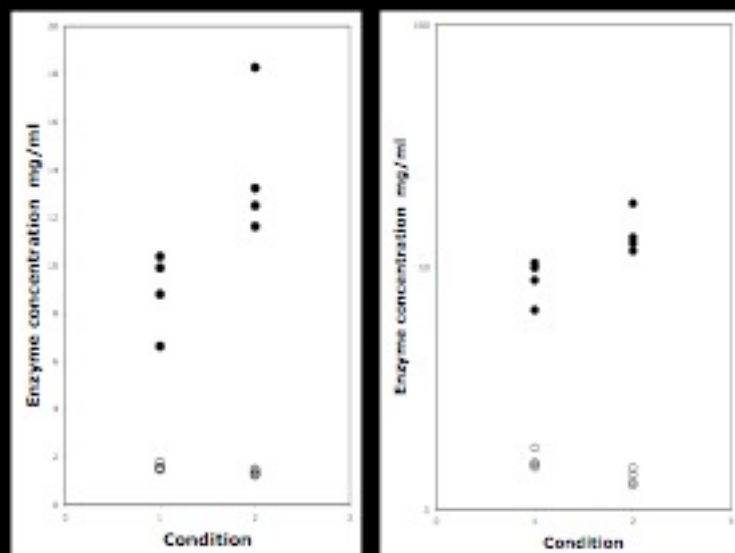
	Eaten	Not eaten
Uninfected	1	49
Lightly Infected	10	35
Highly Infected	37	9



Hamster	Heart	Brain	Condition
4	1.49	6.625	Long days
1	1.525	10.375	Long days
7	1.555	9.9	Long days
2	1.79	8.8	Long days
3	1.385	12.5	Short days
8	1.485	11.625	Short days
5	1.255	18.275	Short days
6	1.285	13.225	Short days



From: Cobb, Introduction to design and analysis of experiments. Springer

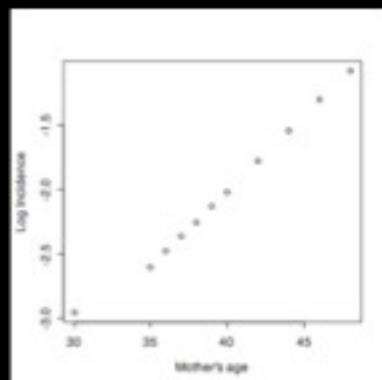
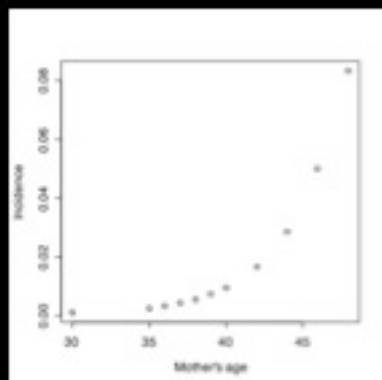


From: Cobb, Introduction to design and analysis of experiments. Springer

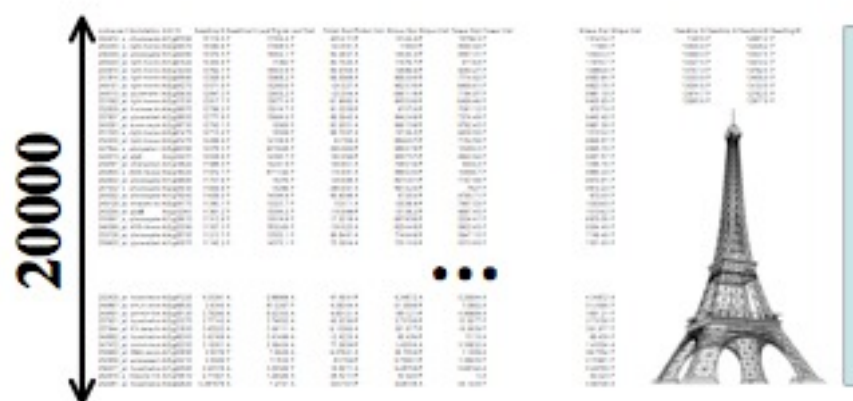
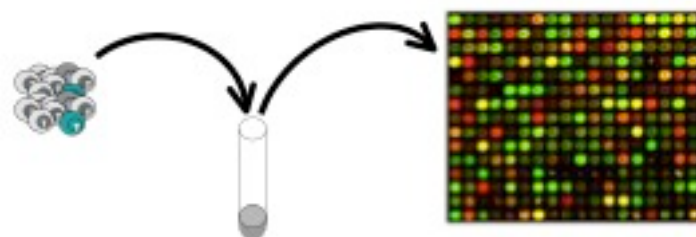
"Down syndrome, the most common genetic birth defect associated with mental retardation, occurs equally across all races and levels of society. The effects of the disorder on physical and mental development are severe and are expressed throughout the life span. The individual's family is also affected emotionally, economically, and socially."

Mother's age	Incidence of Down syndrome	Mother's age	Incidence of Down syndrome
under 30	less than 1 in 1000	39	1 in 135
30	1 in 900	40	1 in 105
35	1 in 400	42	1 in 60
36	1 in 300	44	1 in 35
37	1 in 230	46	1 in 20
38	1 in 180	48	1 in 12

Source: <http://www.afonweb.com/cs/downsyndrome/index.htm?page=hescox.html>

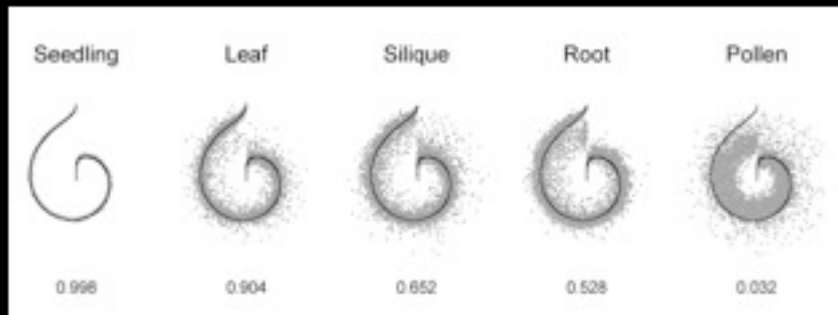


Doesn't it look suspicious ?



A Table about 400 meters tall ?

## How to plot the data ?



*Arabidopsis thaliana*

Formal description of data:

Proportion, location and spread

Measurements of location and spread

Mean

Variance

Standard deviation

Median

Standard error

Mode

Interquantile/Interquartile range

Estimating mean from sample

Estimating the mean from the sample's histogram

### Quick Formulas

Mean

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$$

$$\bar{Y} = \sum_{i=1}^n Y_i f(Y_i)$$

Variance

$$s^2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n - 1}$$

Standard deviation

$$s = \sqrt{s^2}$$

Standard error (of the mean)

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}}, \quad SE_{\bar{Y}} = \frac{s}{\sqrt{n}}$$

### Quick Formulas

Proportion

$$\hat{p} = \frac{X}{n}$$

$$Mean(\hat{p}) = p$$

$$\sigma(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$$



We never have access to  
the **true** probability distribution of population  
neither  
the **true** sampling probability distribution

**We can only get estimates !**

**What is a good estimator ?**

What are the good estimators for the mean, the  
standard deviation, or for the proportion ?