How to present results of genomic studies in an intelligible form

Henner Simianer

Animal Breeding and Genetics Group, Department of Animal Sciences, Georg-August-University Göttingen, Germany

Purpose of a scientific talk at a conference

- Convey a message to the audience
- The audience should be able to understand your reasoning
- Provide evidence that your empirical results are sound and reliable
- The audience should be convinced by your arguments

Reports on an empirical study

- Description of your experiment should ideally enable others to repeat the experiment
- Due to restrictions (mainly in time) often not fully achievable
- But still try to be as precise and comprehensive as possible

Reports on an empirical study

Step 1: State your research question

Ideally as formal hypothesis

If you want to test whether a certain factor F (a marker, a gene, a breed) has an effect on a variable y (growth, health):

H_o: F does not affect y

H_1: F does affect y (one-sided or two sided)

Reports on an empirical study

Step 2: Describe your experiment

„Material“ and methods

Provide all relevant information on the substrate you worked with (animals, tissues, technologies) in your experiment

Describe the actual experiment (what was done, how often, when were samples taken etc.)

Describe editing and quality control of the data (outlier detection, filtering criteria, imputation etc.)

Reports on an empirical study

Step 2: Describe your experiment

„Material“ and methods

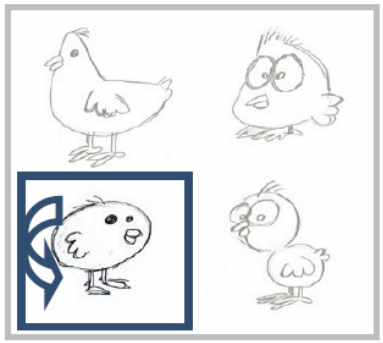
Describe the statistical approaches you use

Just naming a software (we used SAS/R ...) is not sufficient.

Provide the complete models you have used (including the relevant assumptions)

Describe in such detail that a knowledgeable person can follow

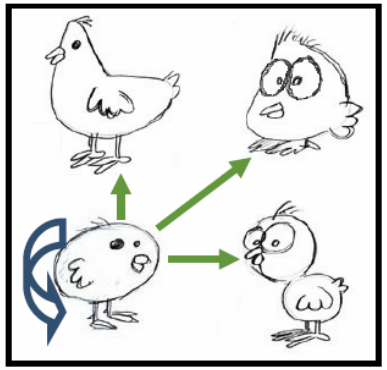
Direct model⇔ Direct-Social model



Direct

$$A_D$$

$$\sigma^2_{A_D}$$



Direct-indirect

$$A_T = A_D + 3A_S$$

$$\sigma^2_{A_T} = \sigma^2_{A_D} + 6\sigma_{A_{DS}} + 9\sigma^2_{A_S}$$

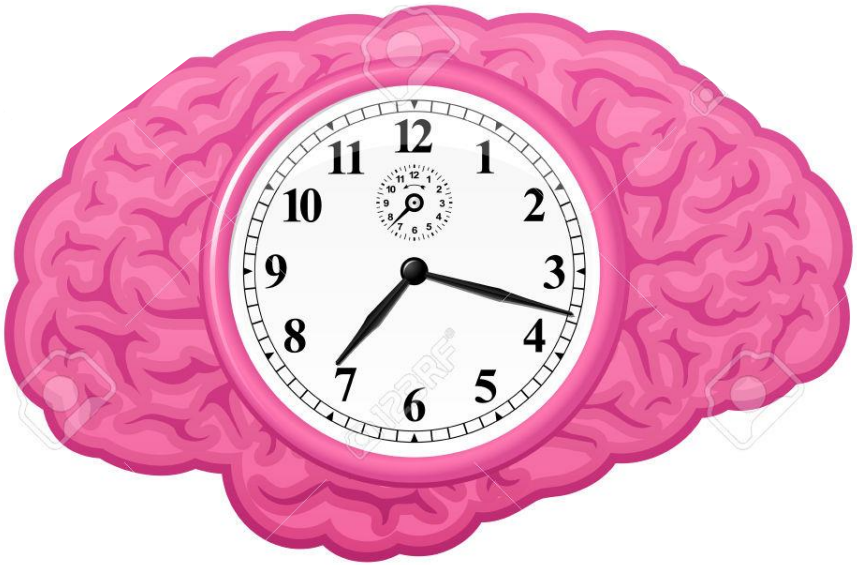
WAGENINGEN UR
For quality of life

Peeters et al., EAAP 2013

Reports on an empirical study

Step 3: Present your results

Graphs are much more intuitive than tables – the human brain is analog



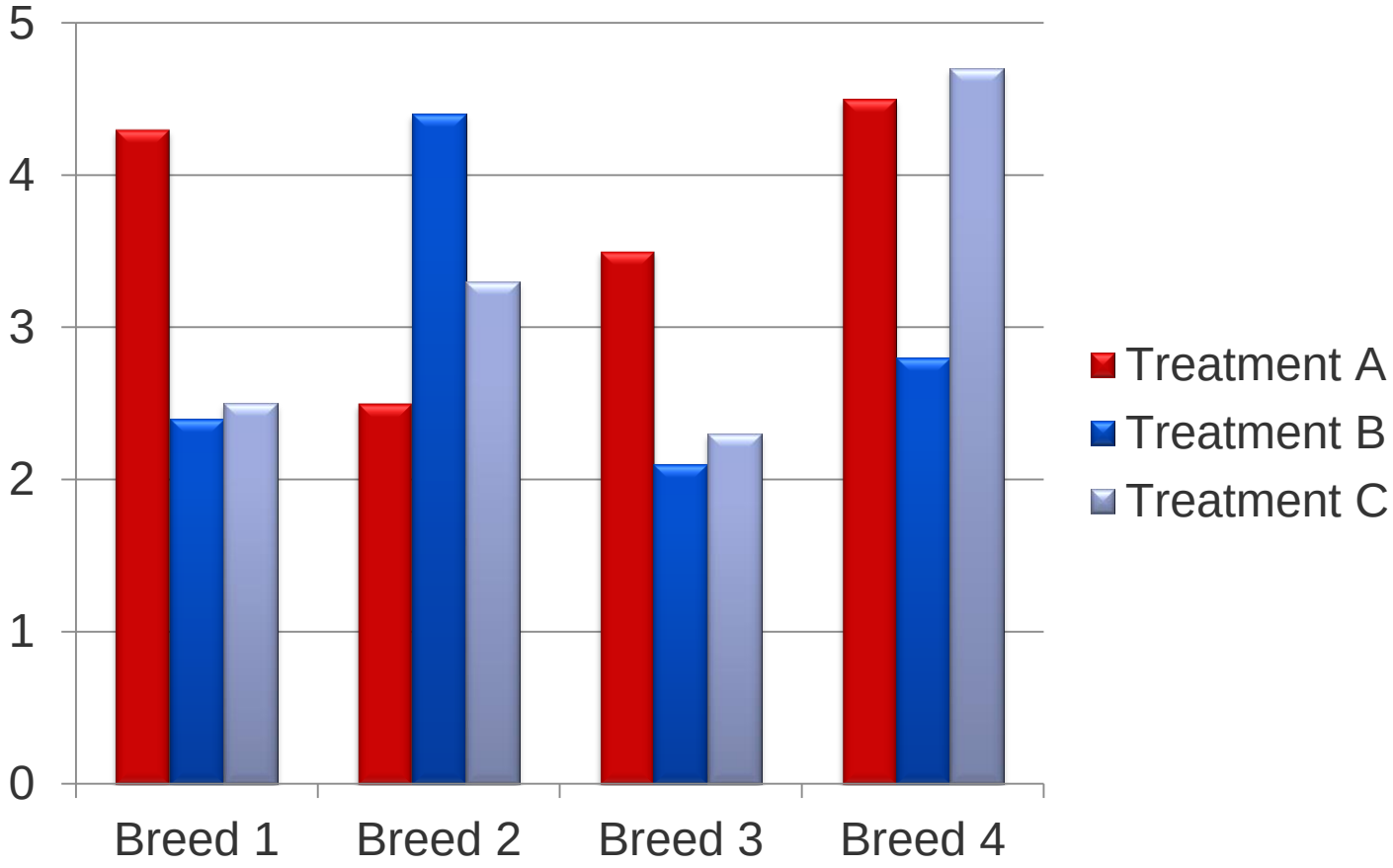
Experiment I: which breed x treatment combination has the lowest performance? 3 seconds

	Treatment A	Treatment B	Treatment C
Breed 1			
Breed 2			
Breed 3			
Breed 4			

Experiment I: which breed x treatment combination has the lowest performance? 3 seconds

	Treatment A	Treatment B	Treatment C
Breed 1	4,3 ± 0,2	2,4 ± 0,1	2,5 ± 0,1
Breed 2	2,5 ± 0,1	4,4 ± 0,2	3,3 ± 0,2
Breed 3	3,5 ± 0,3	2,1 ± 0,4	2,3 ± 0,2
Breed 4	4,5 ± 0,2	2,8 ± 0,2	4,7 ± 0,3

Experiment I: which breed x treatment combination has the lowest performance?



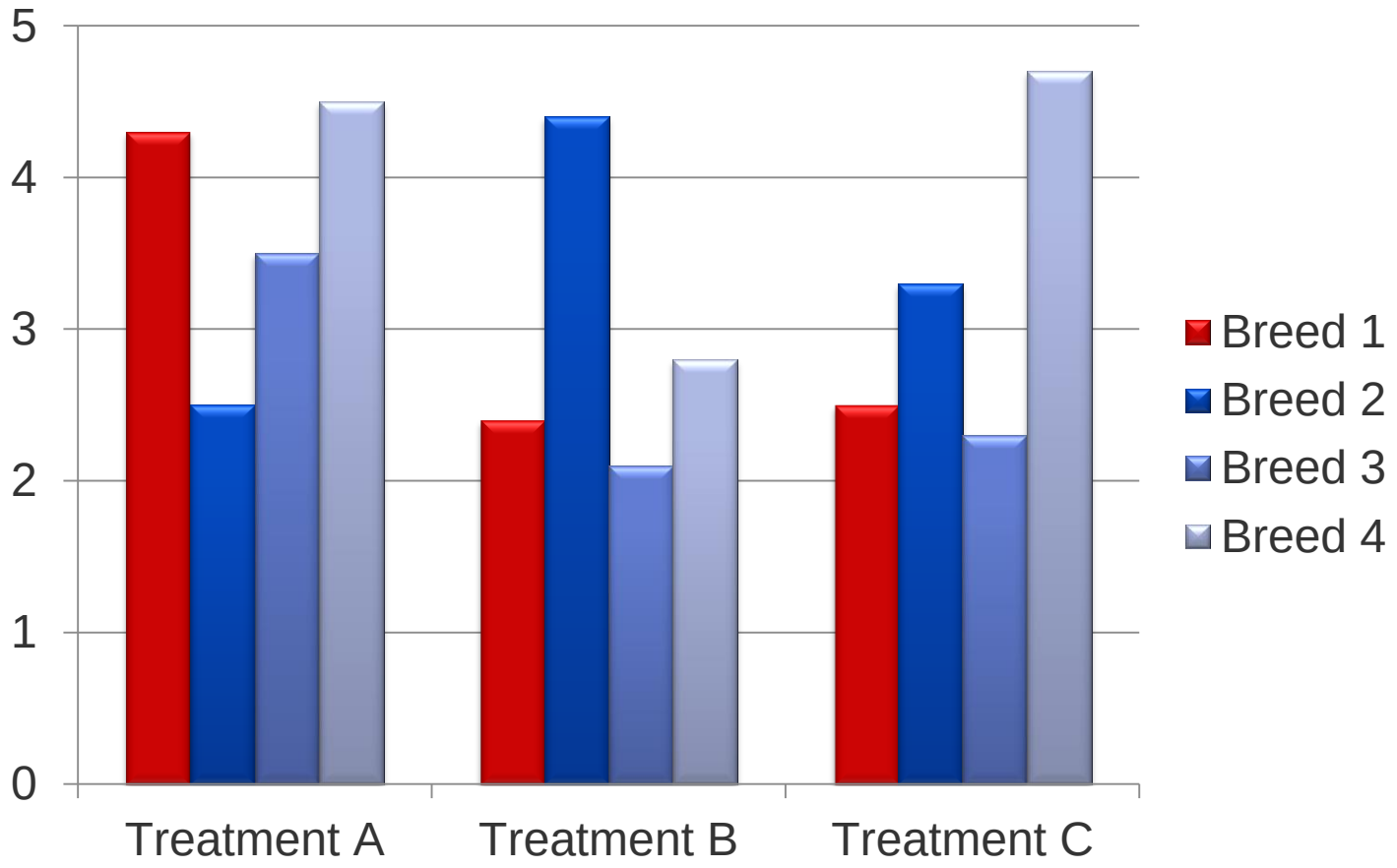
Experiment II: which breeds react similar to the three treatments?
6 seconds

	Treatment A	Treatment B	Treatment C
Breed 1			
Breed 2			
Breed 3			
Breed 4			

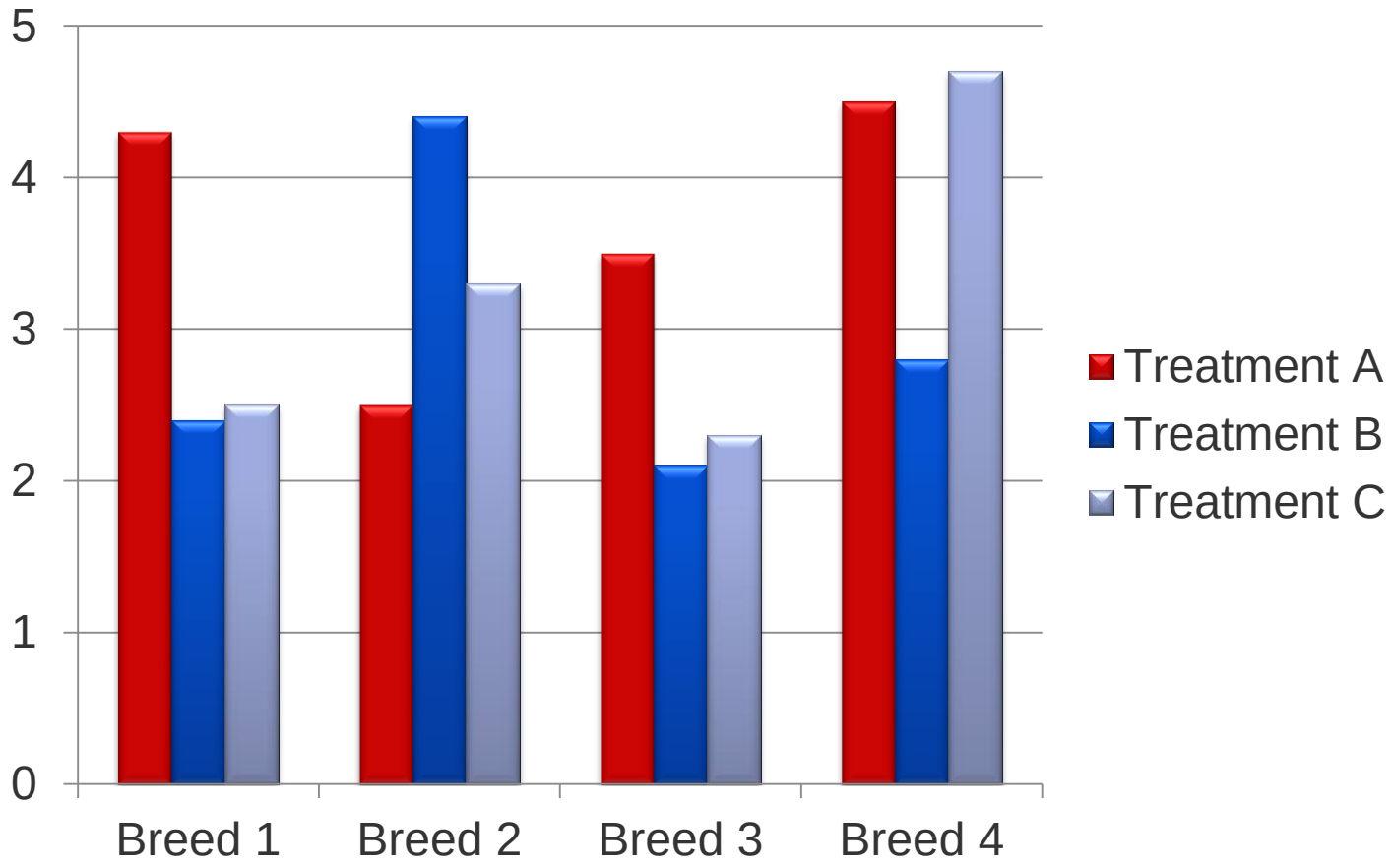
Experiment II: which breeds react similar to the three treatments?
6 seconds

	Treatment A	Treatment B	Treatment C
Breed 1	4,3 ± 0,2	2,4 ± 0,1	2,5 ± 0,1
Breed 2	2,5 ± 0,1	4,4 ± 0,2	3,3 ± 0,2
Breed 3	3,5 ± 0,3	2,1 ± 0,4	2,3 ± 0,2
Breed 4	4,5 ± 0,2	2,8 ± 0,2	4,7 ± 0,3

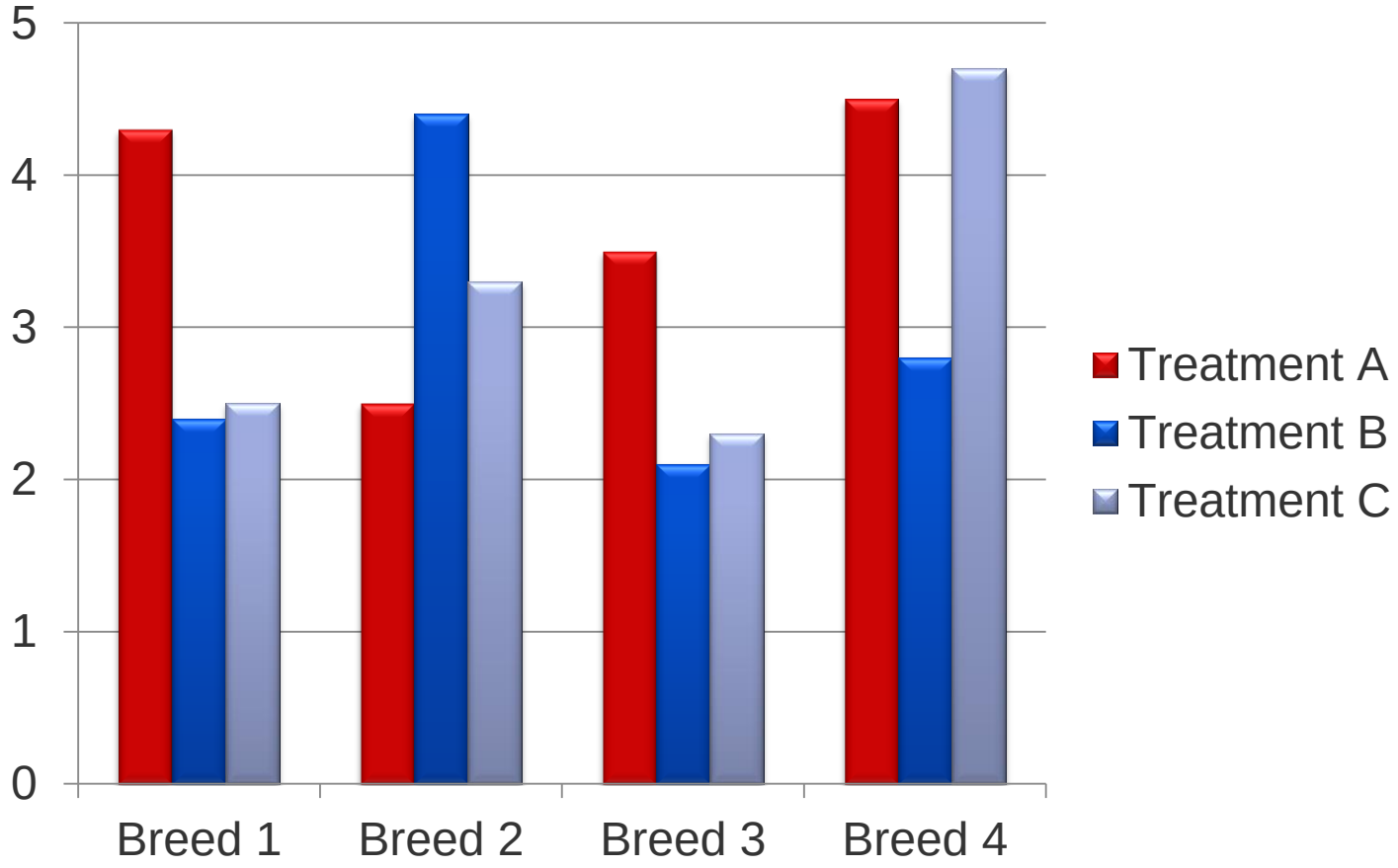
Experiment II: which breeds react similar to the three treatments?



Experiment II: which breeds react similar to the three treatments?



Experiment II: which breeds react similar to the three treatments?

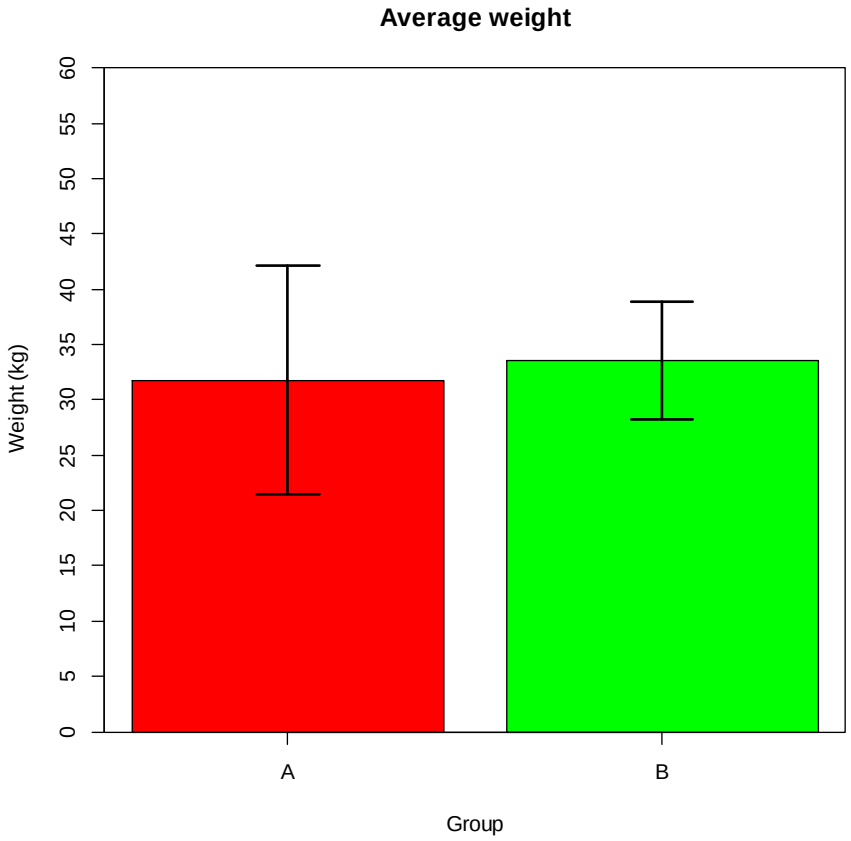


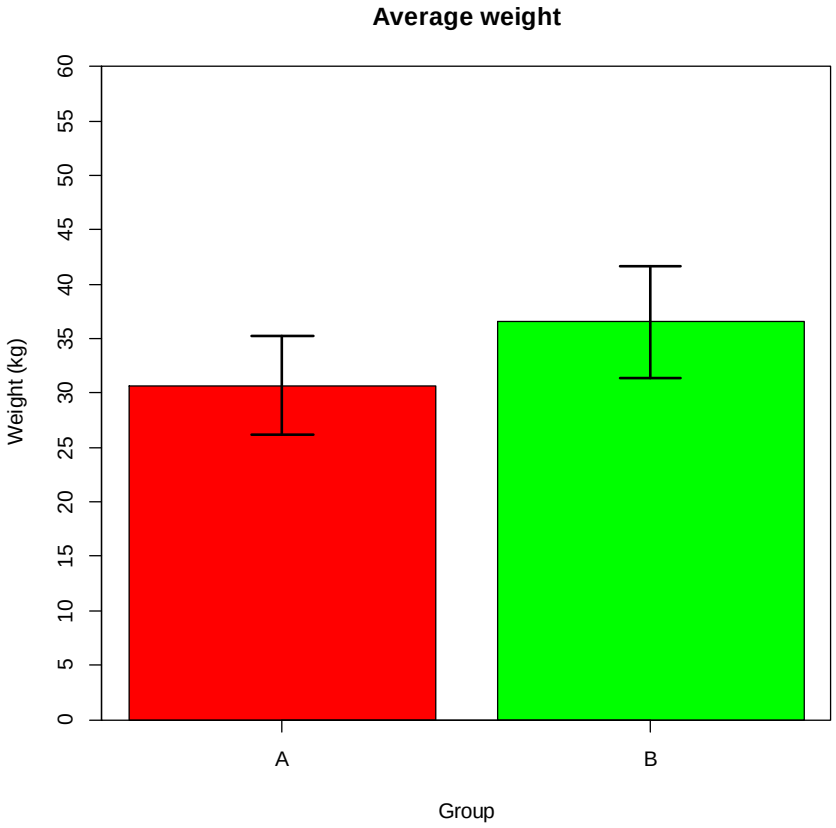
Reports on an empirical study

Step 3: Present your results

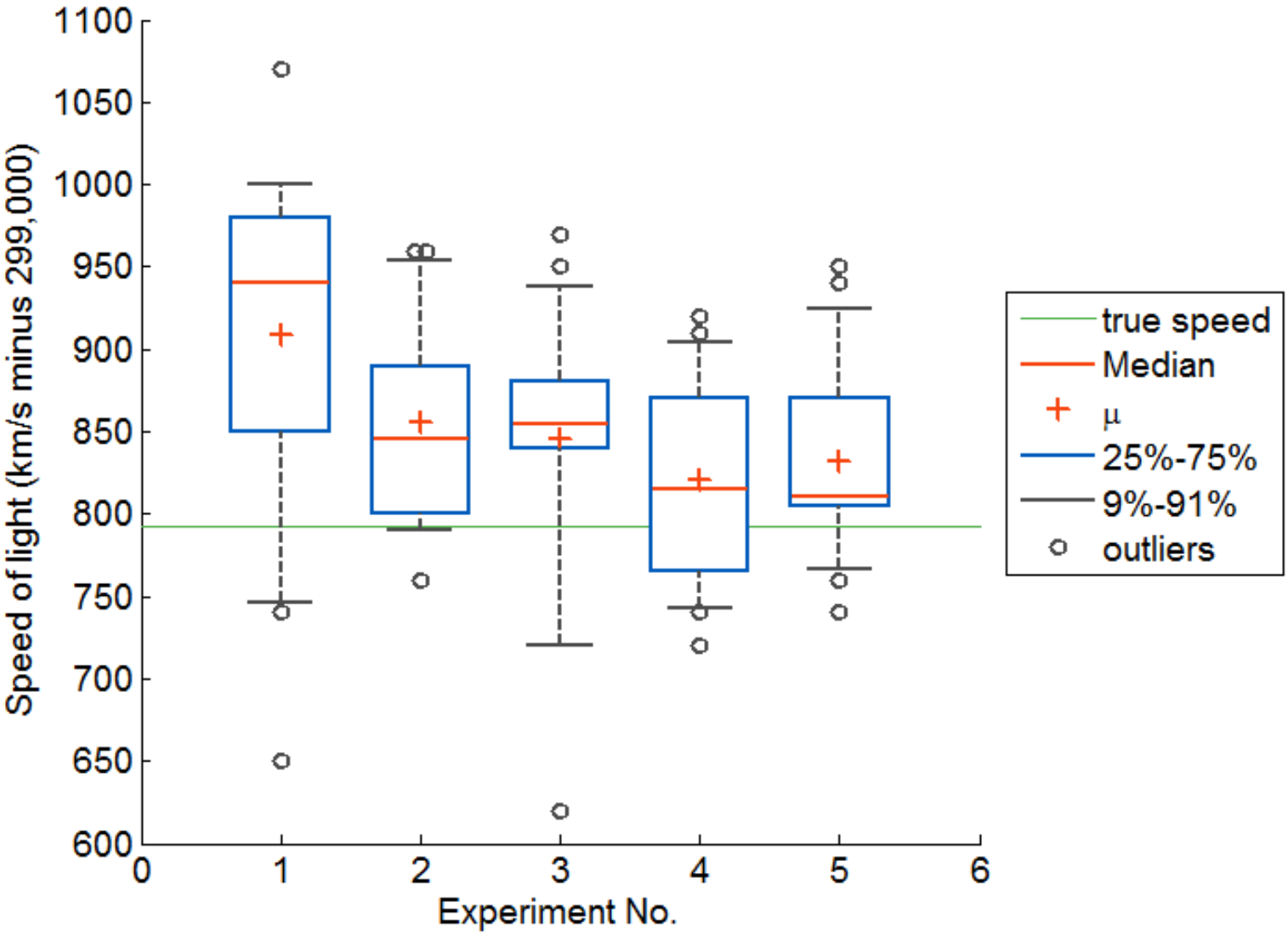
Report your results with some indication of the precision of your estimates

- standard errors
- confidence intervals
- posterior distributions

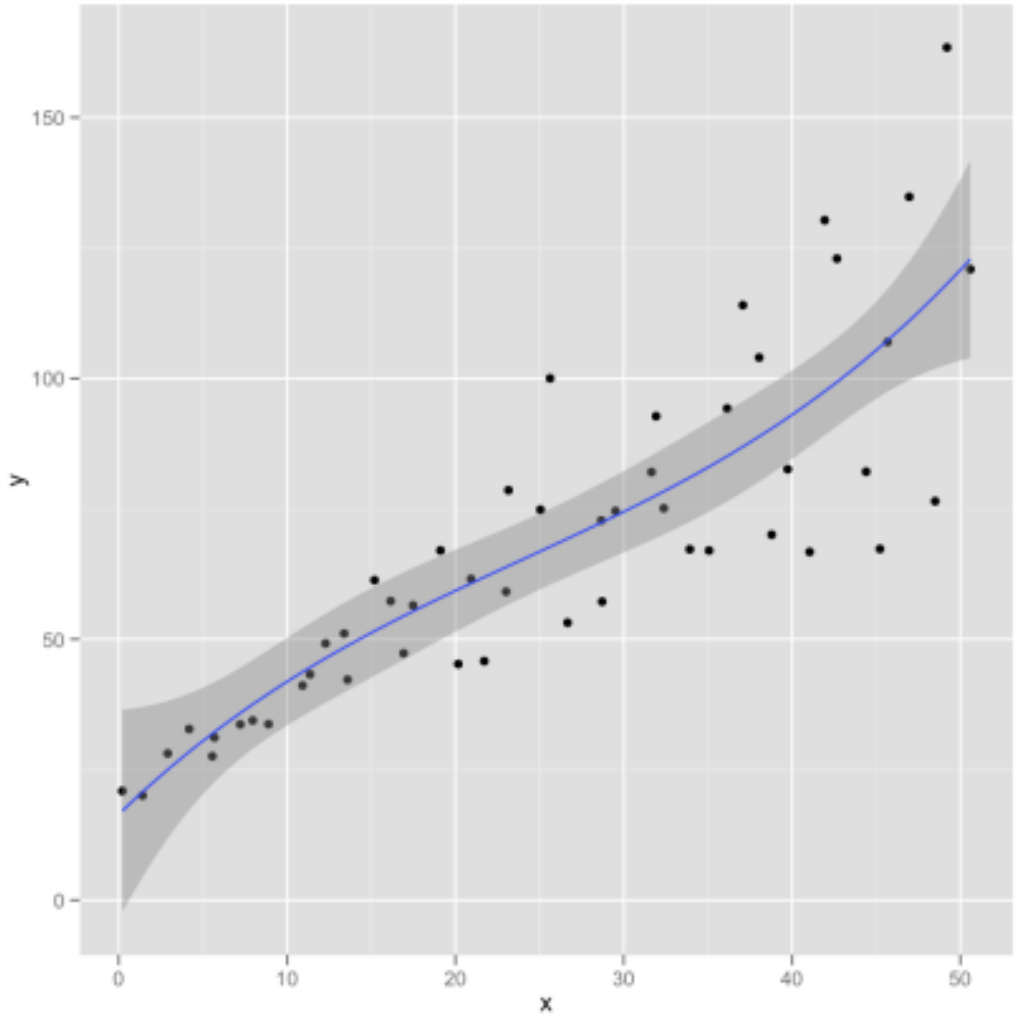

Average weight
Weight (kg)
0
5
10
15
20
25
30
35
40
45
50
55
60
A
B
Group


Average weight
Weight (kg)
0
5
10
15
20
25
30
35
40
45
50
55
60
A
B
Group

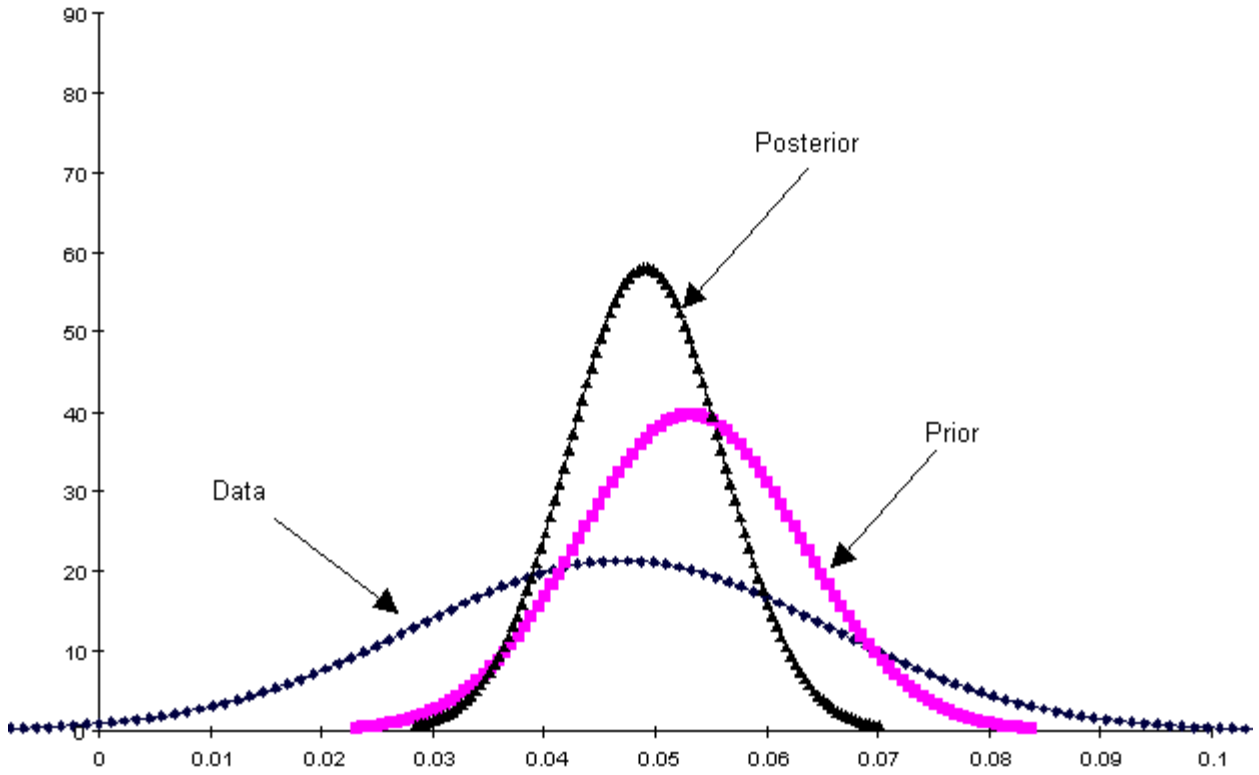
Box plots are a good way of characterising the entire distribution



You can also show the ‘confidence band’ around an estimated (non-linear) regression



Bayesian statistics provide posterior distributions of estimated quantities

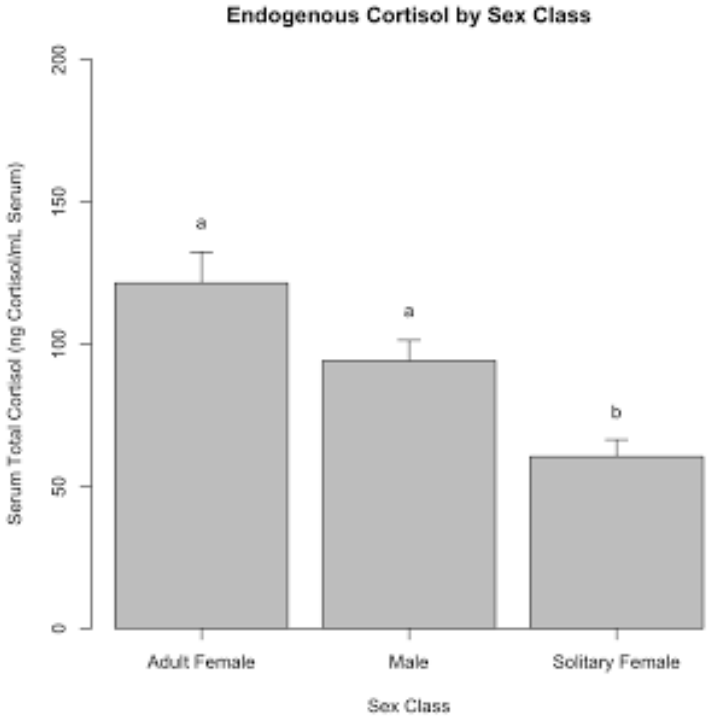


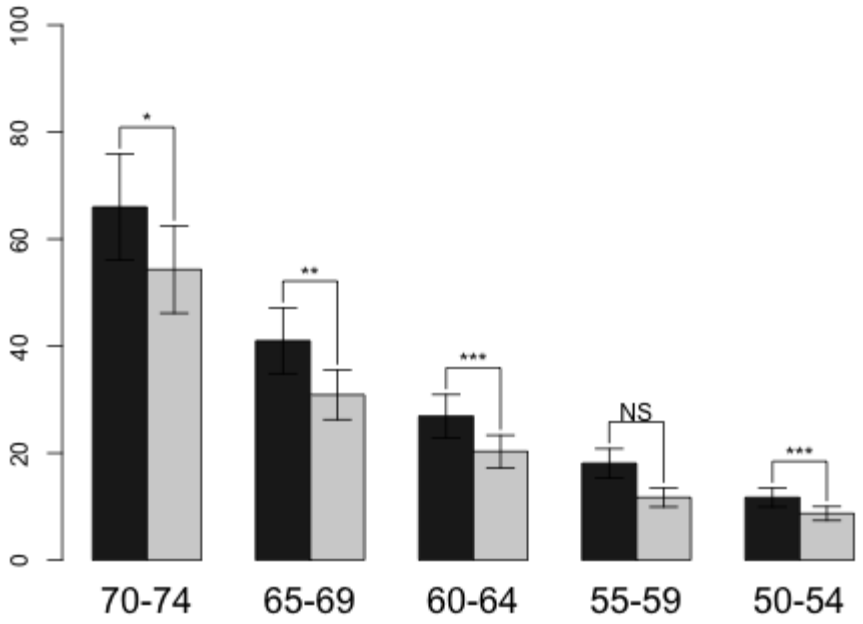
Reports on an empirical study

Step 3: Present your results

Report the statistical significance of your results

- You have stated earlier which hypotheses were tested and which statistical tests were used
- Report p-values or ‚usual' error levels (* = 0.05; ** = 0.01; *** = 0.001)


Endogenous Cortisol by Sex Class
Serum Total Cortisol (ng Cortisol/mL Serum)
0
50
100
150
200
a
a
b
Adult Female
Male
Solitary Female
Sex Class


0
20
40
60
80
100
*
**
***
NS
***
70-74
65-69
60-64
55-59
50-54

Account for multiple testing

Nominal error level $\alpha = 0.05$ (*) means that under the H_0 (no effect) 1 out of 20 tests finds an effect

N = 1000 tests → 50 ‚significant' effects even under the H_0

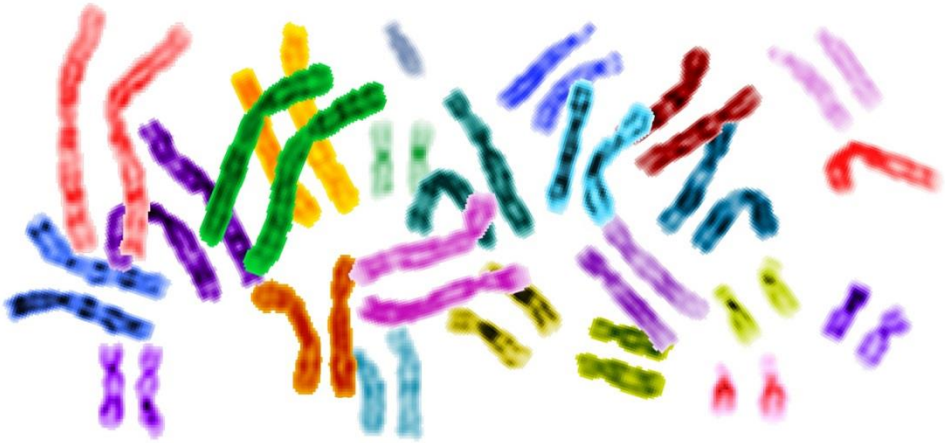
- Bonferroni (1935) correction: use in each test the test level α /N; the global error probability is kept but testing is very (too) conservative – improved versions e.g. by Holm (1979) and Hochberg (1988)
- False Discovery Rate (Benjamini and Hochberg, 1995): Test such, that a proportion α of the significant results are false positive
- Permutation test (Doerge and Churchill, 1996): generate the distribution of the test statistic under H_0 through permutation of the data.

Genome-wide vs. chromosome-wise testing

With multiple testing the power decreases with the number of tests

- If you can *a priori* restrict the location of a QTL to a subregion (e.g. a chromosome), then chromosome-wise testing is legitimate

- Otherwise you always have to account for all tests you are doing → only genome-wide results should be communicated

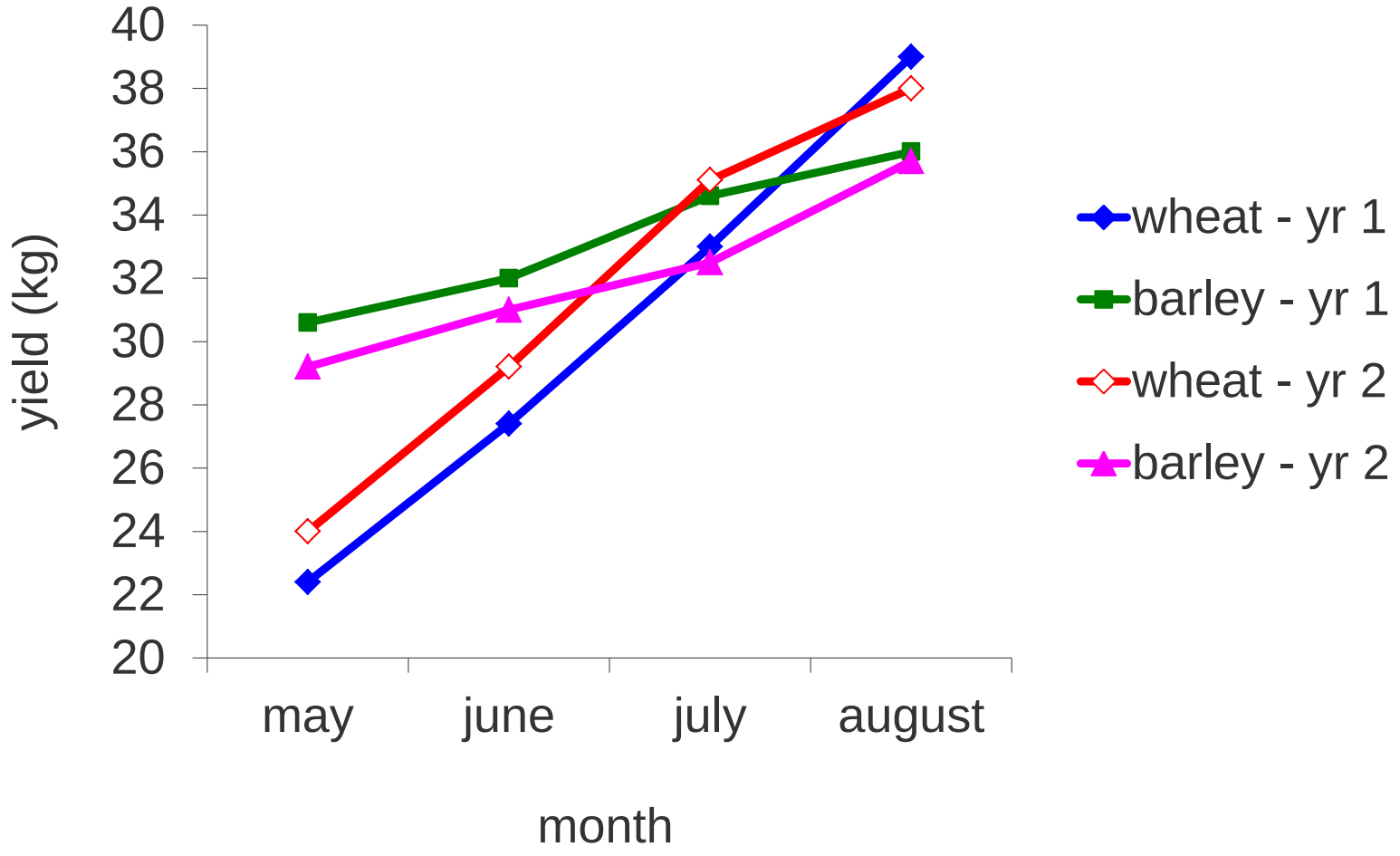


Reports on an empirical study

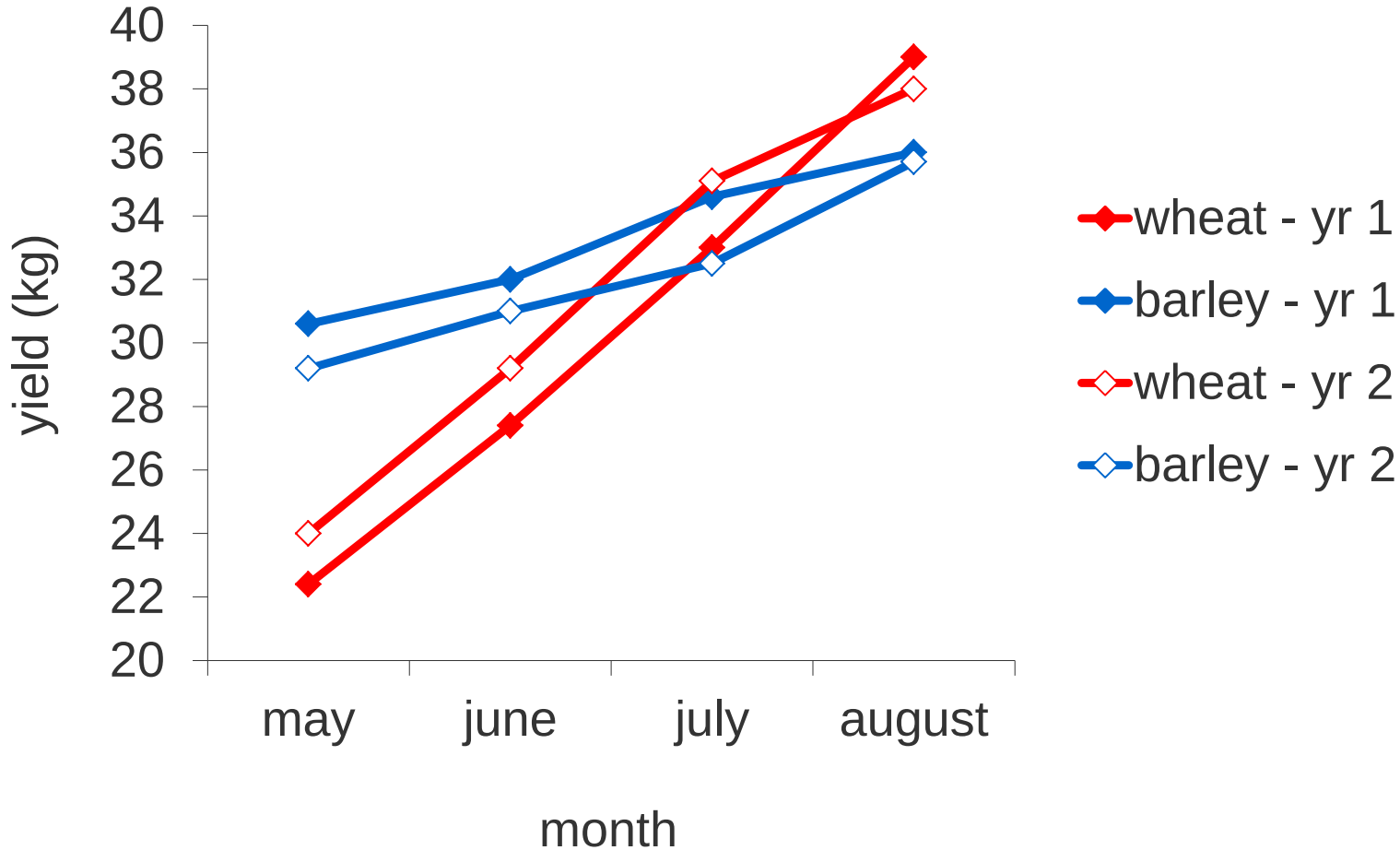
Step 3: Present your results – some basics

- state clearly what is shown on a slide
- put titles on axes in sufficiently large and readable fonts
- add a legend, if necessary
- Use colours and symbols in a systematic way
- Assign colours, symbols, and line types consistently across slides to the same object

Figures – which type?



Figures – which type?



Reports on an empirical study

Step 3: Present your results

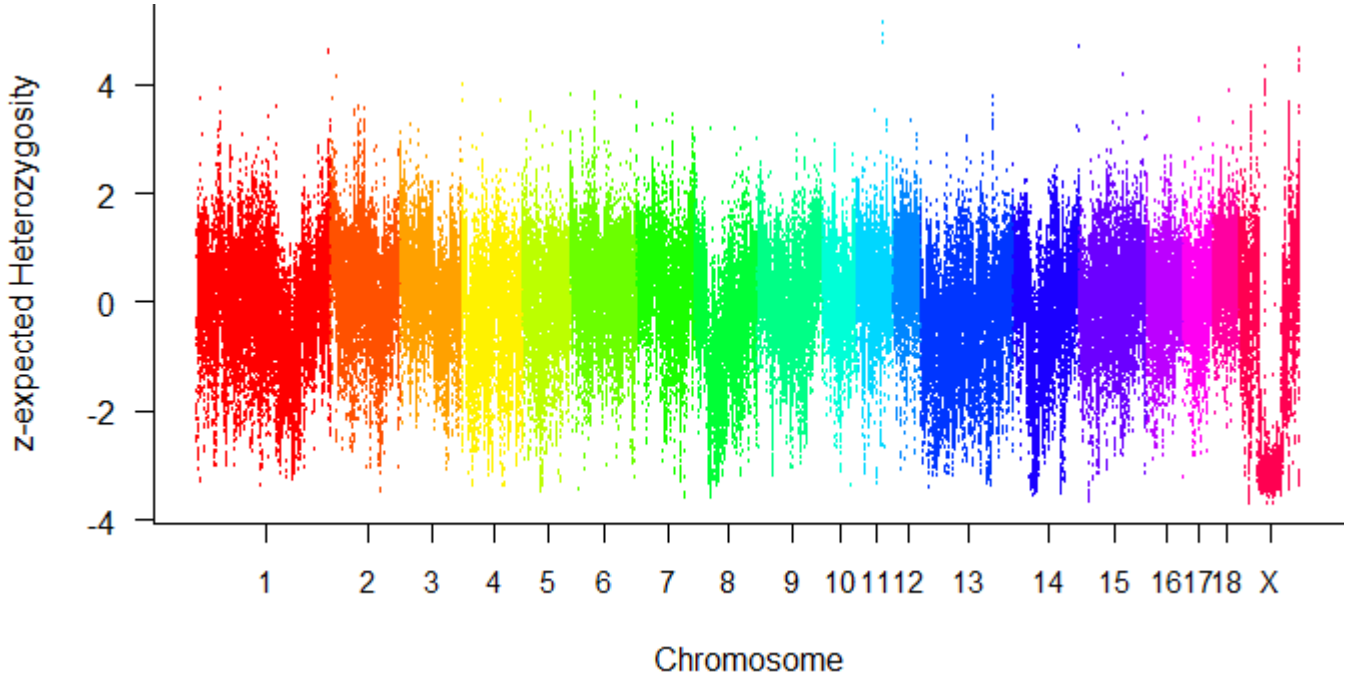
Be creative!

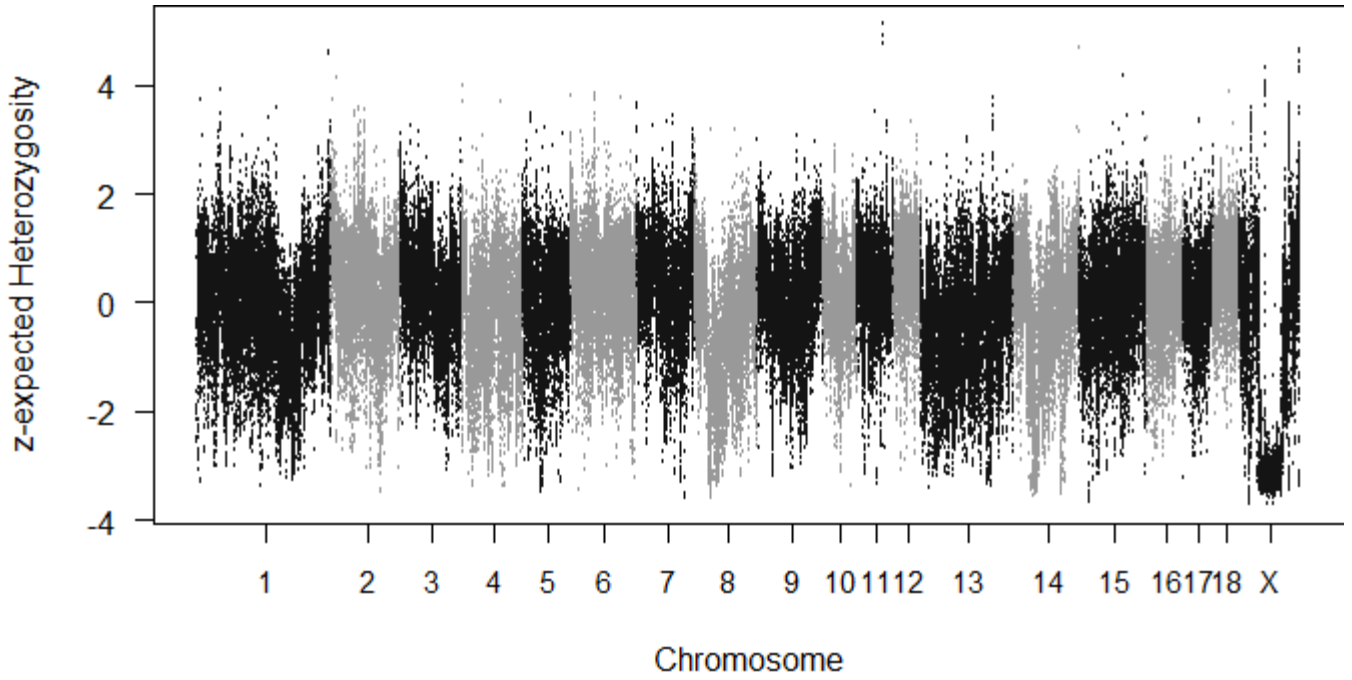
Avoid default settings of standard software

Make use of the great opportunities of modern statistics and graphics programs

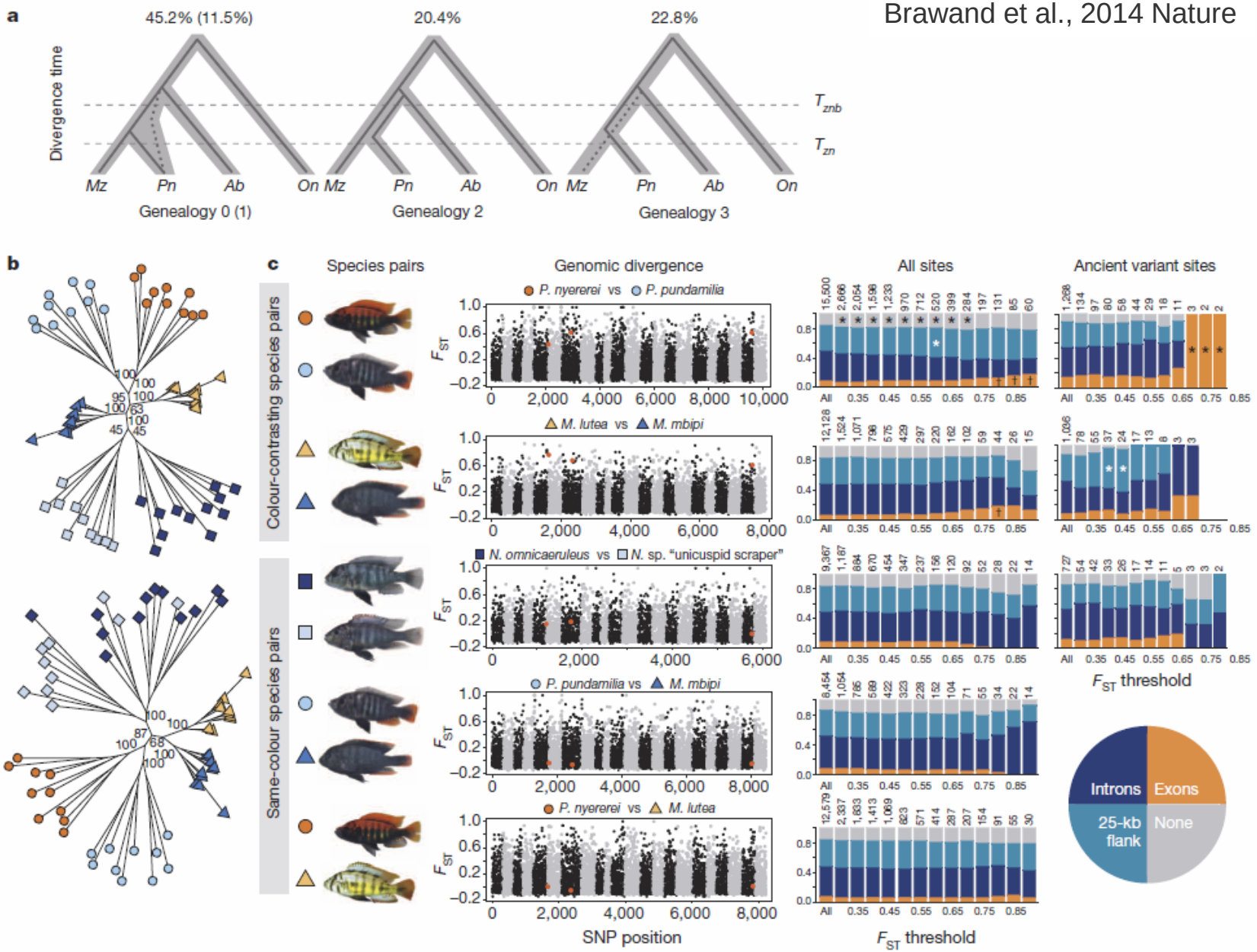
Be inspired by what the leaders in the field are doing

<u>But:</u> Fanciness should not be at the expense of clarity

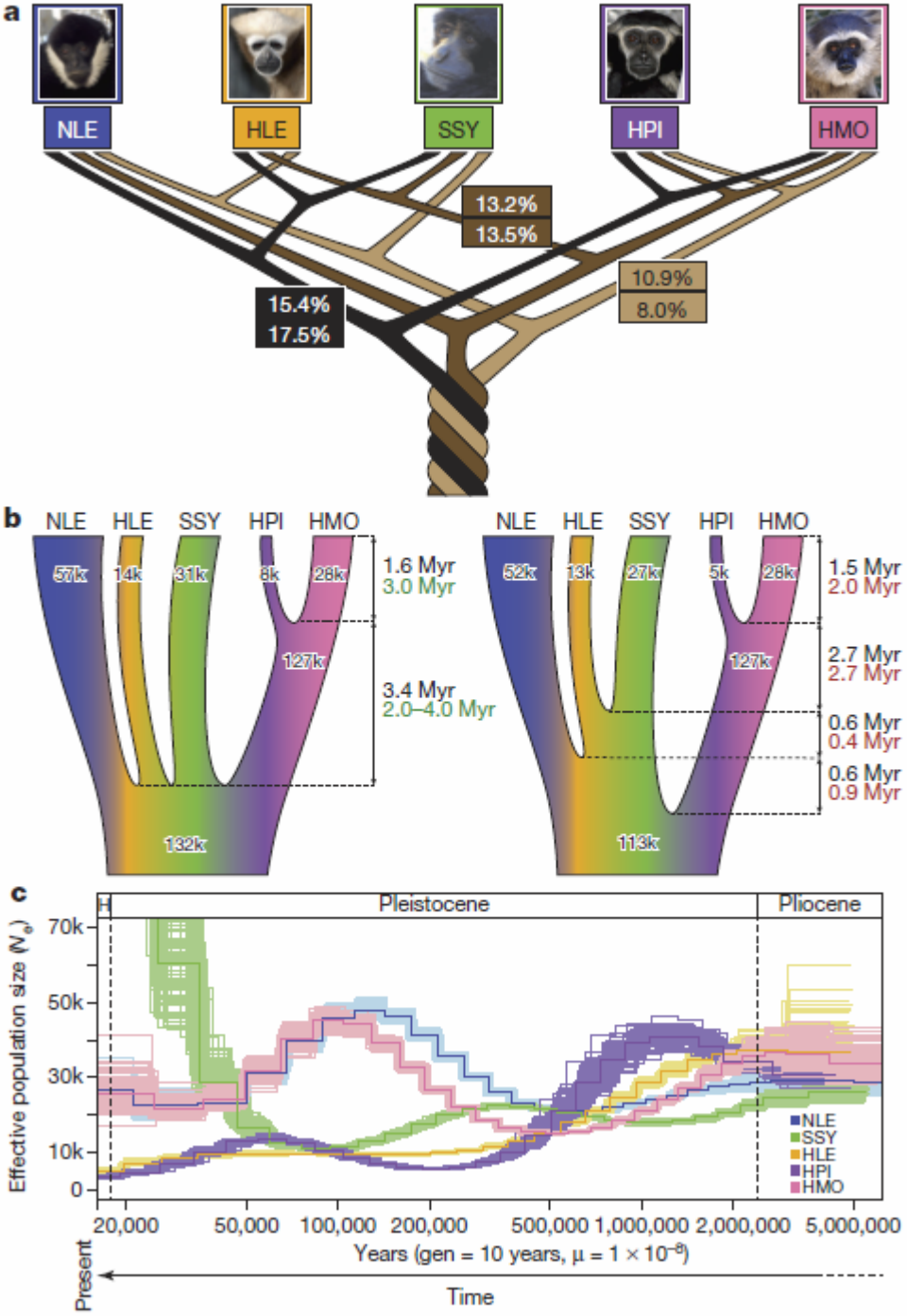

z-expected Heterozygosity
4
2
0
-2
-4
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
X
Chromosome

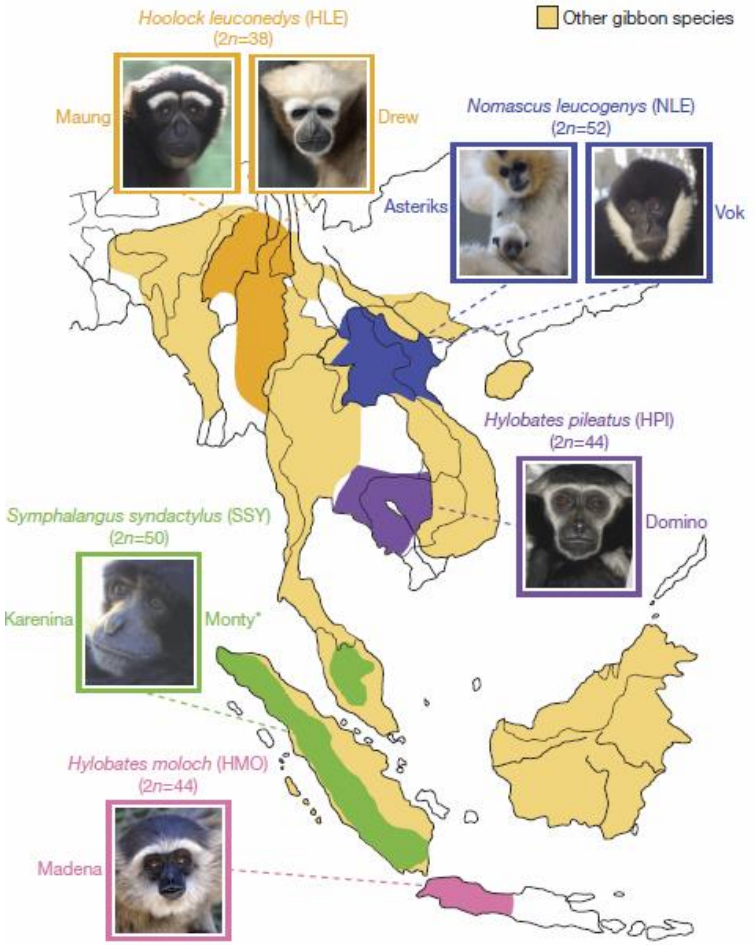

z-expected Heterozygosity
4
2
0
-2
-4
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
X
Chromosome

Brawand et al., 2014 Nature

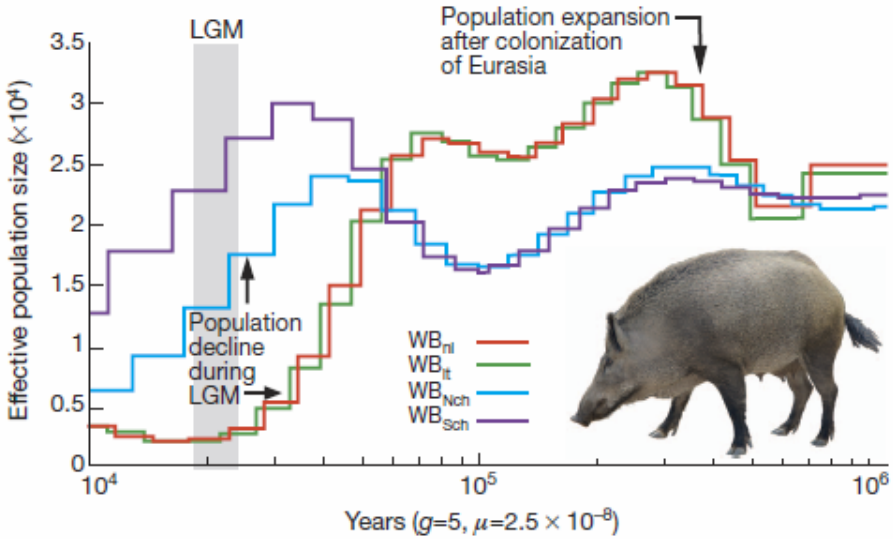


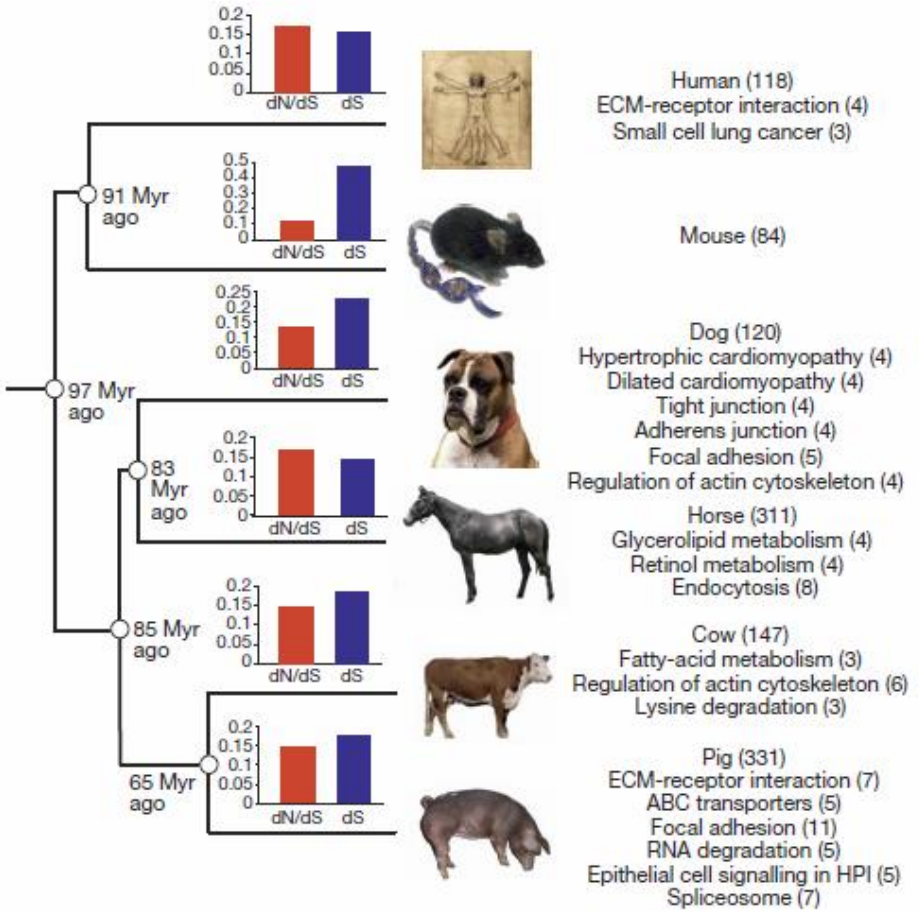
Carbone et al., 2014 Nature





Groenen et al., 2014 Nature

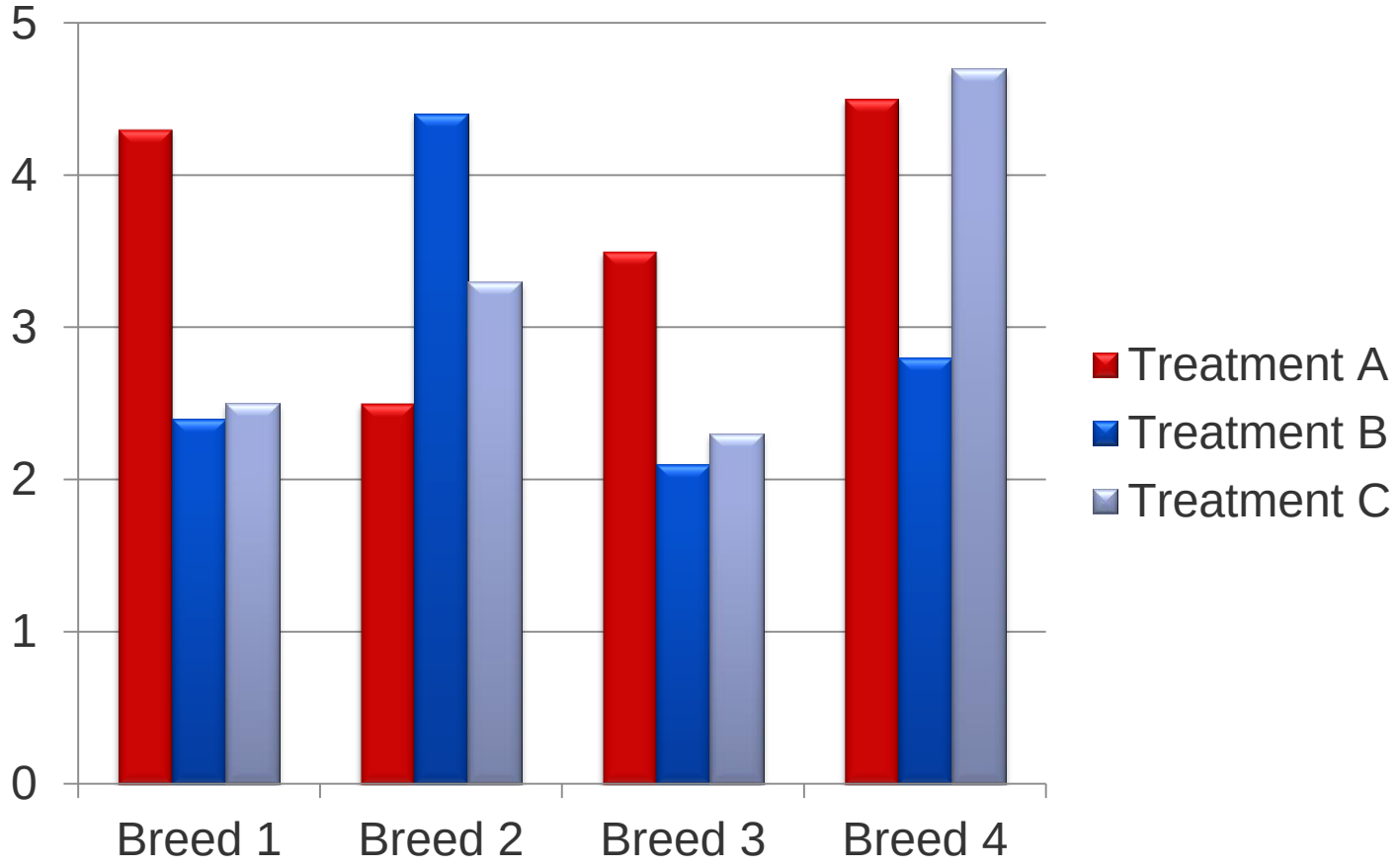




Final remarks

	poor presentation	good presentation
poor science		
good science		

It is worth spending some thought and time on preparing the perfect presentation of your research

Learn the rules so you know how to break them properly
Thank you!

Experiment II: which breeds react similar to the three treatments?