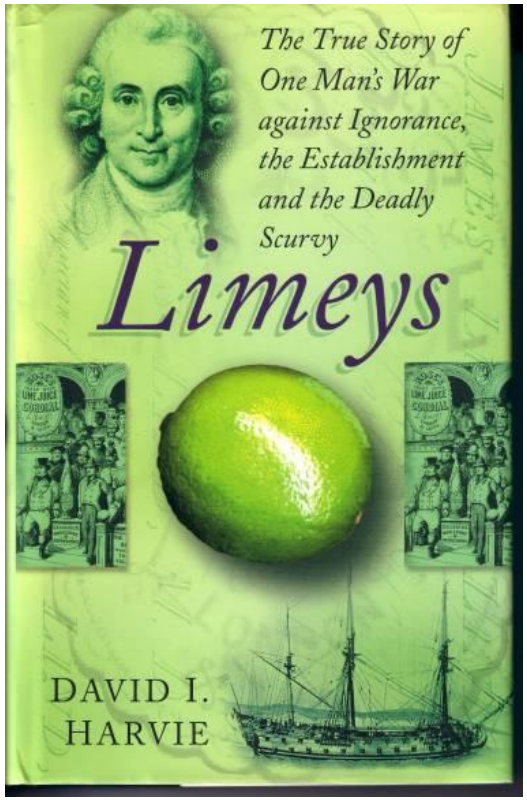
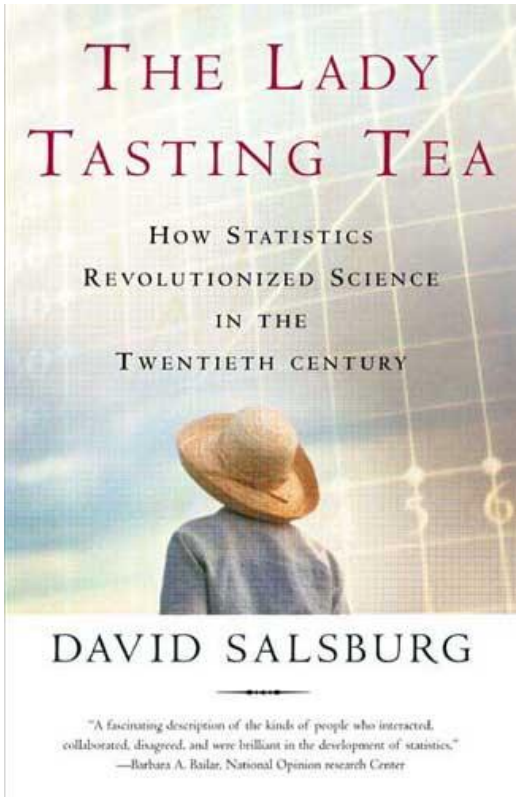
Tom JM Van Dooren
CNRS CR1 researcher

team "Phenotypic Variability and Adaptation" VPA
Institute for Ecology and Environmental Sciences Paris

vandoore@biologie.ens.fr
tomvandooren.eu

- ♠ adaptive dynamics modelling
- ♠ life history evolution and statistical analysis - phenomics
- ♠ annual killifish eco-evo-devo
- ♠ comparative analysis of life history evolution
- ♠ evolution and climate change





The history of experimental design

HOME MY UNIVERSITY

UNIVERSITY OF TWENTE.50

PROSPECTIVE STUDENTS EDUCATION RESEA

Home > Organization > honorary doctorates

UT AWARDS HONORARY DOCTORATES TO MYTHBUSTERS, TWO LEADING SCIENTISTS AND A BUSINESS PIONEER

UNIVERSITY OF TWENTE CELEBRATES 50TH DIES NATALIS IN NOVEMBER

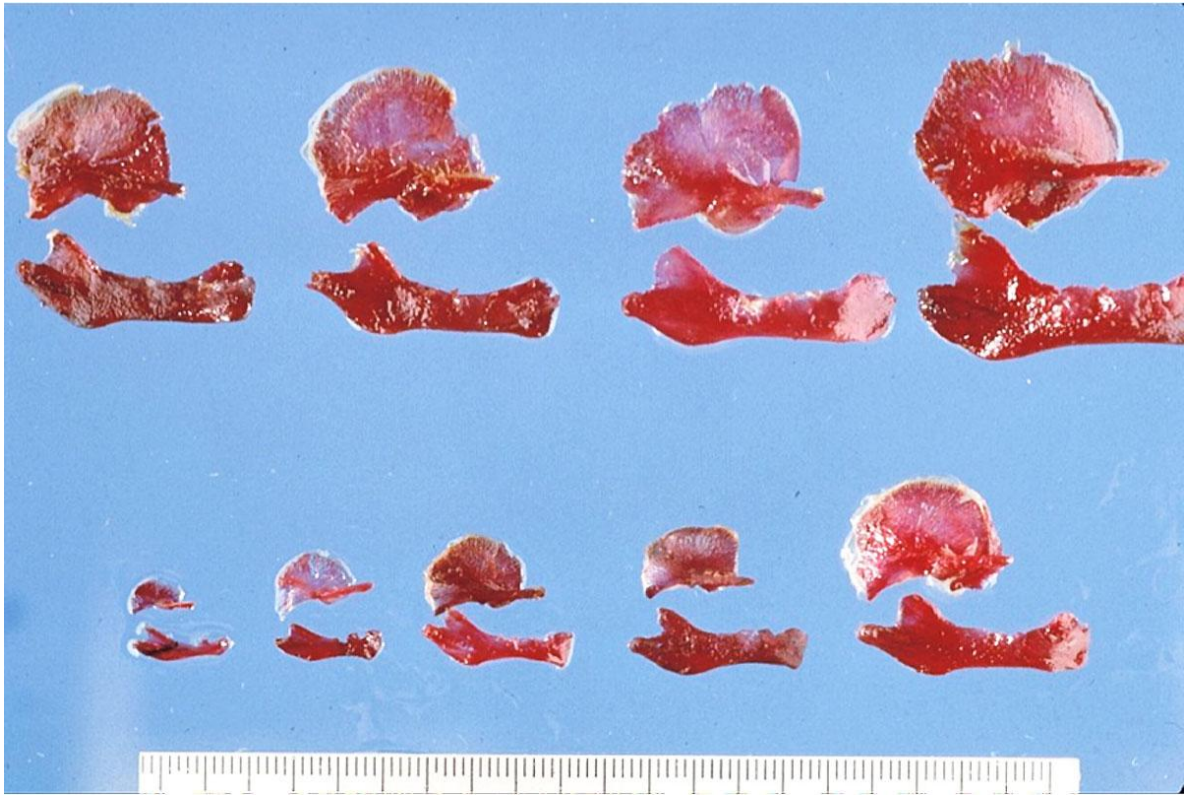
Wednesday, June 29, 2011

At the celebration of its 50th *Dies Natalis*, the anniversary of its Foundation Day in late November of this year, the University of Twente will be awarding honorary doctorates to the leading scientists Prof. Wolfgang Knoll and Prof. Helga Nowotny as well as the inspirational entrepreneur Henri Termeer. For their role in popularising science and technology, a honorary doctorate will also be awarded to the creators of the Discovery Channel's MythBusters, Jamie Hyneman and Adam Savage.

Falsify theories

Design: Randomization - Replication - Factorial combinations

Measure phenotypes

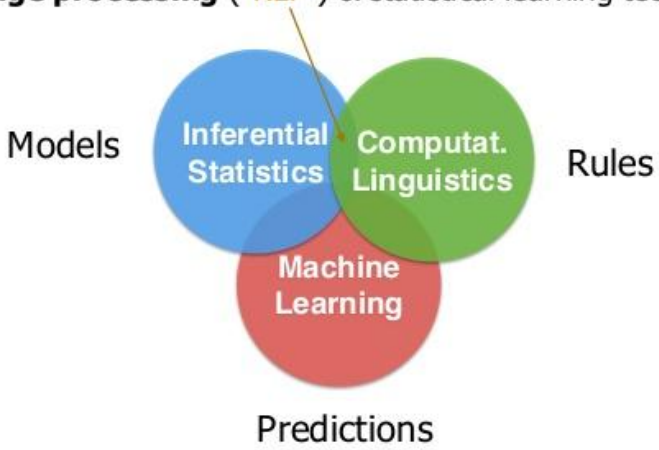


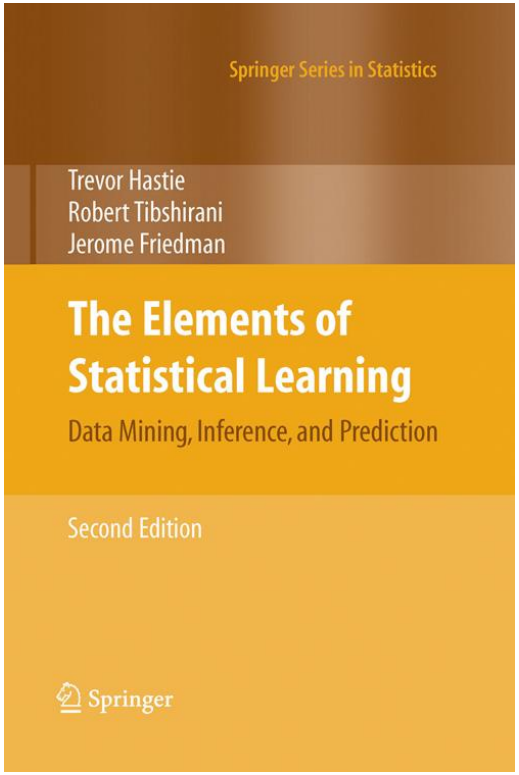
Human fetal temporal bone and mandible

https://embryology.med.unsw.edu.au/embryology/

"Text Mining" or "Text Analytics"

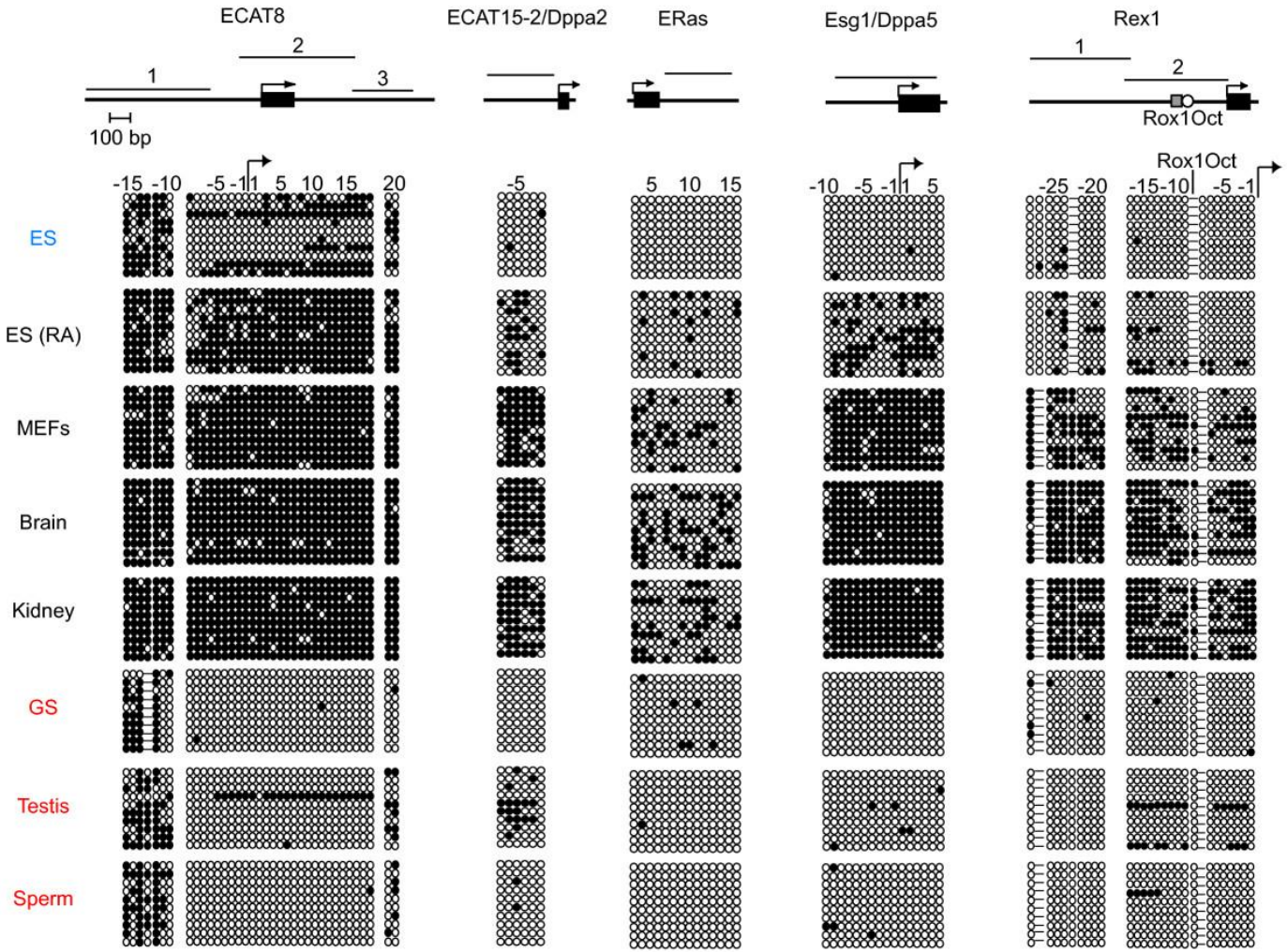
The discovery of {new or existing} facts by applying **natural language processing** ("NLP") & statistical learning techniques.





Discovery and prediction

Measure phenotypes



DNA methylation states of ECATs expressed in GS cells.Imamura *et al. BMC Developmental Biology* 2006 **6**:34 doi:10.1186/1471-213X-6-34

Philosophy of analysis:

o) Formulate the scientific question as sharp as possible, making it answerable

a) Experimental design – Planned observations

b) Reflect on the data generating mechanism

↓

Collect data

↓

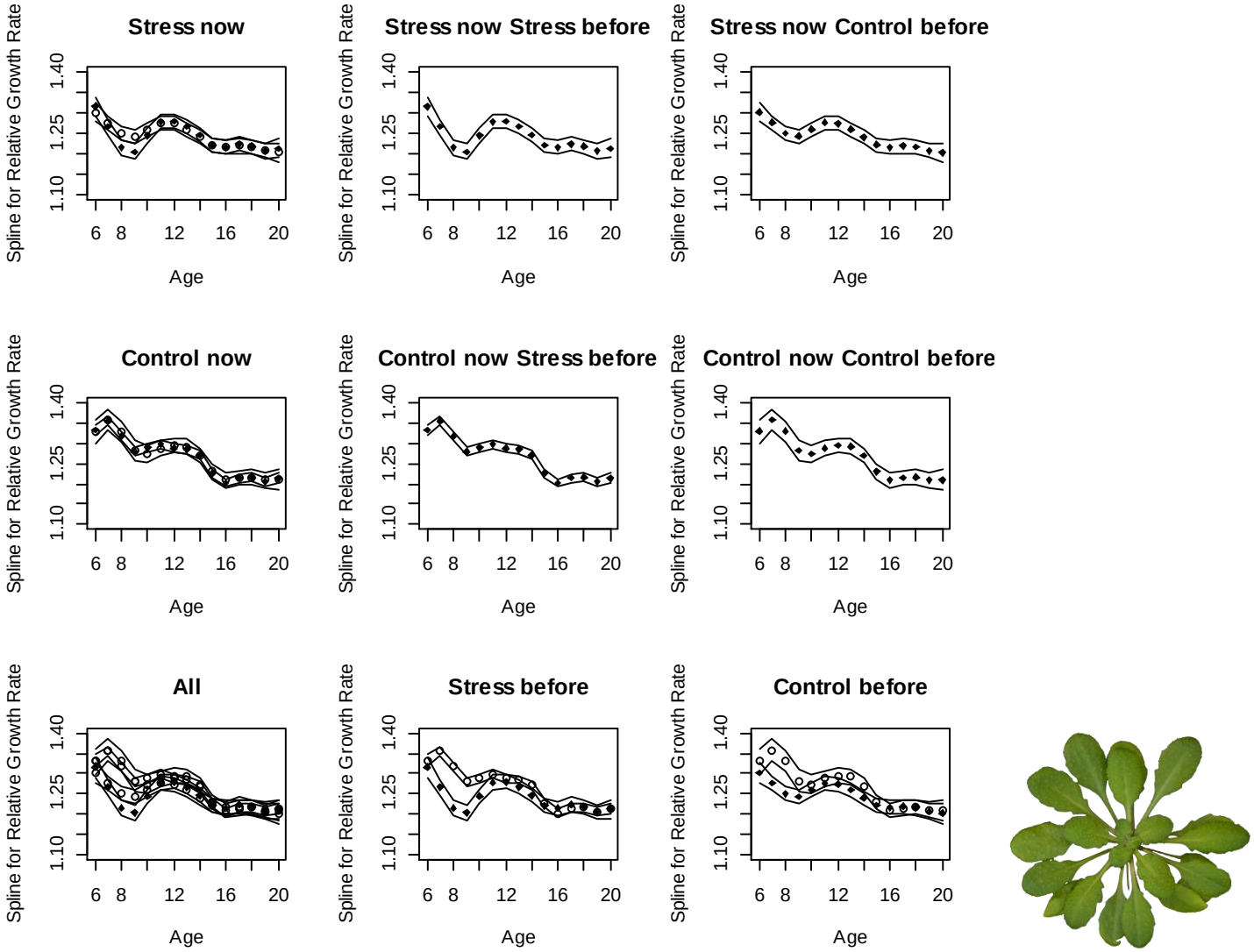
Inspection - Summary statistics - Graphics

↓

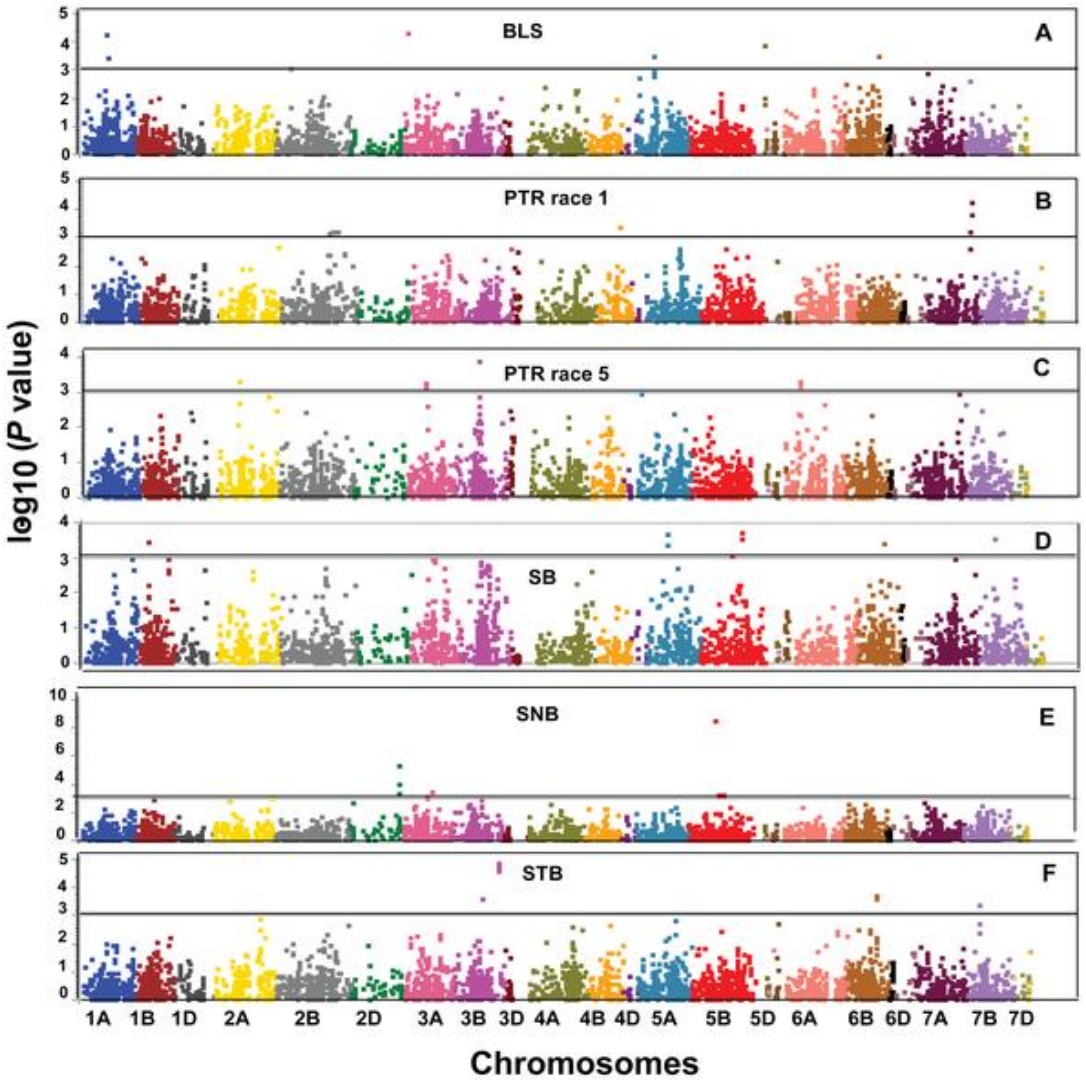
Model fitting – Model selection - Model validation

↓↑

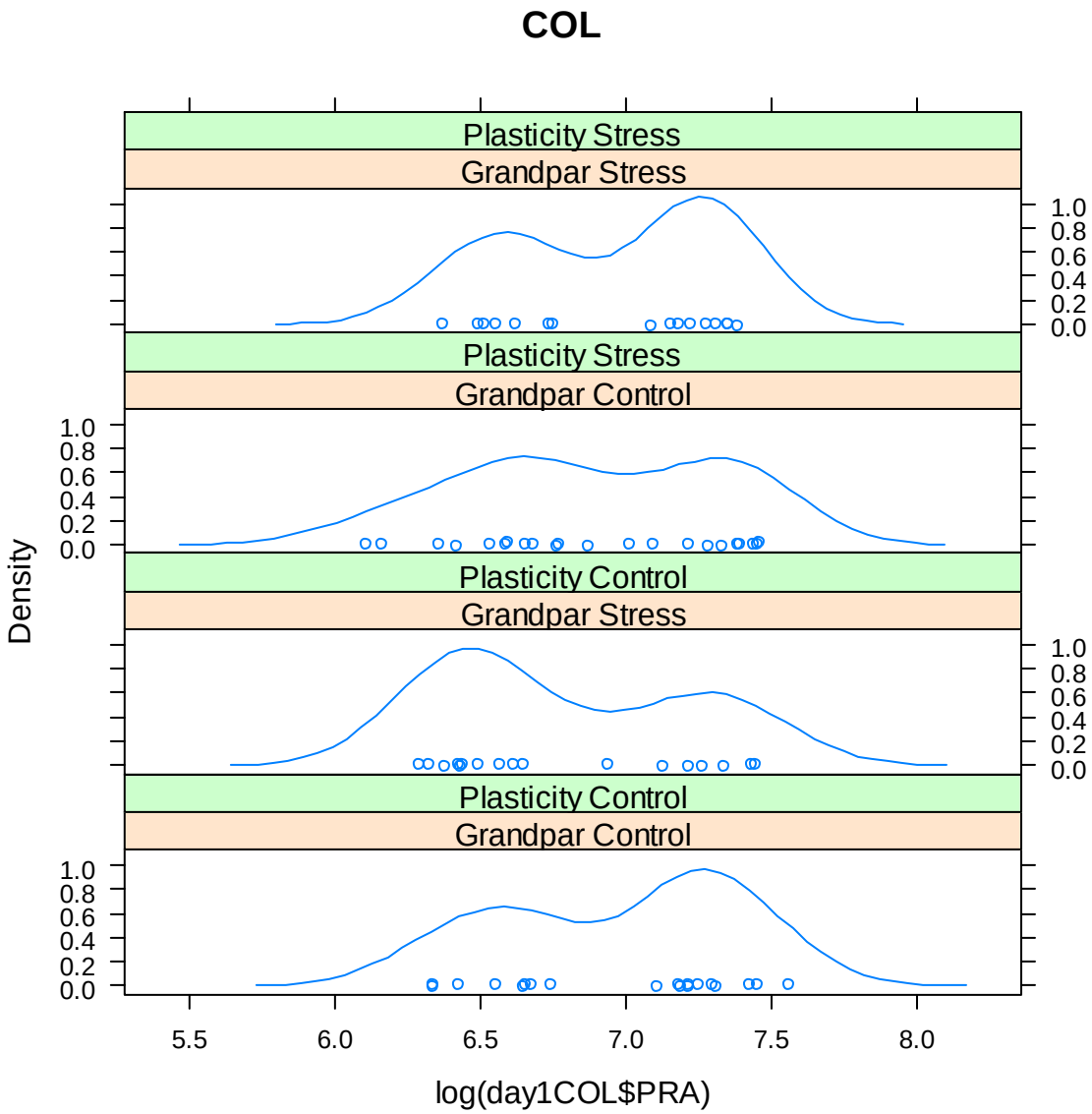
Inference



Arabidopsis plasticity experiment

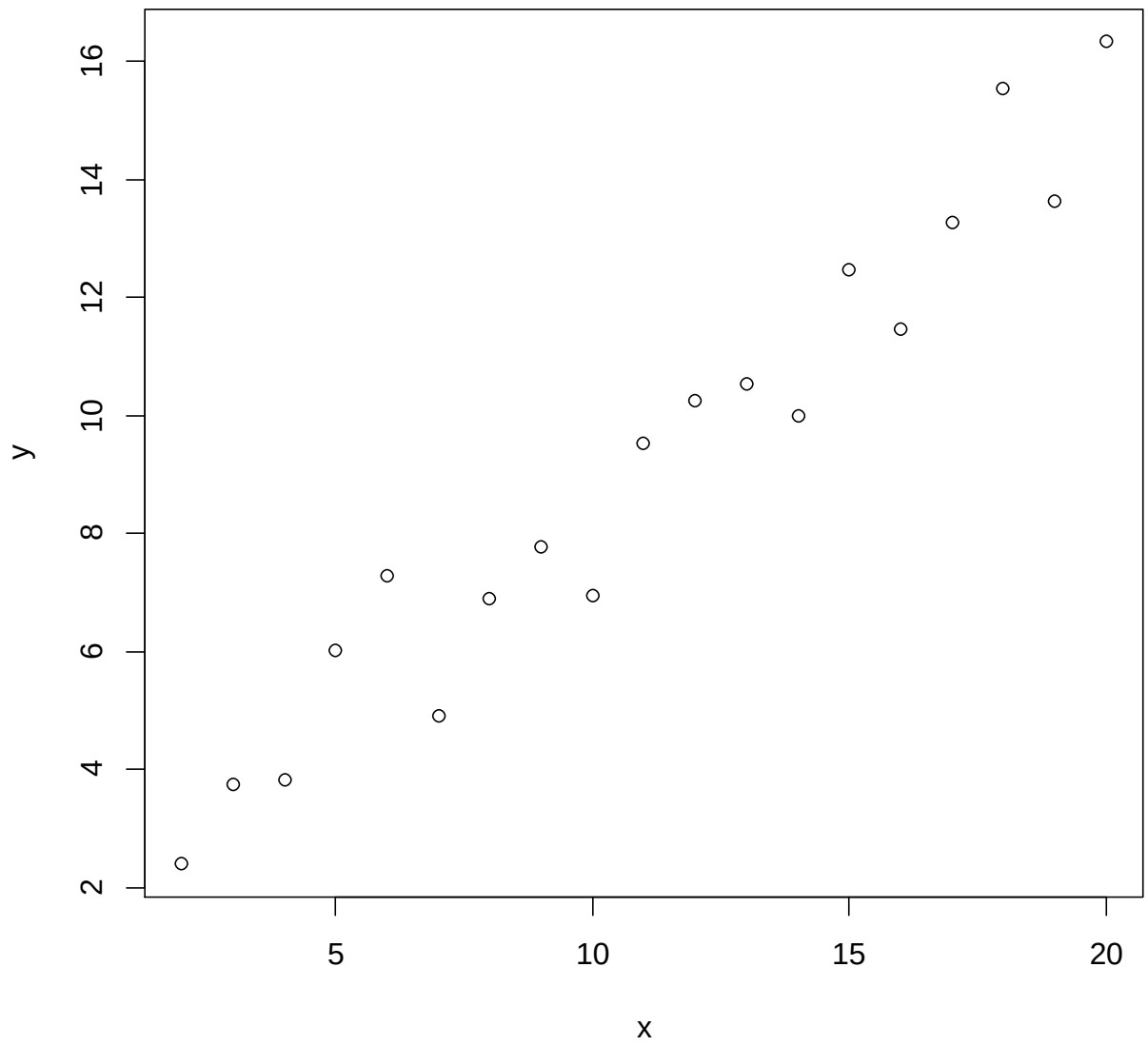


Gurung S, Mamidi S, Bonman JM, Xiong M, et al. (2014) Genome-Wide Association Study Reveals Novel Quantitative Trait Loci Associated with Resistance to Multiple Leaf Spot Diseases of Spring Wheat. PLoS ONE 9(9): e108179.



Initial size distributions

Arabidopsis plasticity experiment


y
2
4
6
8
10
12
14
16
5
10
15
20
x

In this course:

When the *y* (response variable) is from a distribution that is constrained (for example positive or 0-1)
and we still want to use a linear style of model

When we have many explanatory variables

When there are explanatory variables that we didn't measure, but we can see their effects

When we have to carry out many (**MANY**) similar tests

If we want to test the effects of one variable on another, but we need to take many other variables into account

When our data points are not independent measurements

Course structure: Repetitions with variations

	Ecology	**Transcriptomics**	**QTL SNP GWAS**
Response is constrained	TVD	MLMM	
Many explanatory variables	TVD		
Hidden explanatory variables	TVD	MLMM	TMH
Model selection	TVD	MLMM	TMH
Many similar tests		MLMM	TMH
	Glm **Mixed models**	**Lm** **Glm** **Mixtures**	**Lm** **Mixed models**

-

Tom VAN DOOREN

Introduction to R
Probability Distributions
Generalized Linear Models
Inference - Model selection
Mixed Models - Random Effects

Marie-Laure MARTIN-MAGNIETTE

Model-based Clustering: Independent mixture models
Model-based Clustering: Hidden Markov models

Tristan MARY-HUARD

Association studies in Genetics: monogenic and oligogenic models
Association studies in Genetics: polygenic and GWAS models

For me:

- lectures with embedded R exercises
  Try these while I talk, or during exercise time

- Exercises with simulated data, data from books/articles

- Question/Discussion time

During/after course pdf files and scripts available from

tomvandooren.eu/ M1Course2016.html

Helpful background reading:

Michael Crawley's R book

Yudi Pawitan "In All Likelihood"

John Fox mixed models regression models

Doug Bates unpublished book on mixed models

Several wiki's with overviews of methods; R help files

http://www.jstatsoft.org/

♠ French - English dictionary:

http://biol09.biol.umontreal.ca/legendre/Lexique.pdf

Everybody will analyse a dataset for this part of the course.

Search dryad for "generalized linear models", "mixed models"

http://datadryad.org/ → DONE

Choose a dataset.

Work alone or in small groups (<4) on the same paper.

Hand in scripts and short report by January 27, 2016.

Use the exercise time to ask questions and show progress.