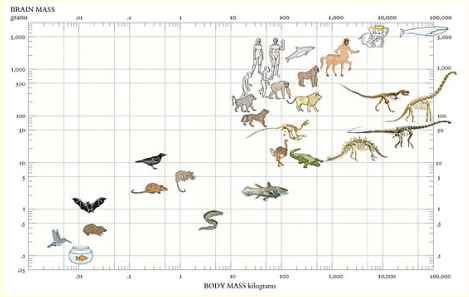
Advancing in R

Intro & Workflow



Learning outcomes

- Improve usefulness of R and RStudio
- Adopt a systematic workflow template
- Improve theoretical knowledge of linear and linearized models
- Appreciate role of coefficients as effect size estimates
- Practice information theory approach to model simplification & selection
- Learn best practice approach to more complex model construction

Course guidelines for using R:

- Always work in an RStudio project
- Avoid typing in the console
- Avoid using pulldown menus
- Avoid attaching data
- Annotate your source file
- Adhere to workflow order (i.e., always plot effects before modelling them)

Suggested workflow

1) Assess structure & quality control
2) Visualize
   a) Distributions of response & predictor(s)?
   b) Collinearity?
   c) Effects? (ADF & predict coefficients)
   d) Possible alternative explanations?
3) Analyses
   a) Model quality (assumptions & fit)
   b) Simplification & selection (if appropriate)
4) Better plots

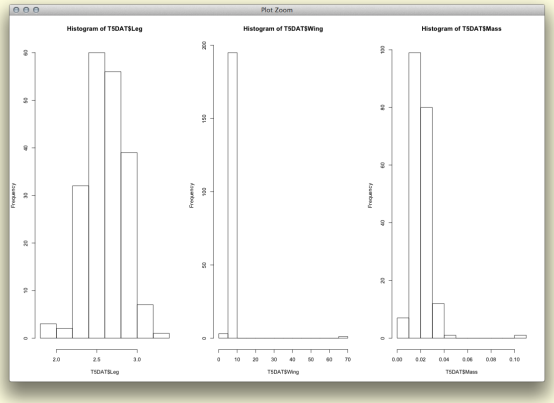
The ‘ADF’ method

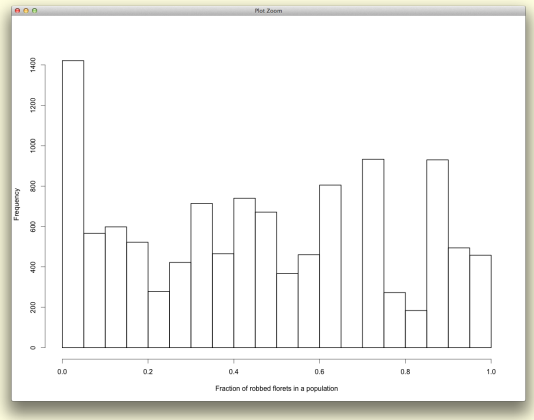
Prof. Glenn Morris
Univ. of Toronto

Any Damned Fool can see the effect

Suggested workflow

1) Assess structure & quality control
2) Visualize
   a) Distributions of response & predictor(s)?
   b) Collinearity?
   c) Effects? (ADF & predict coefficients)
   d) Possible alternative explanations?
3) Analyses
   a) Model quality (assumptions & fit)
   b) Simplification & selection (if appropriate)
4) Better plots





```
> set.seed(6)
> VAR2<-rlnorm(15,2,.5)
> boxplot(VAR2)
```

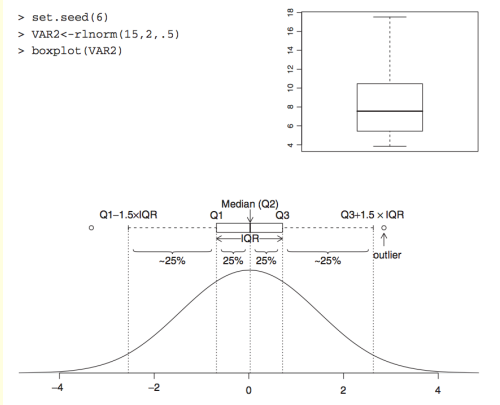
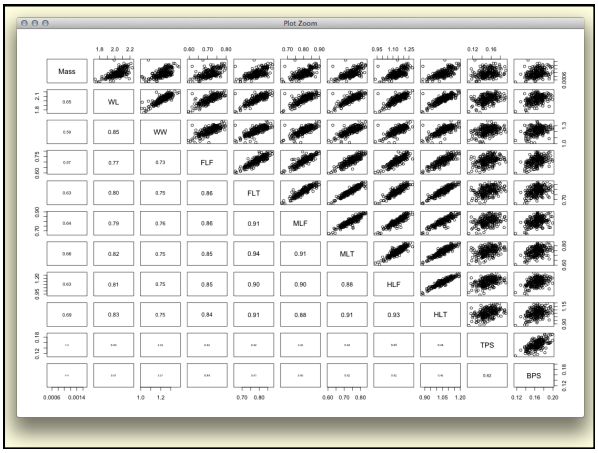
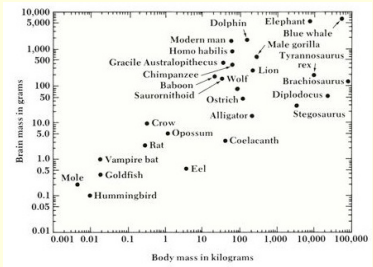


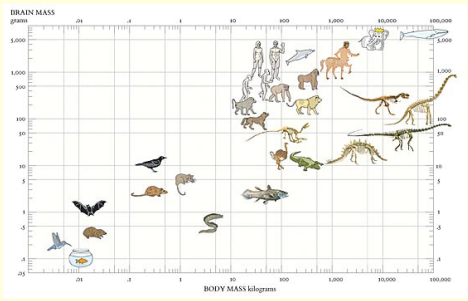
Fig 5.5 Boxplot of a standard normal distribution (mean=0, sd=1).



Visual display of quantitative data

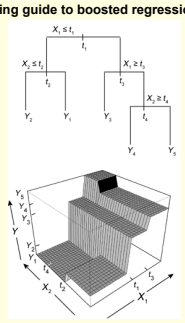


Edward Tufte



The analysis output can be primarily graphical in itself

A working guide to boosted regression trees



Journal of Animal Ecology
Volume 77, Issue 4, pages 802-813, 8 APR 2008 DOI: 10.1111/j.1365-2656.2008.01390.x
http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2656.2008.01390.x/full#f1

Suggested workflow

1) Assess structure & quality control
2) Visualize
   a) Distributions of response & predictor(s)?
   b) Collinearity?
   c) Effects? (ADF & predict coefficients)
   d) Possible alternative explanations?
3) Analyses
   a) Model quality (assumptions & fit)
   b) Simplification & selection (if appropriate)
4) Better plots