

# Modelação Ecológica

## AULA 8

09 November 2019 – 14:00-16:30 – room 2.3.16

Tiago A. Marques

# CORRELATION IS NOT CAUSATION



The invalid assumption that correlation implies cause is probably among the two or three most serious and common errors of human reasoning.

— *Stephen Jay Gould* —

AZ QUOTES

# Spurious correlations



Now a ridiculous book!

- Spurious charts
- Fascinating factoids
- Commentary in the footnotes

Amazon | Barnes & Noble | Indie Bound

<http://tylervigen.com/spurious-correlations>

**W**e all know the truism “Correlation doesn’t imply causation,” but when we see lines sloping together, bars rising together, or points on a scatterplot clustering, the data practically begs us to assign a reason. We want to believe one exists.

Statistically we can’t make that leap, however. Charts that show a close correlation are often relying on a visual parlor trick to imply a relationship.

Tyler Vigen, a JD student at Harvard Law School and the author of *Spurious Correlations*, has made sport of this on his website, which charts farcical correlations—for example, between U.S. per capita margarine consumption and the divorce rate in Maine.

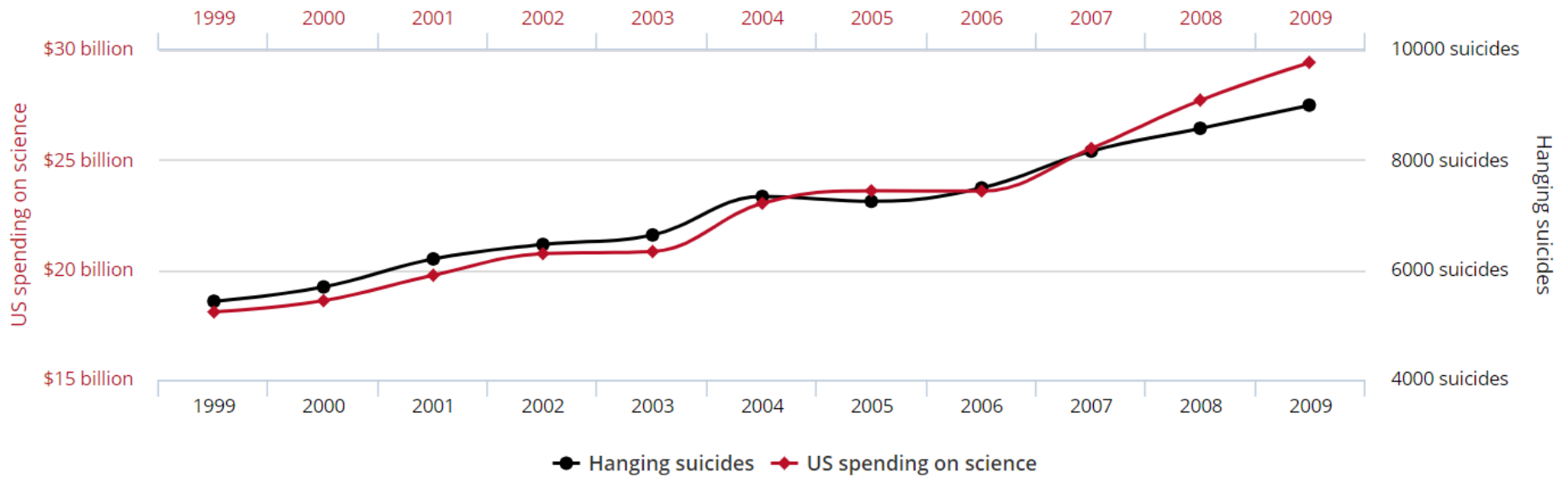
Vigen has programmed his site so that anyone can find and chart absurd correlations in large data sets.



# US spending on science, space, and technology correlates with Suicides by hanging, strangulation and suffocation



Correlation: 99.79% (r=0.99789126)



tylervigen.com

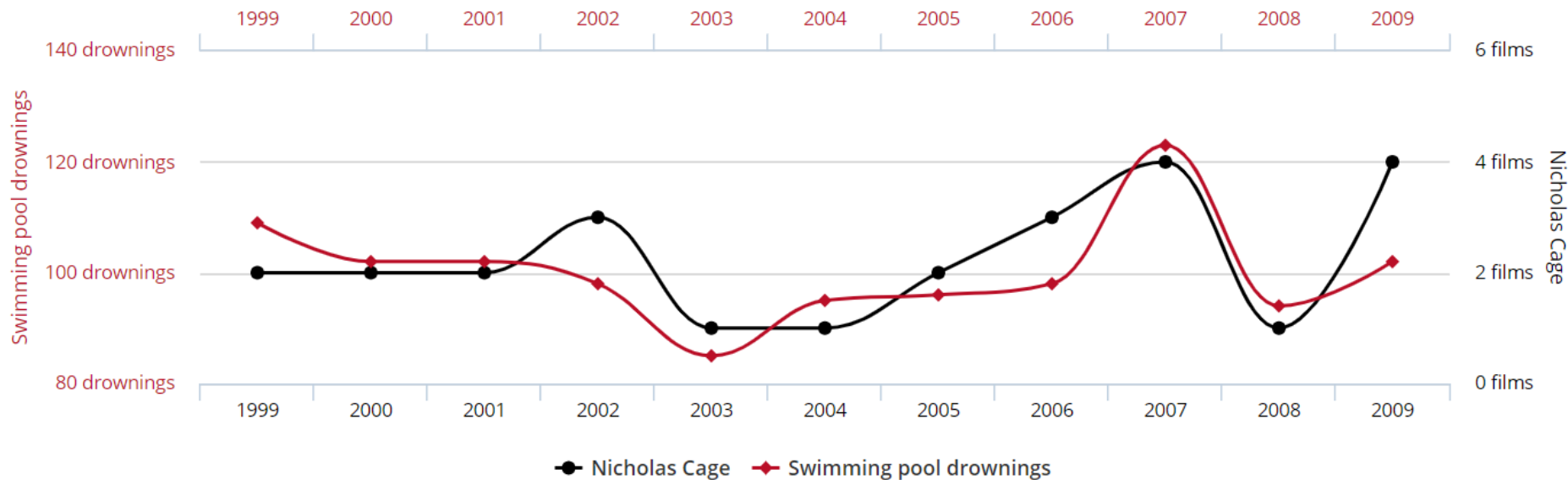
Data sources: U.S. Office of Management and Budget and Centers for Disease Control & Prevention



# Number of people who drowned by falling into a pool correlates with Films Nicolas Cage appeared in



Correlation: 66.6% (r=0.666004)



tylervigen.com

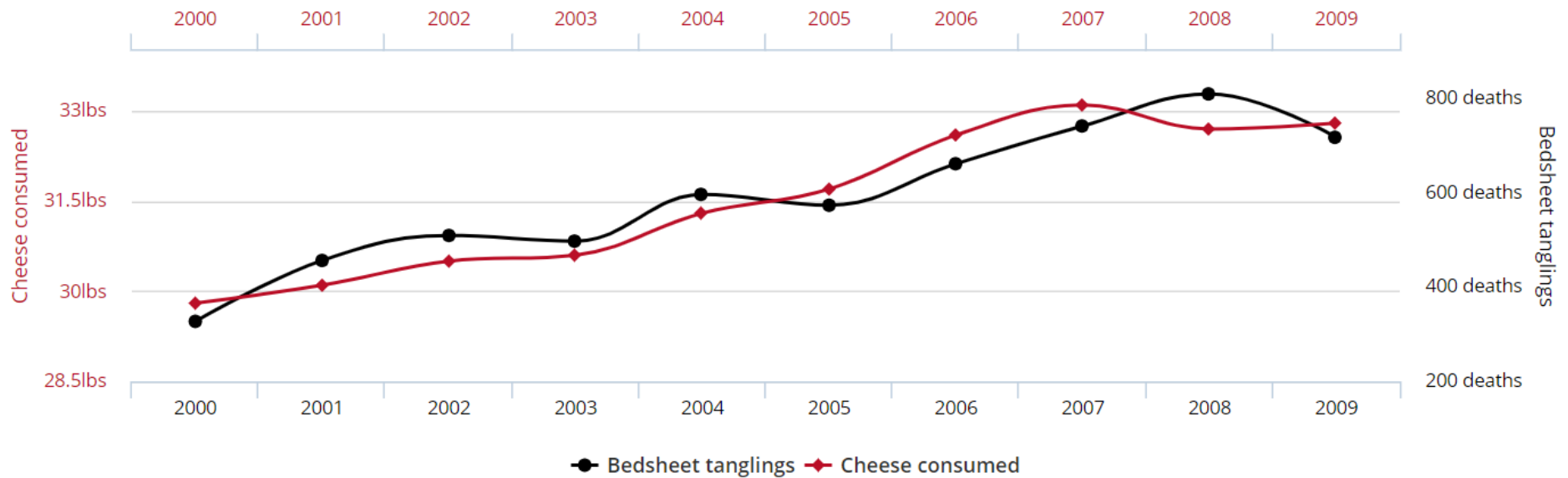


## Per capita cheese consumption

correlates with

## Number of people who died by becoming tangled in their bedsheets

Correlation: 94.71% ( $r=0.947091$ )



tylervigen.com

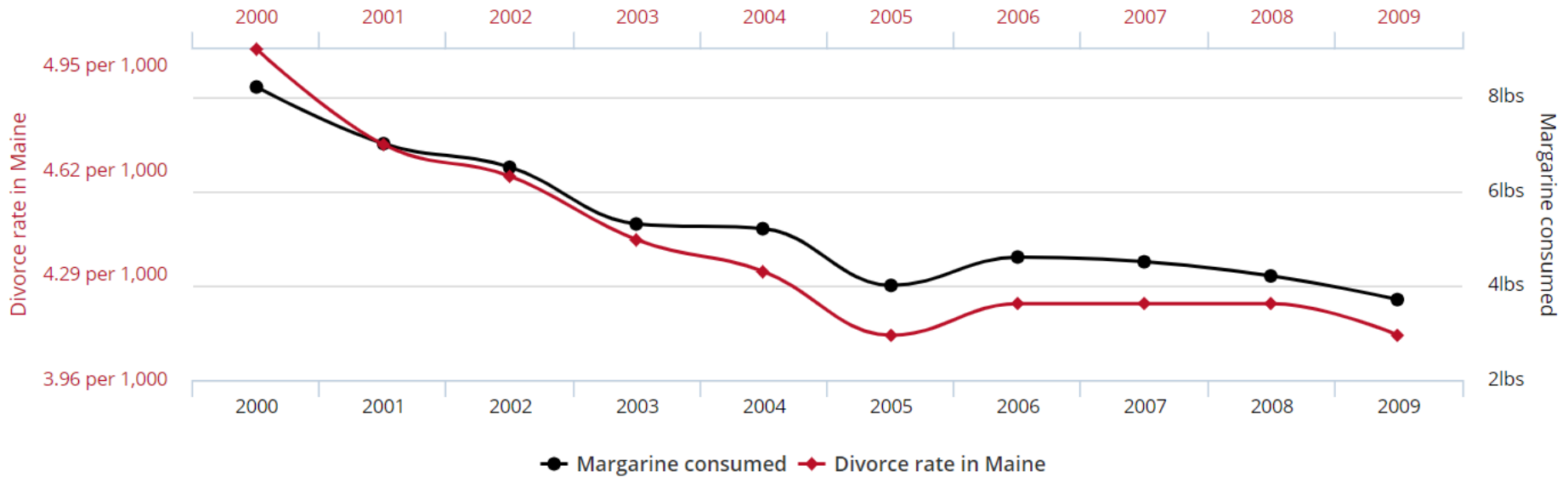
Data sources: U.S. Department of Agriculture and Centers for Disease Control & Prevention



## Divorce rate in Maine correlates with Per capita consumption of margarine



Correlation: 99.26% (r=0.992558)



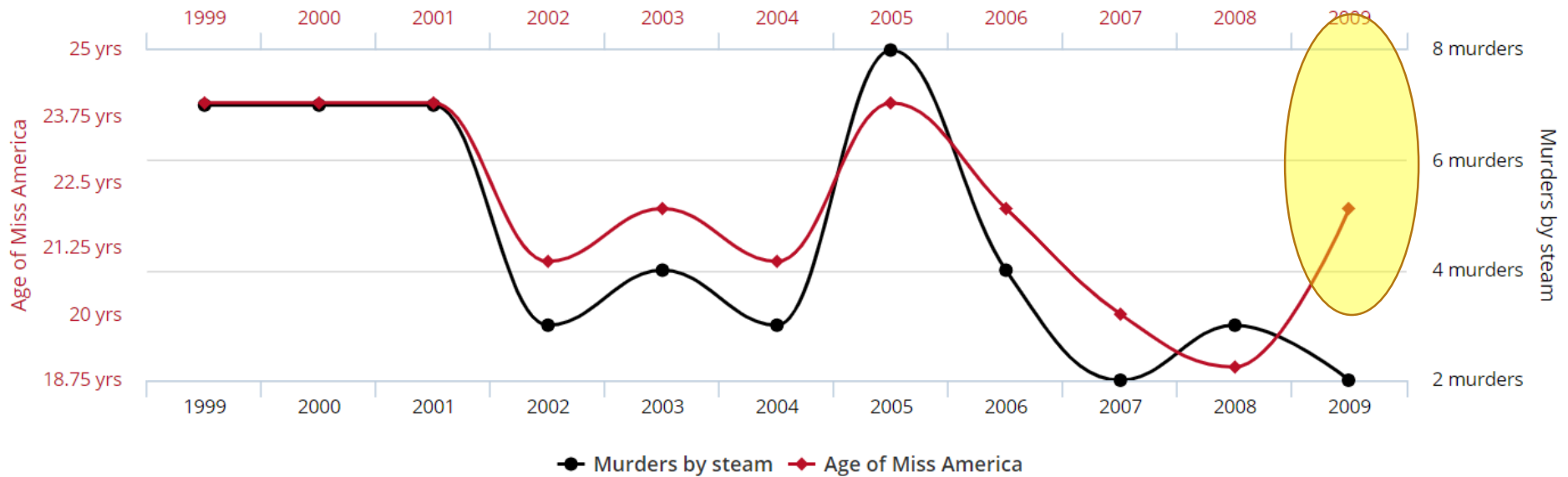




# Age of Miss America correlates with Murders by steam, hot vapours and hot objects



Correlation: 87.01% ( $r=0.870127$ )



tylervigen.com

Data sources: Wikipedia and Centers for Disease Control & Prevention

Remember Slide 6, aula 2 ? – correlações espúrias!

# Exercise solutions

## Gestão de Páginas

- ▼ Modelação Ecológica
  - Modelação Ecológica(Ecologia Marinha)
  - Modelação Ecológica(Ecologia e Gestão Ambiental)
- ▼ Aulas
  - ▼ Aula1
    - PDFs
  - ▼ Aula2
    - An extra set of exercises to practice R - (stolen from numerical
  - ▼ Aula3
  - ▼ Aula4
    - A task about merge
  - ▼ Aula5
    - FT1
    - FT2
  - ▼ Aula6
    - Simulating regression data
  - Aula7 08 10 2019
  - Aula8 09 10 2019
  - Aula9
- ▼ Outros Recursos
  - ▼ PDFs
    - Ecopath
    - Jorgensen&Fath
    - Developments In Ecological Models
  - R Cheat Sheets
  - Propostas de resolução de fichas de trabalho

## Propostas de resolução de fichas de trabalho

Página **Ficheiros 3** Permissões Link

Adicionar Ficheiro

#	Nome
1	Solution to the final task challenge on Tutorial 1 <i>Fina/Task.pdf</i>
2	Solution to the extra set of exercises in Aula 2 <i>Ficha_de_trabalho_1.pdf</i>
3	Solution to FT2 - aula 5 <i>MEFT2.pdf</i>

# Support material on R and statistical models

## Gestão de Páginas

- Modelação Ecológica
  - Modelação Ecológica(Ecologia Marinha)
  - Modelação Ecológica(Ecologia e Gestão Ambiental)
- Aulas
  - Aula1
  - Aula2
  - Aula3
  - Aula4
  - Aula5
  - Aula6
  - Aula7 08 10 2019
  - Aula8 09 10 2019
  - Aula9
- Outros Recursos
  - PDFs**
  - Ecopath
  - Jorgensen&Fath
  - Developments In Ecological Models
  - R Cheat Sheets

+ Criar

## PDFs

Página Ficheiros 9 Permissões Link

Adicionar Ficheiro

#	Nome
1	Modelling ecological systems in a changing world <i>Evans2012.pdf</i>
2	Norberg_et_al-2019 A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels <i>Norberg_et_al-2019-Ecological_Monographs.pdf</i>
3	The importance of stupidity in scientific research <i>Schwartz2008.pdf</i>
4	Ecological Models and Data in R <i>Boiker2007.pdf</i>
5	Numerical Ecology with R <i>Borcardetal2001EcologyUseR.pdf</i>
6	Introduction to Probability and Statistics Using R <i>IPSUR.pdf</i>
7	A Beginner's Guide to R <i>Zuuretal2000useR.pdf</i>
8	Analyzing Ecological Data <i>zuur_2007.pdf</i>
9	Mixed Effects Models And Extensions In Ecology With R <i>Zuur_Mixed-effects-models-and-extensions-in-ecology-with-R.pdf</i>

# R Cheat Sheets

## Gestão de Páginas

- ▼ Modelação Ecológica
  - Modelação Ecológica(Ecologia Marinha)
  - Modelação Ecológica(Ecologia e Gestão Ambiental)
- ▼ Aulas
  - ▶ Aula1
    - Aula2
    - Aula3
  - ▶ Aula4
  - ▶ Aula5
  - ▶ Aula6
  - Aula7 08 10 2019
  - Aula8 09 10 2019
  - Aula9
- ▼ Outros Recursos
  - ▶ PDFs
    - R Cheat Sheets

+ Criar

## R Cheat Sheets

Página **Ficheiros 6** Permissões Link

Adicionar Ficheiro

#	Nome	Permissões
1	base-r.pdf	Público
2	cartography.pdf	Público
3	data-import.pdf	Público
4	rstudio-ide.pdf	Público
5	advancedR.pdf	Público
6	rmarkdown-2.0 (1).pdf	Público

# ASSESSMENT ME 2019/2020

- MINI Practical work with own data – 15% (Groups of 3)
- WORK on a model of your choice, theoretical veined, using an example paper (of students choice, but subject to previous agreement from teacher)) – 20 % (groups of 4, 4 pAGEs + ORAL Presentation 10 mins)
- Practical work mecoco – 20 % (groups of 1) (ASIDE: with the POTENCIAL to write a joint paper, not for the assessment, but certainly a plus)
- PRACTICAL MODELLING WORK (Own data or data from others) – 45 % (groups of 4, 4 pages, Oral Presentation 10 mins)

# TRABALHO PRÁTICO – MODELAÇÃO (45%)

- Ideias:
  - recolher dados de propósito
  - usar dados recolhidos em outras cadeiras
  - falar com professores que possam ter dados interessantes
  - procurar dados na bibliografia
  - fazer um trabalho só com dados simulados,
  - etc...
- Fundamental:
  - Formular uma pergunta ecológica, e que possa ser respondida com os dados,
  - (E não o contrário) tenho uns dados, que pergunta posso responder com eles

# Guião do Trabalho Prático

## Gestão de Páginas

Importa

- ▼ Modelação Ecológica
  - Modelação Ecológica(Biologia da Conse
  - Modelação Ecológica(Ecologia Marinha)
- ▼ Teóricas
  - Aula 1
  - Aula 2
  - Aula 3
  - Aula 4
  - Aula 5
  - Aula 6
  - Aula 7
  - Aula 8
- ▼ Teorico-Práticas
  - ▶ Aula 1
  - Aula 2
  - Aula 3
- ▼ **Material Suplementar**
  - R Cheat Sheets
  - Comunicação por email
  - ▶ Referências Bibliográficas
  - TPC Aula 1

## Material Suplementar

Página **Ficheiros 2** Permissões Link

Adicionar Ficheiro

#	Nome	Permissões
1	Exemplo de um documento obtido por um relatório dinâmico <i>alunos2018MEco.pdf</i>	Professores de Modelação Ecológica (MEco) 1 Semestre - 2018/2019 ou Alunos de Modelação Ecológica (MEco) 1 Semestre - 2018/2019
2	GuiãoTrabalhoPratico.pdf <i>GuiaoTrabalhoPratico.pdf</i>	Professores de Modelação Ecológica (MEco) 1 Semestre - 2018/2019 ou Alunos de Modelação Ecológica (MEco) 1 Semestre - 2018/2019

Em breve!!!

SUBMIT WORK TO CONGRESS?



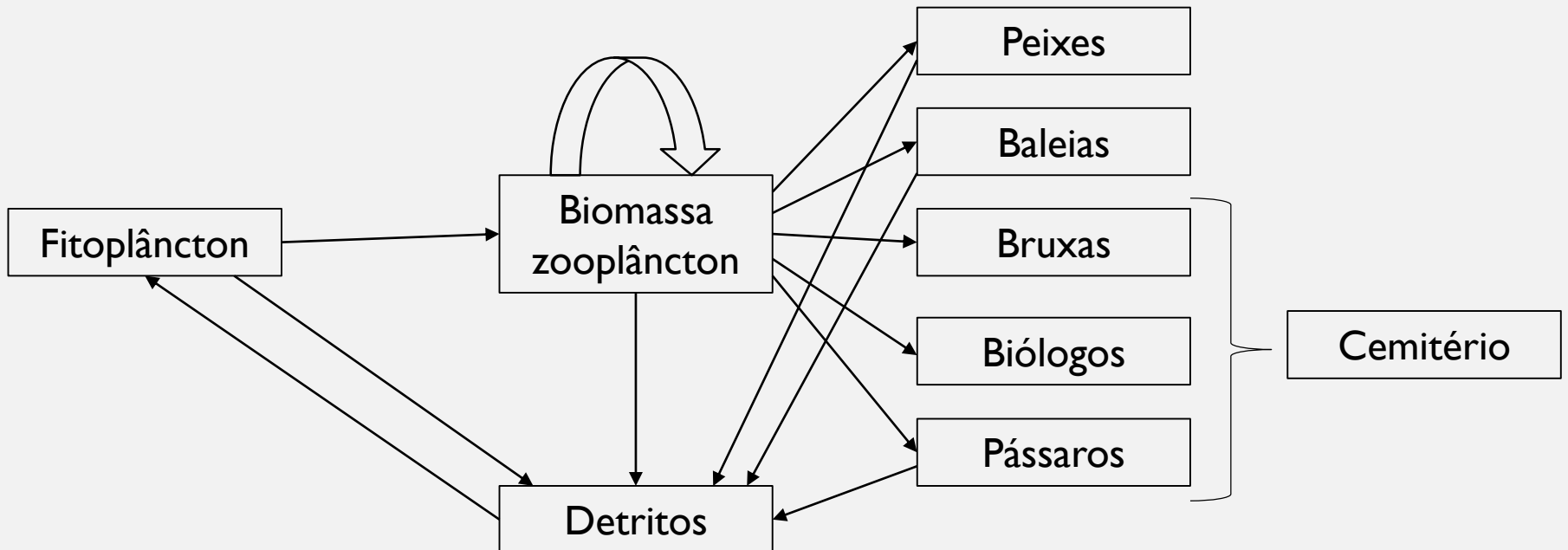
1. Elabore esquemas conceptuais de modelos que representem as seguintes situações:

a. Biomassa de zooplâncton numa zona costeira



1. Elabore esquemas conceptuais de modelos que representem as seguintes situações:

a. Biomassa de zooplâncton numa zona costeira



1. The same system can be represented by a simple model or a very complex model.
2. The scale and detail must be chosen as a function of the objectives of the model.
3. What needs to be accounted for, or not, depends solely on what are the inferences to be made.

1. Elabore esquemas conceptuais de modelos que representem as seguintes situações:

a. Biomassa de zooplâncton numa zona costeira

b. Distribuição de uma espécie de lapa numa zona costeira rochosa

c. Produção de uma unidade de aquacultura em regime intensivo

d. Biomassa de peixes de uma unidade de aquacultura em regime intensivo

e. Funcionamento de um estuário

f. Capturas da pesca de determinado segmento da frota

g. Rendimento dos pescadores de determinado segmento de frota



# Model selection in regression

We often want to select a subset of available covariates to use in a regression model (those more important/relevant to explain the response) and hence obtain a simplified model.

The regression coefficients of the simplified models are different from those of the original model (saturated model, if the model with all variables)

Framework:

- statistical tests
- Information criteria

Method:

- Forward selection
- Backward elimination
- Stepwise selection
- Test all combinations



## Regression e GLM

---

### Some additional topics about regression

- “Dummy” variables (used to code factors, R does it for you! – but is key to understand it – interpreting model coefficients depends on it)
- Interaction between independent variables (multicollinearity – one should only consider variable ecologically relevant and remove those highly correlated, especially those less ecologically relevant)
- Other fitting approaches not MSM (Maximum likelihood, REML)
- Non-linear models

```
#Making up an ANOVA
```

```
#An ANOVA
```

```
set.seed(1234)
```

```
#define sample size
```

```
n=200
```

```
#define treatments
```

```
tr=c("a","b","c","d")
```

```
#how many treatments
```

```
ntr=length(tr)
```

```
#balanced design
```

```
n.by.tr=n/ntr
```

```
#generate some data – not used here!!!
```

```
xs=runif(n,10,20)
```

```
type=as.factor(rep(tr,each=n.by.tr))
```

```
#get colors for plotting
```

```
cores=rep(1:ntr,each=50)
```

```
#define 4 means
```

```
ms=c(1.7,2.6,4.3,5.3)
```

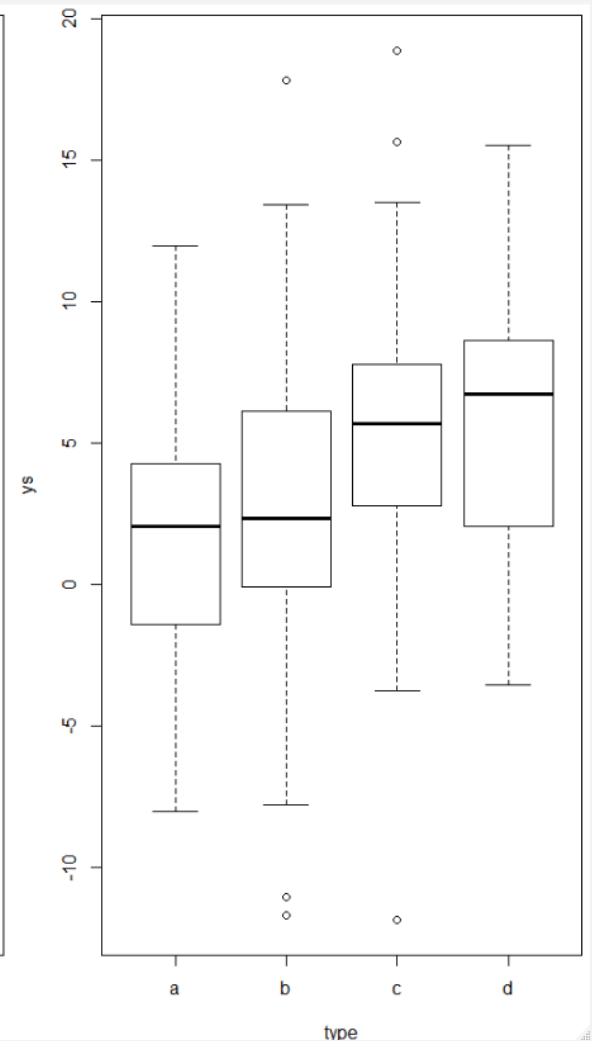
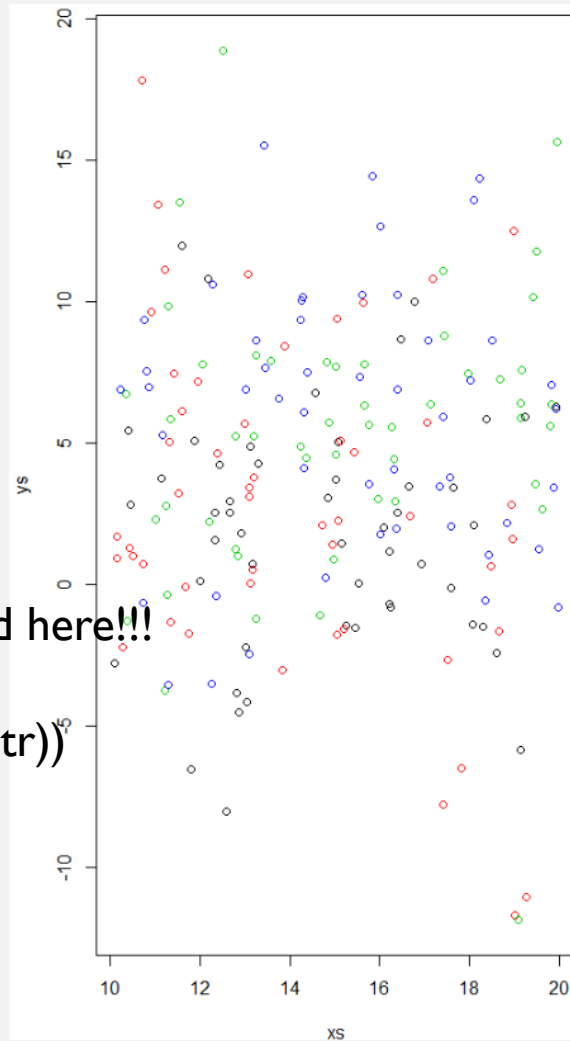
```
#ys, not a function of the xs!!!
```

```
ys=ifelse(type=="a",ms[1],ifelse(type=="b",ms[2],ifelse(type=="c",ms[3],ms[4]))) + rnorm(200,0,5)
```

```
par(mfrow=c(1,2),mar=c(4,4,0.5,0.5))
```

```
plot(xs,ys,col=cores)
```

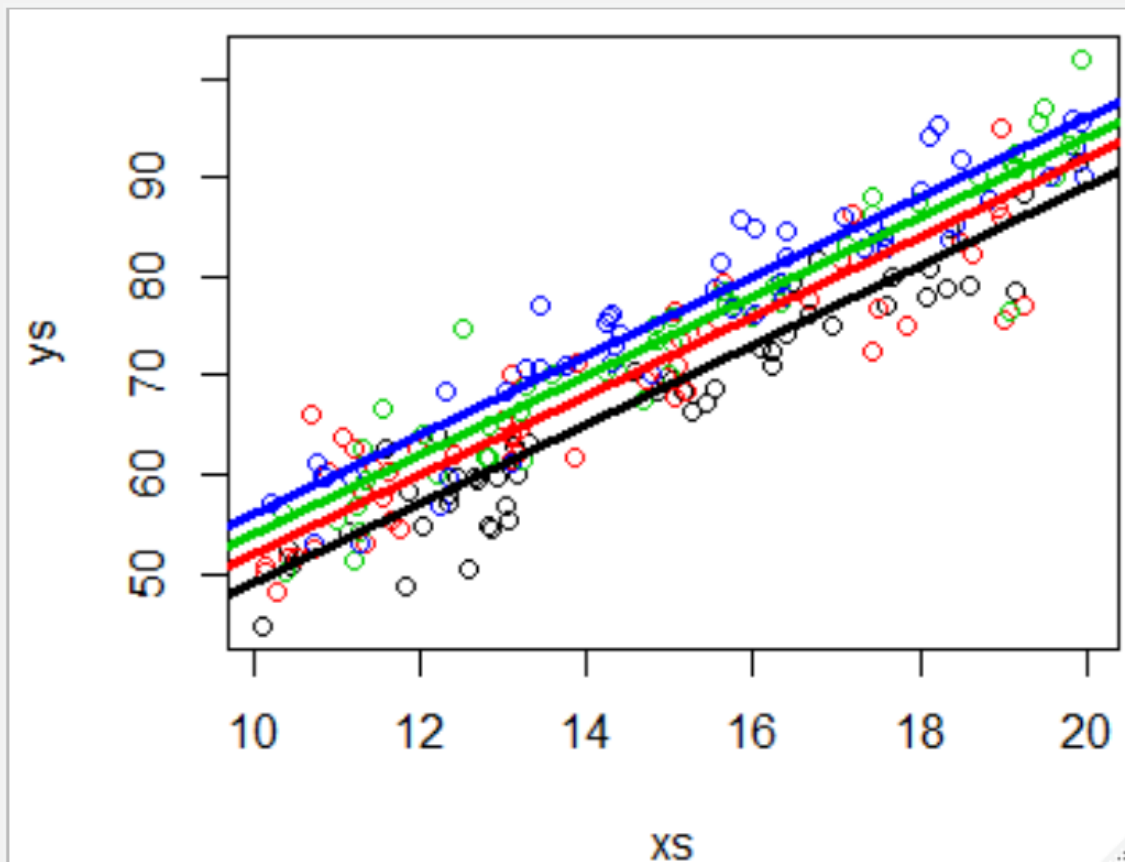
```
plot(ys~type)
```



## A second dummy example – covariance analysis!

```
set.seed(1234)
#create data
xs=runif(200,10,20)
type=rep(c("a","b","c","d"),each=50)
cores=rep(1:4,each=50)
ys=3+4*xs+ifelse(type=="a",5,ifelse(type=="b",8,ifelse(type=="c",10,12)))+rnorm(200,0,4)
#plot the data
plot(xs,ys,col=cores)
#make a regression model
summary(lm(ys~xs+type))

abline(9,4,lwd=3,col=1)
abline(9+3,4,lwd=3,col=2)
abline(9+5,4,lwd=3,col=3)
abline(9+7,4,lwd=3,col=4)
```





See extra material in folder

## Gestão de Páginas

- ▼ Modelação Ecológica
  - Modelação Ecológica(Ecologia Marinha)
  - Modelação Ecológica(Ecologia e Gestão Ambiental)
- ▼ Aulas
  - ▶ Aula1
  - ▼ Aula2
    - An extra set of exercises to practice R - (stolen from numerical
  - Aula3
  - ▼ Aula4
    - A task about merge
  - ▼ Aula5
    - FT1
    - FT2
  - ▶ Aula6
  - Aula7 08 10 2019
  - Aula8 09 10 2019
  - ▼ Aula9
    - On Regression Models**
- ▶ Outros Recursos



## On Regression Models

Página Ficheiros 4 Permissões Link

Adicionar Ficheiro

#	Nome	Permissões
1	data.csv	Público
2	OnRegMods.Rmd	Público
3	OnRegMods.pdf	Público
4	OnRegMods.html	Público

+ Criar