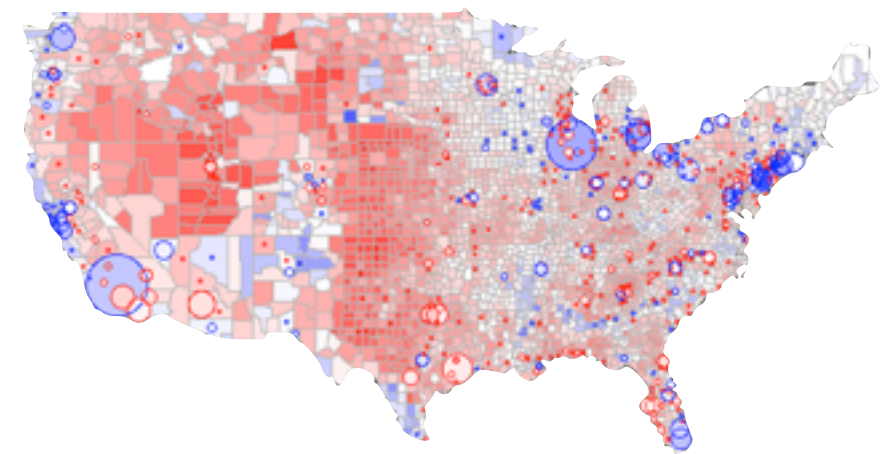
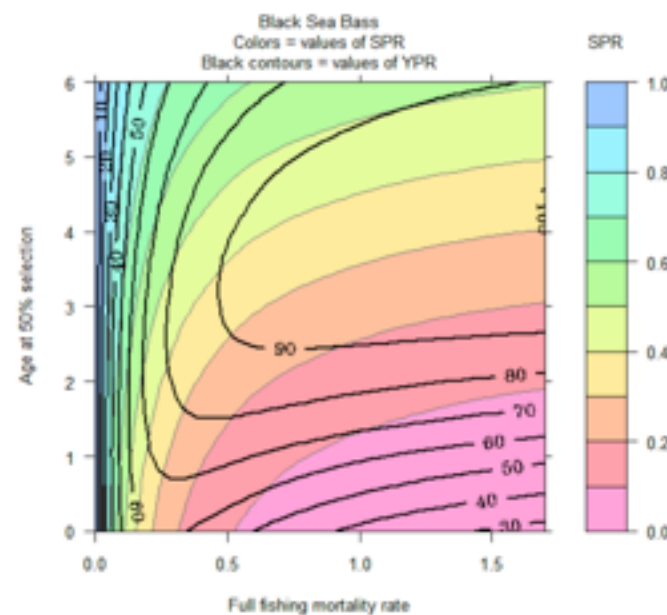


Introduction to R

Owen Jones

jones@biology.sdu.dk



What is R?

“A free programming language and software environment for statistical computing and graphics”

- Manipulate data
- Statistical analysis
- Mathematical modeling, simulation
- Plot figures, animations
- Write scripts and functions for own analysis

Installing R and R-studio



Go to: <http://mirrors.dotsrc.org/cran/>



MS Windows - select “*base*” and “*Download R 3.0.1 for Windows*”. This downloads the installer “*R-3.0.1-win.exe*”. Run this to install the program.



Apple OSX 10.6 and later - Download the installer package “*R-3.0.1.pkg*” and double click it to install the program.

Installing R and R-studio



Go to: <http://www.rstudio.com/ide/download>

Click on: 

Click to download the recommended version.

Follow the instructions!

R Studio

Write code as scripts here.

```
1 library(ape)
2 data(carnivora)
3 #par(mfrow=c(1,2))
4 hist(log(carnivora$FW))
5 hist(log(carnivora$GL))
6 logFW = log(carnivora$FW)
7 logGL = log(carnivora$GL)
8
9 model = lm(logGL~logFW)
10
11 summary(model)
12
13 plot(logFW, logGL, xlab = "Female weight (log)", ylab = "Gestation time (log)", type = "n")
14 points(logFW, logGL, pch = 16, cex = 2, col = "#FF000060")
15 abline(model, lty=2, lwd=2)
16:1
```

Data information here.

Variable	Value
carnivora	112 obs. of 17 variables
logFW	numeric[112]
logGL	numeric[112]
model	lm[13]

Console: code executed and outputs displayed here.

```
> logFW = log(carnivora$FW)
> logGL = log(carnivora$GL)
> model = lm(logGL~logFW)
> summary(model)
```

Call:
lm(formula = logGL ~ logFW)

Residuals:

Min	1Q	Median	3Q	Max
-0.70054	-0.12468	0.01541	0.14293	0.91134

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.92993	0.03877	101.367	< 2e-16 ***
logFW	0.13948	0.01776	7.854	8.64e-12 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

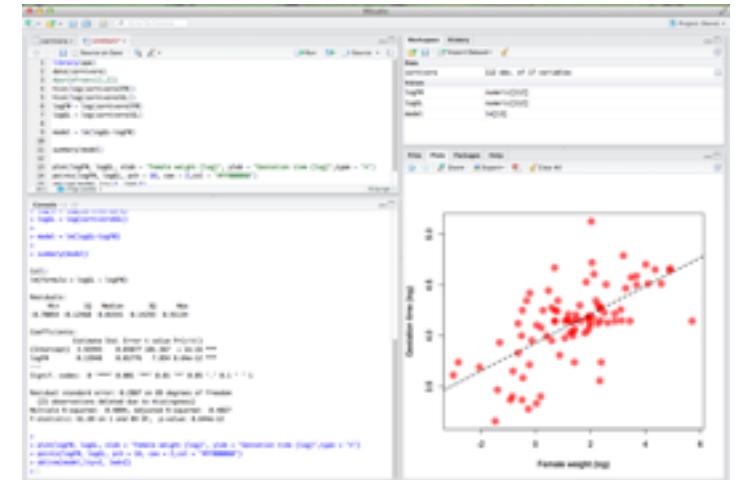
Residual standard error: 0.2887 on 89 degrees of freedom
(21 observations deleted due to missingness)
Multiple R-squared: 0.4094, Adjusted R-squared: 0.4027
F-statistic: 61.69 on 1 and 89 DF, p-value: 8.644e-12

Plots, file explorer, packages, and help here.

The plot shows a positive correlation between log-transformed female weight and log-transformed gestation time. The x-axis is labeled 'Female weight (log)' and ranges from -2 to 6. The y-axis is labeled 'Gestation time (log)' and ranges from 3.5 to 4.5. Red points represent the data, and a dashed black line represents the linear regression model.

R Studio

- Data entry in Excel
- Explore data in R to find errors
- Do analysis and exploration in R, not Excel
- Be lazy and use scripts (don't just type everything into the console!)
- Scripts: programs allowing you to repeat, edit, correct your work



The R language

Object oriented programming language

- Objects - numbers, variables
- Functions - manipulations of objects

The R language

Basic operations

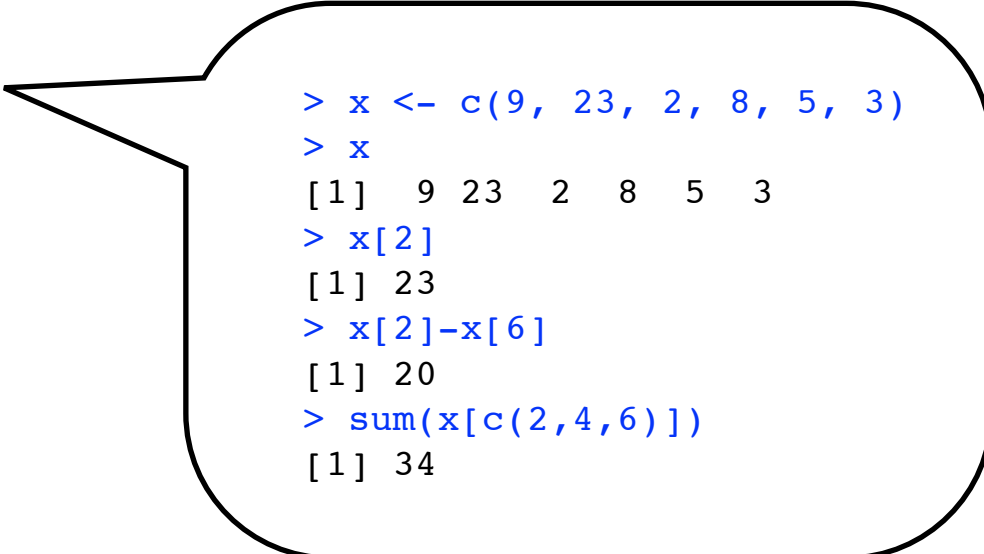
- Arithmetic (+, -, *, /, ^, sqrt)
- Matrix maths (t, *, %*%)
- Assignment (=, <-)

Getting help (e.g. for the “lm” function).

- ?lm, ??lm
- help("lm"), help.search("lm")

“Objects” in R

- Vectors and elements
- Data frames
- Lists
- Matrices



```
> x <- c(9, 23, 2, 8, 5, 3)
> x
[1] 9 23 2 8 5 3
> x[2]
[1] 23
> x[2]-x[6]
[1] 20
> sum(x[c(2,4,6)])
[1] 34
```

“Objects” in R

- Vectors and elements
- Data frames
- Lists
- Matrices

```
> Fruit
      Name Colour Quantity
1  banana yellow        5
2    kiwi  green        8
3 strawberry   red       20
> Fruit$Price = c(10,12,17)
> Fruit
      Name Colour Quantity Price
1  banana yellow        5     10
2    kiwi  green        8     12
3 strawberry   red       20     17
> Fruit$Colour
[1] yellow green  red
Levels: green red yellow
```

“Objects” in R

- Vectors and elements
- Data frames
- Lists
- Matrices

```
> n = c(2, 3, 5)
> s = c("aa", "bb", "cc", "dd", "ee")
> b = c(TRUE, FALSE, TRUE, FALSE, FALSE)
> x = list(n, s, b, 3)
> x
[[1]]
[1] 2 3 5

[[2]]
[1] "aa" "bb" "cc" "dd" "ee"

[[3]]
[1] TRUE FALSE TRUE FALSE FALSE

[[4]]
[1] 3

> x[[2]]
[1] "aa" "bb" "cc" "dd" "ee"
> x[[2]][3]
[1] "cc"
```

“Objects” in R

- Vectors and elements
- Data frames
- Lists
- Matrices

```
> A = matrix(c(5, 9, 10, 2, 5, 7), nrow=3, ncol=2)
> A
      [,1] [,2]
[1,]    5    2
[2,]    9    5
[3,]   10    7
> t(A)
      [,1] [,2] [,3]
[1,]    5    9   10
[2,]    2    5    7
> A * c(1,2,3)
      [,1] [,2]
[1,]    5    2
[2,]   18   10
[3,]   30   21
>
> A[1,]
[1] 5 2
> A[,2]
[1] 2 5 7
> A[1,1]
[1] 5
```

“Classes” in R

- “**Class**” defines the type of object and can influence what functions do.
- Common classes

`data.frame`

`list`

`matrix`

`integer`

`numeric`

`factor`

`character`

Model objects: `lm`, `glm`, `aov` etc.

Importing data from Excel

- Create a folder/directory for analysis and set it to be R's working directory

```
setwd( "PATH" )
```

- Save Excel file out as a *.csv file
- Import using the `read.csv` function

```
A <- read.csv( "PATH", header = TRUE )
```

Manipulating dataframes

- Subsetting data using logical operators
(`==`, `<`, `>`, `<=`, `>=`)

- By row/column

```
x[1:5, ], x[ ,6:10]
```

- `subset` function

```
> Fruit
  Name Colour Quantity Price
1  banana yellow      5    10
2   kiwi  green      8    12
3 strawberry   red     20    17
>
> subset(Fruit, Price<15)
  Name Colour Quantity Price
1 banana yellow      5    10
2  kiwi  green      8    12
```

Export data from R

- Export using the `write.csv` function

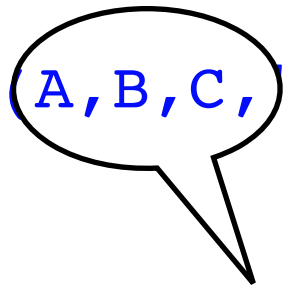
```
write.csv(myData, file = "myData.csv", header  
= TRUE, row.names = FALSE)
```

- This file can be opened in Excel
- Save the entire workspace

```
save.image("myWorkspace.RData")
```

- Save the parts of the workspace

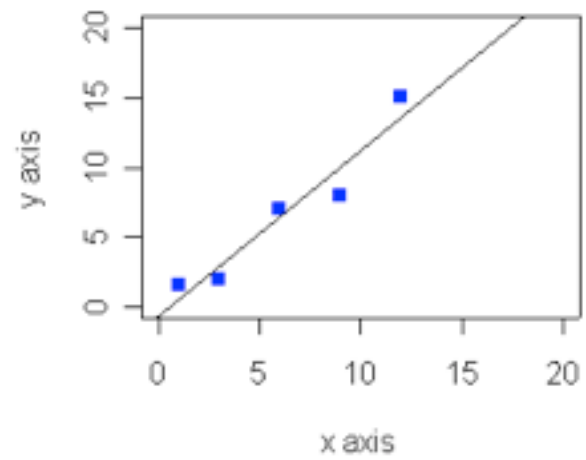
```
save(A, B, C, "myWorkspace.RData")
```



Objects you want to save

Graphics in R

my plot



Autos

