



**University of
Zurich^{UZH}**



MAKERERE UNIVERSITY

Data Analysis with R: Day 1

Sonja Hartnack, Terence Odoch & Muriel Buri

October 2017

Goals of the course

To be able to...

- import data sets to R
- describe data with R
- apply basic statistical tests in R
- some ideas for more advanced statistical tools ...
- simulate a data set similar to own research

Getting to know each other

- My name is ...
- I am doing a Master / a PhD in ...
- I hope to learn in this course how to
- My personal goal for this course is ...

How do we reach these goals

- hands on exercises with R:
 - chickwts
 - ToothGrowth
 - bacteria
 - perulung
 - ... and others.
- interactive discussions
- asking a lot of questions: ask google!
- group work
- short motivational lectures

Get started with data set: chickwts

An experiment was conducted to measure and compare the effectiveness of various feed supplements on the growth rate of chickens.

```
data(chickwts)
head(chickwts)
```

```
##   weight      feed
## 1    179 horsebean
## 2    160 horsebean
## 3    136 horsebean
## 4    227 horsebean
## 5    217 horsebean
## 6    168 horsebean
```

Ideas for plotting the data

Ideas for plotting the data

```
boxplot(weight ~ feed, data = chickwts)
```

Ideas for analysing the data

Ideas for analysing the data

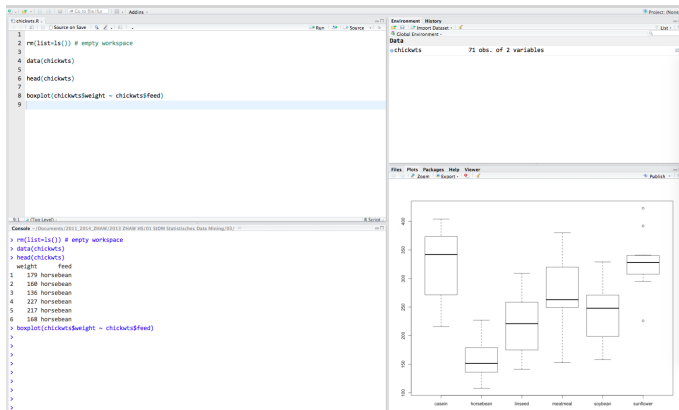
```
summary(aov(weight ~ feed, data = chickwts))
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## feed           5 231129   46226    15.37 5.94e-10 ***
## Residuals     65 195556     3009
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Functionalities in R and RStudio



- Source
- Console
- Environment, History, Files
- Files, Plots, Packages, Help





- Know in which directory you are

```
getwd()  
## [1] "/Users/murielburi/Dropbox/201710_Makerere/02_Lectures"
```

- Set directory path

```
setwd("C:\\Users\\admin\\201710_Makerere\\02_Lectures\\data\\")
```

- Always clean up before starting with new R-Script

```
rm(list=ls()) # empty workspace
```



```
?chickwts
```

Also, have a look at the examples at the end of the help pages.

Exercise 1



A data frame in R: chickwts



chickwts[ROWS , COLUMNS]

	weight	feed
1	179	horsebean
2	160	horsebean
3	136	horsebean
4	227	horsebean
5	217	horsebean
6	168	horsebean
7	108	horsebean
8	124	horsebean
9	143	horsebean
10	140	horsebean
11	309	linseed
12	229	linseed

chickwts[6, 1]

	weight	feed
1	179	horsebean
2	160	horsebean
3	136	horsebean
4	227	horsebean
5	217	horsebean
6	168	horsebean
7	108	horsebean
8	124	horsebean
9	143	horsebean
10	140	horsebean
11	309	linseed
12	229	linseed

chickwts[11, 2]



Values of ...

```
# ... all columns of sixth observation:  
chickwts[6, ]  
  
# ... all columns of sixth to eleventh observation:  
chickwts[6:11, ]  
  
# ... all columns of sixth, eleventh and twentieth observation:  
chickwts[c(6, 11, 20), ]  
  
# ... all rows of first column (weight):  
chickwts[, 1]  
  
# ... all rows of second column (feed):  
chickwts[, 2]  
# or use the "$" sign as a reference to column "feed":  
chickwts$feed
```

What is a data frame in R?



A data frame is used for storing a list of vectors of equal length. For example, the following variable `df` is a data frame containing three vectors `n`, `s`, `b`.

```
n <- c(2, 3, 5)
s <- c("aa", "bb", "cc")
b <- c(TRUE, FALSE, TRUE)
df <- data.frame(n, s, b) # df is a data frame
```

Following are the characteristics of a data frame:

- The column names should be non-empty.
- The row names should be unique.
- The data stored in a data frame can be of numeric, factor or character type.
- Each column should contain same number of data items.

Data frame in R



```
a <- c(1, 2, 3, 4)
```

```
a
```

```
## [1] 1 2 3 4
```

```
data.frame(a)
```

```
##    a
```

```
## 1 1
```

```
## 2 2
```

```
## 3 3
```

```
## 4 4
```

```
b <- c("d", "h", "h", "d")
```

```
dat <- data.frame(a, b)
```

```
dat
```

```
##    a b
```

```
## 1 1 d
```

```
## 2 2 h
```

```
## 3 3 h
```

```
## 4 4 d
```

Data frame in R: How to add a vector variable



```
c <- factor(c("male", "female", "male", "female"),
            levels = c("female", "male"))
dat$c <- c # use "$" to refer to the additional vector variable
dat
```

```
##   a b    c
## 1 1 d  male
## 2 2 h female
## 3 3 h  male
## 4 4 d female
```

```
# What is the dimension (number of rows and columns) of our data frame?
dim(dat) # 4 rows and 3 columns
```

```
## [1] 4 3
```

Exercise 2



How to install a package



```
install.packages("MASS")  
library("MASS")  
data(bacteria)  
?bacteria
```

How to install a package



The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains the R script:

```
1 install.packages("MASS")
2 library(MASS)
3 data(bacteria)
4 str(bacteria)
5
```
- Console:** Shows the output of the script:

```
package 'MASS' successfully unpacked and MD5 sums checked
The downloaded binary packages are in
  C:\Users\admin\AppData\Local\Temp\RtmpIF4M32\downloaded_packages
> library(MASS)
> data(bacteria)
'data.frame': 220 obs. of 6 variables:
 $ y : Factor w/ 2 levels "n","y": 2 2 2 2 2 2 1 2 2 2 ...
 $ ap : Factor w/ 2 levels "a","p": 2 2 2 2 1 1 1 1 1 1 ...
 $ h1o: Factor w/ 2 levels "h1","1o": 1 1 1 1 1 1 1 1 2 2 ...
 $ week: int 0 2 4 11 0 2 6 11 0 2 ...
 $ ID : Factor w/ 50 levels "X01","X02","X03",...: 1 1 1 1 2 2 2 3 3 ...
 $ trt : Factor w/ 3 levels "placebo","drug",...: 1 1 1 1 3 3 3 2 2 ...
> View(lung)
```
- Environment:** Shows the loaded data:
 - bacteria: 220 obs. of 6 variables
 - lung: 636 obs. of 6 variables
- Packages:** Shows the installed packages, including MASS.

Exercise 3



How to google for getting help in R

- Google for **select observations in R**.