



**University of  
Zurich<sup>UZH</sup>**



MAKERERE UNIVERSITY

# **Data Analysis with R:**

## **Lecture Slides (all)**

Sonja Hartnack, Terence Odoch & Muriel Buri

October 2018

# 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

### **Course schedule:**

- Starting at 9:00am / 9:30am (?)
- Tea breaks in between
- Lunch break
- Teaching until 4.30pm (~ 5pm)

### **Optaining a certificate is conditional on ...**

- active participation in class
- attending at least 75 % of the course (lecture & exercises)
- short final exam (format to be defined)

# 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 & presentations of student solutions
- 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.

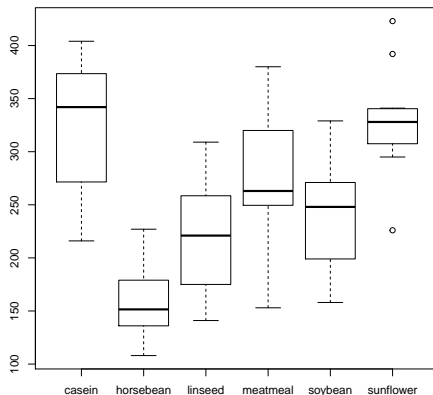
```
data(chickwts)
head(chickwts)
```

```
##      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

## A hands on example

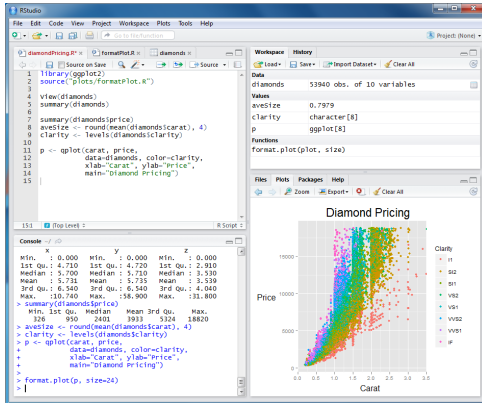


```
x <- c(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
y <- c(20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30)
plot(x, y)
```

# Functionalities in R and RStudio



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





- Define manually a new folder called **RCourse** in your personal documents on your personal computer
- Know in which directory you are

```
getwd()  
## [1] "/home/mburi/ownCloud/git/DataAnalysisWithR/Lectures"
```

- Set directory path

```
# back- and forslash is dependent on the system  
setwd("C:/Users/muriel/Documents/RCourse/")  
setwd("C:\\Users\\muriel\\Documents\\RCourse\\")
```

- Always clean up before starting with new R-Script

```
rm(list=ls()) # empty workspace, delete previously saved variables
```



```
?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   |
| 13 | 181    | 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   |
| 13 | 181    | 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.
- Each column should contain same number of data items.



```
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 variable (var)



```
my.var <- c(1.3, 1.5, 1.8, 2.4)
# use "$" to refer to the additional vector variable
dat$my.var1 <- my.var
dat$my.var2 <- my.var
dat
```

```
##   a b my.var1 my.var2
## 1 1 d     1.3     1.3
## 2 2 h     1.5     1.5
## 3 3 h     1.8     1.8
## 4 4 d     2.4     2.4
```

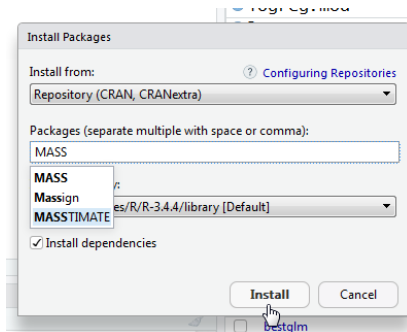
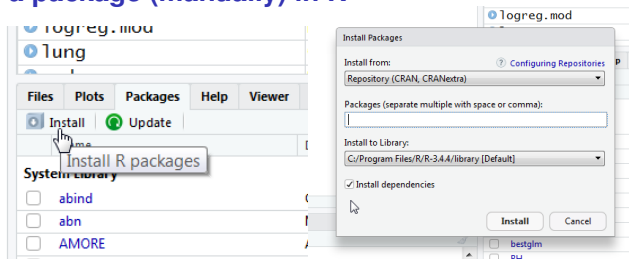
```
# What is the dimension (number of rows and columns) of our data frame?
dim(dat) # 4 rows and 3 columns
```

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

## Exercise 2



# How to install a package (manually) in R



## Using R is like cooking ...

|   |   |
|---|---|
| Get into the kitchen  | Change working directory  |
| Get specialist electric tools into your kitchen (e.g. blender, ice-cream maker, etc.) | Install packages  |
| Switch on your specialist electric tools  | Load packages using the "library" function  |
| Bring in your ingredients   | Import data and save to R data frames   |
| Check your ingredients  | Use the function "summary" and basic tables to check your data for missing or implausible values (e.g. a number in a variable where "yes" or "no" are expected) |
| Chop things up (if required)  | Split or filter data  |
| Cook, using general and specialist tools  | Carry out further descriptive and test statistics   |

# How to install a package in R



```
# INSTALL package (only done ONCE!)
install.packages("MASS")
# LOAD package (whenever you use something from it!)
library("MASS")
data(bacteria)
?bacteria
```



## Exercise 3

