Data Analysis with R:
Day 1 - Lecture Slides

Sonja Hartnack, Terence Odoch & Muriel Buri

October 2018

Lecture Slides for Day 1

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

General remarks

Course schedule:
- Starting at 9:00am / 9:30am (?)
- Tea breaks in between
- Lunch break
- Teaching until 4.30pm ($\sim$ 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 & student's present their own solutions

- ask us a lot of questions but also ask google for help!

- 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.

```
# load data set "chickwts"
data(chickwts)
# the head(...) function shows the first 6 observations
head(chickwts)

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

# FUNCTION - open bracket - DATA SET / VARIABLE - close bracket
```
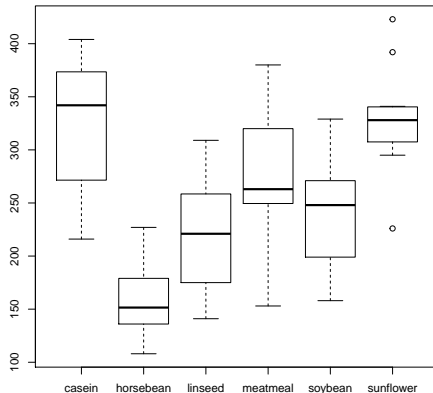
Ideas for plotting the data

Ideas for plotting the data

```
# use x axis to show the categorical variable (feed),
# y axis to represent the continuous variable (weight)
# boxplot (y.cont.variable ~ x.cat.variable, data = dataset)
# ?boxplot
boxplot(weight ~ feed, data = chickwts)
```

Ideas for analysing the data

Ideas for analysing the data

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

##             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

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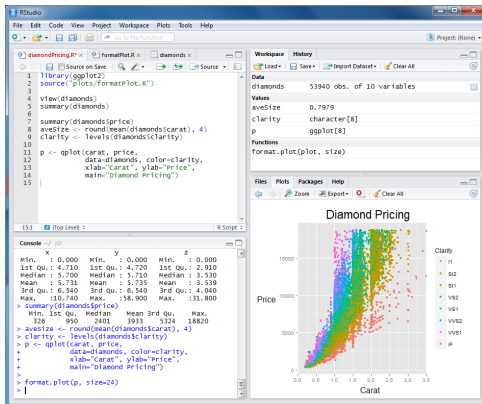
```r
x <- c(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
y <- c(20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30)
plot(x, y)
```

Functionalities in R and RStudio

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

Good housekeeping!

- 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
```

How to get help in R

```
?chickwts
?boxplot
```

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

Exercise 1

A data frame in R: `chickwts`

Rows and columns of a data frame: `chickwts`

R

Values of ...

```r
# ... all columns of sixth observation:
chickwts[6, ]

# ... all columns of sixth to eleventh observation:
chickwts[c(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
```